PROCEEDINGS

ICIST 2016

6TH INTERNATIONAL CONFERENCE ON
INFORMATION SOCIETY AND TECHNOLOGY

6th International Conference on Information Society and Technology

ICIST
2016

Proceedings

Publisher

Society for Information Systems and Computer Networks

Editors

Zdravković, M., Trajanović, M., Konjović, Z.

ISBN: 978-86-85525-18-6

Issued in Belgrade, Serbia, 2016.

6th International Conference on Information Society and Technology
ICIST 2016

Conference co-chairs

prof. dr Zora Konjović, University of Novi Sad
prof. dr Miroslav Trajanović, University of Niš
dr Milan Zdravković, University of Niš

International Programme Committee

Alexis Aubry, Université de Lorraine, France
Miloš Bogdanović, University of Niš, Serbia
Carlos Coutinho, Universidade Nova de Lisboa, Portugal
Žarko Čojbašić, University of Niš
Mariana Damova, Mozaika, Bulgaria
Michele Dassisti, Politecnico di Bari, Italia
Igor Dejanović, University of Novi Sad
Anna Fensel, Semantic Technology Institute (STI) Innsbruck, University of Innsbruck, Austria
Ip-Shing Fan, Cranfield University, UK
Vladimir Filipović, University of Belgrade, Serbia
Stevan Gostojić, University of Novi Sad
Wided Guédria, CRP Henri Tudor, Luxembourg
Nataša Gospić, University of Belgrade, Serbia
Miša Ivković, E-Drustvo, Serbia
Valentina Janev, Institute Mihajlo Pupin, Serbia
Ricardo Jardim-Gonçalves, UNINOVA, New University of Lisbon, Portugal
Dorina Kabakchieva, University of National and World Economy, Sofia, Bulgaria
Zora Konjović, Singidunum University, Serbia
Aleksandar Kovačević, University of Novi Sad, Serbia
Srđan Krčo, University of Belgrade, Serbia
Eduardo Rocha Loures, Pontifícia Universidade Católica do Paraná, Brasil
Vuk Malbaša, University of Novi Sad
Zoran Marjanović, University of Belgrade, Serbia

Miloš Madić, University of Niš, Serbia

Istvan Mezgar, Computer and Automation Research Institute, Hungarian Academy of Sciences, Hungary

Dragan Mišić, University of Niš, Serbia

Branko Milosavljević, University of Novi Sad

Gordana Milosavljević, University of Novi Sad

Néjib Moalla, University Lyon 2 Lumière, France

Siniša Nešković, University of Belgrade, Serbia

Ovidiu Noran, Griffith University, Australia

Hervé Panetto, Université de Lorraine, France

Mirjana Pejić Bach, University of Zagreb, Croatia

Valentin Penca, University of Novi Sad

David Romero, Tecnológico de Monterrey, Mexico

Camille Salinesi, Pantheon-Sorbonne University, Computer Science Research Center (CRI), France

Joao Sarraipa, UNINOVA, Portugal

Jean Simão, Universidade Tecnológica Federal do Paraná, Brasil

Goran Sladić, University of Novi Sad

Jelena Slivka, University of Novi Sad

Richard Mark Soley, OMG, USA

Kamelia Stefanova, Faculty of Applied Informatics and Statistics, University of National and World Economy, Sofia, Bulgaria

Leonid Stoimenov, University of Niš, Serbia

Anderson Szejka, Pontifical University Catholic of Paraná, Brasil

Sašo Tomažić, University of Ljubljana, Slovenia

Miroslav Trajanović, University of Niš, Serbia

Milan Trifunović, University of Niš, Serbia

Milan Vidaković, University of Novi Sad

Nikola Vitković, University of Niš, Serbia

Georg Weichhart, Johannes Kepler Universität Linz, Austria

Jelena Zdravković, Stockholm University, Sweden

Milan Zdravković, University of Niš, Serbia

Martin Zelm, INTEROP-VLab, Belgium

CONTENT

Volume 1

Foreword to the Proceedings of the 6th International Conference on Information Society and Technology

Zora Konjović*, Milan Zdravković**, Miroslav Trajanović**
* Singidunum University, Belgrade, Serbia
** Laboratory for Intelligent Production Systems (LIPS),
Faculty of Mechanical Engineering, University of Niš, Niš, Serbia

zkonjovic@singidunum.ac.rs, milan.zdravkovic@gmail.com,
miroslav.trajanovic@masfak.ni.ac.rs

I. INTRODUCTION

Another successful edition of ICIST conference series has been organized at Kopaonik winter resort, on February 28th – March 2nd, 2016. ICIST is one of the most influential ICT events in the region with a long tradition of academic and industrial participation. Besides provided networking opportunities, in the past few editions it significantly boosted the quality of the presented work, with the outputs to the reputable journals.

In this Foreword, we present the highlights of the recent ICIST edition and introduce a reader to the content of the book of proceedings. Finally, we provide the discussion on the current state of the play in this year's focal area of the ICIST conference. As it was the case last year, the book of proceedings is organized in two volumes. Volume 1 chapters are the papers that have been accepted for presentation in the regular sessions. Volume 2 chapters are the papers that have been presented at the poster sessions.

II. OPEN AND BIG DATA

According to IBM assertions, we create 2.5 quintillion bytes of data per day and as of 2020 we shall reach the total amount of 44 zetabytes. 44 zetabytes is estimated to be almost 60 times the amount of all the grains of sand on all the beaches on earth. Individuals, various organizations, and governments produce huge amounts of data as part of their everyday activity/work – data related to environment, public-transport, health, education, etc. This data is used by governments to improve public-services, by companies to improve businesses, and by individual(s) to improve her/his status. General opinion is that there are many opportunities to use such data beyond the purpose it was originally collected. This is the driving force behind Big Data and Open Data movements.

Gartner IT Glossary[1] defines Big Data as follows: "Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation." The Open Definition[2] provided by Open Knowledge[3] sets out principles that define "openness" in relation to data and content. The definition defines "open" by the statement: "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)". "Open data" is defined by the statement: "Open data and content can be freely used, modified, and shared by anyone for any purpose".

According to "Open Data Barometer Global Report - Second Edition"[4] a global movement to make government "open by default" gained momentum in 2013, when the G8 leaders signed the Open Data Charter[5]. This was followed in 2014 by the G20 largest industrial economies pledging to advance open data as a weapon against corruption, and in 2015 by proposal of the International Open Data Charter[6], which is signed by 19 national, state and city governments worldwide.

There are a number of organization and groups that are driving Big Data, Open Data and Open Government Data (OGD) research, best practice and technologies. Without an ambition to be exhaustive we shall list some of them here. European Commission (Communication on the data-driven economy[7]), United Nations (Global Pulse[8]), many national (The US Big Data Research and Development initiative[9], the Australian Government

[1] http://www.gartner.com/it-glossary/big-data/
[2] http://opendefinition.org/
[3] The Open Knowledge Foundation, trading as Open Knowledge, is a not-for-profit organization. It is incorporated in England & Wales as a company limited by guarantee.
[4] http://opendatabarometer.org/assets/downloads/Open Data Barometer - Global Report - 2nd Edition - PRINT.pdf
[5] UK Cabinet Office, (June 18th 2013) G8 Open Data Charter and Technical Annex, https://www.gov.uk/government/publications/open-data-charter
[6] http://opendatacharter.net/wp-content/uploads/2015/10/opendatacharter-charter_F.pdf
[7] http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc_id=6210; http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc_id=6216
[8] http://www.unglobalpulse.org/
[9] https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf

Public Service Big Data Strategy[10], the UK ESRC Big Data Network[11]) and local governments/agencies, universities across the Globe, the Insight Centre for Data Analytics[12], Govlab[13], the Omidyar Network[14], the Open Data Institute[15], the Open Data Research Network[16], the Open Government Partnership[17], the Open Knowledge Foundation[18], the Sunlight Foundation[19], W3C[20], the World Bank[21], and the World Wide Web Foundation[22] exhibit internationally recognized activities and results in Big Data, Open Data and Open Government Data. Open Government Data in Serbia is still at infancy stage. Directorate for eGovernment of the Ministry of State Administration and Local Self-Government is in charge with Open Government Data in Serbia.

One of the most consequent (with respect to "openness") initiatives related to Open Data and Open Government Data, which is extremely valuable for underdeveloped countries such as Serbia, comes from Sir Tim Berners-Lee. The World Wide Web Foundation established by Sir Tim Berners-Lee in 2009 and the Open Data Institute jointly initiated research called 'Open Data Barometer'. The Barometer was supported by the Open Data for Development (OD4D) program, a partnership funded by Canada's International Development Research Centre (IDRC), the World Bank, United Kingdom's Department for International Development (DFID), and Global Affairs Canada (GAC). Within the 'Open Data Barometer'[23] framework, three documents are created so far containing a comprehensive job already done: 'Open Data Barometer - 2013 Global Report', 'Open Data Barometer Global Report - Second Edition', and 'Open Data Barometer - ODB Global Report 3rd Edition'. As declared by the authors in the document 'Open Data Barometer - 2013 Global Report': "Above all, the Open Data Barometer is a piece of open research. All the data gathered to create the Barometer will be published under an open license, and we have sought to set out our methodology clearly, allowing others to build upon, remix and reinterpret the data we offer. Data collected for the Barometer is the start, rather than the end, of a research process and exploration."

As confirmation of being "a piece of open research", all three reports expose the research methodology in detail, all the collected data, and all the results of the analyses for 77, 86 and 92 countries worldwide. All reports, in addition to an exhaustive analysis, contain a ranking of the countries on open data readiness, implementation, and impact as well as key findings for the observed timeframes (years 2013, 2014, and 2015). Just to mention that of all Ex-Yugoslavia countries only Macedonia was included in the third report.

Yet another initiative, directly bringing valuable results to underdeveloped countries including Serbia, is the World Bank's project ODRA (Open Data Readiness Assessment). The World Bank's Open Government Data Working Group has prepared and published a revised draft 'Open Data Readiness Assessment Tool'[24] aimed to assist in diagnosing what actions a government could consider in order to establish an Open Data initiative. The latest version of ODRA consists of two documents the 'User Guide'[25] and the 'Methodology'[26]. The approach proposed by World Bank was applied so far to 10 countries including Serbia. The assessment for Serbia, which was published in December 2015[27], contains both the overall assessment and suggested list of actions. The action plan was focused on integrating actions at the top, middle and bottom, including the active involvement of civil society and the business community. A first focus of those actions is making open data available where that is easy to do so, and to form pilot groups of government agencies, civil society, business and developers to quickly create a few practical examples of the usage of open data, which can serve as example for further extension of the open data program. It is recognized that for a sustainable and integrated role of open data as part of public service delivery however, the problems with retaining skilled staff and maintaining a sufficient level of IT knowledge across government are a significant obstacle. Nevertheless, by the end of February 2016 several governmental institutions joined the ODG initiative and open some of their data sets (see Table 1). All these datasets can be accessed via eGovernment portal of the Republic of Serbia, link http://data.gov.rs/.

In July 2014, the European Commission outlined a new strategy on Big Data, supporting and accelerating the transition towards a data-driven economy in Europe. The IDC study[28] predicted the Big Data technology and services market to grow worldwide from \$3.2 billion in 2010 to \$16.9 billion in 2015. Wikibon[29] claims that the overall Big Data market grew from \$18.3 billion in 2014 to \$22.6 billion in 2015. The study 'Big Data Analytics: An assessment of demand for labor and skills, 2012-2017'[30] predicts that in the UK alone, the number of big data staff specialist working in large firms will increase by more than 240% over the next five years. There are also several studies that have investigated the value of the Open Data economy. Graham Vickery[31] estimated that EU27 direct public sector information (PSI)[32] re-use

[10] https://www.aiia.com.au/documents/policy-submissions/policies-and-submissions/2013/the_australian_public_service_big_data_strategy_04_07_2013.pdf
[11] http://www.esrc.ac.uk/research/our-research/big-data-network/
[12] https://www.insight-centre.org/
[13] http://www.thegovlab.org/
[14] https://www.omidyar.com/
[15] http://theodi.org/
[16] http://www.opendataresearch.org/
[17] http://www.opengovpartnership.org/
[18] https://okfn.org/
[19] http://sunlightfoundation.com/
[20] https://www.w3.org/
[21] http://www.worldbank.org/en/about
[22] webfoundation.org
[23] http://opendatabarometer.org/
[24] http://opendatatoolkit.worldbank.org/docs/odra/odra_v2-en.pdf
[25] http://opendatatoolkit.worldbank.org/docs/odra/odra_v3.1_userguide-en.pdf
[26] http://opendatatoolkit.worldbank.org/docs/odra/odra_v3.1_methodology-en.pdf
[27] http://www.rs.undp.org/content/serbia/sr/home/library/democratic_governance/ocena-spremnosti-za-otvaranje-podataka/
[28] http://ec.europa.eu/digital-single-market/news-redirect/17072
[29] http://siliconangle.com/blog/2016/03/30/wikibon-names-ibm-as-1-big-data-vendor-by-revenue/
[30] http://ec.europa.eu/digital-single-market/news-redirect/17072
[31] Vickery, G. (2011). Review of Recent Studies on PSI Re-use and Related Market Developments, Information Economics, Paris
[32] Directive 2003/98/EC on the re-use of public sector information5 (the PSI Directive); Directive 2013/37/EU, amending Directive 2003/98/EC

market was of the order of EUR 32 billion in 2010. The aggregate direct and indirect economic impacts from PSI applications and use across the whole EU27 economy are estimated to be of the order of EUR 140 billion annually.

Table 1. Serbian governmental institutions that open datasets by the end of February 2016.

Institution	Available formats	Accessibility restriction
Office of The Commissioner for Information of Public Importance and Personal Data Protection	csv	None
Ministry of Education Science and Technological Development	html, json, csv, xls	None
Ministry of Interior Affairs/None	csv, xls	None
Public Procurement Agency/None	csv	None
Environment protection Agency/None	csv	None
Medicines and Medical Devices Agency of Serbia	csv	Governmental institutions only

The above estimates of direct and indirect PSI re-use are based on business as usual, but other analysis suggests that if PSI policies were open, with easy access for free or marginal cost of distribution, direct PSI use and re-use activities could increase by up to EUR 40 billion for the EU27.

In McKinsey's 2013 report 'Open data: Unlocking innovation and performance with liquid information'[33] authors state: "An estimated $3 trillion in annual economic potential could be unlocked across seven domains. These benefits include increased efficiency, development of new products and services, and consumer surplus (cost savings, convenience, better-quality products). We consider societal benefits, but these are not quantified. For example, we estimate the economic impact of improved education (higher wages), but not the benefits that society derives from having well-educated citizens. We estimate that the potential value would be divided roughly between the United States ($1.1 trillion), Europe ($900 billion) and the rest of the world ($1.7 trillion)".

According to the FP7 project BYTE (Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities)[34], from the beginning of the millennium every three years a new wave of Big Data technologies has been building up: 1) The "batch" wave characterized by distributed file systems and parallel computing, 2) the "ad-hoc" wave of NewSQL characterized by underlying distributed data structures and distributed computing paradigm, and 3) the "real-time" wave, which enables insights in milliseconds through distributed stream processing. Even if it is quite clear what is expected from the Big Data technology today (cope with volume, variety, and velocity), there is neither a single tool nor choice of a platform that could satisfy these expectations. Current solutions are mainly concerned with high-volume and high-velocity issues with two architectural patterns passable: well-known Schema on Read[35] and Lambda Architecture[36] that appeared in 2013 and is considered a kind of consolidation of the Big Data technology stack. A key is that lambda architectural style recognizes the very different challenges of volume and velocity. Hence, this pattern splits data handling into three layers: fast layer for real-time processing of streaming data, the batch layer for cost-efficient persistent storage and batch processing, and the serving layer that enables different views for data usage. Even with these two patterns well established, the third dimension of Big Data, variety, remains unresolved as well as many other challenges such as veracity, actionability, and privacy. These are up-and-coming subjects of Big Data research and development today.

The ultimate vision of Open Government Data heavily relies upon Linked Data because efficient utilization of such data calls for intelligent automated access to globally distributed, heterogeneous data. Hence, the Open Government Data technology basically corresponds to the Linked Data technology, i.e. two technologies that are fundamental to the Web, URIs and HTTP, supplemented by RDF technology, RDFS, and OWL. Unfortunately, current Open Government Data deployments are commonly only isolated, internally linked islands of datasets provided in CSV or spreadsheet formats. We have selected the article "Linked Open Government Data: Lessons from Data.gov.uk"[37] for comments here because it addresses issues that are, by our opinion, of crucial importance for the future OGD. In this paper authors opt for the use of Semantic Web standards in OGD referring to this vision as the Linked-Data Web (LDW). Thereby, they emphasized four research challenges as relevant for representing OGD in RDF: discovering appropriate datasets for applications; integrating OGD into the LDW; understanding the best join points (the points of reference the databases share) for diverse datasets; building client applications to consume the data. As a conclusion, they exposed the lessons learned addressing governments, technical community and citizens as well as the list of bottlenecks in exporting OGD to the LDW. In addition to the unwillingness of public service providers to surrender control of their data, the issues related to discovery of OGD, ontological alignment, interfaces, and consumption are recognized as bottlenecks and, consequently, candidates for future research.

[33] Manyika, J. Chui, M. Groves, P. Farrell, D. Van Kuiken, S. and Doshi, E, A. (2013). Open data: Unlocking innovation and performance with liquid information, McKinsey Global Institute, New York.

[34] http://byte-project.eu/

[35] https://www.techopedia.com/definition/30153/schema-on-read

[36] Marz, Nathan, "Big Data – Principles and best practices of scalable realtime data systems", Manning MEAP Early Access Program, Version 17, no date. http://manning.com/marz/BDmeapch1.pdf

[37] Shadbolt, Nigel, O'Hara, Kieron, Berners-Lee, Tim, Gibbins, Nicholas, Glaser, Hugh, Hall, Wendy and schraefel, m.c. (2012) Linked open government data: lessons from Data.gov.uk. *IEEE Intelligent Systems*, 27, (3), Spring Issue, 16-24. (doi:10.1109/MIS.2012.23).

III. ICIST 2016 KEY FIGURES

In the ICIST event preparation phase, 63 of the distinguished researchers in the field from 18 countries have accepted the invitation to participate in the work of the International Programme Committee (IPC).

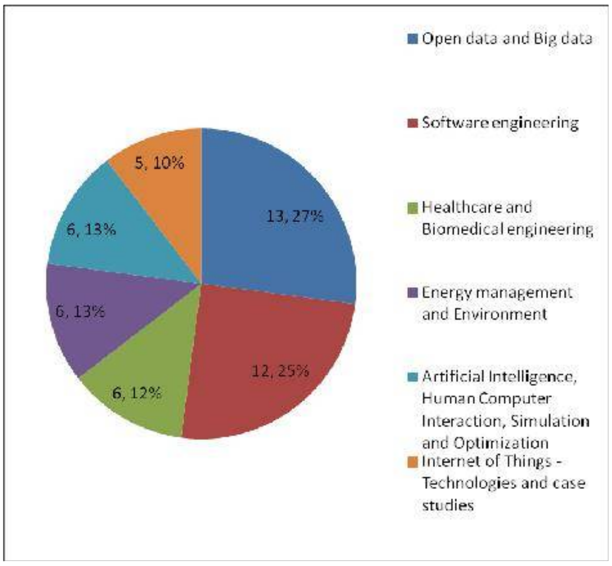


Fig. 1. Distribution of regular papers among main topics

This year, 80 papers were submitted to the conference, out of which the authors of 72 papers were invited to present their research work in regular and poster sessions. Thus, overall acceptance rate was 90%. Based on the reviewers' comments, the IPC has found that 48 papers show higher relevance and the level of scientific contribution, appropriate for presentation at the regular sessions, resulting with the regular paper acceptance rate of 60%. 220 researchers from 20 countries contributed as authors or co-authors to the submitted papers, making this event truly international.

Based on the works presented in the accepted regular papers, the following main topics and corresponding session distribution has been adopted by the IPC co-chairs: Open data and Big data; Software engineering; Healthcare and Biomedical engineering; Energy management and Environment; Artificial intelligent, HCI, Simulation; and Optimization and Internet of Things – Technologies and case studies. The distribution of regular papers among main topics is illustrated on Fig.1.

To the large extent, this distribution corresponds to the topics addressed by the previous three ICIST editions, and it clearly demonstrates the selection of topics of the major interest in the Serbian research community, as it is the most present in the conference series.

IV. SCIENTIFIC PROGRAMME

Since Open data and Big data was previously selected by the co-chairs as the focal topic of this year's ICIST, the largest number of papers addressing this area was submitted. The diverse topics addressed included but weren't restricted to government initiatives and their assessments, open data use in judicial systems, metadata management, transformation and analysis of linked data, social networks data analysis, etc. The scientific programme was complemented with the special round table discussion on the use of open government data in Serbia and presentation of the existing data sets.

The topic of software engineering was addressed mostly in the field of model-based software engineering, demonstrating very large interest of the local scientific community in this area.

Second year in a row, Healthcare and Biomedical Engineering becomes one of the most visited sessions, thanks to highly international participation and two very successful, nationally funded projects in this area. This year's edition addressed the potential of use of the ICT for the resolution of the specific health-related challenges, such as vertigo disorders, imaging issues, information extraction from EHR, implant material selection, and others.

The session related to the topic of energy management and environment dealt with the issues of energy management in multiple carrier infrastructures, energy management in neighborhoods, integration of network analysis system with GIS and tools for urban air quality studies.

Internet of Things remains the topic of high interest. This year's session addressed real-time biofeedback systems, RFID supported healthcare information systems, cloud-based IoT platforms and IoT demonstrations in the domains of agriculture, aquaculture and tourism.

Last, but not the least, one dedicated session presented 6 papers with the recent results of the research in the domains of simulation and optimization, human-machine interaction and artificial intelligence.

A. *Poster sessions*

Traditionally, the poster sessions are venues with the most interesting discussions that take place in less formal environment. The participating authors have presented a number of high-quality technical solutions and methodologies for solving different technical and societal problems.

V. ACKNOWLEDGEMENT

The editors wish to express a sincere gratitude to all members of the International Program Committee and external reviewers, who provided a great contribution to the scientific programme, by sending the detailed and timely reviews on this year ICIST's submissions.

The editors are grateful to the organizing committee of YUINFO conference for providing full logistics and all other kinds of support in setup of exciting scientific and social program of ICIST 2016.

Towards flexible short answer questions in the Moodle Quiz

Milena Frtunic*, Leonid Stoimenov*, Oliver Vojinovic*, Ivan Milentijevic*
* Faculty of Electrical Engineering, University of Niš, Niš, Serbia
milena.frtunic@elfak.ni.ac.rs, leonid.stoimenov@elfak.ni.ac.rs, oliver.vojinovic@elfak.ni.ac.rs, ivan.milentijevic@elfak.ni.ac.rs

***Abstract*— Assessment of student's knowledge and observation of their learning progress by using learning management systems is proven to be a very good solution. Main reason for that is provided possibility for testing a large number of students at the same time and automated evaluation of all quiz attempts. However, there are still a lot of limitations regarding automatic evaluations. In order to reduce one of those limitations in Moodle LMS, this paper will present an improvement of the automatic evaluation for questions with short answers in Moodle LMS system. Within this paper, an algorithm which is based on checking the similarity between two sentences will be introduced as a new solution for assessing short answer questions created in English language.**

I. Introduction

The advancement of technology had a great influence in all aspects of an average person's everyday life. Due to this, a number of new opportunities and possibilities for more efficient learning and informing were introduced in education. These changes have led to the development of different learning management systems (LMS) that offer numerous features in both the administrative and teaching field for both educational institutes and individual tutors.

Teaching field consists from different aspects, starting from storing teaching materials and giving lessons to testing students' knowledge. Learning management systems which offer possibility for testing knowledge provide teaching staff a quick and efficient way to evaluate knowledge of a large number of students at the same time. Such systems offer some form of automatic evaluation of tests, but that automatization usually applies only to questions with offered predefined answers where student have to choose one or more answers he thinks is correct. Some systems offer possibility for automatic evaluation of questions where students write free answer that they find to be correct. Unfortunately, in those situations in order for automatic evaluation to work and question to be evaluated as correct, written answer usually has to be identical to one that is defined to be correct by questions' creator.

In those cases teaching staff use questions with free answers only when answer can be written in a single word or said in one or two correct ways. In all other situations they usually use other types of questions, such as those where student needs to choose answers she/he thinks are correct or make some matching between facts. Although, this way of forming quiz keeps it 100% possible to be automatically evaluated, it leaves space for students to select correct response by choosing randomly. Because of that, it is questionable whether the quiz results are presentation of real students' knowledge. For that reason, teaching stuff is forced to define questions where they will expect longer answers knowing that automatic evaluation probably won't work correctly. In those situations they are obligated to go manually through every test and check whether it is necessary to reevaluate question response and change final points.

The goal of this paper is to present a solution for this problem in form of improving automatic evaluation by including algorithm that will check semantic similarity between written answer and the one defined as correct by question's creator. This research is implemented on Moodle (Modular Object Oriented Dynamic Learning Environment) learning management systems. This choice was made because Moodle is one of the most popular LMS in Europe and it is in everyday use at Faculty of Electrical Engineering, University of Niš. Further, it offers a very rich module for testing students' knowledge but still has previously mentioned limitations.

The proposed solution offers new possibilities within Moodle quiz module and gets this part of the Moodle system to a higher level. The idea is to offer an improvement of automatic evaluation of questions with short answers. The main goal is to create reliable solution for evaluating few word long answers written in English language. In the first stage the proposed solution will be attached to MoodleQuiz Android mobile application [1] and testing of the system will be done on students that use mobile application for attempting Moodle quiz.

In the next part of this paper Moodle system will be further described and special attention will be devoted to the module for attempting and evaluating quizzes. It will be explained how this module operates and what options are available. After that, special attention will be given to short answer questions as they are central part of this research. Later, computational lexicon WordNet will be discussed as it is the base for checking the semantic similarity of answers. Further, original MoodleQuiz Android application will be explained along with all updates that were developed for the sake of implementation of this research. Also, attention will be given to newly developed *MatchingService* service that is one of the core tasks of this project. Its task is to check similarity between two sentences, in this case, answer

that was written by a student and answer that is predefined to be correct.

II. Moodle and Moodle quiz

Moodle (Modular Object-Oriented Dynamic Learning Environment) is one of the most popular open source learning management systems [2]. It is used in 220 countries and has 64041 registered sites of which the majority is in America and Europe [3]. Moodle is a very scalable learning management system and presents a suitable solution for both big universities and organizations with thousands of students and individual tutors with small number of participants. Core of the Moodle system consists of courses with their resources and activities. Besides that, Moodle supports over twenty activities and modules for different purposes and scope.

Moodle quiz module is very rich module and contains number of different possibilities when creating both quizzes and questions. It has functionalities for creating and publishing different types of quizzes that can be used for automatic evaluation of student's knowledge and monitoring student's progress during the course. Moodle quiz can be used not only for establishing students' grade, but also for giving students possibility to test their knowledge while studying material and preparing for exam.

When creating the quiz, creator defines overall setup for quiz like grading method and question behaviors like possibility for multiple attempts. At this point creator chooses individual questions he wants to be used or question groups from which the question can be selected randomly.

This module consists of a large variety of question types [4]. These questions are kept in the question bank when created and can be re-used in different quizzes. Some of the basic question types are: description, true/false, short answers, essay, matching questions, multiple-choice questions, numerical, calculated, embedded questions etc. For each question type there is number of options that can be set. Some of the options are common for all question types, like setting how many points correct answer is worth. Other options depend on the question type and its specifics.

Since Moodle quiz is proven to be very useful module for evaluating students' knowledge, a number of plugins what extend regular question types were created [5]. These plugins offer new question types that extend existing ones or present totally new question types that can be very useful, depending on the area quiz and its' questions are from.

This module supports automatic evaluation of quiz attempts which does not need to be used if teacher wants to do that job manually. Each question type has its' own rules for evaluating question and calculating points. How one question is going to be evaluated is set when question is created. Within one question type creator can choose different settings for different questions based on his wishes.

Results of every quiz attempt course administrator can review in the course administration panel on Moodle system. At that point teaching staff can reevaluate the question by changing points one has won on the question and automatically change total number of point earned on the attempt.

Short answer questions

Short answer questions are questions that are answered by entering free text that can contain different types of characters entered in form of one or more words, phrases or sentence. Depending on the setting, when answering the questions one needs to pay attention on small and capital letters. This option may or may not be set depending on the creators' wish. Creator can set one or more than one correct answer. Further, he can enter more answers and assign how worth in percentage each answer is from a maximal number of points for that question. Fig. 1 presents an example of question with short answer in Moodle system.

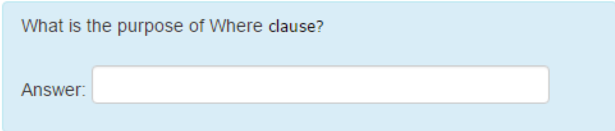


Figure 1. Example of question with short answer in Moodle system

Moodle quiz module offers automatic grading solution for quiz attempts. This works for all question types except essays which have to be evaluated manually and short answer questions in some situations. Automatic evaluation of questions with short answers checks whether answer written by student is identical to one defined by questions' creator. If it's the same, students gets the points, otherwise he doesn't. In situations where there are more than one answers defined for one question and they are worth different percentages of maximum number of points, student receives percentage that corresponds to the answer that is identical to his.

This system of evaluation leaves possibility for answers that are formulated in a different way not to be properly assessed because they are not identical with defined ones. Such situation can be very common since usually same thing can be said in a several ways and one sentence can be phrased differently. In those situations manual reevaluation of such questions is needed and this process can take a lot of time and energy if many students have attempted that quiz.

To avoid manual reevaluation of quiz attempts, question answers have to be specific so that they can be expressed in only one way. Other solution is answers to be very short so that possibility for evaluation errors is minimized. In order to overcome these limitations within this paper a solution for evaluating accuracy of the answers by semantic similarity is proposed. All details of the proposed solution will be explained more closely later in this paper.

III. WordNet

For the purpose of this research it is necessary to include Natural Language Processing (NLP) in order to assure successful sentence comparison. Word Sense Disambiguation (WSD) is considered to be one of the core tasks in Natural Language Processing [6]. Its purpose is to assign for each word in the sentence appropriate sense(s) and for that purpose supervised and unsupervised methods can be used.

A majority of WSD methods use external knowledge sources as central component for performing WSD. There are different knowledge sources such as ontologies, glossaries, corpora of texts, computational lexicons,

thesauri etc. WordNet is one of the external knowledge sources that has been widely used for performing WSD [7]. It is computational lexicon for English language created in 1985 under the direction of Professor George Armitage Miller in the Cognitive Science Laboratory of Princeton University. Over time, many people gave their contribution to WordNet development and today the newest version (3.1) is available on the Internet and consists of 155287 words organized in 117659 *synsets* [8].

WordNet database consists of nouns, verbs, adjective and adverbs grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Each synset represents a structure, which consists of a term, its class, connections with all semantically related terms and brief illustration of the use of the synset members. Most frequently used semantic relations are: hypernymy (kind-of or is–a), hyponymy (the inverse relations of hypernymy), meronymy (part-of) and holonymy (the inverse of meronymy). In figure 2, one example of the hyponym taxonomy in WordNet is presented.

WordNet can be efficiently used in a number of unsupervised methods which introduce semantic similarity measures for performing word disambiguation. In such cases WordNet is used to determine the similarity between words. Rada et al. [9], Leacock and Chodorow [10], Wu and Palmer [11], have successfully used WordNet as a base for creating graph-like structure in order to perform word similarity measurement. Within this paper Wu and Palmer method will be used for measuring similarity between words.

IV. Improvement of Automatic Assessment of Questions with Short Answers

The main purpose of this paper is to propose the improvement of automated assessment of Moodle questions with short answers. The goal is this solution to be used on questions which answers are a few word long sentences. Because of that, this solution won't be used to automate the evaluation of essays in Moodle system. The solution applies only for answers written in English language. This goal is implemented in a Moodle quiz application that was developed for solving Moodle quizzes on Android mobile devices.

In order to implement this improvement, it was necessary to realize two things:

- Create a web service that compares two sentences and returns their similarity and
- Improve MoodleQuiz application in order to support communication with service and develop new mechanism for assessing questions with short answers.

It the next part of this paper both components will be discussed in detail.

A. *Service for sentence similarity measurement*

For the purpose of comparing two sentences, in this case student's answer and correct answer defined by professor, *MatchingService* was created. The service is designed for English language only and relies on WordNet as a resource for English words. As an input, service requires two sentences for which similarity should be calculated. After processing service returns similarity percentage expressed with value from 0 to 1. Algorithm implemented within MatchingService actually consists of two independent algorithms, one for creating similarity matrix for two sentences and the other for calculating sentence similarity.

Algorithm for creating similarity matrix is preformed first. At the beginning it receives two sentences and preforms tokenization. Tokenization presents partitioning sentences into lists of words ($List_a$ and $List_b$) with removing all stop words (frequently occurring, insignificant words). Further, an array with Part of speech (POS) words is created. This array contains all types of words that can be used for comparison: noun, verb, adjective, adverb, unknown and other.

At this point everything is set for creation of similarity matrix which represents the similarity of each pair of words from the arrays ($List_a$ and $List_b$). This means that if the arrays have length m and n, created similarity matrix will have dimensions mxn. For each pair of terms in the arrays, similarity is checked in two ways and final result is better of the two.

Within the first way, the similarity among the characters in the words is checked. In this way the result is 1 (100% similarity) if the words are identical.

Within the second way, WordNet computational lexicon is used as a resource for determination of word type and appropriate sense in the sentence, as well as similarity with other words in the sentence. In order to achieve this, a type from a POS array is assigned to every word in a pair (T_a, T_b) and for such combination a

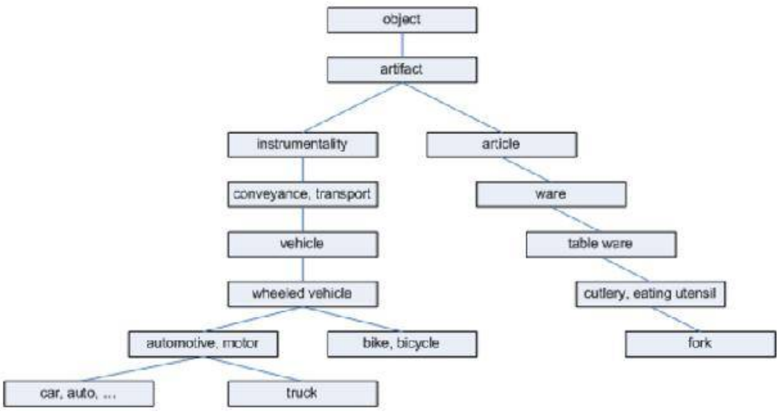


Figure 2. Example of the hyponym taxonomy in WordNet

semantic similarity is checked $sim(T_a, T_b)$. It is done by measuring the path length to the root node from the last common ancestor in the graph-like structure created with WordNet resource for information and relations among terms. After having path lengths calculated Wu and Palmer calculation is used for final determination of similarity for one POS combination for $sim(T_a, T_b)$. The calculation is done by scaling measured distance between the root node and the least common ancestor of the two concepts with the sum of the path lengths from the individual terms to the root.

Such calculation is performed for every combination of pair (T_a, T_b) with elements from POS array and final similarity is the highest result from all combinations.

After both algorithms for word similarity are performed, better result from both algorithms is inserted in similarity matrix. This procedure is repeated for every pair of words in $List_a$ and $List_b$.

After having similarity matrix created, algorithm for calculating similarity between entered sentences is performed. Within this prototype a heuristic method of calculation is used based on the similarity matrix. Following method is used:

$$score = (sumSim\_i + sumSim\_j)/(m+n) ,$$

where:

score – is result of final similarity of similarity matrix and can have value in range of [0,1];

m i *n* – are similarity matrix dimensions;

sumSim_i – sum of maximal elements in matrix per column:

$$sumSim\_i = \sum_{i=0}^{m-1} \max(i) ,$$

sumSim_j – sum of maximal elements in matrix per row:

$$sumSim\_j = \sum_{j=0}^{n-1} \max(j) ,$$

After final calculation of score is finished, the result is sent back to the client.

B. *MoodleQuiz application*

MoodleQuiz is an Android application developed for attempting Moodle quizzes on mobile devices [1]. Application is designed for devices with Android operating system version 2.2 or higher. Application is designed to support four basic types of Moodle questions that are commonly used at Faculty of Electrical Engineering in Niš: true/false questions, questions with short answers, matching questions and multichoice questions. Fig. 3 represents an example of question with short answer in MoodleQuiz application.

MoodleQuiz application is supported by Moodle Web service which is responsible for communication with Moodle system and access to all necessary data. Further, this service does all needed updates in Moodle database in order to assure consistency in Moodle system and assure that no difference is noticed between quizzes attempted in Moodle system and ones attempted on MoodleQuiz application. The whole system is designed to be compatible with Moodle version 2.5. Within this project an update is made, and now the system is compatible with Moodle version 2.8.2.

Within this project no changes were made in user interface of Android application, so users won't notice any changes in their user experience. Most changes were made in part of the application which is executed after the attempt is finished. In original application, quiz submission included assessment of all questions in the quiz and calculation of final score. This was done based on quiz and question setup, limitations set by quiz designer, question designer and official Moodle documentation. After that, final score along with other information is sent to the server and inserted into Moodle database, so that it can be available in administration panel in Moodle system.

Figure 3. Example of question with short answer in MoodleQuiz application

New version of the application made for this research has new actions added in order to support communication with MarchingService service. The service is called in POST method. When it is called, two sentences are forwarded in JSON format, correct answer and answer given by student in the attempt. Since the service returns percentage of the similarity between the send sentences, final score on the question is formed by multiplying the maximal number of points one can score on the question with similarity percentage received from the service.

This service is called during the preparation of the results for sending to Moodle server. When preparing the results, each time question with short answer appears, MatchingService is called and recalculation of point on the question and total score is done. Since each question with short answer can have more than one answer defined, service is called for each defined answer separately and for final result the highest similarity percentage between student's and correct answer is taken.

By testing this method for assessing question with short answers it was concluded that in cases where percentage was lower than 20%, student gave incorrect answer. For

that reason, in those cases correction of the points on that question is introduced and student receives 0 points. When percentage is higher than 20%, total score is calculated like described. After all questions have total marks, final score is calculated and update of the Moodle database is done. After that moment, course administrators can review the attempt normally in Moodle system.

Table 1 presents examples of answers defined by teachers and how some of the answers were evaluated by system and teachers. Given examples have results for two questions. Correct answers in rows with number 1a and 1b are correct answers for one question and row with number 2 for other question. Since number 1a and number 1b belong to the same question, one written answer is compared with both correct answers.

As it can be seen, results given by algorithm proposed in this paper are pretty accurate for question which answers is marked with number 2. For other answers the results were not that accurate. However, since both correct answers belong to the same question better percent will be taken for calculation of the points for that question. Having that in mind, final calculation is not too imprecise. Nevertheless, answer “*to filter database entries*” was badly evaluated in both combinations. This indicates that proposed semantic similarity evaluation can’t be taken for granted and in order to work question’s creator has to offer few combinations of word choices.

TABLE I.
ILLUSTRATION OF THE CORRECT AND WRITTEN ANSWERS AND RESULTS RETURNED FROM THE SYSTEM

No.	Correct answer	Written answer	Systems result [%]	Teacher evaluation [%]
1a	Where clause specifies conditions in the query	to filter database entries	43	100
		purpose is to specify conditions	67	90
		Where clause limits rows	86	100
		specifies conditions in the query	83	100
		Where clause specifies conditions in the query	100	100
1b	Where clause limits which rows will be returned	to filter database entries	51	100
		purpose is to specify conditions	54	90
		Where clause limits rows	95	100
		specifies conditions in the query	62	100
		Where clause specifies conditions in the query	83	100
2	Virtualization refers to the act of creating a virtual (rather than actual) version of something, including virtual computer hardware platforms, operating systems, storage devices, and computer network resources.	virtualization creates virtual version of something	69	70
		creates virtual version of something like computer resources	80	80
		creation of virtual computer platform, operating system, storage device or resources	89	90
		lorem ipsum lorem ipsum lorem ipsum lorem ipsum lorem ipsum	0	0
		virtualization refers to creating virtual version of hardware platform, operating system, storage devices, computer network resources	96	100

V. CONCLUSON AND FURTHER WORK

In this paper an improvement of automatic assessment of Moodle questions with short answer is presented. Offered solution provides much more flexibility while creating questions with short answers and testing students' knowledge. This primarily refers to the ability to define question with short answer whose answer can be a short sentence and have a confidence that it will be correctly automatically evaluated. Offered solution increases usability where this type of questions can be used and still assures that whole quiz can be totally automatically evaluated the moment it is submitted. Proposed solution reduces the necessity to use different types of question which offer answers in order to maintain 100% automatic evaluation of the whole quiz and save time.

Solution presented in this paper is currently in test phase and checking the reliability of comparing sentence similarity. Based on the results from table 1 it can be concluded that proposed algorithm made progress in comparison with current Moodle system's evaluation algorithm. However, at this point in order to expect better results question's creator should enter more correct answers with different word choices in the answers.

The plan is to test system's reliability on course Database at Faculty of Electrical Engineering, University of Niš. Based on the results of the mass testing, the algorithm will be further improved in order to provide better results. At this point, the goal is to assure correct evaluation of given answers that contain few words and minimize the possibility of error and the need to go manually through tests and evaluate each question separately. After that, the aim is to define the rules that should be followed when formulating questions and answers in order to get the best possible results.

If the system proves to be reliable in the next phase it will be transferred on Moodle system directly. With this step it will become available on Moodle system and dependence on MoodleQuiz application will no longer exist. In this way all Moodle system users will have the possibility of using this solution and it won't be limited on Android users only.

ACKNOWLEDGMENT

The research presented in this paper was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia as part of the project "Infrastructure for electronically supported learning in Serbia" number III47003.

REFERENCES

[1] M. Frtunić, L. Stoimenov, and D. Rančić, "Incremental Development of E-Learning Systems for Mobile Platforms," *ICEST 2014*, vol. 1, pp. 105–108, 2014

[2] J. Hant, "Top 100 Tools for Learning 2013," *7th Annual Learning Tools Survey*, September 2013

[3] "Moodle Statistics," 2015 [online] https://moodle.net/stats/

[4] "Moodle Questions," 2015. [online] https://docs.moodle.org/25/en/Question_types

[5] "Question Types, " 2015 [online] https://moodle.org/plugins/browse.php?list=category&id=29

[6] R. Navigli, "Word sense disambiguation: A survey," *ACM Computing Surveys (CSUR) Surveys*, vol. 41 issue 2, article no. 10, USA, 2009

[7] C. Fellbaum, "WordNet: An Electronic Lexical Database, Language, Speech, and Communication," *The MIT Press*, 1998.

[8] "WordNet statistics," 2015 [online] http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html

[9] R. Rada, H. Mili, E. Bicknell and M. Blettner, "Development and application of a metric on semantic nets," *IEEE Trans. Syst. Man Cybern*. 19 (1) (1989) 17–30.

[10] C. Leacock and M. Chodorow, "Combining local context and WordNet similarity for word sense identification," C. Fellbaum (Ed.), WordNet: An Electronic Lexical Database, *MIT Press*, pp. 305–332, 1998

[11] Z. Wu and M.S. Palmer, "Verb semantics and lexical selection," *Proceedings of the 32th Annual Meeting on Association for Computational Linguistics*, pp. 133–138, Las Cruces, New Mexico, 1994

The Business Process Transformation Framework Implementation through Metamodel Extension

Vladimir Maruna[*], Tom Mercer[**], Igor Zečević[***], Branko Perišić[***], Petar Bjeljac[***]
[*] MD&Profy, Beograd, Serbia
[**] Value Chain Group, Lawrenceburg, Ohio, USA
[***] Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
vladimir.maruna@mdprofy.com, tommercer@value-chain.org, {igor.zecevic, perisic, pbjeljac}@uns.ac.rs

***Abstract* — The interoperability levels at which complex modeling frameworks may cooperate directly constraint the ease of bidirectional transformation of model artifacts. A particularly challenging approach is to extend targeting modeling framework with another, previously external modeling framework, through metamodel transformations. In this paper, the context and rationales for selecting the metamodel extending approach while embedding The Business Process Transformation Framework methodology implementation into SAP PowerDesigner modeling framework, is described. The paper is focused solely on the first release of the solution, whose main mission was to: define the scope of the concepts belonging to the fundamental methodology dimensions; support the conversion of Value Chain Group Business Process Transformation Framework content from the previous ValueScape implementation to SAP PowerDesigner metamodel extension; and further improvements and extensions of data migrated from the ValueScape to SAP PowerDesigner modeling environment.**

I. Introduction

The Business Process Transformation Framework (BPTF) is a business processes management methodology whose main mission is to improve the alignment between: business strategy, business processes, people, and systems necessary to support that strategy [1]. BPTF is a consequence of research activities focused on switching from the traditional, document based business process design, to the contemporary model based/model driven paradigm [2].

In order to build and maintain a specification of business process design or transformation activities, the model based approach utilizes the predefined library of formally specified building blocks, that may be assembled into arbitrary complex structures representing the information flow among business processes, regardless their scale or abstraction level [3].

The Value Chain Group (VCG) [4] has specified and developed the BPTF methodology with the corresponding Information Model (BPTF IM) [5] which serves as a foundation for all of the methodology concepts definitions. Different implementations of BPTF methodology are provided through mapping of the BPTF IM to: logical database design (for example the entity relationship model) [6]; object-oriented models [7, 8]; or metamodels that are accessible through the BPM/BPA modeling tools [9]. Depending on the implementation of metamodel it is possible to reuse the existing or create a completely new framework that enables: the scope definition, design, versioning and analysis of the digital forms of BPTF artifacts. Prior to the implementation of BPTF with SAP Power Designer integrated modeling environment (PD) [10], VCG has used ValueScape [11], an application that was specifically developed to support the utilization of BPTF IM.

The main focus in this paper is the VCGBPTFwPD integrated modeling environment, which implements and automates the selected use cases of the VCG BPTF methodology on top of SAP PowerDesigner integrated modeling tools. The concrete solution is implemented based on the standard extending mechanism of PD metamodel [12]. The PD Extension, as the central component of VCGBPTFwPD, implements BPTF IM using the ontology mapping approach.

The destination concepts, to which all of the BPTF IM source concepts are mapped, are specified in the form of an object oriented model (the metaclasses model). It represents a domain specific model that is translated into the necessary extensions of PD metamodel.

The main reasons for selecting SAP PowerDesigner among the variety of BPM supporting tools were:

- SAP PowerDesigner has already been in use by the targeted stakeholder groups;
- The existence of extending mechanisms within the PD methodology that enable resolving the potential structural conflicts that may arise when converting a domain model to the corresponding metamodel extensions;
- The existence of standard functions that support data import from external files into PD models, enabling easy and fast creation of the initial BPTF models;
- The existence of PD scripting languages that enable programmatic support for the implementation of selected BPTF methodology procedures.

As a consequence, several beneficial results emerged :

- The possibility of establishing, sharing, and upgrading underlining domain knowledge through modeling process and the utilization of a modeling framework;
- The efficient standardization of methods, activities, artifacts and other elements of VCG's BPTF;
- The possible integration of VCG BPTF methodology artifacts with other development frameworks used either at the enterprise level in context of arbitrary

Enterprise Architecture (EA) project [19], or within the frame of individual development projects.

The VCGBPTFwPD integrated modeling environment is planned to be incrementally implemented through several versions/releases. In this paper, the first incremental version is presented. Its main mission was to: define the scope of the concepts belonging to the fundamental dimension of BPTF IM building blocks; support the conversion of VCG BPTF content from ValueScape to PowerDesigner; and enable further improvements and extensions of data migrated from ValueScape to PowerDesigner integrated modeling environment.

II. THE BUSINESS PROCESS TRANSFORMATION FRAMEWORK FUNDAMENTALS

BPTF is expressed through three main dimensions, which are uniformly applicable to arbitrary enterprise (business) systems: Business Building Blocks, Value Chain Segmentation, and Continuous Improvement Programs [1]. Altogether they constitute the core transformation pool.

The Business Building Blocks dimension enables the declaration of basic executable/functional elements that constitute the BPTF. When combined together, they form (build): the Value Streams, the Process Flows, and the Activity Flows, together with the concepts that belong to the Value Chain Segments.

The Value Chain Segmentation dimension defines business processes in the form of patterns or prescriptions that belong to some particular industrial hierarchy, and are qualified for transformation enhancement in order to upgrade the enterprise performance. In BPTF, business processes are defined as linked collections of Value Streams, Process Flows, and Activity Flows.

The Continuous Improvement Programs dimension describes the business process improvement methods that attach the time dependent dynamical behavior to the Value Streams and Business Building Blocks [13, 14]. BPTF models, via the standard building block interconnections, express the value chain segments and their contents that constitute a particular enterprise state in the arbitrary instance of time.

A. The Business Process Transformation Framework Information Model (BPTF IM)

BPTF IM [5] is a VCG specification that describes all of the BPTF concepts, together with the assigned attributes (slots) and their associations. It allows the utilization of the existing BPM/BPA tools for capturing, designing, versioning, and analyzing of the BPTF artifacts in digital form. Different versions of BPTF implementation methodology can be achieved by mapping the BPTF IM into different metamodels.

BPTF IM includes 52 information concepts that are explicitly divided into three groups, which correspond to the previously described BPTF's dimensions. Each BPTF concept is defined in the tabular form through the application of ontology descriptors.

B. The Business Building Blocks

The Business Building Blocks dimension defines standard building blocks of BPTF methodology. Standard building blocks have a normalized definition - each block is defined exactly once in a uniform manner, regardless of the number of referenced value streams. According to that, building blocks can be organized hierarchically and mutually associated according to the current dictionaries ontology.

The BPTF methodology building blocks definitions and business processes vocabulary are developed through model driven approach, which is based on two reference models: the Value Reference Model (VRM) and the eXtensible Object Reference Model (XRM).

VRM is an industry independent, analytical work environment that establishes a classification scheme of business processes using the hierarchical levels and the process connections, which are established through their inputs/outputs [15]. Additionally, this model establishes the contextual links with best practices and metrics, which are referenced in the classification process, to determine the criticality level of company processes. VRM represents the frameworks' spot from which the design of industry-specific processes, i.e. XRMs, begins [15].

XRM represents the domain specific extensions of VRM models. XRM models are industry-specific dictionaries that have to be created by further decomposition of business processes defined in the VRM models. XRM establishes a domain-based view of a company and enables the analysis of industry processes that are specific for that industry branch. XRM also provides a structure for housing private/protected knowledge, while maintaining a standard vocabulary using VRM.

III. VCG BPTF WITH POWERDESIGNER

The VCGBPTFwPD integrated modeling environment is composed of five components: BPTF IM Extension, Libraries, Methodology, BPTF Administration, and The Framework Management.

The BPTF IM Extension Component implements VCG BPTF IM in the form of PD metamodel extensions. It allows the representation and management of all BPTF methodology dimensions and includes: necessary extensions of visual and descriptive notation; and procedures that automate certain BPTF methodology segments.

The Libraries Component encapsulates the set of Business Process Models that were created through the BPTF ValueScape tools content migration to the PowerDesigner integrated modeling environment. Migration is performed by expanding the PD metamodel with the developed BPTF IM Extension component. The model reflects the hierarchical structure of the existing VCG library.

The Methodology Component enables the application of the VCG Transformation Lifecycle Management (TLM) methodology [15] through PowerDesigner integrated modeling environment. Relaying on the BPTF IM Extension and Library components, it defines and automates the activities and procedures needed for business process transformation.

The BPTF Administration Component contains procedures that support the administration of user defined and directed development/upgrade of all BPTF element models. It includes the incremental definition and promotion of BPTF methodology changes.

The Management Component supports the VCGBPTFwPD version control mechanisms.

A. The BPTF IM Object Oriented Model (OOM)

BPTF IM Object Oriented Model (OOM) is an object representation of the domain specific model that defines syntax and semantics specific for BPTF concepts representation. All of the BPTF IM elements are mapped to a specified OOM. The ontology mapping process is performed according to the predefined set of rules presented in Table I.

The association rules, defined within the BPTF, are established in OOM through the Business Rules Objects [16]. For each rule there is a corresponding Business Rule Object that encapsulates textual description of the constraints applicable to the instances of attributes or associations.

TABLE I.
BPTF IM - PD OOM ONTOLOGY MAPPING

BPTF IM 1.0	Object Oriented Model
BPTF IM 1.0	Object Oriented Model
BPTF Concept	OOM Class
BPTF Concept Attribute	Corresponding Class as Class Attribute
BPTF Concept Description	Class Comment
BPTF Concept Attribute Description	Class Attribute Comment
BPTF Concept Attribute Format	Class Attribute Data Type
BPTF Concept Attribute Option	Class Attribute Standard Checks >> List of Values
BPTF Concept Association	Class Association

In Fig. 1, a high level package model of the OOM is presented. The high level packages follow the BPTF IM dimensions. The individual packages contain only those BPTF IM concepts that belong to the particular VCG BPTF dimension. Each package may reference concepts defined in other packages. The referenced concept appears in the form of a shortcut within the referring package.

Figure 1. The OOM high level package model

Following the prescribed mapping rules, concepts that belong to the Business Building Blocks dimension in BPTF IM release 1.0 are represented by the PD Class Diagram [16]. This class diagram is located in the Building Block Concepts Package of specified OOM. When creating the BPTF IM Extension, BPTF IM OOM was the initial reference to support the transformation of defined concepts to elements of PD metamodel.

B. BPTF IM Extension

1) The General Characteristics of BPTF

Every BPTF concept contains a unique attribute identifier (ID) and an attribute name that carries the semantics. For the purposes or implementing BPTF IM extensions, the root metaclasses of PD metamodel are used. Attribute ID is mapped to the meta attribute ObjectID of the IdentifiedObject metaclasses. The abstract metaclasses IdentifiedObject is inherited by all of the PD metamodel metaclasses. The attribute name is mapped to the meta attribute Name of NamedObject abstract metaclass. Meta attributes: Comment and Description from the NamedObject metaclass are used for further description of BPTF concepts.

2) BPTF Category and BPTF SubCategory

Three BPTF parra-categories and three subcategories belonging to the BPTF IM Extension are mapped to the metaclass ExtendedObject stereotypes with corresponding names: IO Category, IO SubCategory, Metric Category, Metric SubCategory, Practice Category, and Practice SubCategory (Fig. 2). Each of the BPTF subcategories is associated through the ExtendedAttribute to the appropriate BPTF category. This attribute is used in all subcategories to establish and maintain the "belongs to" relationship, which is directed from the subcategory to the corresponding category. The direction of this relation is changeable, meaning that the model user may choose the appropriate model category to which the subcategories belong.

The implemented category and subcategory concepts are not visually presented in diagrams (instances of these stereotypes do not have symbolic views). Their purpose is solely to classify other BPTF concepts, and they need to be present in that context only.

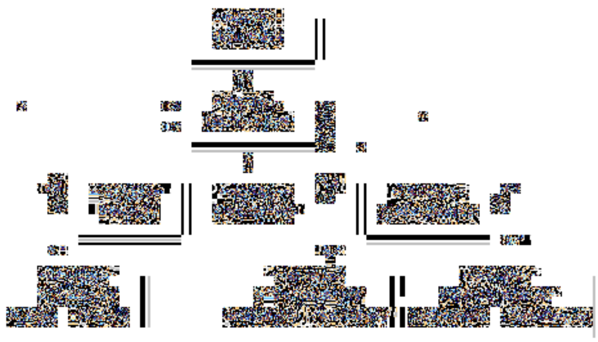


Figure 2. The BPTF Category/SubCategory concepts

3) BPTF IO, Metric, Practice and Capability

IO, Metric, Practice and Capability concepts are mapped to the corresponding stereotypes of Data metaclass, that are extended with necessary metaattributes in correspondence to the BPTF IM OOM. This metaclass enables the type of the information that has to be exchanged between business processes at the conceptual level of BPM [17]. It is focused on the information semantics rather than the technical aspects. Each instance of the above BPTF Data stereotypes (IO, Metric, and

Practice) may be associated with exactly one BPTF subcategory of the appropriate type (IO SubCategory, Metric SubCategory, and Practice SubCategory). ExtendedAttribute created in the Data stereotypes is used to establish and maintain the "belongs to" relationship between Data stereotype and the appropriate subcategory.

The additional extensions within BPTF IM Extension were necessary in order to implement the appropriate links between Practice and Capability concepts (Fig. 3). The association between these stereotypes is implemented using a standard PD concept, the ExtendedCollection. This collection, embedded in Practice stereotype, defines the standard set of functions (add, new, remove) that facilitate handling connections between one instance of Practice and a large number of Capability instances. The other side of this association is implemented through CalculatedCollection, enabling the selected Capability object to receive a list of all Practice objects previously associated through the ExtendedCollection.

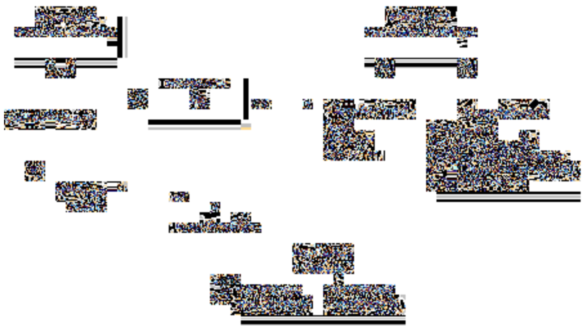


Figure 3. The Association of Practice and Capability Concepts

4) BPTF Priority Dimension

The PriorityDimension concept is mapped on to the ExtendedAttributeType metaclass of PD metamodel. Priority Dimension has a predefined set of values: Asset, Brand, Cost, Innovation, Reliability, Sustainability and Velocity. The reason for this mapping type is the relatively static nature of the PriorityDimension concept. Its values can only be changed at VCG level, while at the users level, the only possibility is to make the selection from a predefined list of values.

5) BPTF VRM /XRM processes

All of the VRM and XRM processes are implemented as corresponding stereotypes of Process metaclass: VRMLevelOne, VRMLevelTwo, VRMLevelThree, XRMLevelOne, XRMLevelTwo, and XRMLevelThree. Process metaclass is the specialization of the main activity of BPM and enables the creation of entities that deliver a set of services [17]. The standard process decomposition, supported by the PD BPM, completely supports the decomposition of VRM and XRM processes defined in BPTF IM. It is possible to decompose an arbitrary process through the hierarchical structure with corresponding dependences.

6) Mapping the BPTF VRM to XRM Processes

Within VCGBPTFwPD, an XRMLevelOne concept provides a standard solution for decomposition, synchronization, and association of VRM and XRM models. This concept is used to connect a generic BPTF environment (VRM) with its implementation, a specific, part (XRM). Processes on the first XRM level are created as synchronized copies of the VRM third level processes.

The associations between operational processes (level three of VRM) and a group of level one XRM processes in the BPTF Extension are established by adding the ExtendedAttribute to XRMLevelOne stereotype. The value of this attribute is automatically assigned in the initial phase of creating XRMLevelOne process instances by cloning the VRMLevelThree processes. XRMLevelOne object inherits all features of the original VRMLevelThree process, including comments, descriptions, and associations with IO, Metric, and Practice objects.

7) The BPTF Rules

In addition to the associated slots and associations concepts, BPTF IM 1.0 includes a set of rules that need to be met in order to obtain valid BPTF models. Within BPTF IM Extension, these rules are implemented by the Custom Check, a standard PD metamodel extension mechanism. Custom Check allows the definition of additional model content syntax and semantic validation rules [12]. Business logic, encapsulated in the Custom Check objects, is implemented with custom scripts written in VBScript language [18]. The editing and execution of these scripts are integral functions of PD. Each Custom Check is created as an extension of exactly one metaclass or a metamodel stereotype whose instances are validated.

IV. The Framework Libraries

The Library Component of VCGBPTFwPD is an ordered set of Business Process Models. These Models are the consequence of BPTF content migration from the ValueScape tool to the PowerDesigner modeling environment. The model architecture reflects the hierarchical structure of the existing VCG libraries.

The concepts, belonging to the Business Building Block dimension, are classified into two groups by the BPTF IM. The first group encapsulates categories and subcategories of: Input/Output, Metric, Practices, and Capabilities concepts.

The second group consists of concepts that are used to represent the value flow and /or value changes. This group is also implicitly divided into two categories of processes, which use the concepts defined in the first group. The first category covers the VRM processes, and other processes categories consist of XRM processes. In Fig. 4, a decomposition of the BPTF model is presented. It explicitly states that:

- Following the described internal organization of The Framework, it is obvious that There is exactly one BPTF Dictionary Model;
- There is exactly one BPTF Value Reference Model, which encapsulates generic and abstract experiences and the business process enhancement recommendations;
- There is exactly one BPTF eXtensible Reference Model per problem domain which contains domain specific recommendations, best practices and embedded knowledge. For the end user, XRM is the starting point from which the business process upgrade activities begin.

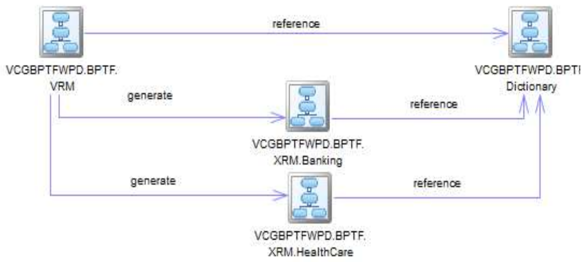


Figure 4. BPTF model decomposition

A. Load and Transform Procedures

Before the automation of ValueScape tool contents could start, it was necessary to implement the import procedures that are compliant with the standard PD data import mechanism (PD Import) [12]. In our case, PD Import is extended with two procedures:

- The „Load and Transform Association“ procedure, which supports building BPTF metamodel instances associations with 1: n cardinality for VRM/XRM processes, IO, Metric, Practice, Category and SubCategory objects; and n: m cardinality for Practice-Capability;
- The „Load and Transform Business Process" procedure structures the VRM and XRM processes in a tree hierarchy according to the descriptions of parent and corresponding collection of child processes.

B. The Dictionary Model

The Dictionary Model is implemented as a PD BPM with corresponding BPTF IM Extensions. The model contains all of the BPTF methodology element descriptions: Input/Output, Metric, Practice, Category and SubCategory classifiers, as well as the Capability concepts catalog. Objects from the Dictionary Model are created through previously described Load and Transform procedures, and are accessible outside the Dictionary Model. According to the principle of good localization embedded in BPTF concepts, the access to these objects is mitigated by the referencing mechanism represented by the inter-model and cross-model shortcuts. This allows an immediate automatic propagation of each modification to all of the referencing models.

C. The Value Reference Model (VRM)

The Value Reference Model is implemented as PD BPM with BPTF IM Extensions. The modeled processes are expressed in three hierarchical segments of VRM Industrial (Plan, Govern and Execute). The VRM32XRM1 Procedure

The VRM32XRM1 procedure generates first level XRM processes based on third level VRM processes, and supports the synchronization of corresponding VRMLevelThree-XRMLevelOne pares. It is implemented as a BPTF IM Extension procedure. In the preparation phase, The Framework user selects one of the possible XRMs that have to be synchronized with the corresponding VRM. In the execution phase, the procedure compares all of the three level VRMs with all of the first level processes of the selected XRM. If the corresponding processes differ, either in attribute values or in the associated objects, the XRMLevelOne process is automatically synchronized. Third level VRM processes without the corresponding copies of XRMLevelOne processes are automatically cloned at the first level of the XRM. The first level XRM processes that do not possess the originating process at the third level of VRM are removed.

D. The eXtensible Reference Model (XRM)

In VCGBPTFwPD, the initial eXtensible Reference Model, that contains the three level hierarchy of XRM processes decomposition, is implemented as PD BPM with BPTF IM Extension. It represents the starting point of domain specific technology needs of the particular user.

In the first phase, based on the already formed third level VRM process, the first level XRM's processes are created. For each VRMLevelThree, the VRM32XRM1 procedure has already created the copy of objects in the XRM, with the corresponding stereotype. In the second (content formation) phase, relying on standard PD Import from external files, the second and third level XRM processes are created. In the third phase, the relations between all levels of XRM processes are associated to the objects defined in the Dictionary Model. This is performed by the “Load and Transform Association” procedure. In the final phase of XRM model creation, with the support of the "Load and Transform Business Process", the complete XRM process tree is automatically generated.

E. The Check Model

In the model check phase, the validation of all defined constraints is performed at one place, thereby simplifying the use and reusing existing mechanisms to which the PD integrated modeling environment users are already familiar to. In addition to pre-defined rules (for example object name uniqueness), the list of rules may include user-defined rules that are established through the Custom Check object metamodel extensions.

All of the rules that are defined in the BPTF IM Extension through the Custom Check mechanism are appended to the existing ones, via embedded validity checks that exist within PD Business Process Model. The additional checks are available through the Check Model procedure for each extended BPM. The Check Model function is used for all of the syntax and semantic checks for all the automatically generated members of BPTF Libraries component.

V. CONCLUSIONS

In this paper, the VCGBPTFwPD, an integrated modeling environment that implements a selected subset of the Value Chain Group Business Process Transformation Framework (VCG BPTF) methodology use-cases via extending the SAP PowerDesigner (PD) integrated modeling tool metamodel, is presented.

VCG BPTF is a model based approach to design/transformation of business processes, which is specified by the corresponding Information Model (BPTF IM). It assumes that concepts and constraints that are described by BPTF IM may be mapped to the metamodel of existing BPM/BPA tools and extended to support the capturing, design, versioning and analysis of BPTF artifacts in digital form. The main component of the

described solution is the BPTF IM Extension, implemented through the standard SAP PowerDesigner (PD) extending mechanisms.

The extension is implemented by applying a systematic approach to developing a UML profile based on a domain model or metamodel. In the first phase, an object oriented model (OOM) describing the mapping of BPTF ontology is created. The second phase included a systematic translation of all of specified OOM elements to the appropriate extensions of PD metamodel. The created BPTF IM Extension is applied to the Business Process Model to enable automated artifacts generation from the former VCG ValueScape to the newly created framework in order to form the initial library of BPTF models. All models that were created in such a way were validated over a set of user-defined syntax and semantic checks.

The first version of The Framework is currently in VCG BPTF Training and Certification program, assumed to gradually replace the existing ValueScape application afterwards. The results that we expect to obtain during the testing and certification phases will be used for further evaluation and improvements of the developed framework. The advanced version of The Framework is planned to enrich the BPTF IM Extension by additional mechanisms that would capture the rest of the BPTF IM's dimensions.

Acknowledgment

The paper origins from the results of VCG BPTF with Power Designer project launched by the Value Chain Group in cooperation with the MD&Profy Company and the personal participation of employees at the Computing and Control Department, Faculty of Technical Sciences, University of Novi Sad.

References

[1] T. Mercer, D. Groves and V. Drecun, “BPTF Framework, Part I”, Business Process Trends, September 2010, available at: http://www.bptrends.com/publicationfiles/09-14-10-ART-BPTF%20Mercer%20et%20al-Final.pdf [accessed on 29 January 2016].

[2] T. Bellinson, “The ERP software promise, Business Process Trends”, July 2009, available at: http://www.bptrends.com/publicationfiles/07-09-ART-The%20ERP%20Software%20Promise%20-Bellinson.doc-final.pdf [accessed on 29 January 2016].

[3] T. Mercer, D. Groves and V. Drecun, “Part II - BPTF Architectural Overview”, Business Process Trends, Octobar 2010, available at: http://www.bptrends.com/bpt/wp-content/publicationfiles/THREE%2010-05-10-ART-BPTF%20Framework-Part%202-final1.pdf [accessed on 29 January 2016].

[4] Value Chain Group, http://www.value-chain.org [accessed on 29 January 2016].

[5] Value Chain Group, Bussines Process Transformation Framework (BPTF) Information Model - Building Blocks Segment, http://dx.doi.org/10.13140/RG.2.1.4252.4325.

[6] V. C. Storey, “Relational database design based on the Entity-Relationship model”, Data & knowledge engineering, 7(1), 47-83, (1991), http://dx.doi.org/10.1016/0169-023X(91)90033-T.

[7] J. Rumbaugh, M. Blaha, W. Premerlani., F. Eddy, W.E. Lorensen, “Object-oriented modeling and design”, (Vol. 199, No. 1). Englewood Cliffs: Prentice-hall (1991).

[8] P. Shoval, S. Shiran, “Entity-relationship and object-oriented data modeling - an experimental comparison of design quality”, Data & Knowledge Engineering, 21(3), 297-315 (1997), http://dx.doi.org/10.1016/S0169-023X(97)88935-5.

[9] M. J. Blechar, J. Sinur, “Magic quadrant for business process analysis tools”, Gartner RAS Core Research Note G, 148777 (2008), available at: http://expertise.com.br/eventos/cafe/Gartner_BPA.pdf [accessed on 29 January 2016].

[10] SAP PowerDesigner DataArchitect, available at: http://www.sap.com/pc/tech/database/software/powerdesigner-data-architect [accessed on 29 January 2016].

[11] R. Burris, R. Howard, “The Bussines Process Transformation Framework, Part II”, Business Process Trends (May 2010), available at: http://www.bptrends.com/publicationfiles/05-10-ART-BPTFramework-Part2-Burris&Howard.doc1.pdf [accessed on 29 January 2016].

[12] SAP, Customizing and extending power designer. PowerDesigner 16.5, available at: http://infocenter.sybase.com/help/topic/com.sybase.infocenter.dc38628.1650/doc/pdf/customizing_powerdesigner.pdf [accessed on 29 January 2016].

[13] T. Mercer, D. Groves, V. Drecun, “Part III – Practical BPTF Application”, Business Process Trends, November 2010, available at: http://www.bptrends.com/publicationfiles/FOUR%2011-02-10-ART-BPTF%20Framework--Part%203-Mercer%20et%20al%20--final1.pdf [accessed on 29 January 2016].

[14] G. W. Brown, “Value chains, value streams, value nets, and value delivery chains”, Business Process Trends, April 2009, available at: http://www.ww.bptrends.com/publicationfiles/FOUR%2004-009-ART-Value%20Chains-Brown.pdf [accessed on 29 January 2016].

[15] Value Chain Group, Process Transformation Framework, available at: http://www.omgwiki.org/SAM/lib/exe/fetch.php?id=sustainability_meetings&cache=cache&media=vcg_framework_green_09-09.pdf [accessed on 29 January 2016].

[16] SAP, Object-Oriented Modeling. PowerDesigner 16.5. http://infocenter.sybase.com/help/topic/com.sybase.infocenter.dc38086.1650/doc/pdf/object_oriented_modeling.pdf [accessed on 29 January 2016].

[17] SAP, Business Process Modeling. PowerDesigner 16.5. http://infocenter.sybase.com/help/topic/com.sybase.infocenter.dc38088.1653/doc/pdf/business_process_modeling.pdf [accessed on 29 January 2016].

[18] A. Kingsley-Hughes and D. Read, “VBScript programmer's reference”, John Wiley & Sons, 2004.

[19] H. Peyret, G. Leganza, K. Smillie and M. An, “The Forrester Wave™: Business Process Analysis”, EA Tools, And IT Planning, Q1 (2009), available at: http://www.bps.org.ua/pub/forrester09.pdf [accessed on 29 January 2016].

[20] American Productivity and Quality Center: "Process classification framework", available at: www.apqc.org/free/framework.htm [accessed on 29 January 2016].

USING CONTEXT INFORMATION AND CMMN TO MODEL KNOWLEDGE-INTENSIVE BUSINESS PROCESSES

Siniša Nešković*, Kathrin Kirchner**
*Faculty of Organizational Sciences, University of Belgrade, Jove Ilića 154, Belgrade, Serbia
**Berlin School of Economics and Law, Alt-Friedrichsfelde 60, Berlin, Germany
sinisa.neskovic@fon.bg.ac.rs, kathrin.kirchner@hwr-berlin.de

Abstract—**Knowledge-intensive business processes are characterized by flexibility and dynamism. Traditional business process modeling languages like UML Activity Diagrams and BPMN are notoriously inadequate to model such type of processes due to their rigidness. In 2014, the OMG standard CMMN was introduced to support flexible processes. This paper discusses the main benefits of CMMN over BPMN. Furthermore, we investigate how context information of process instances can be used in CMMN to allow runtime flexibility during execution. The proposed technique is illustrated by an example from the healthcare domain.**

I. INTRODUCTION

Business Process Management (BPM) often deals with well-structured routine processes that can be automated. But in last years, the number of employees that intensively use non-routine analytic and interactive tasks increased [1]. Knowledge workers like managers, lawyers and doctors need experience with details of the situation to carry out their work processes. While software is provided to support routine tasks, this is less the case for knowledge work [2]. Knowledge-intensive processes have a need of human expertise in order to be completed. They integrate data in the execution of processes and require substantial amount of flexibility at run-time [3]. Thus, activities might require a different approach during each process execution. Case-based representations of knowledge-intensive processes provide a higher flexibility for knowledge workers. The central concept for case handling is the case and not the activities and their sequence. A case can be seen as a product which is manufactured. Examples of cases are the evaluation of a job application, the examination of a patient or the ruling for an insurance claim. The state and structure of any case is based on a collection of data objects [4].

Standards for business process modeling, e.g., UML Acitivity Diagrams [5] or BPMN 2.0 [6], usually abstract from flexibility issues [7]. BPMN is used for modeling well-structured processes. However, in BPMN 2.0, ad-hoc processes can be used to model unstructured models. Elements in an ad-hoc sub-process can be executed in any order, executed several times or are even omitted. No rules are defined for the task execution in the sub-process, so the person who executes the process is in charge to make the decision.

Case Management Model and Notation (CMMN) is an OMG modeling standard for case modeling introduced in 2014 [8]. Although a new version 1.1 in December 2015 increased the clarity of the specification and corrected features to improve the implementability and adoption of CMMN, the language still needs to be adapted by process modelers and prove that it is capable to support different knowledge-intensive processes with lots of flexibility.

One important aspect of flexible business processes is the context in which a process is executed. Many practical situations exist where the workflow of activities depends heavily on surroundings of the particular process instance. For example, in healthcare domain patients are treated depending on their particular health conditions, and routine fixed treatment cannot be prescribed in advance.

The context plays an important role in several application areas such as natural languages, artificial intelligence, knowledge management, and web systems engineering, ubiquitous computing and Internet of things computing paradigm. In the domain of business process modelling, context awareness is a relatively new field of research. Context aware self-adaptable applications change their behavior according to changes in their surroundings [9]. CMMN has no direct support to model context aware processes.

However, certain mechanisms exist which can be used to model a context based flexible process. In this paper we propose a technique how this can be achieved.

Our paper is structured as follows: Section II introduces CMMN as a means to support flexible processes. In Section III, we give a motivational example from healthcare. Section IV provides our idea to integrate contextual information into CMMN models. We discuss related work in section V and summarize our findings in section VI.

II. CMMN AS A MEANS TO SUPPORT FLEXIBLE PROCESSES

Case management requires models that can express the flexibility of a knowledge worker. This can be covered by CMMN [8]. CMMN provides less symbols than BPMN, and might be therefore easier to learn. Since CMMN is a relatively new standard, a brief introduction is given. CMNN is a graphical language and its basic notational symbols are shown in Figure 1.

A Case in CMMN represents a flexible business process, which has two main distinct phases: the design-

time phase and the run-time phase. In the design-time phase, business analysts define a so called Case model consisting of two types of tasks: tasks which are always part of pre-defined segments in the Case model (represented by rounded rectangular), and “discretionary” (i.e. optional, marked as rectangular with a dotted line) tasks which are available to the Caseworker, to be performed in addition, to his/her discretion. In the run-time phase, Caseworkers execute the plan by performing tasks as planned, while the plan may dynamically change, as discretionary tasks can be included by the Caseworker to the plan of the Case instance in run-time.

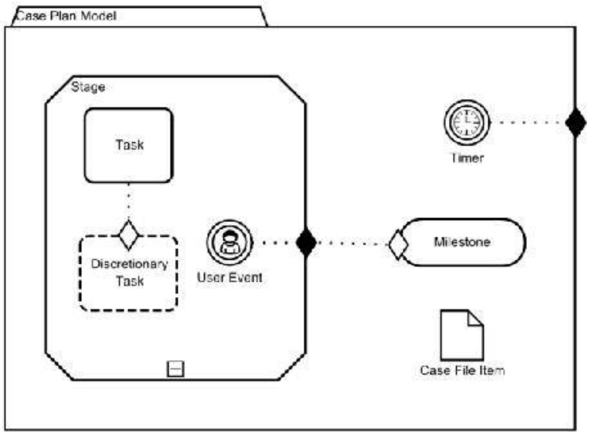


Figure 1. Basic CMMN 1.0 notational symbols

A Stage (rectangles with beveled edges) groups tasks and can be considered as episodes of a case with same pre- or postconditions. Stages can be also planned in parallel.

Sentries (diamond shapes) define criteria as pre- or postconditions to enter or exit tasks or stages. Entry conditions are represented by the white diamond, whereas exit conditions are designated by black diamond. Conditions are defined by combinations of events and/or Boolean expressions.

Milestones (rounded rectangles) describe states during process execution. Thus, a milestone describes an achievable target, defined to enable evaluation of the progress of a case.

Other important symbols are events (circles) that can happen during a course of a case. Events can trigger the activation and termination of stages or the achievements of milestones. Every model is described as case plan model, that implicitly also describes all necessary data. In order to make the data more explicit, a so called `CaseFileItem`, represented as document symbol, can be used.

Additional conditions for the execution can be described in a planning table. Connections are used to express dependencies but no sequential flows. Connections are optional visual elements and do not have execution semantics.

III. Motivational Example

In this paper, we use an example for a CMMN model in a health care scenario [10]. During a patient treatment, or especially in an emergency scenario, medical doctors and nurses need to be free to react and make decisions based on the health state of the patient. Deviations in a treatment process are therefore frequent. Figure 2 shows a CMMN model for the evaluation of a living liver donor.

A person that is considering donating a part of her liver is first medically evaluated to ensure that such a surgery can be carried out. Each evaluation case needs to be started with performing the required task of *Draw blood for lab tests and perform physical examination*. Afterwards, *Perform psychological evaluation* must be performed before the milestone *Initial examination performed* is reached. Thus, the execution of the two stages *Med/Tech investigations* and *Mandatory referrals* is enabled.

Examinations that are performed only sometimes according to medical requirements (depending on the decision of the medical staff) are modeled as discretionary tasks, e.g., *Perform lung function test*. The tasks within a stage can be executed in any sequence.

The stages *Med/Tech Investigations* and *Mandatory referrals* are prerequisites for the milestone *Results available.* If the milestone is reached, the task *Analyze Results* can be executed. According to this analysis, further investigations might be conducted in the stage *Further Investigations*. In this stage, all tasks are discretionary and can be planned according to their need during case execution for a specific patient. The `CaseFileItem`s *Patient Record* and *Patient Analysis Result* contain important information for the decision about executing tasks. A potential donor can be considered as non-suitable at every stage of the evaluation, as shown in the model by the described event.

IV. Embedding Contextual Information into CMMN Models

Business processes are almost never executed isolated, but instead have an interaction with other processes or the external environment outside of their business system. In other words, processes usually execute within a context. Hence, modeling flexible processes requires modeling of the context as well as expressing how contextual information influence the process execution.

In the literature there exist many different approaches for context modeling [11] and how context-aware and self-adaptable applications can be developed [9]. However, these approaches are not possible directly to employ since CMMN is strictly defined standardized modelling language.

However, in CMMN exist several mechanisms for expressing flexibility, which can be used for modeling context aware processes. Here we propose an approach which is based on concepts of `CaseFileItems` and `ApplicabilityRules` of `PlanningTables`.

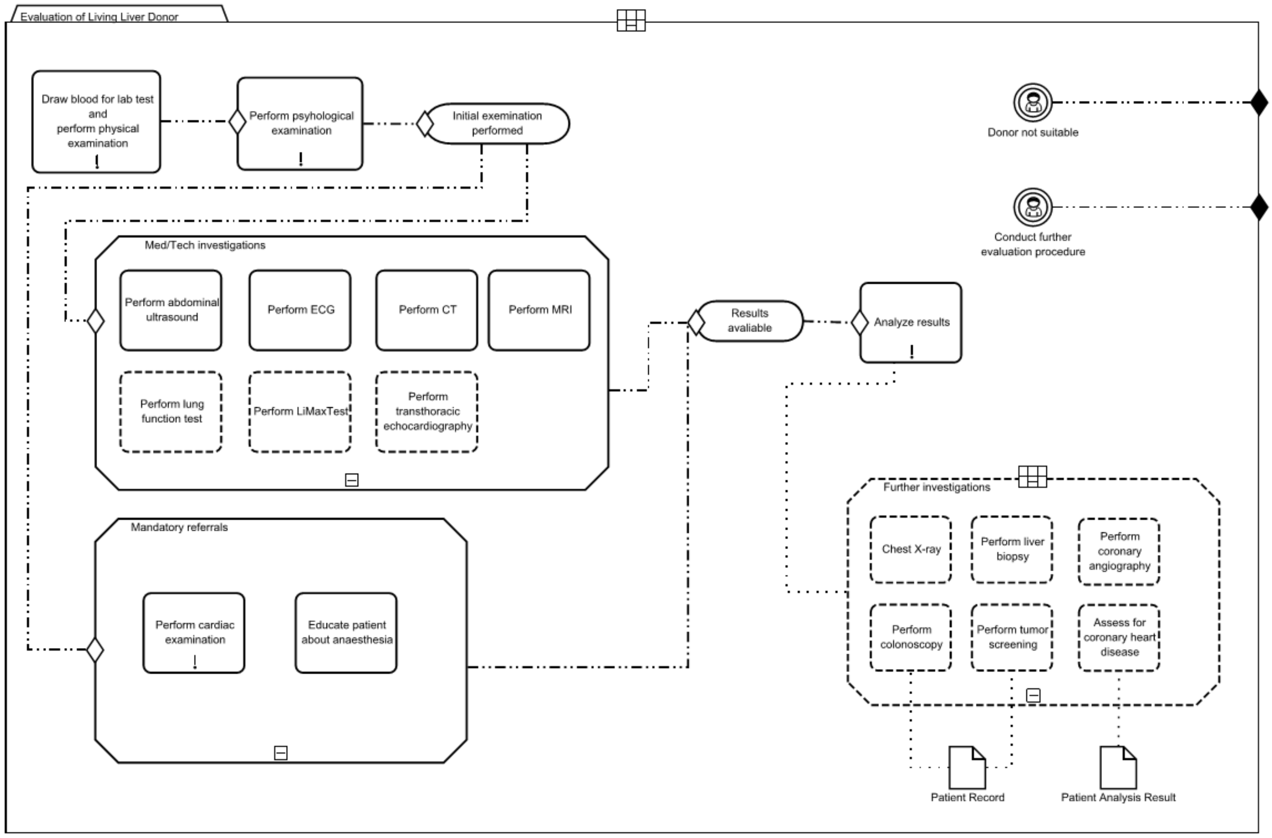


Figure 2. CMMN example for a living liver donor evaluation (adapted from [10])

`CaseFileItem` is essentially a container of information (i.e. document), which is mainly used in CMMN to represent inputs and outputs of tasks. But, it can be also used to represent the context of a whole process or particular process stage. According to the CMMN standard, information in the `CaseFileItems` are also used for evaluating `Expressions` that occur throughout the process model. This allows that contextual information stored in a `CaseFileItem` can be directly used in declarative logic for changing behavior of the process.

CMMN does not prescribe the modelling language for data structures and expressions. Here we propose to use UML class diagrams [12] for modelling `CaseFileItem` structures, and Object Constraint Language (OCL) as a language for modelling expressions [13], [14].

In Figure 3 a simplified example of patient records is shown which can be used as a context for the stage *Further investigations* of the living liver donor evaluation process from Figure 2.

In addition to the basic personal data about a patient (class *Patient Record*), the context also includes analysis results (class *Patient Analysis Result*), each consisting of concrete values for various analysis parameters (class *Analysis Parameter*).

How contextual information influence the process execution is expressed using `Planning tables` and `Applicability rules`. A `Planning Table` is used to define a scope of the context, i.e. which parts (i.e. process stages and tasks) of the case model are dependent on the context. Stages and tasks which have a planning table are decorated in the CMMN graphical notation with a special table symbol (⊞).

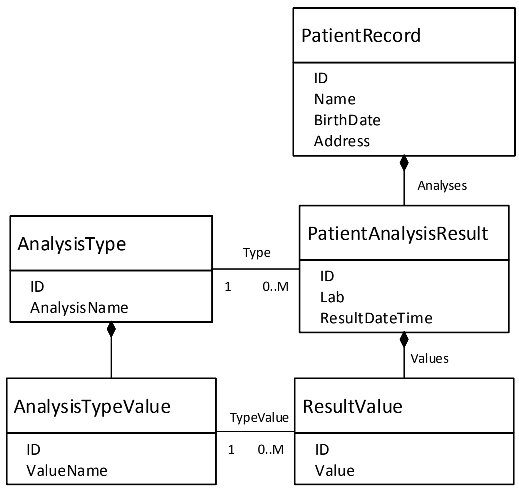


Figure 3. Patient Record - Context model for patients

Each table consists of so called Discretionary Items, which represent tasks or other (sub)stages which are context dependent. Each Discretionary Item can be associated with an Applicability Rule. Rules are used to specify whether a specific Discretionary Item is available (i.e. allowed to perform) in the current moment, based on conditions that are evaluated over contextual information in the corresponding Case File Item.

As an simplified illustration, in Table 1 is given a part of the planning table for the stage Further investigations, which consist of three Discretionary Items (other rules are omitted due to space limitation).

Discretionary Item	**Applicability Rule**
`Perform colonoscopy`	`Over50`
`Perform tumor screening`	`Over50`
`Assess for coronary heart diseases`	`HighBloodPress`

Table 1. Planning Table for *Further Investigation* Stage

The first two `Discretionary Items` are tasks `Perform colonoscopy` and `Perform tumor screening`. These two tasks should be available to the caseworker for execution, when a particular patient case comes to the stage for further medical investigation, only if rule `Over50` is evaluated to true. In other words, these two types of further investigations of a patient should be done only when he/she is older than 50. The rule is specified over context using OCL syntax as follows:

```
Rule Over50:
context: PatientRecord
inv: self.Age >= 50
```

The rule is based on a very simple OCL expression which examine the value of age attribute of patient record.

The third `Discretionary Item` is the task `Assess for coronary heart diseases` which should be available only if rule `HighBloodPressure` is evaluated to true. This rule is based on somewhat more complex OCL expression, which is specified as follows:

```
Rule HighBloodPressure:
context: PatientAnalysisResult
inv: self.Values()->exist(v|
     v.Parameter.Name = ”systolic” and
     v.Value> 140)
```

This rule expression evaluates to true if there exist a patient analysis containing parameter for systolic blood pressure with value bigger than 140. In other words, `Assess for coronary heart diseases` of a patient should be done only when he/she have a higher blood pressure.

V. RELATED WORK

Process models and process-aware information systems need to be configurable, be able to deal with exceptions, allow for changing the execution of cases during runtime, but also support the evolution of processes over time [15].

In the literature, several approaches exist already to support the modeling and execution of flexible processes. Declarative Process Modeling is activity-centered [16]. Constraints define allowed behavior. During runtime, only allowed activities are presented to the knowledge worker, and he decides about the next executed activity. Provop [17] allows the configuration of process variants by applying a set of defined change operations to a reference process model. Configurable Event Process Chains similarly allow the explicit specification of configurations [18].

Regarding execution of flexible processes, Proclets allow the division of a process into parts [19]. These parts can be later executed in a sequence or interactively. ADEPT allows to make changes during the execution time of a process [20].

The research into context-awareness is a well-known research area in computer science. Context-awareness is related to ability of a system to react to the changes in its environment. Many researchers have proposed definitions of context and explanations of different aspects of context-aware computing, according their specific needs [21], [9], [22]. One of the most accepted definition is given in [9]: “*Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.*”

There exist many different context modelling approaches, which are based on the data model and structures used for representing and exchanging contextual information. Strang and Linnhoff-Popien [11] categorized the most relevant context modelling approaches into six main categories:

- *Key-value*: context information are modeled as key-value pairs in different formats such as text files or binary files
- *Markup schemes*: Markup schemas such as XML are used
- *Graphical modeling*: Various data models such as UML class diagrams or ER data models are used
- *Object based modelling*: Objects are used to represent contextual data.
- *Logic based modelling*: Logical facts, expressions and rules are used to model contextual data.
- *Ontology based modelling*: Ontologies and semantic technologies are used.

Since CMMN is open regarding representation language used to model Case File Items, many of the above categories can be used. The approach used in this paper belongs to the graphical approach.

Context-awareness systems typically should support acquisition, representation, delivery, and reaction according to [23]. From this point of view, except for reaction, CMMN has no adequate concepts to support this typical functions. Developers must rely on some external mechanism in order to support these functions.

In addition, according to [19], there are three main abstract levels for modeling and building context-aware applications:

1. No application-level context model: All needed logic needed to perform functions such as context acquisition, pre-processing, storing, and reasoning must be explicitly modeled within the application development.
2. Implicit context model: Applications can be modeled and built using existing libraries, frameworks, and toolkits to perform contextual functions.
3. Explicit context model: Applications can be modeled and built using a separate context management infrastructure or middleware solutions. Therefore, context-aware functions are done outside the application boundaries. Context management and application are clearly separated and can be developed and extend independently.

From this point of view, CMNN supports the second level only partially. Namely, functions for reasoning and reaction are supported with the concepts of `Planning Table` and `Applicability Rules`, while all others must be explicitly modeled and developed each time.

VI. CONCLUSION

Context aware self-adaptable applications, also in the field of business process management, are becoming very popular in recent years [24]. Especially for knowledge-intensive processes, that are typically very flexible, the context plays an important role during the execution of a process instance. In this paper, we investigated how context information of process instances can be used in CMMN to allow runtime flexibility during process execution.

We have shown that using the specified planning table and applicability rules, it is possible to model a very often situation in the health care domain, where a specific workflow of patient treatment is heavily influenced by results of patient state and effects of already performed (i.e. applied to the patient) activities within the same process case. Similarly, this approach is applicable to any other business domain which requires process flexibility based on contextual information.

However, CMNN is lacking support for many typical functions of context-aware systems. Like other business process modeling languages, CMMN is intended to model only high level of business logic, whereas other details are left to be specified and developed at the level of specific applications. However, due to importance of context-awareness in today's business processes, an extension of CMMN with concepts for an explicit support for modelling context-aware functions as well as integration with external context management infrastructures would be needed for modeling context-aware knowledge-intensive processes.

REFERENCES

[1] D. Auer, S. Hinterholzer, J. Kubovy, and J. Küng, "Business process management for knowledge work: Considerations on current needs, basic concepts and models," in *Novel Methods and Technologies for Enterprise Information Systems. ERP Future 2013 Conference, Vienna, Austria, November 2013, Revised Papers*, 2014, pp. 79–95.

[2] H. R. Motahari-Nezhad and K. D. Swenson, "Adaptive Case Management: Overview and Research Challenges," in *2013 IEEE 15th Conference on Business Informatics (CBI)*, 2013, pp. 264–269.

[3] C. Di Ciccio, A. Marrella, and A. Russo, "Knowledge-Intensive Processes: Characteristics, Requirements and Analysis of Contemporary Approaches," *J. Data Semant.*, vol. 4, no. 1, pp. 29–57, Apr. 2014.

[4] W. M. P. Van Der Aalst, M. Weske, and D. Grünbauer, "Case handling: A new paradigm for business process support," *Data Knowl. Eng.*, vol. 53, no. 2, pp. 129–162, 2005.

[5] Object Management Group, *OMG Unified Modeling Language (OMG UML) Version 2.5*. 2015.

[6] Object Management Group, *Business Process Model and Notation 2.0*. 2011.

[7] H. Schonenberg, R. Mans, N. Russell, N. Mulyar, and W. Van Der Aalst, "Process flexibility: A survey of contemporary approaches," *Lect. Notes Bus. Inf. Process.*, vol. 10 LNBIP, pp. 16–30, 2008.

[8] Object Management Group, *Case Management Model and Notation 1.1 (Beta 2)*. 2015.

[9] A. K. Dey and G. D. Abowd, "Towards a better understanding of context and contextawareness," in *Workshop on the What, Who, Where, When and How of Context-Awareness, affiliated with the ACM Conference on Human Factors in Computer Systems*, 2000.

[10] N. Herzberg, K. Kirchner, and M. Weske, "Modeling and Monitoring Variability in Hospital Treatments: A Scenario Using CMMN," in *Business Process Management Workshops*, 2014, pp. 3–15.

[11] T. Strang and C. Linnhoff-Popien, "A Context Modeling Survey," in *First International Workshop on Advanced Context Modelling, Reasoning and Management*, 2004.

[12] Object Management Group, *UML 2.4.1 Superstructure Specification*. 2011.

[13] Object Management Group, *OCL 2.3.1 Specification*. 2010.

[14] J. Warmer and A. Kleppe, *The Object Constraint Language: Getting Your Models Ready for MDA*. Addison Wesley, 2003.

[15] M. Reichert and B. Weber, *Enabing Flexibility in Process-Aware Information Systems*. Springer Berlin Heidelberg, 2012.

[16] M. Pesic and W. M. P. van der Aalst, "A Declarative Approach for Flexible Business Processes Management," *Bus. Process Manag. Work.*, pp. 169 – 180, 2006.

[17] A. Hallerbach, T. Bauer, and M. Reichert, "Capturing variability in business process models: the Provop approach," *J. Softw. Maint. Evol. Res. Pract.*, vol. 22, no. 6–7, pp. 519–546, 2010.

[18] M. Rosemann and W. M. P. van der Aalst, "A configurable reference modelling language," *Inf. Syst.*, vol. 32, no. 1, pp. 1–23, 2007.

[19] W. M. van der Aalst, R. S. Mans, and N. C. Russell, "Workflow Support Using Proclets: Divide, Interact, and Conquer," *IEEE Data Eng. Bull.*, vol. 32, no. 3, pp. 16–22, 2009.

[20] M. Reichert and P. Dadam, "Enabling adaptive process-aware information systems with ADEPT2," in *Handbook of Research on Business Process Modeling*, 2009, pp. 173–203.

[21] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the Internet of Things: A survey," *IEEE Commun. Surv. Tutorials*, vol. 16, no. 1, pp. 414–454, 2014.

[22] O. Saidani and S. Nurcan, "Towards context aware business process modelling," in *8th Workshop on Business Process Modeling, Development, and Support (BPMDS'07), CAiSE*, 2007.

[23] J. Andersson, L. Baresi, N. Bencomo, R. de Lemos, A. Gorla, P. Inverardi, and T. Vogel, "Software engineering processes for selfadaptive systems," in *Software Engineering for Self-Adaptive Systems II*, 2013.

[24] U. Kannengiesser, A. Totter, and D. Bonaldi, "An interactional view of context in business processes," in *S-BPM ONE-Application Studies and Work in Progress*, 2014, pp. 42–54.

Extendable Multiplatform Approach to the Development of the Web Business Applications

Vladimir Balać*, Milan Vidaković**
* Faculty of Technical Sciences/Computing and Control Engineering, Novi Sad, Serbia
** Faculty of Technical Sciences/Computing and Control Engineering, Novi Sad, Serbia
sermasint@gmail.com, minja@uns.ac.rs

***Abstract*—We present the OdinModel framework - an extendable multiplatform approach to the development of the web business applications. The framework enables the full development potential of the application's model through the platform independent abstractions. Those abstractions allow the full code generation of the application from a single abstract model to the multiple target platforms. The model covers both the common and the platform specific development concepts of the different target platforms, which makes it unique. In addition, we can extend the model with any existing development concept of any target platform. The framework with such model provides both the generic and custom modeling of the complete Model, View and Controller parts of the application. OdinModel uses existing development tools for the implementation of the application from the model. It does not force any development technology over some other. Instead, the framework provides a hub from which the web application developers can choose their favorite approach. Currently, the framework covers the development for Java, Python and WebDSL platforms. Support of these three platforms and the extendibility of the framework guarantee the framework support for any development platform.**

I. Introduction

The development of the Model-View-Controller (MVC) web business applications with the general-purpose programming languages, like Java and Python, means that the realization of the domain problem solution is on the level of the programming language details. This means that, if we want to develop the same application on the Java and Python platforms, we outline the one same solution, yet write the code for it twice, for each platform.

Java and Python platforms have various development support tools that simplify the development as much as possible by doing the grunt work for the developers. These tools, however, produce the code that covers specific application's components, while the code outside those components the developers write manually. These tools also produce the code only for their target platform.

The development with the general-purpose programming languages generally has two main parts [1, 2]. In the first part, the developers create the models of the applications in textual and diagram forms. In the second part, the developers deal with the programming implementation of the models from the first part i.e. code writing. In practice, the developers give more importance to the code, than to the models [1, 2]. Consequently, when there are new changes to the application, the developers make the changes to the code, but not to the models, thus leaving the models inconsistent with the implementation of the application. This renders models practically useless for the further development cycle, and the invested work to make the models in the beginning a waste of time and effort [3].

To use the full development potential of the models, the developers may adopt one of Model-Driven Engineering (MDE) paradigms for the application development [2]. MDE offers higher levels of models abstraction, code writing automation, portability, interoperability and reusability than the programming languages [4]. MDE development principles propose use of the models as formal, complete and consistent abstraction of the applications [5]. From those models, the developers can generate the target application's code automatically. The abstraction improves the development process by allowing the developers to shift their focus from the programming languages to the models of the problem domain [2, 6]. The generation of the complete code of the application removes manual writing of the code during the implementation, hides complexity of the development and improves the quality of the application and its code [7, 8, 9].

We propose an extendable multiplatform MDE approach to the development that we call the OdinModel framework. What sets apart our framework from other solutions is that it encapsulates common features of three application's parts in a platform independent manner. We can develop the Model, the View and the Controller part of the application through the one platform independent model that we call Odin model. Our framework currently encapsulates common features of Java, Python and WebDSL [10] applications.

With the OdinModel framework, we write only the solution specific code, from which the accompanying Java, Python or WebDSL code the framework automatically generates. The OdinModel framework produces the complete code and eliminates the manual code writing. By using automation through the generators, we avoid direct work with any tool on any platform. However, the OdinModel framework recognizes that the use of the programming languages and their respective supporting development tools increases the developer's productivity by four hundred percent [11, 12]. In the light of that, the OdinModel framework combines MDE principles and use of proven

development tools with goal to improve the overall productivity, quality and amount of time needed for the development process.

With the OdinModel framework, we aim to make the application development more efficient with the right level of abstraction. We want to describe solutions naturally and to hide unnecessary details, as stated in [13]. Since there are many details in the application's code, we try to automate everything that is not critical, without loss of the expressiveness. We show that this concept can work for Java, Python and WebDSL platforms. Support of these three platforms and the extendibility of the framework, guarantees the framework support for any development platform.

II. RELATED WORK

Model-Driven Architecture (MDA) is MDE paradigm where the developers rely on the standards, primary Unified Modeling Language (UML) and Meta-Object Facility [14, 15]. At first glance, the approaches that adopt MDA are very similar to the OdinModel framework. Analyzing works such as MODEWIS [4, 11], UWA [3], UWE [16, 17], MIDAS [18], ASM with Webile [19], Netsilon [20] and Model2Roo [1], we recognize the same ideas as in the OdinModel framework. However, in MDA approach, the developers use UML to define the three distinct abstract platform independent application's models according to Meta-Object Facility principles [11, 15]. With the OdinModel framework, the developers use a custom modeling language to define one platform independent model. This is the key difference between MDA and OdinModel approach.

Another difference is that UML is not a domain-specific modeling language [21]. Since UML is not a domain-specific, the developers manually program the missing domain-specific semantics or use UML profiles, limited extensions of the language [5]. With the OdinModel framework, the abstract concepts are domain-specific.

With UML, the models and the underlying code are on the same level of abstraction [21]. The same information is in the model and the code i.e. visual and textual presentation. In contrast, OdinModel's modeling language has a higher level of abstraction and each symbol on the model is worth several lines of the code.

Domain-Specific Modeling (DSM) is MDE paradigm where the primary artifact in the application development process is the one abstract platform independent model and the full application's code generation from that model is obligatory [2]. In DSM approach, the focus is on the development in the one specific domain and the developers specify the domain problem solution using the domain concepts. In other words, the modeling language takes the role of the programming language. A modeling language, which directly represents the problems in the specific domain, is a Domain-Specific Language (DSL) [13, 15]. DSL is the integral part of DSM approach, along with the domain-specific code generator and the domain framework [2]. The OdinModel framework adopts DSM paradigm. We recognize the related works that adopt DSM paradigm in the next approaches: DOOMLite [22], WebML [23, 24], and WebDSL.

The main difference between our OdinModel framework and the related DSM approaches is that OdinModel provides the full code generation for the multiple platforms from the start. Odin model abstracts not just common features of the applications on a single platform, as the related approaches do, but common features of the applications with the different underlying platforms. Therefore, our model abstracts and covers both similarities and differences of the different platforms, which, to our knowledge, makes it unique. Other significant differences between OdinModel and the related DSM approaches we present in Table 1.

III. ODINMODEL SPECIFICATION

The key of OdinModel specification is Odin meta-model. It provides the specification of the abstractions of the features needed for the development of the Model, the View and the Controller parts of the applications. These abstractions are the result of the analysis of all the development concepts that the developers must define for each application's part separately. Odin meta-model is, essentially, union of these separate abstractions. Since we focus on multiplatform development, the abstractions cover both intersection and complement sets of the development features from different platforms. These two sets of features are a foundation for the specification of Odin DSL.

The OdinModel framework adopts the four-layered architecture of Meta-Object Facility standard. Essentially, this standard is a specification for definition of DSL [13]. Table 2 shows OdinModel's four-layered architecture. Odin DSL, in this stage, provides concepts that are abstractions of Java, Python and WebDSL features. There are two types of the features: common for all three platforms, and the platform specific. The Platform specific features are important because they allow customization and do not force the use of the generic solutions. However, we offer the generic solutions too.

The Model part of the application manages data access and persistence. With the OdinModel framework, we encourage the use of the tools, which automatically manage most of the database persistence. This means that the Model part of our meta-model only needs to cover the specification of entities, their attributes and their relations. Fig. 1 shows our definition of the *Entity* class. The meta-model class *EntityModel* is the root class and it contains the main domain classes i.e. all the other elements of the meta-model.

TABLE I.
Comparison of DSM approaches

Approach	Multiple target platforms	Custom user code	Custom user interface	Visual editor	Full MVC model
DOOMLite					
WebML		*	*	*	*
WebDSL		*	*		*
OdinModel	*	*	*	*	*

TABLE II.
OdinModel's four-layered architecture

Level	Layer	Implementation
M3	Meta-meta-model	Ecore meta-model
M2	Meta-model	Odin meta-model (Odin DSL)
M1	Model	Odin model
M0	Real world objects	Instance of Odin model

The meta-model classes *NumberField*, *StringField*, *EmailField*, *DateField* and *Fields* represent the attributes of the programming language classes i.e. entity fields. Fig. 2 shows the four types of the fields that OdinModel framework currently supports. Since the all field classes have some common attributes, we define those as the attributes of the super class *Fields*. We define the specific attributes of each field type in their own the meta-model class.

The Class *NumberField* defines the fields with the numerical values. This class can define the three types of the field: ordinary number, primary key and interval of numbers. When we define the primary key field, we can also define the type of the primary key generation through the attribute *generationType*. The same goes for the interval field where we can also define the type of interval through the attribute *intervalType*.

The Class *StringField* defines the fields with a string of characters as a value. It has five specific attributes, where two of them represent the constraints, and the other three define the combo box, a special type of the textual field.

The meta-model classes *OneToMany*, *ManyToOne*, *OneToOne*, *ManyToMany* and *Relations* specify the all four possible types of the relations between entities. The super class *Relations* specifies the common attributes of the relations.

The View part of the applications manages visual presentation. Through the code generation, the OdinModel framework provides the default Create-Read-Update-Delete (CRUD) user interface forms. The entities are the base for the generation of the CRUD forms. The CRUD forms contain the entity attributes as the input or the output form fields. The OdinModel framework provides navigation between these CRUD forms, through the default application's menu.

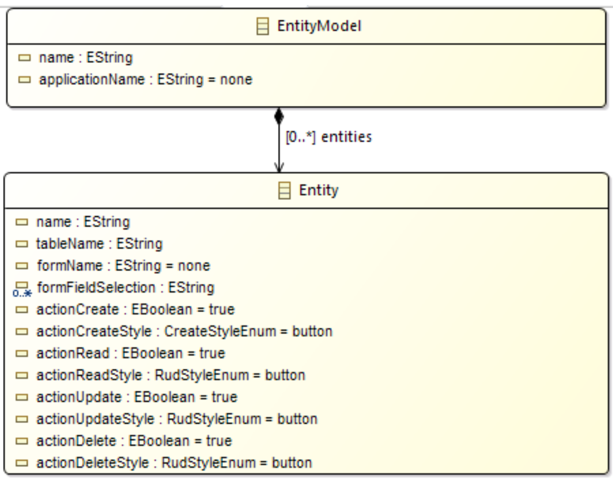


Figure 1. Entity class

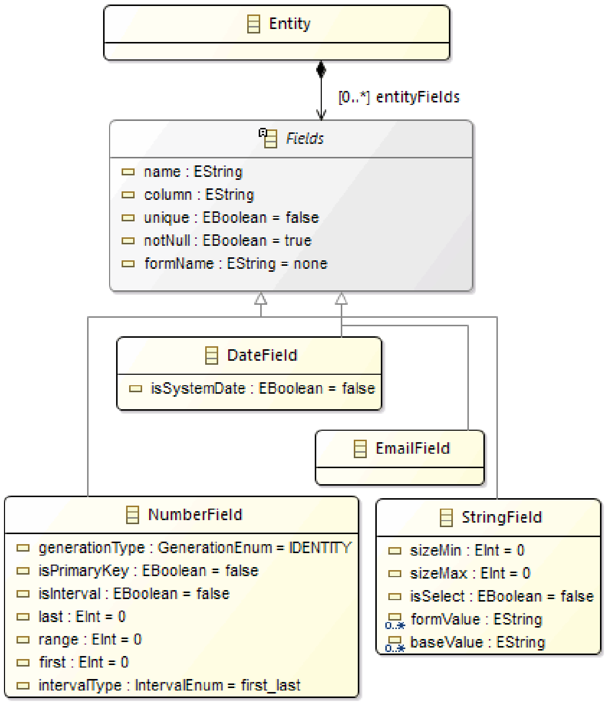


Figure 2. Field classes

The OdinModel framework also allows the customization of the content and the visual presentation of the CRUD forms and the application's menu. We can customize which CRUD operations will be visible on the forms and their visual style. The visual style covers the combinations of buttons, links, tables and fields. We define two meta-model classes with purpose to enable the menu customization, which we present in Fig. 3.

The Controller tier of the applications manages the page navigation, the input validation and the operations. The Odin meta-model specifies two sets of the operations. One set includes CRUD operations. The other set, which extends the CRUD set, includes the user's custom operations. We define the custom operations through the custom method classes. Through those classes, we define the control flow of the operation. We can declare the variables, assign the values to the variables, define IF conditions and define WHILE loops.

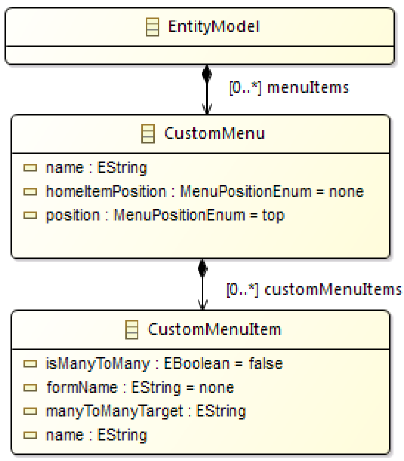


Figure 3. Custom menu classes

IV. ODINMODEL IMPLEMENTATION

The OdinModel framework provides the development environment, which contains Odin DSL, a visual editor for Odin DSL and the code generators. The developers through the visual editor use DSL to create the platform independent Odin model, which specifies the application. The code generators produce the complete application's code from the Odin model.

We now present the implementation of the OdinModel framework through the case study. In Fig. 4, we display the Odin model of a Sport center, which has five persistence objects and covers all four possible types of the relations between those objects.

The use of the OdinModel framework reduces the developer's work to the modeling of the domain concepts that exist in the Odin DSL. The Sport center model has all that is necessary for the specification of the Sport center application. Behind this model, there is an Extensible Markup Language (XML) code, not the programming language code. The code generators use that XML syntax to produce the application's code. In Fig. 5, Fig. 6, and Fig. 7, we present the generated code for the Member entity on the all target platforms.

OdinModel generators produce the code from the symbols, the arguments and the values of the symbols, and the relations between the symbols. If we make changes in the model, those generators apply changes to the all generated files. The generators are extendable. This means that whenever we define, for example, some new Java or Python domain concept in the meta-model, we adapt the corresponding generator. Java generator ignores Python and WebDSL specifics and vice-versa. In other words, if we specify the model with Java specifics, and then choose Python generator, the generator will generate Python application without problems. The generated application is ready-to-deploy. In Fig. 8, Fig. 9, and Fig. 10, we present the CRUD forms, which correspond to the generated codes.

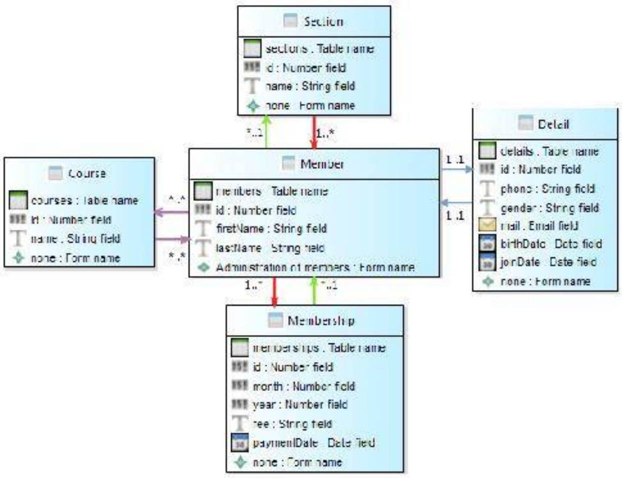


Figure 4. Sport center Odin model

```
… left out code …
@Entity
@Table(name = "members")
public class Member implements Serializable{

   @Id
   @GeneratedValue(strategy= GenerationType.IDENTITY)
   @Column(name = "id")
   private int id;

   @NotNull
   @Size(min = 3, max = 30)
   @Column(name = "first_name")
   String firstName;

   @NotNull
   @Size(min = 3, max = 30)
   @Column(name = "last_name")
   String lastName;

   @ManyToOne(cascade={CascadeType.REFRESH})
   public Section section;

   @ManyToMany(cascade={}, fetch=FetchType.EAGER)
   private Collection<Course>courses = new ArrayList<Course>();

   @OneToMany(cascade = {CascadeType.ALL},
   fetch=FetchType.EAGER, mappedBy = "member")
   @Fetch(value = FetchMode.SUBSELECT)
   public Collection<Membership> memberships;

   @OneToOne(cascade = {CascadeType.ALL})
   public Detail detail;
… left out code …
```

Figure 5. The generated Java class for Member entity

```
class Member(models.Model):
   id = models.AutoField(primary_key=True)

   first_name = models.CharField('First name',
   validators=[RegexValidator(regex='^.{3}$',
   message='Length has to be 3 ', code='nomatch')], max_length=30)

   last_name = models.CharField('Last name',
   validators=[RegexValidator(regex='^.{3}$',
   message='Length has to be 3 ', code='nomatch')], max_length=30)

   detail = models.OneToOneField('Detail')

   section = models.ForeignKey(Section)

   courses = models.ManyToManyField(Course)

   class Meta:

      db_table = "members"
… left out code …
```

Figure 6. The generated Python class for Member entity

```
… left out code …
  entity Member{
   firstName :: String(length = 3)
   lastName :: String( )
   name :: String := " " +firstName +" " +" " +lastName +" " +" "
   //1-1 relation
   detail <> Detail
   //m-m relation
   courses -> Set<Course> (inverse=Course.members)
   //m-1 relation
   section -> Section
   //1-m relation
   memberships -> Set<Membership> (inverse=Membership.member)
  }
… left out code …
```

Figure 7. The generated WebDSL class for Member entity

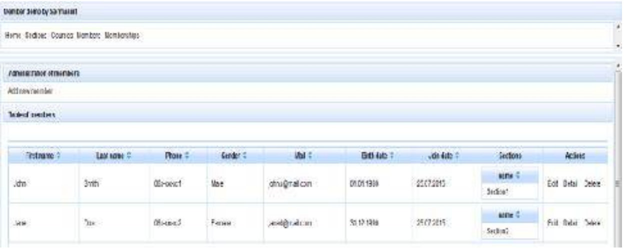


Figure 8. Java CRUD form Member

Figure 9. Python CRUD form Member

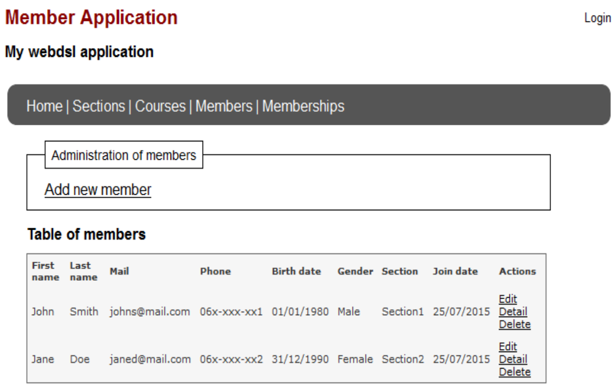


Figure 10. WebDSL CRUD form Member

V. Conclusion

The OdinModel framework improves productivity, portability, maintainability, reusability, automation and quality of MVC web business applications development process. The framework provides the visual modeling of the abstractions of the domain concepts in an original DSL. It provides the full multiplatform code generation from a single abstract model. We validate Odin model's level of abstraction by generating Java, Python, and WebDSL applications, directly from the model,. The incorporation of the proven development technologies shows openness and the extensibility of the approach. In addition to the default code generation, we provide the modeling of the custom user operations and the modeling of the custom user interface. The OdinModel framework does not force any development technology and approach over some other. Instead, it provides a hub from which the developers can choose their favorite development approach. Since the development tools incorporate best practices to produce code, we build on them.

The uniqueness of the OdinModel framework lies in its DSL, which covers both the similarities and the differences of the different target platforms. More precisely, it covers the common and the specific features of the target platforms relevant to the development. Odin DSL does not discard the specifics, but it does not force them either, which makes the Odin model platform independent, as well as composite. The DSL provides the developers with the abstract concepts and the platform specific details.

References

[1] Castrejón, Juan Carlos, Rosa López-Landa, and Rafael Lozano. "Model2Roo: A model driven approach for web application development based on the Eclipse Modeling Framework and Spring Roo." In Electrical Communications and Computers (CONIELECOMP), 2011 21st International Conference on, pp. 82-87. IEEE, 2011.

[2] Kelly, Steven, and Juha-Pekka Tolvanen. Domain-specific modeling: enabling full code generation. John Wiley & Sons, 2008.

[3] Distante, Damiano, Paola Pedone, Gustavo Rossi, and Gerardo Canfora. "Model-driven development of web applications with UWA, MVC and JavaServer faces." In *Web Engineering*, pp. 457-472. Springer Berlin Heidelberg, 2007.

[4] Fatolahi, Ali, and Stéphane S. Somé. "Assessing a Model-Driven Web-ApplicationEngineering Approach."*Journal of Software Engineering and Applications* 7, no. 05(2014): 360.

[5] Voelter, Markus, Sebastian Benz, Christian Dietrich, Birgit Engelmann, Mats Helander, Lennart CL Kats, Eelco Visser, and Guido Wachsmuth. *DSL engineering: Designing, implementing and using domain-specific languages*. dslbook. org, 2013.

[6] Calic, Tihomir, Sergiu Dascalu, and Dwight Egbert. "Tools for MDA software development: Evaluation criteria and set of desirable features." In Information Technology: New Generations, 2008. ITNG 2008. Fifth International Conference on, pp. 44-50. IEEE, 2008.

[7] Tolvanen, Juha-Pekka. "Domain-specific modeling for full code generation." *Methods & Tools* 13, no. 3 (2005): 14-23.

[8] Hemel, Zef, Lennart CL Kats, Danny M. Groenewegen, and Eelco Visser. "Code generation by model transformation: a case study in transformation modularity." Software & Systems Modeling 9, no. 3 (2010): 375-402.

[9] Rivero, José Matías, Julián Grigera, Gustavo Rossi, Esteban Robles Luna, Francisco Montero, and Martin Gaedke. "Mockup-Driven Development: Providing agile support for Model-Driven Web Engineering." Information and Software Technology 56, no. 6 (2014): 670-687.

[10] Groenewegen, Danny M., Zef Hemel, Lennart CL Kats, and Eelco Visser. "Webdsl: a domain-specific language for dynamic web applications." In Companion to the 23rd ACM SIGPLAN conference on Object-oriented programming systems languages and applications, pp. 779-780. ACM, 2008.

[11] Fatolahi, Ali. "An Abstract Meta-model for Model Driven Development of Web Applications Targeting Multiple Platforms."PhD diss., University of Ottawa, 2012.

[12] Iseger, Martijn. "Domain-specific modeling for generative software development." *IT Architect* (2005).

[13] Kosar, Tomaž, Nuno Oliveira, Marjan Mernik, Varanda João Maria Pereira, Matej Črepinšek, Cruz Daniela Da, and Rangel Pedro Henriques. "Comparing general-purpose and domain-specific languages: An empirical study." *Computer Science and Information Systems* 7, no. 2 (2010): 247-264.

[14] Selic, Bran. "The pragmatics of model-driven development." *IEEE software* 20, no. 5 (2003): 19-25.

[15] Cook, Steve. "Domain-specific modeling and model driven architecture." (2004).

[16] Kroiss, Christian, Nora Koch, and Alexander Knapp. Uwe4jsf: A model-driven generation approach for web applications. Springer Berlin Heidelberg, 2009.

[17] Kraus, Andreas, Alexander Knapp, and Nora Koch. "Model-Driven Generation of Web Applications in UWE." MDWE 261 (2007)

[18] Cuesta, Alejandro Gómez, Juan Carlos Granja, and Rory V. O'Connor. "A model driven architecture approach to web

development." In *Software and Data Technologies*, pp. 101-113. Springer Berlin Heidelberg, 2009.

[19] Corradini, F., D. Di Ruscio, and A. Pierantonio. "An ASM approach to Model Driven Development of Web applications.", 2004.

[20] Muller, Pierre-Alain, Philippe Studer, Frédéric Fondement, and Jean Bézivin. "Platform independent Web application modeling and development with Netsilon." Software & Systems Modeling 4, no. 4 (2005): 424-442.

[21] Perisic, Branko. "Model Driven Software Development–State of The Art and Perspectives." In *Invited Paper, 2014 INFOTEH International Conference, Jahorina*, pp. 19-23. 2014.

[22] Dejanovic, Igor, Gordana Milosavljevic, Branko Perišic, and Maja Tumbas. "A domain-specific language for defining static structure of database applications." Computer Science and Information Systems 7, no. 3 (2010): 409-440.

[23] Wimmer, Manuel, Nathalie Moreno, and Antonio Vallecillo. "Systematic evolution of WebML models by coupled transformations." In *Web Engineering*, pp. 185-199. Springer Berlin Heidelberg, 2012.

[24] Ceri, Stefano, Piero Fraternali, and Aldo Bongio. "Web Modeling Language (WebML): a modeling language for designing Web sites." Computer Networks 33, no. 1 (2000): 137-157.

A code generator for building front-end tier of REST-based rich client web applications

Nikola Luburić, Goran Savić, Gordana Milosavljević, Milan Segedinac, Jelena Slivka
University of Novi Sad, Faculty of Technical Sciences, Computing and Control Department
{nikola.luburic, savicg , grist, milansegedinac, slivkaje}@uns.ac.rs

***Abstract* – The paper presents a code generator for creating fully functioning front end web application, as well as a JSON-based DSL with which to write models which the generator uses as input. The DSL is used to describe both the data model and the UI layout and the generated application is written in AngularJS, following the current best practices. Our goal was to produce a code generator which is simple to create, use and update, so as to easily adapt to the climate of technologies which are prone to frequent updates. Our code generator and DSL are simple to learn, offer quick creation of modern, feature rich, web applications with customizable UI, written in the currently most popular technology for this domain. We evaluate our solution by generating two applications from different domains. We show that the generated applications require minor code changes in order to adapt to the desired functionality.**

INTRODUCTION

In recent years classical desktop applications have been replaced by internet-based applications where a server provides core functionalities that are accessed from client applications. In software engineering the terms “front end” and “back end” are distinctions which refer to separation of concerns between a presentation layer and a data access layer respectively. A recent trend in the implementation of internet-based applications is to separate the logic in two independent applications, a back end application which runs on a remote server and a front end application which runs on the browser.

Communication between client applications and the server is mostly done over HTTP, based on the REST software architecture [1]. REST-based services provide clients with a uniform access to system resources, where a *resource* represents data or functionalities, where each resource is identified by its uniform resource identifier (URI). A client interacts with such services through a set of operations with predefined semantics. REST-based services typically support CRUD operations which, in the context of internet-based applications, map to HTTP verbs.

In the recent years there has been an expansion of new technologies for developing front end applications and from year to year the growth of available frameworks and libraries is exponential [2-4]. Likewise, the vast majority of those technologies are volatile and tend to differ significantly from version to version. While few frameworks and libraries have proven to be more than hype, like AngularJS and Bootstrap, even those are prone to upgrades that break legacy software (which is rarely more than a year old). One thing to note is that while the backend technologies are constantly improving, the improvements made in this field revolve around performance or making the developer’s lives easier, while the features, visual appeal and ease of use of front end applications are what bring users in and influence profit [5]. This, in turn, means that there is more to be gained from updating the user interface, than the underlying server application.

Regardless of the technology being used, most information systems contain a standard set of features and functionalities. Features like CRUD (*create*, *read*, *update*, *delete*) forms or authentification are part of most applications, which is why it is possible to create tools which will generate these features automatically. One thing to keep in mind is that the tool would need to change as frequently as the underlying technology, and when talking about technologies which are prone to change and frequent updates, both the generator and its input need to be simple enough in order to be useful.

This paper presents a code generator for generating the front end tier of rich client web applications that rely on REST services as a back end technology. Our generator uses a simple DSL based on JSON as input, which is presented as well. The DSL describes a business application and the generator uses that description to generate an application. We use this code as a starting point of the implementation, giving a head start to the development process. The generated client application is written in AngularJS, using the current best practices [6]. In order to evaluate our solution we have performed two case studies. Using our code generator we have automatically generate implementations for two applications from different domains: a registry of cultural entities and a web shop for a local board game store. We show that the generated applications need minimal modifications in order to be customized according to the specific requirements. While the DSL is technology agnostic, the AngularJS framework [7] was chosen for the generator as it is currently the most popular framework for developing front end web applications. The reason behind its popularity lies in its ability to extend HTML through directives, while offering dependency injection in JavaScript which reduces the number of lines of code written. It also provides two way data binding between the view (HTML) and model (JavaScript) and offers many more features that increase the quality of web applications while minimizing the amount of written code.

The paper is organized as follows. Work related to this paper is presented in the next section. The section „Input DSL“ presents the DSL we use to create our input model for the generator. The section „Code Generator“ describes our code generator. The section after that, titled „Case

Study", presents two applications which were created using our generator as a starting point and shows the amount of code that was generated which required no modification, as well as the amount of code which required some modification or which had to be manually writen.

RELATED WORK

Before developing our solution we considered both the current research in the scientific community and the current industry standards (the popular open-source tools).

In [8] Dejanović presents a complex DSL which is used for the generation of a full stack web application using the Django framework. The DSL covers many different scenarios in order to automatically generate as much code as possible, including defining corner case constraints, custom operations, etc. The resulting DSL is a complex language that requires time and effort to learn. Code generators based on this language are complex and require a lot of time to be implemented if every part of the language is to be covered, which means that such a solution can't be used in a climate where every new project works with a different technology, or at least a significantly different version of the same technology. Paper [9] presents a mockup driven code generator, which is easier to use and requires less effort on the part of the developer, while also offering a significant amount of configuration as far as the user interface is concerned. However, the tool is also far too complex for the group of technologies being examined. The likely scenario is the long development of the code generator itself. With fast changing technologies this has a consequence in generation of already deprecated code. It should be noted that both [8] and [9] are aimed at generating enterprise business applications, which require mature and stable technologies. It should be noted that while both solutions take a MDE (model-driven engineering) approach, our generator focuses on creating a starting point for the implementation of an application.

With regard to code generators for front end web applications the Yeoman scaffolding tool [10] is a popular tool for this area. This tool provides developers with an improved tooling workflow, by automatically taking care of the many tasks that need to be done during project setup, like setting up initial dependencies, creating a suitable folder structure and generating the configuration for the project build tool. Most modern code generators use the Yeoman tool for the first step of building a front end application.

Some solutions that build on the Yeoman tool focus on expanding the scaffolding process, initially creating more folders and files based on some input. While no real business logic is generated, the files are formed with sensible defaults and best practices. The angular generator [11] by the Yeoman Team and Henkel's generator [12] are the most popular solutions from this group of generators. While the use of these tools is easy and usually requires only strings as input, the resulting code isn't runnable, as it's mostly boilerplate code.

A second group of solutions that build on the Yeoman tool try to produce fully functioning applications based on some input, and this is where our generator and DSL fit in. The most popular tool in this area by far is JHipster [13]. Using the command line or XML as input, JHipster creates a fully functioning application (both back end and front end) written using Spring Boot and AngularJS. While JHipster does offer a lot of useful features (built-in profiling, logging, support for automatic deployment, etc.) and the back end application is well implemented, the disadvantage of this tool is the lack of a fully developed front end application. The generated front end application lack important features (GUI elements for many-to-one relationships and lacks any customization of the layout during code generation, which means that every generated application looks exactly the same.

Aforementioned solutions are either too simple and generate only boilerplate code and/or only take the data model into account when building the user interface and/or are too complex for building cutting edge front end web applications. When comparing our solution with solutions produced by the scientific community, we found that the DSLs and code generators were far too complex for our problem domain of generating applications in technologies prone to frequent updates. Using our DSL users can describe not only the required model properties but also the layout which the generator uses to create a custom, well-designed GUI. In the next chapters we present our DSL and the code generator that uses it as input. Our goal here was to create a tool which solves the problems that the previously mentioned solutions have.

INPUT DSL

When constructing our DSL we aimed for simplicity. With that in mind we created a DSL which uses JSON as the underlying format primarily because our assumption is that a front end developer must know JSON and therefore doesn't need to spend time learning the syntax of the DSL.

Our DSL needs to meet the following requirements:

- It describes browser-based applications that receive/send data through the network from/to RESTful web services contained within the server-side application. It describes not only the entities that the application handles (data model), but also the user interface (layout, components, etc.)
- It is simple so that developers can learn it quickly and its associated code generators can be developed efficiently, as we are targeting an area of software engineering known for its many frameworks which change rapidly [2-4]
- There is no redundancy in the description, known as the DRY (*don't repeat yourself*) principle
- It is extensible, so that domain specific UI components can be built and used in the generating process.

Since an instance of our DSL is actually a JSON object, we can describe the constraints of our DSL using the JSON Schema [14].

Our code generator creates two components, a page for viewing multiple entities of a given type (list view, fig. 3), which supports paging, filtering and sorting of the list, and a form for creating a new entity, or viewing and possibly editing an existing one (detail view, fig. 4). The generator takes a JSON document for each entity that we want to generate components for. The list and detail view are generated for each such entity, as well as the entire underlying infrastructure needed for the aforementioned views to work and retrieve data from the server.

We take several factors into account when describing our views – will the table have standard pagination or will it use infinite scroll, will our complex forms be segmented by collapsible sub-forms, using a wizard-like UI with next and finish or have no segmentation, etc. Listing 1 shows the part of the schema that describes the entity object. Note that only the identifier (`id`) and the `groups` element is required, while the rest are either generated using the identifier (`label`, `plural`), or have predefined values (`pagination`, `groupLayout`).

```
"title": "Entity",
"type": "object",
"required": ["id", "groups"],
"properties": {
  "id": { "type": "string" },
  "label": { "type": "string" },
  "plural": { "type": "string" },
  "pagination": {
   "enum": ["default", "infiniteScroll"]
  },
  "groupLayout": {
   "enum": ["collapsible", "wizard",
            "none"]
  },
  "groups": {
   "$ref": "#/definitions/groups"
  }}
```

Listing 1. JSON schema for entity object

The `groups` attribute is an object which contains an `id`, optionally a `label` and an array of `attributes` and optional `subgroups`. By grouping attributes and subgroups we can separate complex forms into smaller, more manageable sub-forms.

An item of the attributes array contains information regarding the identity of the item (`id`, optional `label`), attributes which describe the type of UI component (`type`, `ref`, `object` and `extension`), and information regarding the list view (`table`).

Listing 2 shows the part of the schema related to defining the type of UI component for the given attribute of the entity. None of the listed attributes are required and if the `type` is missing it will default to *string*. If the `type` of attribute is one of the last three listed (*ref, object, extension*), another attribute is needed (which has the same name as the value of the type attribute) which will further describe the UI component. In case the `type` is *ref* the `ref` object describes a relationship with another entity, as well as how to form the UI component (which is defined in the `presentation` attribute). In case our REST back end returns an object that contains an inner object, we use the `object` attribute along with setting the `type` to the previously defined *object* we need. Finally, we use `extension` attribute to define custom components, by supplying a simple string which the generator will process in its own way. As far as our generator is concerned, strings listed in the `extension` attribute are angular directives.

The **`table`** object, not presented in the listings, is related to the functionality and UI of the list view. Three flags are placed in this object, `show` which signifies whether the attribute should be displayed in the table, `search` and `sort` which enable/disable the search and sort functionality of the table for the given attribute.

```
"type": { "enum": [
  "string", "textArea", "number", "email",
  "date", "ref", "object", "extension"
]},
"ref": {
  "type": "object",
  "required": ["entity", "relation"],
  "properties": {
    "entity": { "type": "string" },
    "relation": {
      "enum": ["oneToMany", "manyToOne",
               "oneToOne"]
    },
    "independant": { "type": "boolean" },
    "presentation": {
      "enum": ["inline", "link", "none"]
    }
  }
},
"object": {
  "type": "object",
  "required": ["id", "attributes"],
  "properties": {
    "id": { "type": "string" },
    "attributes": {
      "$ref": "#/definitions/attributes"
    }
  }
},
"extension": { "type": "string" }
```

Listing 2. JSON Schema for attribute type definition

CODE GENERATOR

Our code generator uses instances of our JSON schema presented above to create components for a fully functional front end web application.

The code generator uses the Freemarker template engine. The generator uses the input model writen using our DSL and template files to produce the resulting application. The template files are closely related to the chosen technology, and our templates are created for the latest version of the popular AngularJS framework, along with the Bootstrap CSS library in order to provide a rich, responsive modern front end web application.

Apart form the code generator, a framework was developed which acts as the infrastructure for the generated code. Other than a few input strings (e.g. application name, remote server location) the framework doesn't require any additional information. The generated applications are written using the current best practices for project structure and angular coding styles, contains inbuilt support for internationalization and use visually appealing and intuitive UI components. If a new entity needs to be added into the system, one would only have to supply the generator with the appropriate JSON and copy the resulting folder into the components folder. Fig. 1 shows the folder structure of a generated application with three entities. The content of the `components` folder is what the code generator produces, while everything else is part of the framework.

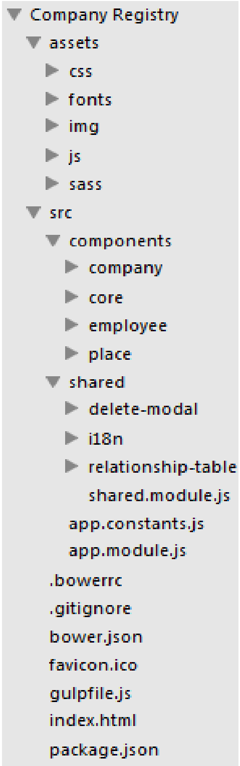


Figure 1. Generated application package structure

The `assets` folder contains resources which our application uses. This includes external JavaScript files (including angular and its plugins), styling sheets and fonts and images. The `src` folder contains the actual code of the application and it is separated into two subdirectories, the `components` folder and the `shared` folder. The `components` folder contains packages which represent various entities in our application and are, for the most part, independent modules which can be taken out of our application and placed into any other angular application where minimal work would be needed to adapt the module to work in the new context. The `core` package contains items that define the application layout, like the homepage, sidebar, header and footer HTML files and the underlying controllers. The `shared` folder contains directives, services and other components which are used throughout the application. This folder, along with the `core` package of the `components` folder, makes up the infrastructure of our generated application. The components located in `shared` are reusable pieces of code which can be used in other applications with virtually no adaptation required. This package includes support for internationalization, a directive for displaying one-to-many relationships and a modal dialogue which prevents accidental deletion. Finally the generator provides files for tracking bower and npm dependencies, which list dependencies of our application and development tools (gulp for managing the build process, karma and protractor for running unit and end-to-end tests) respectively, and a gulp file which contains a set of commonly used tasks.

The code generator creates specific UI elements for associations. The many-to-one relationship is represented using an autocomplete textbox, which offers a list of results that are filtered using the user input. The one-to-many relationship is displayed using an inline table.

Coming back to the generation process, the second phase of application generation takes all the JSON files written in accordance with our schema to produce modular, independent components, which include a list view page and detail view page described in the previous chapter, underlying angular controllers for both pages, a routing configuration file for application state transitions and a service which communicates to the REST endpoints on the remote server. Fig. 2 shows the content of a folder for one of our entities, where the previously described files are listed.

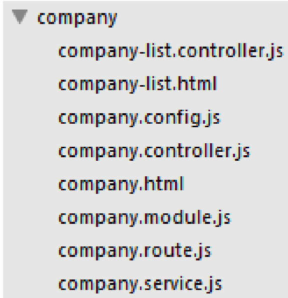


Figure 2. Generated component based on the input DSL

CASE STUDY

This section presents two applications, a registry of cultural entities and a web shop for a local board game store. The first system works with 56 entities, but doesn't require animations or other forms of highly interactive interface. The second system deals with a few entities, but requires a dynamic interface with a lot of animations. Both applications have subsystems which aren't covered by our code generator and have to be implemented manually.

The registry of cultural entities is an information system which records information about cultural institutions, artists, cultural events, domestic and foreign foundations and endowments in AP Vojvodina. Each one of these entities has over ten entities related to them, and some of these entities are shared between the primary five. When speaking in the context of a relational database, the data model consists of over seventy tables, which include tables for recording multilingual information, as well as tables for tracking changes of various entities of the system.

Fig. 3 shows the generated application, localized to Serbian, and its basic layout, where the header and sidebar are mostly static, while the workspace containing the list of institutions changes. The displayed list view is generated for the cultural institution entity. Since this entity has over thirty fields (counting related entities) only a small subset was chosen to be presented in the table. The table offers support for pagination, sorting and filtering based on the displayed attributes. Whenever we want to display multiple entities of a given type we use a table similar to the one displayed in the figure.

By clicking on an item in the table a form for viewing and editing the clicked item is displayed on the workspace, as shown on fig. 4. Most of the form shown in fig. 4 is created using the components the generator provides, like a date picker, an autocomplete textbox for many-to-one relationships and an inline table for one-to-many relationships. A custom UI component which represents a table for adding multilingual data was created as an angular directive (marked with the red square in fig. 4)

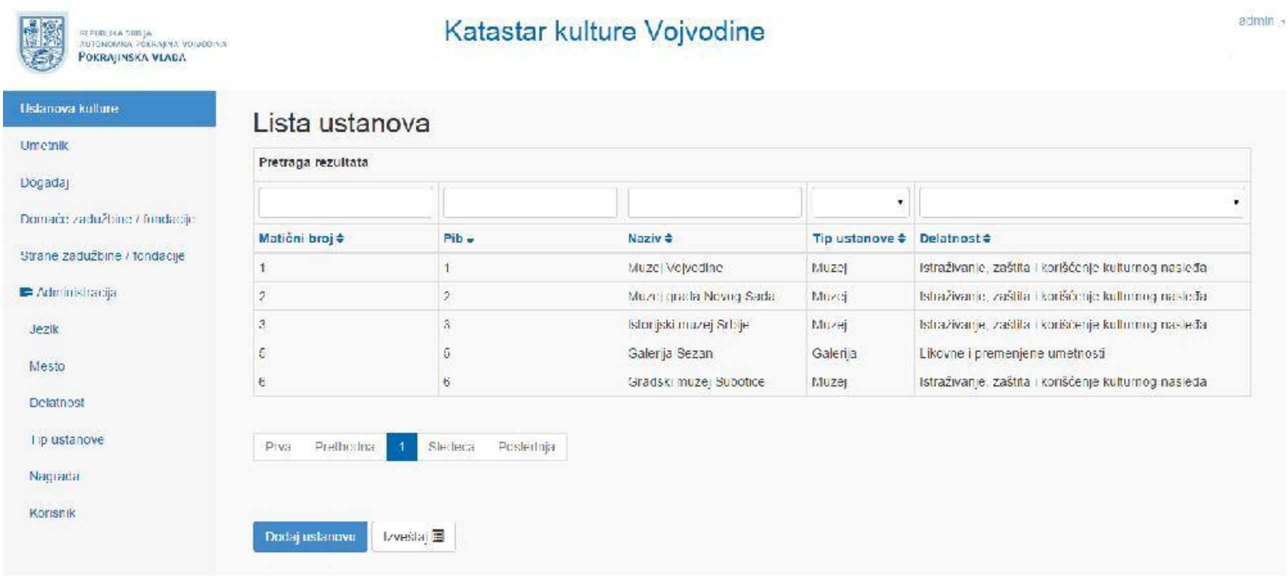


Figure 3. Application layout and list view of cultural entities

and was listed (using the name of the directive) in the input DSL, under `type` *extension*. In this way we have used the extensibility feature of our DSL.

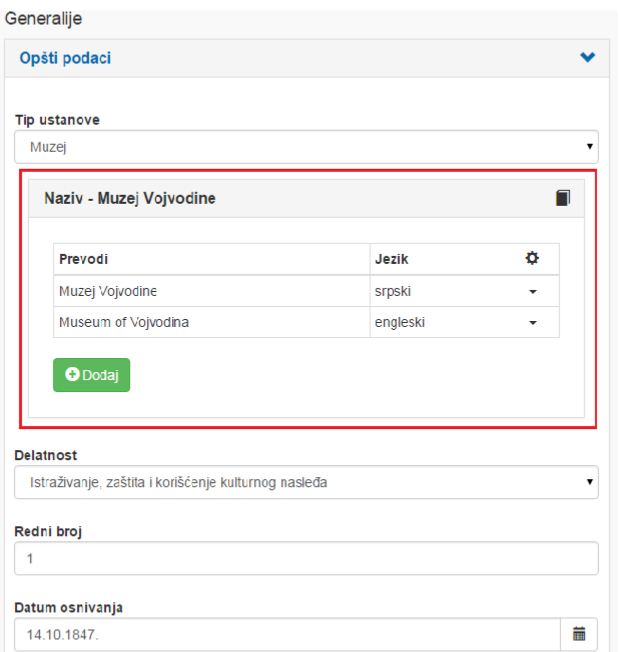


Figure 4. Detail view of a cultural entity

The board game web shop worked with 11 entities but required more work on the UI and UX front. Fig. 5 shows the resulting application. The templates were slightly modified to create a more colourful UI. The detail view has a similar look to the previous application, only the attributes of the entity are displayed as labels and not as input controls and there are fewer fields.

During construction of the listed applications a portion of the code was generated and this code required very little or no modification. A portion of the code was generated but required significant modification, usually by using the generated code as a starting point.

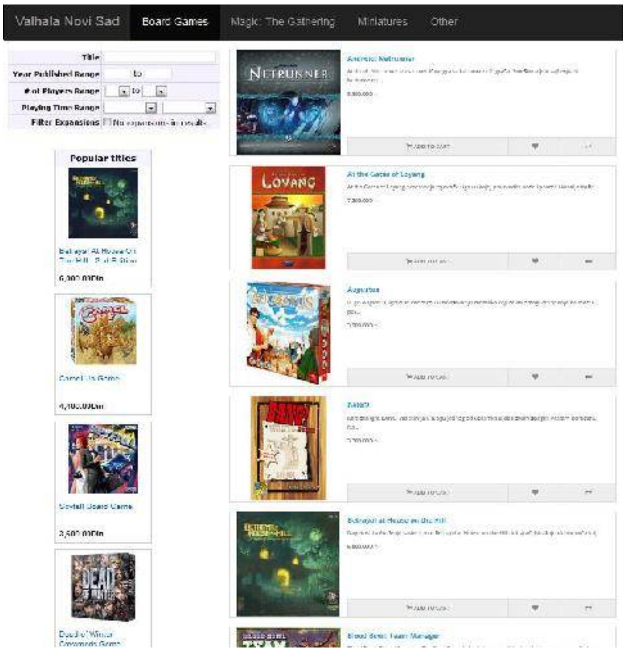


Figure 5. Application layout and list view of web shop

Finally, a portion of the code had to be manually writen so that the system would meet the needed requirements. The percent of generated JavaScript and HTML code that required no or significant modification, as well as the percentile of manualy written code for both applications can be found in table 1.

The front end application for the registry of cultural subjects had about 80% of the code generated which required no or very little modification. Another 10% of the code was generated but required some modification, and this included specific constrains on the forms and custom UI which was different from what our generator provided. The remaining 10% of the code had to be manually written, and this included the subsystem for user authentication, a service for contacting the server to initialize report generation using the WebSocket protocol [15], and a custom homepage.

TABLE I.
PERCENTILE OF GENERATED AND MANUALLY WRITTEN CODE

Application	Registry of cultural subjects	Board game web shop
Number of entities	56	11
Number of pages	19	13
Lines of JavaScript code	7303	2105
Lines of HTML code	5508	985
Generated JavaScript with no or little modification	75%	65%
Generated HTML with no or little modification	90%	65%
Generated JavaScript with significant modification	15%	10%
Generated HTML with significant modification	5%	5%
Manually written JavaScript	10%	25%

The board game web shop required more work on the UI and UX front. While AngularJS does have good support for animations, this wasn't a primary requirement of our code generator, which is the reason why almost no animation and dynamic interface behaviour was generated.

For the front end side of this system about 65% of the code was generated and required no or very little modification, while another 10% required a decent amount of change. The remaining code had to be manually written, and this included the subsystem for making purchases (a shopping cart) and upgrading the UI with animations and graphics.

CONCLUSION

The paper presents a code generator used for creating rich front end web applications, written using the popular AngularJS framework. The generated applications are written following the current best practices for project structure and angular coding styles, contain inbuilt support for internationalization and use visually appealing and intuitive UI components. As input, our code generator uses a model writen in our simple DSL, based on JSON, in order to be easy to learn and in order to avoid overly complicated code generators which take more time to develop than a new version of the technology used in the generated code. Our DSL supports description of both the data model and the user interface layout and components in a concise manner. The code generator uses instances of this DSL as simple JSON objects to construct fully functional applications build with AngularJS and the Bootstrap CSS library.

The DSL and the code generator have been evaluated by creating two applications from different domains. Compared to other similar solutions the generator was either more flexible, by allowing the developer to define both the data model and the layout of the application and/or was easier to use, avoiding corner case constraints and details in implementation and/or was more complete, by generating a full application rather than just project scaffolding.

Our current solution only takes into account the user interface layout and the data model from the REST endpoints. An important part of data driven applications are constraints on user input, and this is something that our DSL and code generator currently don't support. Furthermore, our DSL only takes into account the data model from the REST endpoints and makes the assumption that all applications follow the REST-full pattern for managing entities using a limited set of operations with predefined semantics. This could be a limitation for the practical use of our generator since many systems rely on REST-like web services which do not have a predefined set of methods for manipulating entities.

The future plans for the DSL and code generator include:

- Separating the current DSL into separate logical units, one for defining the data model and its constraints and one for describing the UI layout
- Extend the generator to support REST-like services
- Extend the code generator to support generation of different types of applications, like mobile and desktop and in different technologies

ACKNOWLEDGMENT

Results presented in this paper are part of the research conducted within the Grant No. III-44010, Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] R. Fielding, R. Taylor (2002), Principled Design of the Modern Web Architecture, *ACM Transactions on Internet Technology* (*TOIT*) (New York: Association for Computing Machinery) 2 (2), pp. 115–150, ISSN: 1533-5399

[2] A. Gizas, S. Christodoulou, T. Papatheodorou (2012), Comparative Evaluation of Javascript Frameworks, *Proc. of the 21st International Conference on World Wide Web*, pp. 513-514

[3] D. Graziotin, P. Abrahamsson (2013), Making Sense Out of a Jungle of JavaScript Frameworks, *Product-Focused Software Process Improvement, 14th International Conference (PROFES)*, pp. 334-337, ISSN: 0302-9743

[4] A. Domański, J. Domańska, S. Chmiel (2014), JavaScript Frameworks and Ajax Applications, *Communications in Computer and Information Science (CCIS)* 431: 57-68, ISSN: 1865-0929

[5] C. Kuang, Why good design is finally a bottom line investment, http://www.fastcodesign.com/1670679/why-good-design-is-finally-a-bottom-line-investment, retrieved: 23.11.2015.

[6] J. Papa, Angular Style Guide, https://github.com/johnpapa/angular-styleguide, retrieved: 23.11.2015.

[7] N. Jain, P. Mangal, D. Mehta (2014), AngularJS: A Modern MVC Framework in JavaScript, *Journal of Global Research in Computer Science (JGRCS)* 5 (12): 17-23, ISSN: 2229-371X

[8] I. Dejanović, G. Milosavljević, B. Perišić, M. Tumbas (2010), A domain-specific language for defining static structure of database applications, *Computer Science and Information Systems* 7 (3), pp. 409-440, ISSN: 1820-0214

[9] G. Milosavljević, M. Filipović, V. Marsenić, D. Pejaković, I. Dejanović (2013), Kroki: A mockup-based tool for participatory development of business applications, *Intelligent Software Methodologies, Tools and Techniques (SoMeT), 2013 IEEE 12th Inter'l Conference on*, pp. 235-242, ISBN: 978-1-4799-0419-8

[10] Yeoman Team, Yeoman, http://yeoman.io/, retrieved: 23.11.2015.

[11] Yeoman Team, AngularJS generator, https://github.com/yeoman/generator-angular, retrieved: 23.11.2015.

[12] Tyler Henkel, AngularJS Full-Stack Generator, https://github.com/DaftMonk/generator-angular-fullstack, retrieved: 23.11.2015.

[13] JHipster, http://jhipster.github.io/, retrieved: 23.11.2015.

[14] Internet Engineering Task Force, JSON Schema, Internet Draft v4, http://json-schema.org/, retrieved: 25.11.2015.

[15] Internet Engineering Task Force (2011), The WebSocket Protocol, *RFC 6455*, https://tools.ietf.org/html/rfc6455, retrieved: 26.11.2015.

ReingIS: A Toolset for Rapid Development and Reengineering of Business Information Systems

Renata Vaderna, Željko Vuković, Gordana Milosavljević, Igor Dejanović
University of Novi Sad, Faculty of Technical Sciences, Chair of Informatics, Serbia
{vrenata, zeljkov, grist, igord}@uns.ac.rs

***Abstract*—ReingIS is a set of tools for rapid development of client desktop applications in Java. While it can be used to develop new applications, it is primary intended use is for reengineering existing legacy systems. Database schema is extracted from the database itself and used to generate code for the client side. This remedies the fact that most existing systems do not have valid documentation.**

I. Introduction

Enterprises often use their information system (IS) for a long time. Maintaining such systems can become hard and expensive. They may use libraries, frameworks or even languages that are no longer actively maintained. This can also lead to security threats. Developers who are fluent in technologies used to develop the IS may become scarce. These are some of the reasons why reengineering of existing legacy systems may be necessary. The new information system must take into account existing data structures, persisted data, business processes and flows modeled by the legacy system.

An ideal starting point for developing a new IS would be the technical documentation of the old one. However, such documentation can often be incomplete, not properly maintained (describing the initial version of the legacy system without reflecting changes that have been performed over time) or even non-existent. Therefore, some reverse engineering is usually needed first. Developers may also seek information from the user documentation if it is available. Users themselves can also participate in the process. The goal of these efforts is to replicate functionality of the legacy system while preserving or migrating its contained data.

ReingIS toolset consists of a database schema analyzer, a code generator and a framework. The database analyzer extracts the schema information (tables, columns, types, constraints, etc.) from the database itself. A graphic user interface then allows the developer to define information that could not be read from the schema, like labels and menu structure. Afterwards, the code generator generates components on top of the generic enterprise application framework. The result is an application the users can run straight away in order to inspect it and note necessary changes. These changes are then made using the generator GUI and the process is repeated until satisfactory results are achieved. The toolset also provisions inserting manually coded components and modifications in such a way that subsequent code generation will not overwrite manual changes.

ReingIS toolset facilitates quick introduction of new team members or in-house programmers who maintained the legacy system. The problem with object-oriented technologies is that they are too sophisticated – successful designing and programming using objects takes well-educated and mentored developers, not novices. Classes in class libraries serving as building blocks are too small so the novice has no support. With coarse-grained components and tools built upon the knowledge and experience of senior team members, a novice gets enough support to almost immediately be productive, with the opportunity to gradually master the secrets of modern technologies.

II. Related Work

JGuiGen [1] is a Java application whose main purpose is generation of forms which can be used to perform CRUD (Create, Read, Update, Delete) operation on records of relational database tables. Similarly to our application, JGuiGen can work with a large number of different databases. On top of that, code generated by JGuiGen handles usage by multiple users. Information regarding the database tables is not entered manually. The database is analyzed and descriptions of its tables and their columns are stored in a dictionary, optionally accompanied by comments added by the user with the intention of describing certain elements in more detail, as well as the database schema change history.

When defining a form, it is possible to choose a table contained by the previously mentioned dictionary which will be associated with it. One graphical user interface component is generated for each column of the table and its properties can be customized. Furthermore, JGuiGen also puts emphasis on localization, input validation, accessibility standard [2], and ease of generation of documentation. On top of that, it enables creation of simple reports, which are usually quite significant to business applications.

However, unlike our solution, it does not provide a way in which a user would be able to specify positions of user interface components. They are simply placed one below the other. Additionally, the number of components which can be contained by one tab cannot be higher than 3, while our solution does not enforce this limitation. Associations between two forms cannot be specified using JGuiGen, which means that this feature would have to be implemented after all forms are generated. Similarly, there is no support for calling business transactions.

In [3] the authors use a domain specific language (DSL) to describe tables and columns of a relational database and how they are mapped to user interface components, such as textual fields, lists, combo boxes etc. Description of columns should also contain instructions on how to lay these components out – their vertical and horizontal positions and lengths for components which have it. The generator then uses this information to generate fully

functional forms which can perform various operations on records of previously specified tables. The authors prefer textual notation to a visual one stating better support for big systems as the reason. However, it can be noticed that this solution, just like previously described one, does not support associations between forms, although it is a quite important concept for all, and especially more complex business application. Furthermore, it is not possible to describe and generate activation of business reports and transactions. Finally, as mentioned, this solution demands manual description of tables and columns instead of analyzing the database meta-schema, making the whole process more time consuming and error-prone.

Module for generating application prototypes of the IIS* Case tool [4] is another interesting project which generates fully functional applications which satisfy previously defined visual and functional requirements, allowing records of database tables to be viewed and edited. The process of generation of these applications includes generation of UML (User Interface Markup Language) documents which specify visual and functional aspects of the applicative system and their interpretation which uses Java Renderer. This interpreter transforms UML specifications into Java AWT/Swing components. Furthermore, the module contains Java classes which provide the ability to communicate with the database and pass parameters and their values between forms. Visual properties of the applicative system can be defined by the user by choosing one of the available user interface templates and specifying visual attributes and groups of fields. This is done using another module of the tool. Generation of subschemas of types of forms, whose results provide information which can be used to create SQL queries, is done before the application is generated.

The main difference between this module and our solution lays in the fact that the IIS* Case module is supposed to be used when developing new systems, while ReingIS is optimized to be used when reengineering existing projects with the desire of keeping already existing database schema.

III. IMPLEMENTATION

ReingIS was developed using Java programming language and enables generation of a fully functional client side of a business application based on the meta-schema of an existing database. The architecture of the system is shown in Fig. 1. The application for specifying user roles and permissions is referenced as "security". Each component of ReingIS will be described in more detail in the upcoming sections.

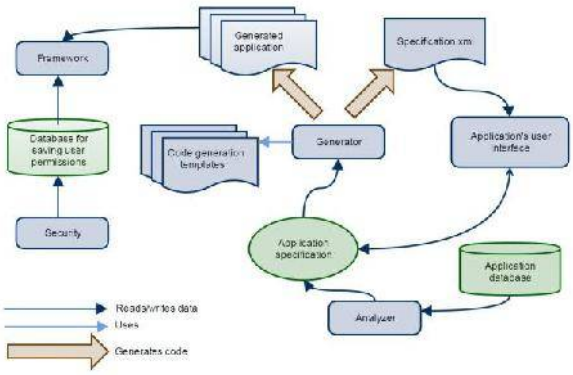


Figure 1. Architecture of the system's framework

The framework provides a generic implementation of all basic concepts of business applications: standard and parent-child forms, data operations (viewing, adding, editing, deleting and searching data), user interface components which enable input validation, activation of reports and stored procedures. For this reason, the framework makes development of business applications easier and quicker. Since all of the important elements were already implemented and tested, that does not need to be done when creating each specific form. Implementation of application elements within the framework follows our standard for user interface specification [5].

1) User interface standard of business applications

The most important elements of our standard, supported by the framework are: standard and parent-child forms and form navigation. The complete description can be found in [5, 6].

Standard form was designed with the intention of making all data and operations which can be performed on them visible within the same screen. Standard operations (common for all entities) are available through icons located in the upper part of the form (toolbar), while specific operations (reports and transactions) are represented as labeled buttons and located on the right side of the form.

Navigation among forms includes zoom and next mechanisms. Zoom mechanism enables invocation of the form associated with the entity connected with the current one by association, where the user can pick a value and transfer it back to the form where zoom was invoked. On the other hand, next mechanism provides the transition from the form associated with the parent entity to the form associated with the child entity in a way that the child form shows only data which was filtered according to the selected parent.

Parent-child form is used to show data which has a hierarchical structure, where every hierarchy element is modeled as an entity in the database and is shown within its standard panel. Panel on the n^{th} level of the hierarchy filters its content based on the chosen parent on the $(n-1)^{th}$ level.

2) Implementation of generic standard and parent-child forms

The core component of the framework is the generic implementation of the standard form (Fig. 2), which allows creation of fully functional specific forms by simply passing description of the table associated with the form in question, its columns and links with other tables, as well as the components which will be used to input data.

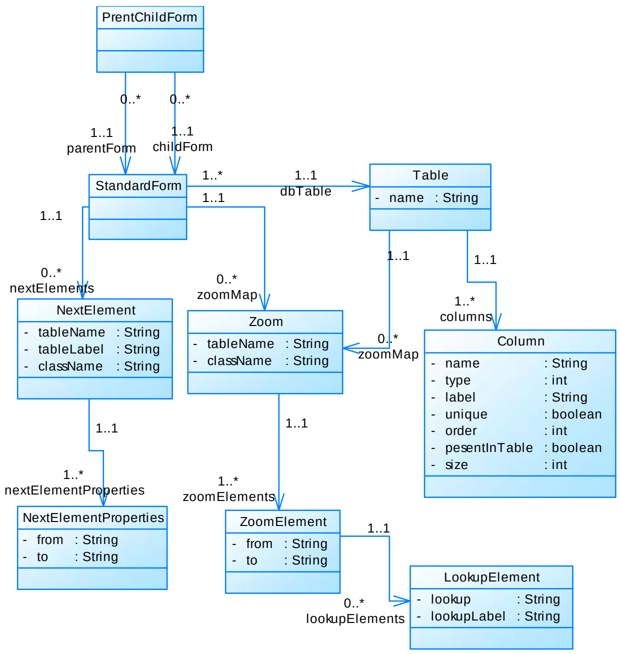


Figure 2. Class diagram of standard and parent-child forms, the framework's core components

The information about the tables and its columns which are passed to the generic forms are represented with classes `Column` and `Table`. Attributes of these classes are used during the GUI construction phase, as well as for dynamic creation of database queries and retrieving their results. This eliminates the need to write any additional code for communicating with the database.

The description of the table's links with other tables is necessary for generic zoom and next mechanisms and is represented by the following classes: `Zoom`, `ZoomElement`, `LookupElement`, `NextElement` and lastly `NextElementProperties`. Classes `NextElement` and `Zoom` contain data related to the tables connected with the current one, as well as names of Java classes which correspond to forms associated with those tables (attribute `className`). The name of a class is all that is needed to instantiate it using reflection. Classes `ZoomElement` and `NextElementProperties` contain information regarding the way in which the columns of one table are mapped to the columns of the other one. This is important for automatic retrieval of the chosen data when zoom mechanism is activated and for automatic filtering when a form is opened using next mechanism. If additional columns of tables connected through the zoom mechanism with the current one should be shown (for example, name and not just id of an entity), their names and labels should be specified as well. Class `LookupElement` is used for this reason.

Validation of the entered data can be enforced on the form itself by using specially developed graphical user interface components. The query is not sent to the database unless all validation criteria is met, which reduces its workload. These components are:

- `ValidationTextField` – represents a textual field which can be supplied with validation data, such as the minimal and maximal possible length of the input, indicator if the field can only contain digits or other characters as well, the minimal and maximal value for numerical input, indicator if the field is required, and, finally, patterns, i.e. regular expressions that the input value is checked against (for example, this can be used to validate an e-mail address).
- `DecimalTextField` –field which is used to enter decimal values. The input is aligned to the right side and the thousands separator is automatically shown when needed. Maximal length of the number and the number of decimals can be specified.
- `TimeValidationTextField` – field used to input time as hours, minutes and seconds.
- `ValidationDatePicker` – component which extends `Jcalendar` component, which is licensed under *GNU Lesser General Public* license.

The generic parent-child form was implemented as two joined standard forms linked through the next mechanism. Creation of a specific form of this type only requires two previously constructed standard forms to be passed.

3) Reports and transactions

Calling previously created Jasper [7] reports and stored procedures can be done through a menu item of the application's main menu, as well as inside standard forms. It is necessary to define parameters needed by the procedure or a report, if there are any. Everything else is done automatically. Framework also provides a generic parameters input dialog, which needs to be extended and supplied with specific input components.

B. *Analyzer*

The analyzer uses the appropriate Java Database Connectivity (JDBC) driver and establishes connection with the database which needs to be analyzed and, using `java.sql.DatabaseMetaData` class, finds the information regarding its tables and their columns, relations and primary and foreign keys. The end user doesn't need to know which JDBC driver and Database Management System (DBMS) are used, which means that a large number of different databases can be analyzed.

Based on the retrieved information, the analyzer creates an initial, in-memory, specification of the business application (i.e. instances of the `StandarForm` class are created), using the following transformation rules:

- Every table is mapped to a standard form
- Every column of the database tables is mapped to an input component contained by the form
- Names of the columns and tables are used as labels of components and forms, in that order
- Types of the input components corresponding to the columns are determined based on the types of those columns. A textual field is added when the column is of a textual type (char, varchar), a date input component is added when the column is of a date type, a textual field which automatically enforced validation which only enables numbers to be entered is added when the column is of a numerical type

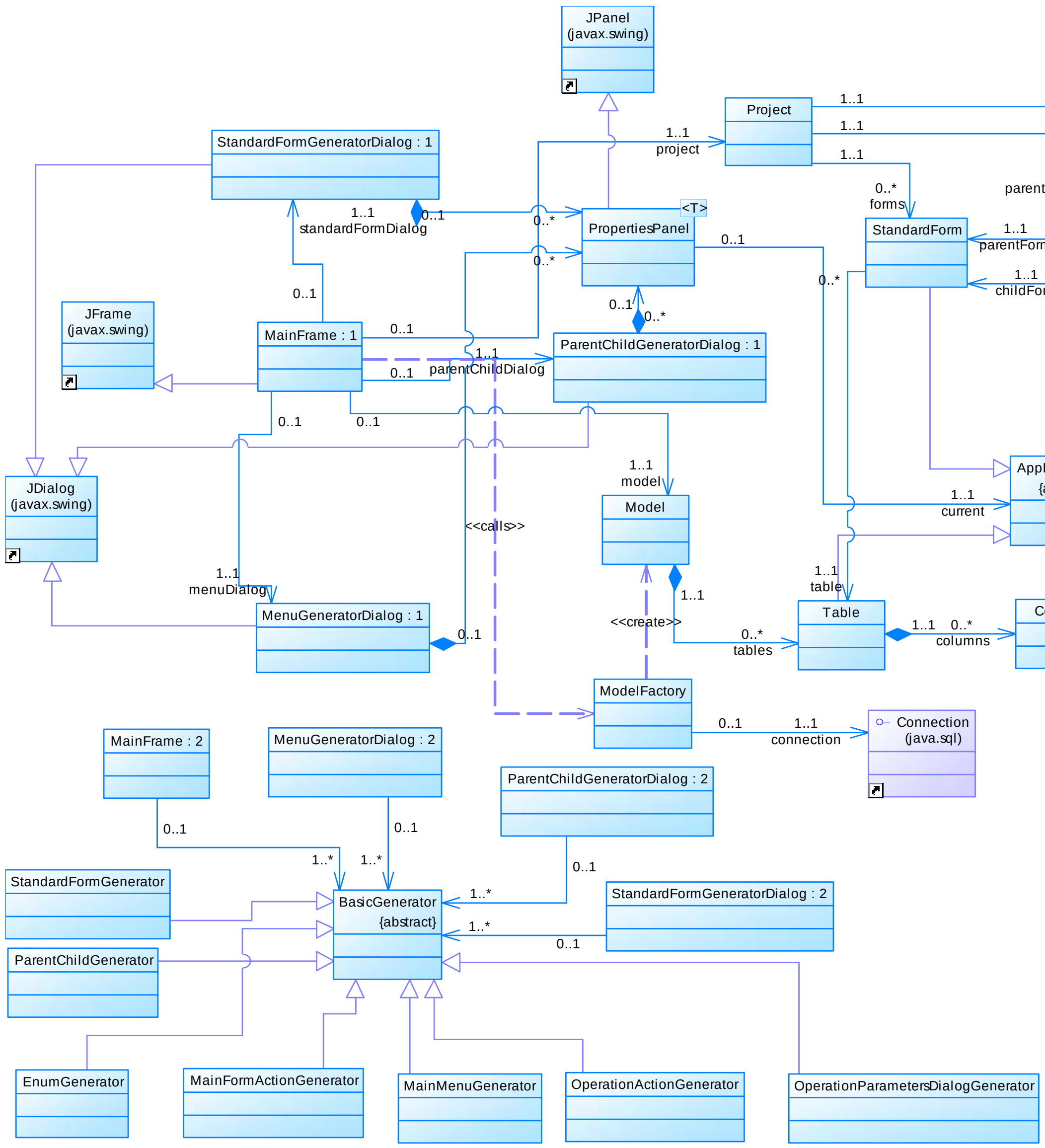


Figure 3. Class diagram of the analyzer, generator and its user interface

- Length of a textual field, as well as width of the corresponding column in the form's table are determined based on the length of the database column

Analysis of the database meta-schema is done when the application is run for the first time or when it is connected to a different database. The acquired data, regarding found database tables and their columns, is saved to a XML file, which is then loaded when the application is started again. Therefore, the time consuming database analysis process is avoided unless it is necessary. If needed, the user can activate the analysis from the application at any time.

The class which provides the mentioned functionality is `ModelFactory`, while the class `Model` represents the database meta-schema, as shown in Fig. 3. This class contains a list of tables discovered during the analysis. Each of them is described by class `Table`, which contains a list of columns represented by class `Column`.

This default specification created by the analyzer can later be changed and enhanced through the generator's components, grouping of fields into tabs and panels, creation of zoom, next and lookup elements and parent-user interface by the users. The mentioned application enables additional specification of labels of forms and child forms etc.

C. *Code Generator User Interface*

The user interface of the generator application consists of three dialogs represented by classes `StandardFormGeneratorDialog` (for specifying standard forms), `ParentChildGeneratorDialog` (for specifying parent-child forms) and `MenuGeneratorDialog` (for specifying menus) – Fig. 3. These dialogs are activated through the main form of the generator application.

The mentioned dialogs rely on instances of classes `StandardForm`, `ParentChildForm` and `Menu` to store data needed to generate forms and menus, such as sizes, titles and input components of forms and names and structures of menus.

When a form is first created, default settings are set (for standard forms, they are based on the analysis results). Therefore, the generator can generate usable program code straight away. These settings can me modified by the users through certain panels contained by the mentioned dialogs. These panels are instances of class `PropertiesPanel` – a parametrized class which enables

various properties of an element associated with it to be changed (class `AppElement`). Class `AppElement` is extended by all classes which represent an element of a business application or one of its parts. In order to enable work to be saved and edited on a later occasion, the term project is introduced. It is represented by class `Project`, which contains lists of defined standard and parent-child forms and menus.

1) Specifying standard forms

Dialog for specifying standard forms enables modification of default form settings set by the analyzer i.e. based on the database meta-schema. The following properties of a form can be adjusted: label, size, associated database table, initial mode (add, search, view/edit), allowed data operations, grouping of contained components into tabs and panels, links with other forms through zoom and next mechanism, specific operations activated from the form – reports and transactions. One database table can be associated with multiple forms (Fig. 4). Additionally, the following properties of form components can be set: label, indicator if its content can be edited, position – specified using MigLayout [10] constants, validation rules etc.

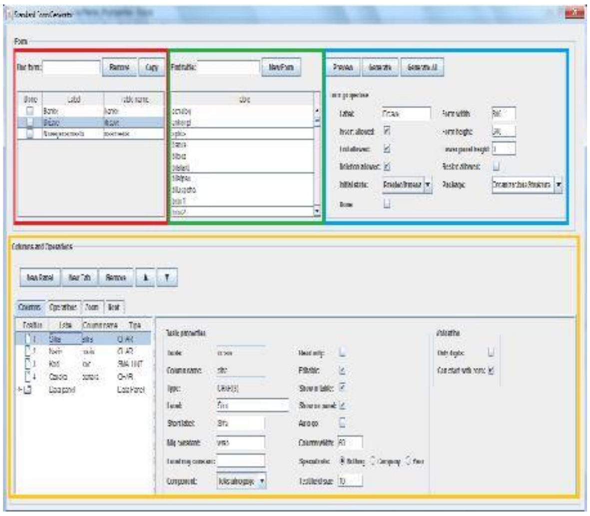


Figure 4. Dialog for specifying standard forms

The dialog consists of:

- A part for choosing associated database table – within green borders
- A part for selecting and searching previously created forms – within red borders
- A part for setting basic form properties – within blue borders
- A part for specifying links with other forms, component groups and component properties – within orange border

Fig. 7 shows an example of a generated standard form.

2) Specifying parent-child forms

Dialog for specifying parent-child forms (Fig. 6) enables users to choose two standard forms, one of which will be the parent, while the other one will be the child. After this step is performed, a new parent-child form is created and default initial values of its properties, such as title and size, are set. Therefore, the corresponding Java class can be generated at any moment and the form's current appearance can be previewed if so desired.

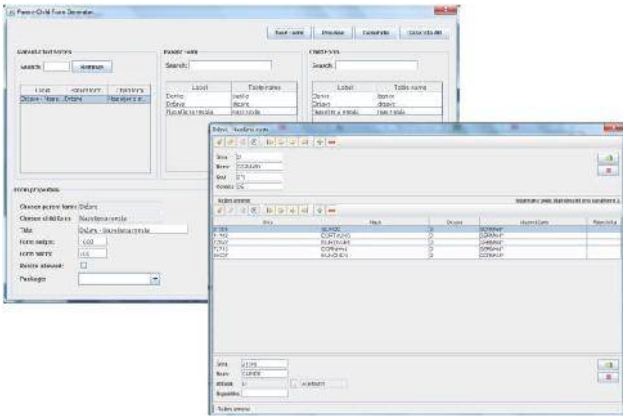


Figure 6. Dialog for specifying parent-child forms and the resulting form

3) Specifying menus

Menus of business applications can be specified using another dialog of the generator application. It is possible to create menus with submenus, which can contain additional submenus as well. The following properties can be defined for menu items: name, shortcut, description (mapped to tool tip text), form or report which will be activated by clicking on the item. If the menu item activates a report, it is necessary to specify parameters which will be passed to it. This dialog is shown in Fig. 5.

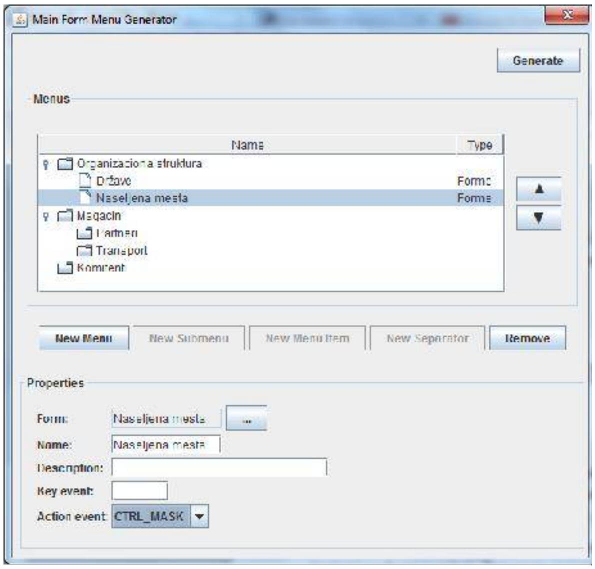


Figure 5. Dialog for specifying application menus

D. Code Generator

Using the database metadata and additional information given by the developers, code is generated (using Freemarker [9] template engine) for the framework components. This includes the main form and its menus, standard forms and parent-child forms. Generated form classes extend generic classes that are a part of the framework. It can be noted that the generator supports synchronization with changes made to the database after the specification of the forms was started. If some columns were added or removed from a database table, the corresponding components are automatically added to or removed from the appropriate form. Similarly, when a database table is deleted, the form associated with it is also deleted.

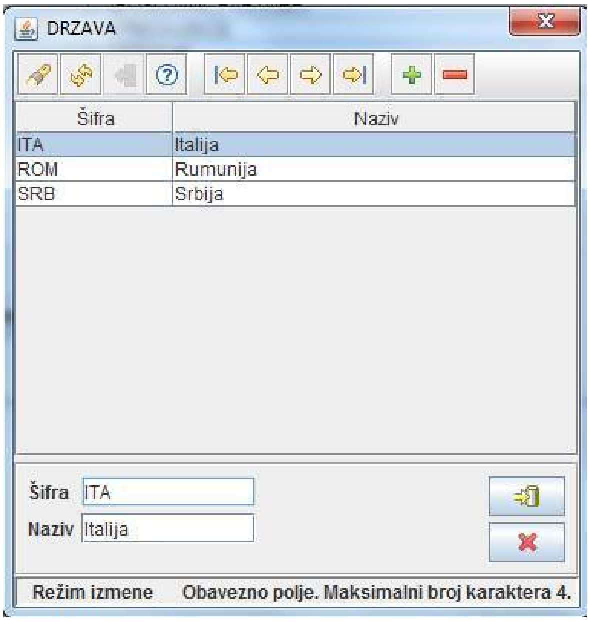


Figure 7. A generated standard form

E. Security

Handling of security concerns for the generated application is based on the Apache Shiro framework [8]. Shiro is a Java framework that supports authentication, authorization, cryptography and session management. Within the framework each operation is given an identity string. Based on the currently active user and the operation identity it can be determined if the operation is allowed. A user interface was implemented in order to allow managing users and groups and assigning them access rights. It is also possible to import existing users from the legacy system.

In order to facilitate construction of the user interface for the end application, a set of classes was implemented. They extend the standard Swing classes and make them aware of the security context. When these components are used, user interface elements are automatically disabled or hidden if the current user is not allowed to perform the action that they are associated with. `SecureAction` is an abstract class which extends Swing's `AbstractAction`. The `actionPerformed()` method is redefined and made final. This method uses Shiro to authorize the action and then calls the `action()` method. This method is to be implemented by classes extending the `SecureAction` class. The identity string of the action is returned by `getActionIdentity()` method, which is also to be implemented by subclasses. Classes `SecureJButton`, `SecureJMenu` and `SecureJMenuItem` extend classes `JButton`, `JMenu` and `JMenuItem` respectively and are to be used with `SecureAction`. All classes register themselves as an `AuthenticationListener` in Shiro. This enables them to react interactively when the current user is switched.

IV. CONCLUSION

The toolset presented here enables reengineering of legacy enterprise information systems. It uses existing database structure as a basis for replicating functionality and preserving data contained in the original system. After adding user interface details that could not be extracted from the schema (labels, menus, etc.) developers can run the code generator which produces generated code on top of the generic framework, resulting in a runnable application. This application can then be presented to the users who can verify its functionality. Since it is possible to repeat the code generation while maintaining all settings and customization, the process also supports forward engineering and incremental development.

The process could be further improved if we were able to extract user interface elements in addition to the database structure. In order for the approach to remain applicable for a wide variety of applications, this requires development of a generic UI element extractor. Each plugin could provide extraction facilities for one technology, e.g.: COBOL screens, .NET forms, Swing frames, web pages, etc.

REFERENCES

[1] JguiGen, http://jguigen.sourceforge.net

[2] A11y standard, http://a11y.me

[3] Lolong S, Kistijantoro A, Domain Specific Language (DSL) Development for Desktop-based Database Application Generator, International conference on Electrical Engineering and Informatics, Bandung, Indonesia, 17-19 July 2011.

[4] Ristic S., Aleksic S., Lukovic I., Banovic J., Form-Driven Application Development, Acta electrotechnica et Informatica, Vol. 12, No. 1, 2012, pp. 9–16, DOI: 10.2478/v10198-012-0002-x

[5] Milosavljevic G., Ivanovic D., Surla D., Milosavljevic B., Automated construction of the user interface for a CERIF-compliant research management system, The Electronic Library, Vol. 29 Iss: 5, pp.565 - 588

[6] Perišić, B., Milosavljević, G., Dejanović, I., Milosavljević, B., UML Profile for Specifying User Interfaces of Business Applications, Computer Science and Information Systems, Vol. 8, No. 2, pp. 405-426, DOI 10.2298/CSIS110112010P, 2011.

[7] Jasper Reports, http://jaspersoft.com/

[8] Apache Shiro framework, http://shiro.apache.org/

[9] Freemarker Java Template Engine, http://freemarker.incubator.apache.org/

[10] MigLayout - Java Layout Manager for Swing, SWT and JavaFX2, http://www.miglayout.com

Assessing Cloud Computing Sustainability

Valentina Timčenko, Nikola Zogović, Miloš Jevtić, Borislav Đorđević
University of Belgrade/Institute Mihailo Pupin, Belgrade, Serbia
{valentina.timcenko, nikola.zogovic, milos.jevtic, borislav.djordjevic}@pupin.rs

Abstract— **In this paper we deal with the issue of providing a suitable, comprehensive and efficient sustainability assessment framework for cloud computing technology, taking into consideration the multi-objectivity approach. We provide the comparison methodology for Sustainable Development Goals models, and apply it to the proposed multi-objective cloud computing sustainability assessment model and the general United Nations (UN) framework, taking into consideration the emerging issue of open data.**

I. Introduction

Cloud computing represents an innovative computing paradigm designed with an aim to provide various computing services to the private and corporate users. As it provides a wide range of usage possibilities to the users, along with the sustainability, it has become one of the most promising transformative trends in business and society. As such, this trend imposes the need of proposing a model for assessing cloud computing sustainability. The extensive reference research indicates that there were several attempts to proceed with this idea, but there is still no unified approach. The sustainability approach is a qualitative step forward when compared to other methodologies. Taking all this into consideration, we have proposed a new model which is still in the research phase. The basics of our concept are presented in [1].

The framework development becomes more challenging with taking into consideration the need of integrating the issue of open data to the framework proposal. The Open Data phenomenon is initiated by the Global Open Data Initiative (GODI) [2]. The goal is to present an idea of how governments should deal with the open data accessibility, raise awareness on open data, support the growth of the global open data society and collect, increase, and enlarge the databases for open data. Different countries started to gradually accept the idea of open data and are taking the initiative for the introduction of adequate legislation. The national and international laws related to the free access to information of public importance constitutionally guarantee human rights and freedom, and form an integral part of numerous international documents which set standards in this area. E.g. Serbian Government regulates the right to free access to information with a special law constituted in 2006. It constitutionally guarantees and regulates the right of access to information, and in addition to access to information of public importance held by public authorities, it includes the right to be truthfully, completely and timely informed on issues of public importance [3]. This law establishes the Commissioner for Information of Public Importance, as an autonomous state body which is independent in the operation of its jurisdiction.

The foundation idea of our framework is to encompass the four different aspects that are highly influenced by the trends in cloud computing development, and provide a comprehensive multi-objective (MO) model for assessing sustainability. Such a MO perspective is foreseen to take into account how cloud computing affects economy, business, ecology and society. This methodology provides flexibility in allowing all participants to support objectives that they found relevant for their needs, eliminating the necessity to find a way to fit to any of the existing constraints, which is typical for a pure sustainability approach [4]. The named areas are of the primary interest as the consumers are becoming heavy users of cloud computing services satisfying their needs for social communication, sensitive data exchange, or networking, all in compliance with the rights stated in Universal Declaration of Human Rights (UDHR) [5]. This trend also strongly influences the economical development, strengthens the business communities [6] and significantly raises the environmental awareness of the society [7].

The goal of this paper is to further elaborate proposed model, proceed with the comparison to the state of the art in this area, and positioning of our model. The research of the current state of the art in the area of cloud computing sustainability assessing models leads only to the United Nations (UN) general model, thus it will serve as the foundation for initial consideration and reference for comparison [4].

The UN model relies on the introduction of Sustainable Development Goals (SDG) defined by Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) which motivates international community to put additional attention to the indicator framework and associated monitoring systems. The first guidelines for SDG establishment were given in 2007 [8]. The named document provides the set of Indicators of Sustainable Development and presents recommendations on the procedures for adapting them at national level, in accordance to national priorities and needs. More recently, in 2015, UN report on "Indicators and a Monitoring Framework for the Sustainable Development Goals" was published as a response to the need for contribution in support of the SDGs implementation. It outlines a methodology of establishing a comprehensive indicator framework in a way to support the goals and targets proposed by the Open Working Group (OWG) on the SDGs [4].

The framework for the sustainability assessment heavily depends on the access to the open data, which should be available under no condition. Moreover, the availability of the data is the necessary condition for assessing the sustainability, as building a special, dedicated system for collecting such an amount of data is unprofitable. The Inter-agency and Expert Group on

SDGs (IAEG-SDGs) have organized a set of meetings in Bangkok, during October 2015, where the main topic was the development of an indicator framework which purpose is to monitor the goals and targets of the post-2015 development agenda. As it is emphasized in Global Policy Watch report, it was agreed that the UN framework in its final version is to be presented to UN Statistical Commission in March 2016 [9]. Until then, it is of importance to find a proper agreement on the suggested indicators for each defined goal, being aware that indicators alone cannot be sufficient for measuring the advancement of the development of the goal.

In this paper we first introduce the comparison methodology for SDG models. Then, we apply it to the UN and to the proposed MO cloud computing sustainability assessment model, taking into consideration the open data initiative principles. Finally, we conclude with some remarks related to the provided comparison.

II. Cloud Computing Sustainability models

The models for sustainability assessment can be classified as general and specific. Alternatively, the models can be territorially (geographically) classified as global, regional, local and national. For the needs of comparison and evaluation of the cloud computing sustainability, as a general model we chose the one proposed by UN, and compare it to the MO framework.

The UN framework relies on 100 sustainable development indicators defined in conjunction with 17 SDGs [4]. Our aim is to try to provide the mapping of sustainable development indicators to our MO framework. Figure 1 presents UN framework principles for global monitoring indicators.

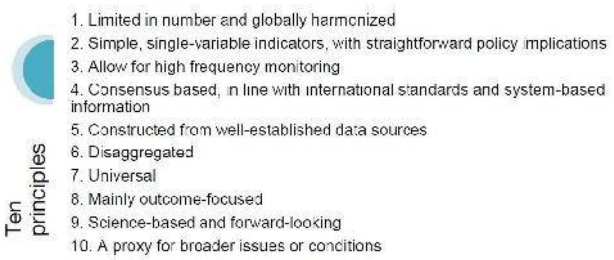


Figure 1. Ten principles for Global Monitoring Indicators [4]

The UN framework SDGs are listed in Figure 2.

Taking into account defined goals and list of UN indicators [4] we provide a mapping of the indicators to the areas covered by the MO framework. It is performed taking into account the definition of the indicators, without a specific policy and rules for mapping. Figures 4, 5, 6, and 7 provide the corresponding mapping of the indicators (represented in form of the numbers, as they appear in [4]).

Goal 1. End poverty in all its forms everywhere
Goal 2. End hunger, achieve food security and improved nutrition, and promote sustainable agriculture
Goal 3. Ensure healthy lives and promote well-being for all at all ages
Goal 4. Ensure inclusive and equitable quality education and promote life-long learning opportunities for all
Goal 5. Achieve gender equality and empower all women and girls
Goal 6. Ensure availability and sustainable management of water and sanitation for all
Goal 7. Ensure access to affordable, reliable, sustainable and modern energy for all
Goal 8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all
Goal 9. Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation
Goal 10. Reduce inequality within and among countries
Goal 11. Make cities and human settlements inclusive, safe, resilient and sustainable
Goal 12. Ensure sustainable consumption and production patterns
Goal 13. Take urgent action to combat climate change and its impacts
Goal 14. Conserve and sustainably use the oceans, seas and marine resources for sustainable development
Goal 15. Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss
Goal 15. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels
Goal 17. Strengthen the means of implementation and revitalise the global partnership for sustainable development

Figure 2. Seventeen UN framework SDGs [10]

Figure 3 presents the general overview of the proposed MO Assessment Framework for cloud computing, showing the first two layers of the model. Each of the shown branches is further layered in accordance to specific area characteristics.

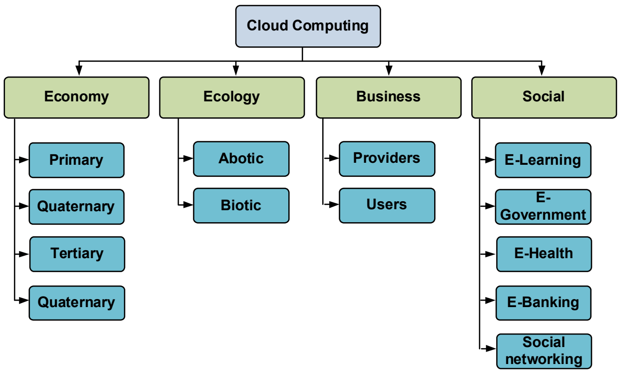


Figure 3. General overview of a proposed Multi-objective Assessment Framework for Cloud Computing

Figure 4 represents the mapping within the social aspects area. It covers the provisioning to the users the set of e-services, taking into consideration the fulfilment of rights claimed in Universal Declaration of Human Rights (UDHR) and legislative of certain country [5]. The set of e-services can be grouped into: e-Learning/e-Education, e-Government, e-Health, e-Banking, and social networking. These basic services can be further analysed through benefits and issues/risks. All of these subcategories have a set of common characteristics, and some of the most important are the privacy and security of the data which is shared among different user categories, and awareness that there is a need for developing services to help users with disabilities to efficiently satisfy their special needs.

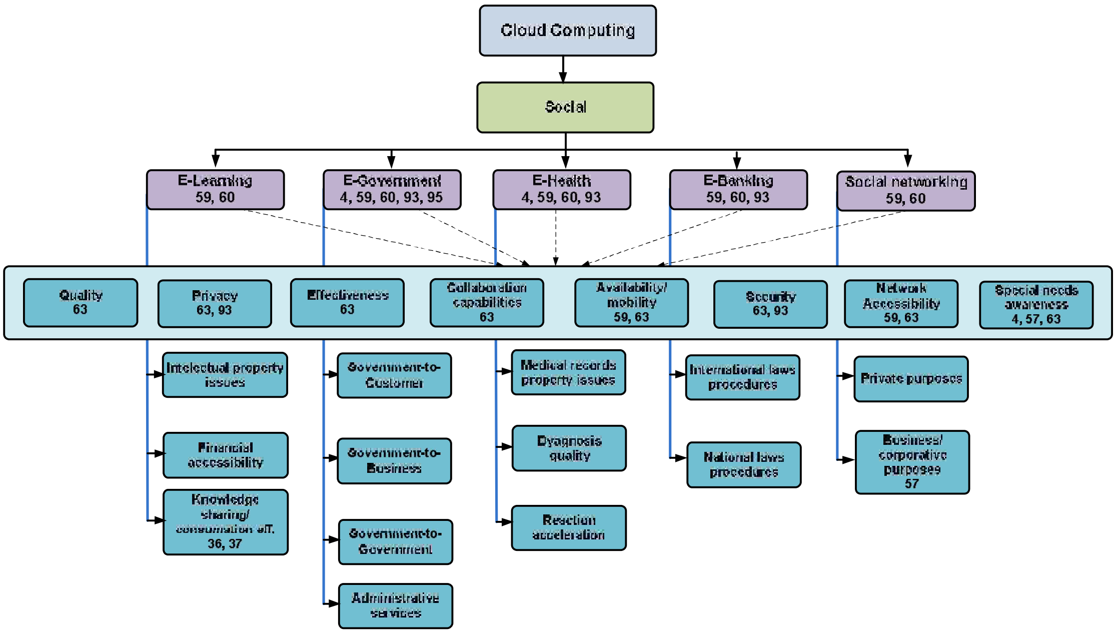


Figure 4. Social aspects for framework assessment

Figure 5 shows the ecology branch of the framework. Ecology objectives can be classified according to general ecology factors into abiotic and biotic [11]. The abiotic branch deals with the issue of pollution generated during the use of cloud computing resources in different life-cycle phases, and with carbon footprint that is typical for each cloud computing component. Biotic branch considers impact of cyber physical systems to the reestablishment of degraded homeostasis or to the process of keeping the existing homeostasis.

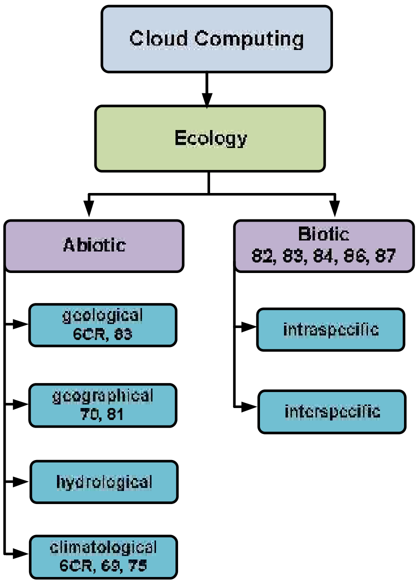


Figure 5. Ecology assessment framework

Figure 6 provides mapping within the economy area.

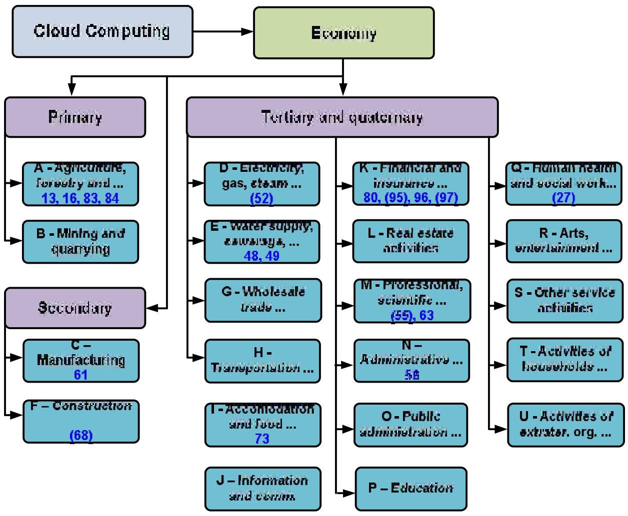


Figure 6. Economy aspects for framework assessment

The service economy (SE) concept [12] is analysed as the integrated part of Information and Communications Technologies (ICT) domain. As such, it fits well within the MO cloud computing framework [1]. Actually, the economy sector is treated based on the widely known concept of the four economic sectors – primary, secondary, tertiary and quaternary [13] and classification of economic activities is obtained from the UN International Standard Industrial Classification (ISIC) hierarchy with 21 categories of activities labelled with letters A to U [14].

Special effort is put into covering the business area, as that sector is not covered by UN model. The focus is on the business objectives related to the interests of cloud computing business users (Figure 7).

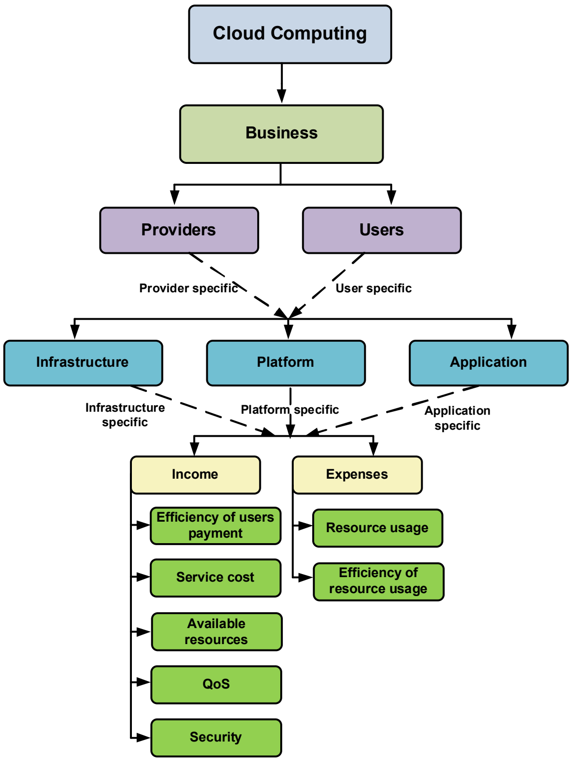


Figure 7. Business aspects for framework assessment

This area highly depends on Open Data initiative implementation as, for example, it can multiply educational and job opportunities and can help people achieve greater economic security.

We have put an effort into allocating all the UN indicators to the defined MO framework sectors. There are some sectors with no indicators assigned which shows that the UN framework for sustainability has not considered that all specified activities are equally important for sustainability. On the other hand, some of the indicators that we have previously identified as associated with e.g. some economic section (figure 6), latter could not be assigned to any ISIC section. The very same situation appears within other considered areas (figures 4 and 5), and especially for the business area (figure 7) as we could not find indicators that can cover it successfully. The frameworks are assessed based on following:

1. Control cycles phases defined for the chosen model
2. Choice of the target user
3. Principles for determining indicators and targets
4. Number of indicators
5. Readiness of the framework
6. Areas covered by sustainability assessment models

where this list of points is a cornerstone for further comparison procedure and evaluation.

III. Methodology

Cloud Computing is the concept designed with an aim to satisfy many different flavours, and it is tailored toward different users and companies. The main expectations of most cloud services are at least to allow self monitoring and self-healing, existence of the proper service level agreement, advanced automation, pay-per-used service, and high level of reliability and availability [15]. From the standpoint of control systems, an important role in this comparison plays the understanding of the purposes which have initiated the application of the sustainability assessment procedure. The theory of control systems relies on: control cycles, multi-objective optimization and dynamic control. Dynamic control theory is founded on the need for allowing a controlled transition from some specific state to the desired state of the system. The Multi-Objective Optimization (MOO) encompasses the formulation of the issue based on the vector of the objectives, as the approach relying on the single objective may not satisfactorily represent the considered problem. Dynamic control of the system should allow the most efficient combination of the MOO and adaptive control, in a way to keep transitions slight, without dislocating the system to the undesirable regions. The idea is to allow transition from the system state oriented assessment framework to the system control oriented framework, where it is important to provide dynamic MO control of the system and keep the homeostasis in desired state.

IV. Comparison

The comparison of the MO and UN frameworks is provided taking into consideration the aspects listed in previous chapter. When making MO framework comparison to the UN framework, several observations can be made.

1. When considering **the control phases**, both models provide a set of specific phases, where some of them coincide.

The UN framework does not rely on the real control cycle but on a set of phases: Monitoring, Data Processing, and Presentation. Unlike the UN framework, the proposed MO framework relies on the full control cycle. Figure 8 represents the comparative overview of the defined UN phases versus the MO framework cycle.

The **Monitoring phase - UN** relies on the list of Global Monitoring Indicators (GMI) whose progress is supervised on defined time basis taking into consideration local, national, regional, and global level of monitoring. MO framework considers this first phase as **Data Collecting - MO** phase, as it basically relies on that process. For the process of monitoring/data collection there is a need to cover a wide range of data types. The UN framework, with 100 indicators and a set of sub indicators targeting different levels (global, regional, local and national), requires an enormous monitoring system that would process the huge amount of collected data. As it is evident that creating such a system would be a time consuming and costly task, the monitoring/collection of data should rely on existing systems and in particular to those owned by the State. Therefore, the UN framework relies mostly on a national level data monitoring, while the idea of the MO is to collect open data and private data.

Data processing phase - UN is assumed to be realized by specialized UN agencies and other international organizations that are connected to national statistical offices (NSO), companies, business and civil society organizations. They put efforts into determining the standards and systems for collecting and processing data.

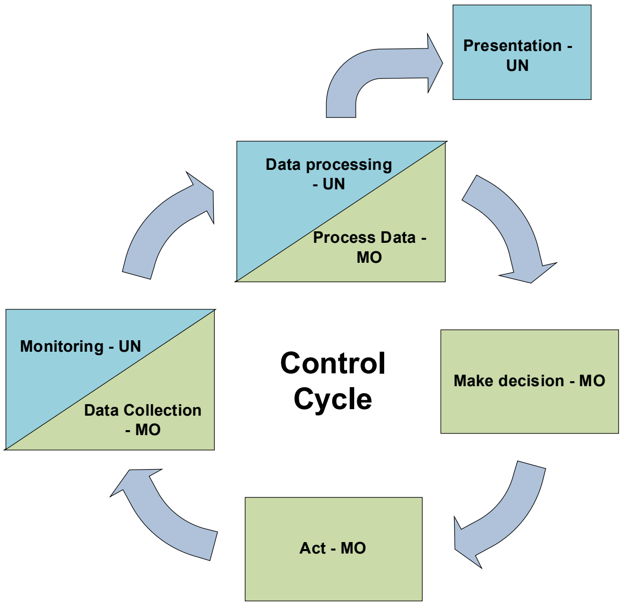


Figure 8. MO versus UN SDG framework

Presentation and analysis - UN is the final phase of the UN model. It is performed through generation of different reports, and organization of workshops and conferences. In contrast, the MO framework considers the **Process Data - MO** phase as the input for the **Make Decision - MO** phase. In this phase the decision makers are offered the possibility to use information generated based on the MO optimization in order to make the decisions. On the bases of the taken decisions it is further possible to proceed to the **Act - MO** phase which corresponds to the operational part of the MO framework.

2. The target user/group represents important difference between this two frameworks. The UN framework sees the final user as a target, while all the data are publicly available. In contrast, the MO framework is primarily designed for corporate users, who take part in managing the processes based on the specific technology. Based on the profound research related to this aspect, we have realized that there is a high need to raise the awareness of the necessity to incorporate to the framework the fact that the technology forms a great part in every day's life of personal and corporate users. We have noticed that the UN framework lacks the indicators/sub-indicators that would properly indicate the level of the exploitation of the latest technology trends.

3. The high level consideration is the adopted set of **principles for determining indicators and targets**. The UN model relies on 10 principles defined towards fulfilment of the idea of an integrated monitoring and indicator framework (Figure 1). The basic principle of MO framework is to provide a multi-objective dynamic control system. The indicators must give real time information, and it must be made available before the defined time limit.

4. When thinking about the **number of indicators**, UN framework encompasses 100 indicators and 17 groups of the goals (defined on global, regional, local and national levels), whereas MO model is still in development, and aims to encompass companies grouped by size (global, regional, local) and ownership structure (public, private, combined). It is noticed that there is a lack of proper indicators for the area of business sector. It is also of great importance to provide proper set of indicators that would cover technological development, science and academia.

5. The **readiness of the framework:** the UN framework is a long year's process documented by, so far, two editions. The third edition is expected to be shortly published, and it is foreseen that it will encompass the business aspects as well. On the other hand, the MO framework is still in research and development phase.

6. The main discussion topic is the **areas covered by sustainability assessment models.** The MO framework is dominant as it covers areas of economy, business, society, ecology, while the UN framework still lacks the business indicators. The UN framework claims the necessity of covering this area as well. The major contribution to this initiative is claimed to be on several stakeholders and organizations supporting sustainability development, whereas the ultimate goal is to align the business metrics to the defined SDG indicators. For guaranteeing the best possible mapping, it is important to identify the crucial business indicators which can successfully track the business factors and their relation to SDGs.

In MO framework we consider business area from the very start. We cover both the service providers and end users. The framework encompasses used infrastructure, platform type, and used applications. When considering the infrastructure provider objectives it is important to consider those related to income maximization (efficiency of users payment, service cost, available resources, quality of service (QoS), and security) and the other related to expense minimization (resource usage, efficiency of resource usage, etc.). QoS in cloud computing depends on performance, fault tolerance, availability, load-balancing, scalability, while security aspects can be analysed through the security at different levels, sensitivity of security systems, and determination of security systems. Security objectives are usually in divergence with performance and energy efficiency. Moreover, the open data would seem to be a necessary condition for the implementation of our framework at full capacity. The initiative for the opening of the government data has to deal with the need to provide transparency, participation of different interested sides, and to stimulate development of the new services related to the proper and safe data usage.

At the national level the data is often non accessible, thus there is a need for open data initiative. The UN framework considers the use of open public data while MO framework relies on the use of both open public data and private data (Figure 9). Although the UN framework has not launched the open data initiative it will use it for its functioning. MO framework also needs a huge amount of diverse data, mostly referring to the open data which is held by the state. The accessibility to it depends on the existence of the laws that regulate the open data concept. E.g. in Serbia it is regulated with the Law on Free Access to Information of Public Importance [3].

The open data combined with cloud computing can facilitate development of the innovative approaches, in a way that the companies are using open data to make use of market gaps and recognize prominent business opportunities, develop novel products and services and create new business models.

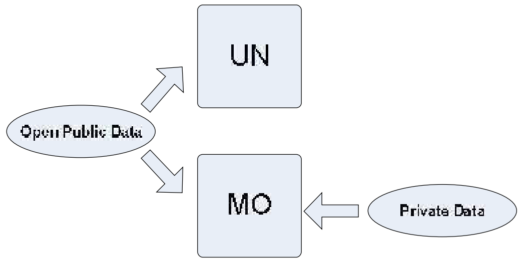


Figure 9. MO versus UN SDG framework control cycles

The publishment of the open data can increase the data supply, engage larger number of industrial and private users and allow business insight for government employees. Figure 10 shows an overview of cloud computing and Open Data relationship [16].

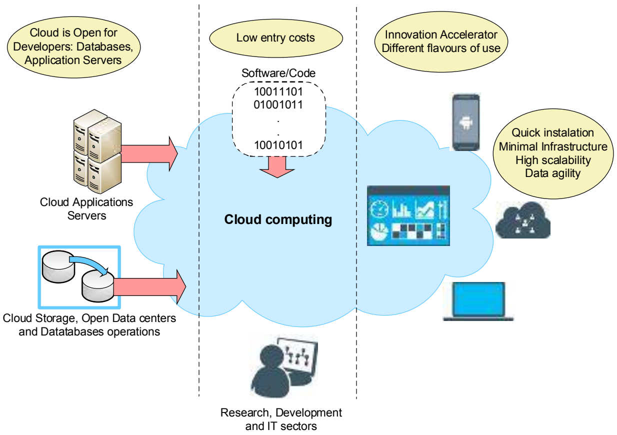


Figure 10. Cloud Computing and open data relationship

The open data is important part of overall government information architecture. It should be enabled to be meshed up with data from different sources (operational systems, external sources) in a way that is easy to be consumed by the citizens/companies with different access devices. Data in all formats should be also available for the use of the developers thus making them easier the process of developing new applications and services. The cloud computing platforms are ideal for encouraging and enabling the business value and business potential of open data. The government agencies are using this data and usually combine it with other data sources. Cloud enables new applications and services to be built on those datasets and enables data to be easily published by governments in very open way, independent of the used access device or software. Cloud allows high scalability for such use, as it can store huge amounts of data, process the millions of transactions and serve large number of users. Additionally, cloud computing infrastructure is driving down the cost of the development of the new applications and services, and is driving ability of access by different devices and software. Still, it is of great importance to consider the possibilities of integrating higher security and privacy concerns when dealing with the use of open data in cloud computing [17].

V. CONCLUSION

In this paper we first present two sustainability assessment frameworks, United Nations Sustainable Development Goals framework and our proprietary Multi-Objective model for assessing sustainability framework. We have explained the applied methodology and provided the qualitative frameworks comparison. It is clearly pointed out the necessity of having available the open data for both UN and MO frameworks. The general conclusion is that the research and development community still has to invest more time and resources into the development of the cloud computing applications that would help the efficient use of the data, improve services, and stimulate public and corporate innovations.

ACKNOWLEDGMENT

The work presented in paper has partially been funded by the Ministry of Education, Science and Technological Development of Republic of Serbia: V. Timcenko by grants TR-32037/TR32025, N. Zogović and B. Djordjevic by grant III-43002, and M. Jevtic by grant TR-32051.

REFERENCES

[1] Nikola Zogović, Miloš Jevtić, Valentina Timčenko, Borislav Đorđević, "A Multi-objective Assessment Framework for Cloud Computing," TELFOR2015 Serbia, Belgrade, pp. 978 - 981, 2015.

[2] The Global Open Data Initiative (GODI) Online: http://globalopendatainitiative.org/

[3] "Law on Free Access to Information of Public Importance," "Official Gazette of RS" No. 120/04, 54/07, 104/09 and 36/10 (in Serbian)

[4] "Indicators and a monitoring framework for the sustainable development goals – Launching a data revolution for the SDGs," A report to the SG of the UN by the LC of the SDSN, 2015

[5] Assembly, UN General. "Universal declaration of human rights," UN General Assembly, 1948

[6] Litoiu, Marin, et al. "A business driven cloud optimization architecture," Proc. of the ACM Symp.on Applied Computing. ACM, 2010

[7] Lambert, Sofie, et al. "Worldwide electricity consumption of communication networks," Optics express, 20.26, 2012

[8] "Indicators of Sustainable Development - Guidelines and Methodologies," UN – Economic & Social Affairs, 3rd ed, 2007

[9] Barbara Adams, "SDG Indicators and Data: Who collects? Who reports? Who benefits? ," Global Policy Watch, November 2015

[10] Mairi Dupar, "New research shows that tackling climate change is critical for achieving Sustainable Development Goals," The Climate and Development Knowledge Network (CDKN), 2015

[11] V. Pesic, J. Crnobrnja-Isailovic, Lj. Tomovic, "Principi Ekologije," University of Montenegro, 2010, (in Serbian).

[12] Fuchs, Victor R. "The service economy," NBER Books, 1968

[13] Z. Kennesey, "THE PRIMARY, SECONDARY, TERTIARY AND QUATERNARY SECTORS OF THE ECONOMY," Review of Income and Wealth, Vol. 33, Issue 4, pp. 359–385, December 1987.

[14] "International Standard Industrial Classification of All Economic Activities," Rev. 4, United Nations, New York, 2008

[15] Dimitris N. Chorafas, "Cloud Computing Strategies," CRC Press, 2011

[16] Mark Gayler, "Open Data, Open Innovation and The Cloud," A Conference on Open Strategies - Summit of NewThinking, Berlin Nov 2012.

[17] S. Pearson, A. Benameur, Privacy, "Security and Trust Issues Arising from Cloud Computing," 2nd IEEE Int. Conference on Cloud Computing Technology and Science, pp. 693-702, 2010.

Dataflow of Matrix Multiplication Algorithm through Distributed Hadoop Environment

Vladimir M. Ćirić, Filip S. Živanović, Natalija M. Stojanović, Emina I. Milovanović, Ivan Z. Milentijević
Faculty of Electronic Engineering, University of Niš, Serbia
{vladimir.ciric, filip.zivanovic, natalija.stojanovic, emina.milovanovic, ivan.milentijevic}@elfak.ni.ac.rs

Abstract — **Increasing of processors' frequencies and computational speed with components scaling is slowly reaching its saturation with current MOSFET technology. From today's perspective, the solution lies either in further scaling in nanotechnology, or in parallel and distributed processing. Parallel and distributed processing have always been used to speedup the execution further than the current technology had been enabling. However, in parallel and distributed processing, dependencies play a crucial role and should be analyzed carefully. The goal of this paper is the analysis of dataflow and parallelization capabilities of Hadoop, as one of the widely used distributed environment nowadays. The analysis is performed on the example of matrix multiplication algorithm. The dataflow is analyzed through evaluation of the execution timeline of Map and Reduce functions, while the parallelization capabilities are considered through the utilization of Hadoop's Map and Reduce tasks. The implementation results on 18-nodes cluster for various parameter sets are given.**

I. Introduction

The current projections by the International Technology Roadmap for Semiconductors (ITRS) say that the end of the road on MOSFET scaling will arrive sometime around 2018 with a 22nm process. From today's perspective, the solution for further scaling lies in nanotechnology [1]. However, parallel and distributed processing have always pushed the boundaries of computational speed, through history of computing, further than it had been enabled by the current chip fabrication technology. Two promising trends nowadays, which enable applications to deal with increasing computational and data loads, are cloud computing and MapReduce programming model [2].

Cloud computing provides transparent access to the large number of compute, storage and network resources, and provides high level of abstraction for data-intensive computing. There are several forms of cloud computing abstractions, regarding the service that is provided to users, including Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) [3].

MapReduce is currently popular PaaS programming model, which supports parallel computations on large infrastructures. Hadoop is MapReduce implementation, which has attracted a lot of attention from both industry and research. In a Hadoop job, Map and Reduce tasks coordinate to produce a solution to the input problem, exhibiting precedence constraints and synchronization delays that are characteristic of a pipeline communication between Maps (producers) and Reducers (consumers) [4]. In distributed processing in general, as well as in the MapReduce, the crucial problems that lie in front of designers, are data dependency and locality of the data. While data dependency influence is obvious, data locality has indirect influence on execution speed in distributed systems due to the communicational requirements. One of the roles of the Hadoop is to automatically or semi-automatically handle the data locality problem.

There are several models and simulators that can capture properties of MapReduce execution [2], [5]. The challenge to develop such models is that they must capture, with reasonable accuracy, the various sources of delays that a job experiences. In particular, besides the execution time, tasks belonging to a job may experience two types of delays: (1) queuing delays due to the contention at shared resources, and (2) synchronization delays due to the precedence constraints among tasks that cooperate in the same job [4].

The goal of this paper is the analysis of dataflow and parallelization capabilities of Hadoop. The analysis will be illustrated on the example of matrix multiplication algorithm in Hadoop, proposed in [6]. The dataflow will be analyzed through evaluation of the execution timeline of Map and Reduce functions, while the parallelization capabilities will be considered through the utilization of Hadoop's Map and Reduce tasks. The results of the implementation for various parameter sets in distributed Hadoop environment consisting of 18 computational nodes will be given.

The paper is organized as follows: Section 2 gives a brief overview of MapReduce programming model. In Section 3 dataflow of MapReduce phases for matrix multiplication algorithm is presented, and data dependencies are discussed. Section 4 is devoted to the analysis of the parallelization capabilities of the matrix multiplication algorithm, as well as to the implementation results, while in Section 5 the concluding remarks are given.

II. Background

The challenge that big companies are facing lately is overcoming the problems that appears with big amount of data. Google was the first that designed a new system for processing such data, in the form of a simple model for storing and analyzing data in heterogeneous systems that can contain many nodes. Open source implementation of

this system, called Hadoop, became an independent Apache project in 2008. Today, Hadoop is a core part of a lot of big companies, such as Yahoo, Facebook, LinkedIn, Twitter, etc [7].

The Hadoop cluster consists of collection of racks, each with 20-30 nodes, which are physically close and connected. The cluster consists of three types of nodes depending on their roles: (1) Client host - responsible for loading data into the cluster, forwarding MapReduce job that describes the way of processing data, and collecting the results of performed job at the end; (2) Master node - in charge of monitoring two key components of Hadoop: storage of big data, and parallel executions of computations; (3) Slave node - used for performing actual data storage and computing.

There are two main components of Hadoop system: (1) Distributed File System - Hadoop DFS (HDFS), used for big data storage in cluster; (2) MapReduce - framework used for computing big data stored in HDFS.

The HDFS lies as a layer above existing file system of every node in the cluster, and its blocks are used for storing input data in the form of the input splits (Figure 1). Large files can be split into a group of small parts called blocks, which have default size of 64MB. The size of these blocks is fixed, due to the simplification of indexing [9]. Usually, HDFS workflow consists of 4 parts: (1) transferring input data from Client host to HDFS, (2) processing data using MapReduce framework on the slave nodes, (3) storing results by Master node on HDFS, and (4) reading data by Client host from HDFS.

There are two transformations in MapReduce technique that can be applied many times on input files: Map transformation, which consists of M_T Mappers or Map tasks, and the Reduce transformation, which consists of R_T Reducers or Reduce tasks. The parameters M_T and R_T are specified in system configuration of Hadoop, R_T explicitly, and M_T implicitly through specification of the blocksize. In the Map transformation, each Map task processes one small part of the input file and forwards the results to the Reduce tasks. After that, in the Reduce transformation, Reduce tasks gather the intermediate results of Map tasks and combine them to get the output, i.e. the final result, as shown in Figure 1.

The Mappers, during theirs execution, executes M_F Map functions to perform required computations. One Map function transforms input data, according to input (key_{in}, $value_{in}$) pairs, into the set of intermediate (key_{im}, $value_{im}$) pairs (Figure 1). Let us note that the number of executed Map functions M_F is equal to the number of different keys key_{in}, and that this number doesn't need to be equal to the configured number of Map tasks M_T.

In the phase between Map and Reduce, called Shuffle and Sort, all intermediate data with the same key key_{im} are grouped and passed to the same Reduce function (Figure 1). The number of executed Reduce functions R_F is equal to the number of different keys key_{im}. It doesn't need to be equal to the configured number of Reduce tasks R_T. In the end, all data from the Reduce tasks are written into separate output.

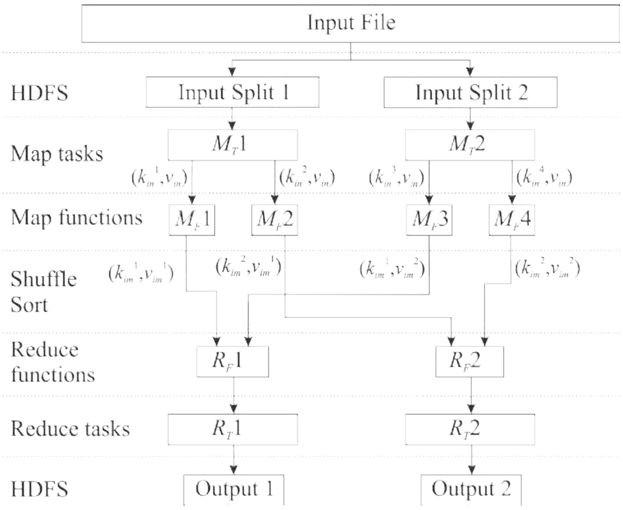


Figure 1. MapReduce data and process flow

MapReduce inherits parallelism, fault tolerance, data distribution and load balancing from Hadoop system itself [8]. As mentioned before, it consists of two main phases, namely, Map and Reduce, each one implemented by multiple tasks (M_T+R_T) running on multiple nodes (N) [4].

Figure 2 shows a simple example of a timeline representing the execution of a Hadoop job composed of M_T=2 Mappers and R_T=1 Reducer, running on N=3 nodes. The number of Map functions in algorithm shown in Figure 2 is M_F=4, and the number of Reduce functions is R_F=4. There is one additional Reducer R_M that collects outputs from all Reduce functions. The notation used for particular Map functions within Map tasks is M_F^i, where i represents the number of the function. The order at which the Reduce functions R_F^i, i=1,2,3,4, begin their execution is defined by the order at which the Map functions M_F^i, i=1,2,3,4, finish theirs. Precisely, Reduce function R_F^i should start as soon as Map function M_F^i finishes and the node that executes Reduce task is idle. At the end, the merge task (R_M) can start only after all Reduce tasks finish.

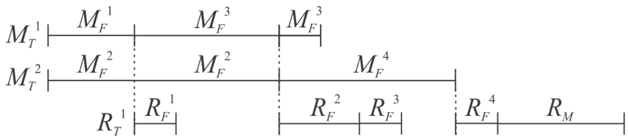


Figure 2. Execution timeline of Hadoop job

In Figure 2, Map tasks are denoted as M_T^i, where i denotes the number of particular Mapper. As shown in Figure 2, two Map tasks, M_T^1 and M_T^2, start execution immediately at the beginning of the job execution, on separate nodes, while the Reduce task (R_T^1) is blocked and, therefore, waits. As soon as the first Map function (M_F^1) finishes, the first Reduce function (R_F^1) can begin its execution. Also, another Map function M_F^3 is assigned to the task M_T^1 that was executing M_F^1. This point in time is shown in Figure 2 with a dotted vertical line. It also represents a synchronization point when the set of functions executing in parallel changes. From this point in time, M_F^3, M_F^2 and R_F^1 are executing. Since M_F^3 starts

executing only after M_F^1 finishes, there is a serial precedence between them.

The execution of a Hadoop job represents a set of synchronization points, and each one of them delimits the parallel execution of different sets of functions. In order to maximize performance due to the synchronization characteristic of Hadoop system, and to utilize parallelization capabilities of Hadoop, the number of Map tasks, Map functions, Reduce tasks and Reduce functions should be carefully planned in accordance to the available number of computational nodes.

III. Dataflow of Matrix Multiplication in Hadoop Environment

We will illustrate parallelization capabilities of Hadoop on the example of matrix multiplication algorithm proposed in [6]. Let us briefly discuss the dataflow timeline of the algorithm from [6], and allocation of computations onto Map and Reduce functions M_F and R_F.
Let A and B be matrices of order IxK and KxJ, respectively, and let C be their product as

$$c_{i,j} = \sum_{k=1}^{K} a_{i,k} \cdot b_{k,j} \,. \qquad (1)$$

According to the matrix multiplication algorithm proposed in [6], the value of the key key_{in} that distinguishes the Map functions is common index k from (1). In this case, the total number of Map functions M_F, that are executed by Map tasks, is equal to M_F=K, i.e. to the number of columns in matrix A and the number of rows in matrix B. Map function M_k obtains all partial products $c_{i,j}^k = a_{i,k} \cdot b_{k,j}$, where i=1,2,...,I, and j=1,2,...,J.
The example of the multiplication of matrices A and B of order 2x3 and 3x4, respectively, is shown in Figure 3 and Figure 4.

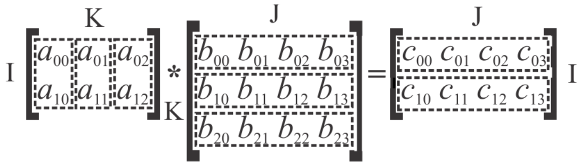


Figure 4. The example of Matrix multiplication $C_{2,4}=A_{2,3}\cdot B_{3,4}$

According to (1), all elements of the first column of the matrix A, i.e. a_{00} and a_{10} in Figure 4, are needed for the multiplication with all elements of the first row of the matrix B, b_{00}, b_{01}, b_{02} and b_{03}. The same holds for other columns of the matrix A, and the rows of the matrix B, as it is shown with dashed lines in Figure 4.
From the above, every Map function M_k will get k-th column of matrix A and k-th row of matrix B, as shown in Figure 3. Within each Map function M_k, every element $a_{i,k}$, i=1,2,...,I, of matrix A will be multiplied with every element $b_{k,j}$, j=1,2,...,J, of matrix B, producing partial products $c_{i,j}^k = a_{i,k} \cdot b_{k,j}$. For example, within Map function M_1, the element a_{00} will be multiplied by b_{00}, producing partial result c_{00}^0, as denoted with gray circles and arrows in Figure 3. The same stands for all other elements from M_1. As a result, Mapper M_1 will produce first intermediate results for all elements in the resulting matrix C. On the other hand, while the Mappers are responsible for multiplying, Reducers are responsible for summarizing intermediate results $c_{i,j}^k$ for every element $c_{i,j}$ in the resulting matrix C. In the example given in Figure 3, c_{00}^0, c_{00}^1 and c_{00}^2 are summarized into c_{00}.
According to the computations allocation of the particular matrix multiplication algorithm, there is no data dependency between Map functions, and all Map functions can be executed in parallel. The same holds for the Reduce functions.
On the other hand, each Reduce function can start its execution only when all Map functions finish their computations. Therefore, in this algorithm, there is no overlapping between Map and Reduce phase (Figure 3).

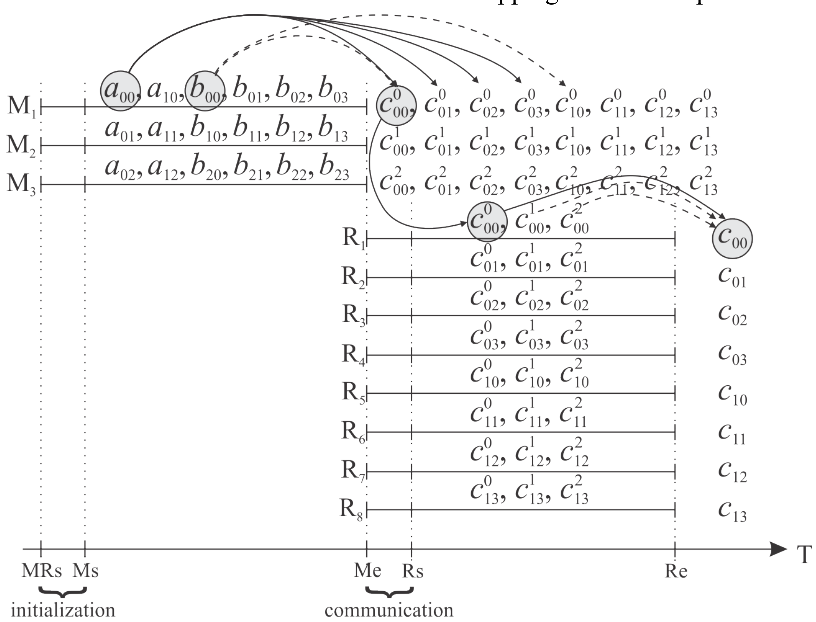


Figure 3. The dataflow timeline of the matrix multiplication algorithm in the MapReduce distributed environment

IV. Implementation Results

In the previous section it was shown how the partial computations are allocated to Map and Reduce functions. As mentioned before, the numbers of Map and Reduce functions are parameters of the algorithm, while Map and Reduce tasks are configured according to the capabilities of the cluster.

For this particular matrix multiplication algorithm, all Map functions can start in parallel at the point denoted with Ms on the T axis in Figure 3. Ideally, the number of nodes N, and the number of the Map tasks M_T should be equal to the number of required Map functions M_F. However, as the number of Map functions M_F is equal to the dimension K of matrices A and B, this number will always in practice overcome the number of available nodes N in the cluster. Therefore, one Map task will execute many Map functions. The same holds for the Reduce functions. All Reduce functions can start in parallel at the point of time denoted as Rs in Figure 3 and last until Re. The number of available Reduce tasks R_T will limit the parallelization in this case, as well.

The algorithm is implemented and executed on the Hadoop cluster consisting of N=18 nodes. The characteristics of nodes are the following: Intel(R) Core(TM)2Duo, CPU E4600@2.40GHz, RAM: 1GB.

We executed the algorithm for two scenarios: (1) fixed number of Reduce tasks, equal to the number of nodes (R_T=N=18), and various number of Map tasks ($1 \le M_T \le 2 \cdot N$=36), and (2) fixed number of Map tasks, equal to the number of nodes (M_T=N=18), and various number of Reduce tasks ($1 \le R_T \le 2 \cdot N$=36). Let us note that in both cases square matrices of order 1.500x1.500 were considered. Thus, the number of Map functions is M_F=1.500, and the number of Reduce functions is R_F=2.250.000.

The obtained results for the MapReduce algorithm for described scenarios are graphically presented in Figure 5. Let us note that for each result shown in Figure 5 there are M_T+R_T tasks configured. Thus, the minimum number of tasks for the first scenario is 1+18=19, and the maximum is 36+18=54, which are executed on 2·18=36 cores. From the results given in Figure 5 it can be seen that the parallelism is underutilized if the total number of tasks is less then 36 (value M/R=18 in Figure 5), due to the fact that there are unused cores. If the number of tasks is greater then the number of cores (Figure 5), there is additional overhead for synchronization that slows down the execution. Due to the characteristic of the matrix multiplication algorithm, the optimal cluster utilization is when the total number of tasks is equal to the number of cores (Figure 5).

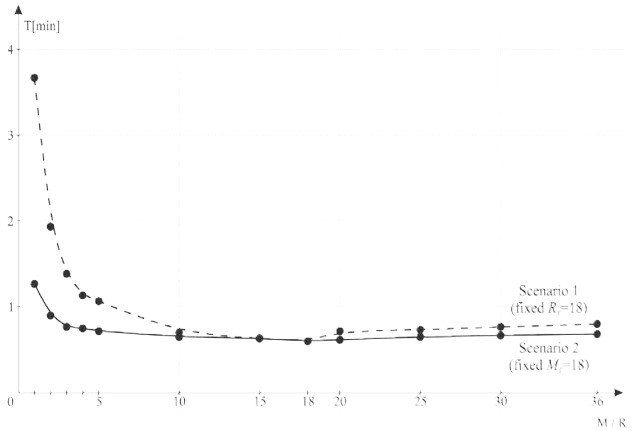


Figure 5. Execution time of MapReduce algorithm for matrix multiplication

V. Conclusion

In this paper the analysis of dataflow and parallelization capabilities of Hadoop is illustrated on the example of matrix multiplication algorithm. The dataflow is analyzed through evaluation of the execution timeline of Map and Reduce functions, while the parallelization capabilities are considered through the utilization of Hadoop's Map and Reduce tasks. The results of the implementation for various parameter sets in distributed Hadoop environment consisting of 18 computational nodes are given. It is shown that the optimal cluster utilization is when the total number of tasks is equal to the number of cores.

VI. Acknowledgment

The research was supported in part by the Serbian Ministry of Education, Science and Technological Development (Project TR32012).

VII. References

[1] Ciric, Vladimir, et al. "Tropical algebra based framework for error propagation analysis in systolic arrays." Applied Mathematics and Computation 225 (2013): 512-525.

[2] Wang, Guanying, et al. "Using realistic simulation for performance analysis of MapReduce setups." Proceedings of the 1st ACM workshop on Large-Scale system and application performance. ACM, 2009.

[3] Rakumar Buyya, James Broberg, Andrzej Goscinski, "Cloud Computing: Principles and Paradigms", Willey, 2011.

[4] Vianna, Emanuel, et al. "Analytical performance models for MapReduce workloads." International Journal of Parallel Programming 41.4 (2013): 495-525.

[5] Ganapathi, Archana. "Predicting and optimizing system utilization and performance via statistical machine learning." (2009).

[6] Živanović S. Filip, Ćirić M. Vladimir, Stojanović M. Natalija, Milentijević Z. Ivan. "Optimized One Iteration MapReduce Algorithm for Matrix Multiplication". IcETRAN, 2015.

[7] Lam, Chuck. Hadoop in action. Manning Publications Co., 2010.

[8] Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: a flexible data processing tool." Communications of the ACM 53.1 (2010): 72-77.

[9] White, Tom. "Hadoop: The definitive guide." O'Reilly Media, Inc., 2012.

Open Government Data Initiative : AP2A Methodology

Milan Latinović*, Srđan Rajčević*, Zora Konjović**
* Agency for Information Society of Republic of Srpska, Banja Luka, Republic of Srpska, Bosnia and Herzegovina
** University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Republic of Serbia
milan.latinovic@live.com, srajcevic@gmail.com, konjovic.zora@gmail.com

ABSTRACT - This paper proposes new and innovative methodology called "Action Plan To Applications", a.k.a. AP2A methodology. As indicated inside methodology name, it's scope/roadmap on how to handle OGD from first phase of action plan, trough data gathering, publishing, molding, all the way to first useful applications based on that data. This methodology keeps in mind lack of infrastructure and all database challenges that could affect Balkan countries, and aims to create roadmap on how to accomplish two very important tasks. First task is ofcourse implementation of OGD concept, and second task is building up informational infrastructure (databases, procedures, process descriptions etc.) which is usually bottleneck for every development initiatives. General idea is actually simple, to do these two tasks parallel, within defined process but still flexible enough to allow modifications from actor to actor, institution to institution, data owner to data owner, and ofcourse government to government.

Keywords: *open data, electronic government, methodology, semantics, context*

I. INTRODUCTION

In order to successfully implement OGD in countries with lower level of development (such as Balkan countries, compared to UK, Germany and Estonia ofcourse), there is an urgent need of new, customized, specialized methodology for OGD implementation. This can be done only by learning current methodologies and concepts (UK and Estonia in particular in this paper), and molding it into scope of Balkan countries.

As already mentioned, many public organizations collect, produce, refine and archive a very broad range of different types of data in order to perform their tasks. The large quantity and the fact that this data is centralized and collected by governments make it particularly significant as a resource for increased public transparency.

There is huge list of positive aspects of data openness, as follows:

• Increased data transparency provides basis for citizen participation in decision making and collaboration to create a new citizen oriented services.

• Data openness is expected to improve decision making of governments, private sector and individuals.

• Public is expected to use government data to improve quality of life, for example, through accessing specific databases via their mobile devices, to inform better before they make certain choices, etc.

• Also, OGD is defined as very valuable resource for economic prosperity, new forms of businesses and social participation and innovation.

As described in [1] there are two important society movements that are campaigning for greater openness of information, documents and datasets held by public bodies. The first is "Right to Information" movement and the second is "Open Government Data" movement / initiative.

Right to Information movement can be explained trough Right to Information Act (RTI) which is an act of the Parliament of India related to rights of citizens to ask for a data and get response to their query. This is closely related to existence of some form of Low on Freedom To Access Information. Existence of this law or equivalent seems to be one of the prerequisites for any kind of Open Data initiatives.

OGD movement presents free usage, re-usage and redistribution of data produced or commissioned by government or government controlled entities. This is closely related to government transparency, releasing social and commercial value, participatory governance. As stated on Open Government Data main portal it's about making a full "read-write society, not just about knowing what is happening in the process of governance nut being able to contribute to it.

Having initiative for data openness presupposes existence of digital data in first place. This means existence of valid databases with data which has new value for citizens or consumers. Sometimes this is called "Repository Registry" or "Registry of Repositories" beautifully described in [2]. This problematics deals with registry characteristics, metadata issues, data gathering practices and workflows, issues related to registry updates and future registries.

After existence of digital data is verified, there is completely different issue about deciding if this data is applicable to be open data or not. Having that in mind, owners of data have tough decision to make, regarding which data is eligible to be publically presented and in what form. There is interesting Open Data Consultancy Report made for Scottish Government [3] which aims to resolve this issue and present examples of Government Open Data repositories. Also, remarkable resource of real world examples of Open Data repositories can be found at "Research project about openness of public data in EU local administration" [4].

After polishing and publishing data repositories, a.k.a. data sets, citizens should use this data, either in raw form or by consuming it through applications built upon open data. These applications should respect licenses related to open data defined by data owner. Yes, it's important to point out

that although open data is free for use, this usage can be defined by specific open data licence, such as Creative Commons CCZero (CCO), Open Data Commons Public Domain Dedication and Licence (PDDL), Creative Commons Attribution 4.0 (CC-BY-4.0), Open Data Commons Attribution License (ODC-BY) or other described in [5] and [6]. Also, it's important to note that process of reading of defining licence for that matter, should follow definitions described in RFC 2119, a.k.a. "Key words for use in RFCs to Indicate Requirement Levels", [7].

Governments make Action Plans, but if these plans are just generically copy pasted from other countries without understanding of specific system and infrastructure, than all of the mentioned steps will not happen. Main goal of this work is to write roadmap, framework or even a methodology which describes how to implement functional OGD concept specialized for Balkan countries. This methodology would describe process from Creation of Action Plan to Creation of Application for end user, so we'll call it "Action Plan to Application Methodology" or AP2A methodology. It's understandable that once defined in this paper, this methodology should be tested in real government systems, evaluated and optimized.

II. ANALYSIS OF OGD IMPLEMENTATIONS

This section analyses several e-government systems which includes OGD implementations. Each of these governments are considered to be advanced in compare to Balkan countries. That's why it's important to review their efforts, activities to realize how they spent their time and other resources. Only after finding out more details about these systems we can compare their use cases with our future use cases (use cases and methodologies aimed on Balkan countries).

This section will consider three different countries and their OGD efforts:

• The Netherlands - huge OGD efforts and lots of publically available materials and related services. Basis for this part of research will be Open Government Partnership Self-Assessment Report, The Netherlands 2014 [8]

• Estonia - included into this research as country with state of the art e-government implementations, indicated in e-government report for year 2015 described in [9]

• United Kingdom - will be analyzed for their action plains presented on their Open Government portal [10]. Key idea is to examine 2011-13 UK Action Plan (I), 2013-15 UK Action Plan (II) and current version on 2016-18 UK Action Plan (III). Together with "Implementing an OGP Action Plan" and "Developing an OGP Action Plan" guidelines

A. The Netherlands and OGP

I had a privilege to listen to lectures from Mr Tom Kunzler, project manager at Open State Foundation (Amsterdam Netherlands). This analysis is based on his presentation, combined with examination of [8] with idea to locate interesting initiatives and services related to Dutch Government.

Open State Foundation promotes digital transparency by unlocking open data and stimulates the development of innovative and creative applications.

The Open Government Partnership (OGP) is a multilateral initiative that aims to secure concrete commitments from governments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance. In the spirit of multi-stakeholder collaboration, OGP is overseen by a Steering Committee including representatives of governments and civil society organizations. [11]

To become a member of OGP, participating countries must:

• Endorse a high-level Open Government Declaration

• Deliver a country action plan developed with public consultation, and commit to independent reporting on their progress going forward.

The Open Government Partnership formally launched on September 20, 2011, when the 8 founding governments (Brazil, Indonesia, Mexico, Norway, the Philippines, South Africa, the United Kingdom and the United States) endorsed the Open Government Declaration. Currently, OGP has 67 state members.

After analyzing available materials these are conclusions:

1. Open data is defined as information gathered with public funds. This data should be accessible for everyone without copyright restrictions or any kind of payment for this data.

2. Open data should be presented in an open data standard (without commercial standards such as .xsl, but with usage of .csv, .json or .xml data). Of course this is not mandatory, but it is preferred (according to the national open data portal, data.overheid.nl).

3. it’s preferable that data is machine readable, but it's not mandatory.

Also, there is significant consideration about quality of data in terms of separating public information from open data. It's clearly stated that public information is indeed publically presented, but it doesn't assure required quality to be presented as open data.

This issue has been addressed by Tim Berners Lee, the inventor of the Web and Linked Data initiator, who suggested 5-star deployment scheme, as presented within [12]. This scheme proposes five levels of validating quality of specific open data / open database / open dataset etc.

Image 2.1.1. - 5-star deployment scheme by T. B. Lee [12]

1-star data defines publically available data on the web with an open licence (i.e. PDF file)

2-star data defines data which is available in a structured form (i.e. XLS)

3-star data defines data which is available in a structured form with usage of open standards (i.e. CSV)

4-star data defines usage of Uniform Resource Identifiers (URIs) to identify data (i.e. RDF)

5-star data defines linking your data (defined in previous levels) with someone else's data (i.e. with Wiki data)

Related to examples of data that can be made opened and applications that could be created with this data, resources point to several specific aspects of usage:

1. Public transportation data - Making public transportation data open can lead to creation of variety of applications widely used in everyday life, tourism etc. and also makes sure that there will be some healthy competition with the 'official' public transport apps. And that apps will be better because consumers can choose and they have to compete with each other.

2. Open Decision Making on a local level - The municipalities Amstelveen, Den Helder, Heerde, Old IJsselstreek and Utrecht are the first five municipalities in the Netherlands that release meetings minutes, agenda's and other relevant documents as open data. This is the outcome of a pilot done by Open State Foundation with the Ministry of Interior, the Association of Municipalities and the clerks of these five municipalities. This is an important step to make local politics transparent and accessible to citizens, journalists and councilors. [13]

3. Openspending - All financial information of Dutch municipalities and provinces available as open data at www.openspending.nl. It's possible to compare and benchmark budgets and realization and aggregate data per inhabitants, households and surface of government. Web site is used by councilors, citizens, journalists, consultants and regional governments.

B. *Estonia e-Government*

Based upon the report [9] it's possible to reconstruct road map that Estonia accomplished since year 2001 till now. In last fifteen years Estonia positioned as one of the fastest growing e-government and research environments, which makes it interesting for this analysis.

The development of e-government and information society in Estonia can be summarized as follows, taking the key points of development and key initiatives:

• 2001 - Implementation of X-Road system (est. "X-tee") which represents middle layer for exchanging data between different key points within public administration. These activities are followed by creation of eDemocracy portal which encourages citizens to involve in public debates and decision making.

• 2002 - Implementation of national ID cards, which represent digital identity of citizen, and can be used for business, public administration and private communication. (Comment: Below, will be obvious that this project is the basis for all further activities)

• 2003 - Finland and Estonia sign agreement on harmonizing communications using digital certificates , the project " OpenXAdES " which is an open initiative that promotes "universal digital signature". Also, the same year was created portal www.eesti.ee that represents a "one- stop- shop" , i.e. portal of public services administration Estonia.

• 2004 - Adoption of the new Information Society Policies.

• 2005 - Adoption of the Policy Information Security. Likewise , the same year Estonia established the service for voting via the Internet www.valimised.ee , where citizens can vote using ID card (ID Project from 2002)

• 2006 - The introduction of services for future students to apply to universities online through the portal www.sais.ee . Also, this year introduced a Department for the Fight against security incidents related to Internet space Estonia (a.k.a. Computer Emergency Response Team - CERT). Also, this year Estonia presented a framework for interoperability (a.k.a. Estonian IT Interoperability Framework) , version 2.0.

• 2007 - The establishment of electronic service for taxes and subsidies for individuals and legal entities. That same year, Estonia created the portal Osalusveeb www.osale.ee , which allows public debate on draft legislation related to e - government of Estonia. Finally, introduced a web portal for online registration of companies www.ettevotjaportaal.rik.ee which allows registration of the new company within a few hours, with the use of ID cards (Project from 2002). Also, this year introduced the possibility for citizens through e - government portals require the card for online voting (eVoter card), after which citizens no longer receive ballots by mail.

• 2008 - Introducing Cyber Security Strategy. Also introduced is a service for issuing parking permits portal www.parkimine.ee/en also using ID cards from 2002 project. That same year , introduced the service for a refunds www.emta.ee

• 2009 - On 1 October 2009, the Estonian Informatics Centre - EIC opened its Department for Critical Information Infrastructure Protection (CIIP).CIIP aims to create and run the defense system for Estonia's critical information infrastructure. In August 2009, Estonia's largest ICT companies establish the Estonian Broadband Development Foundation with the objective that the basic infrastructure of the new generation network in Estonian rural areas is developed by the end of 2015.

• 2010 - On 1 July 2010, Estonia switches to digital-TV. The Estonian Government approved on 1 April 2010 an amendment bill to the Electronic Communications Act and the Information Society Services Act regulating the use of individuals' electronic contact data for sending out commercial emails. Implementation of 'Diara' also happened. It's open source application that allows public administrations to use the Internet in order to organize polls, referenda, petitions, public inquiries as well as to record electronic votes using electronic identity cards.

• 2011 - A new version of the State Portal 'eesti.ee' goes live in November 2011 based on user involvement and their feedback. Tallinn, the capital of Estonia, is awarded with the

European Public Sector Award 2011 for citizen eServices. There were lots of other activities within this year but most of them were related to evaluations, conferences and awards. Seems like a year of awards and achievements for government of Estonia.

• 2012 - Preparation of new Information Society Strategy 2020. The greatest benefits of this strategy include: good Internet accessibility, the use of services to support the development of state information and security for citizens and businesses, as well as the development of electronic services.

• 2013 - This year Estonia approves the Green Paper on the Organization of Public Services in Estonia, to establish a definition of "public service", identify problems faced by citizens and enterprises in usage of e-government services. Also, prime ministers of Estonia and Finland finalize first digitally signed intergovernmental agreement related to joint development of e-government services, linked to data exchange layer (known as X-Road).

• 2014 - This year seems to be focused on two agendas "Free and secure internet for all" and "Interoperability and intergovernmental relations". Also, eHealth Task Force is set up at the leadership of the Government Office with a goal to develop a strategic development plan for Estonian eHealth until 2020. Also, Estonia starts implementing X-Road like solutions in other countries (outsourcing knowledge and services), such as agreement with Namibia.

This report has only some predicting data related to this year 2015, so it will not be included into this analysis.

After analyzing available materials these are conclusions:

1. Initially, Estonian government focused on two important aspects of information society "Interoperability aspect" and "eDemocracy aspect". It's interesting to realize that Estonia didn't base Interoperability system upon some large scale concept that covers all databases and ministries. Instead they decided to locate several most important databases, to interconnect them, and within next years to build upon that. So, in a terms of data interoperability and ontology concepts, Estonia used "Bottom To Top model" in its cleanest form. This is very interesting since almost every OGP or OpenData or eGovernment initiative proposes already completed solutions and frameworks which are (by nature) based upon "Top To Bottom model" which isn't what most successful countries used.

2. Seems like Estonian primary focus wasn't on Open Data but on Open Services, meaning that most of Estonian initiatives are focused on producing new service (e-democracy, e-voting, e-academy) and only after significant amount of services and high level of interoperability Open Data became interesting in it's pure form.

3. Since most of Estonian services had lots of future versions, revisions and citizen involvement, seems like Estonian "concept" of e-government looks like this:

a. LOCATE ISSUE LARGE ENOUGH TO BE ADDRESSED WITH SERVICE

b. CREATE FIRST VERSION OF ELECTRONIC SERVICE

c. LINK NEW SERVICE TO ID CARD (PROJECT FROM 2002)

d. PROTECT SERVICE WITH ADEQUATE LEGISLATION

e. GET CITIZENS FEEDBACK ON SERVICE AND LEGISLATION

f. CREATE NEW IMPROVED VERSION OF SERVICE

g. OFFER SERVICE TO OTHER COUNTRIES (KNOWLEDGE INDUSTRY)

4. Estonia makes great effort on involving citizens into public debates (legislative and decision making). It's important to realize that Estonian services aren't based on anonymity, but on proven identity of each individual / citizen, which is realized through ID CARD project, and everything is interconnected with X-ROAD.

5. Baseline for all projects is Bottom-To-Top interoperability (created on several most important databases) connected with Digital Identity Management (probably PKI system) a.k.a. National Identity Provider.

C. United Kingdom's Action Plans

United Kingdom's Action Plan I (2011-13) is initial strategy document which follows the idea “Making Open Data Real” and it focuses on Improving Public Services and More Effectively Managing Public Resources. Most interesting part of this Action Plan (related to this research of course) is Annex A, which lists all data sets planned for a release:

• Healthcare related data sets

Data on comparative clinical outcomes of GP practices in England

Prescribing data by GP practice

Complaints data by NHS hospital so that patients can see what issues have affected others and take better decisions about which hospital suits them

Clinical audit data, detailing the performance of publicly funded clinical teams in treating key healthcare conditions

Data on staff satisfaction and engagement by NHS provider (for example by hospital and mental health trust)

• Education based data sets

Data on the quality of post-graduate medical education

Data enabling parents to see how effective their school is at teaching high

Opening up access to anonymized data from the National Pupil Database to help parents and pupils to monitor the performance of their schools in depth

Bringing together for the first time school spending data, school performance data, pupil cohort data

Data on attainment of students eligible for pupil premium

Data on apprenticeships paid for by HM Government

• Crime related data sets

Data on performance of probation services and prisons including re-offending rates by offender and institution

Police.uk, will provide the public with information on what happens next for crime occurring on their streets

• Transport related data

Data on current and future roadworks on the Strategic Road Network

All remaining government-owned free datasets from Transport Direct, including cycle route data and the national car park database

Real time data on the Strategic Road Network including incidents

Office of Rail Regulation to increase the amount of data published relating to service performance and complaints

Rail timetable information to be published weekly

Next Action Plan II (2013-15) is interesting because it reflected to implementation of previous Action Plan I. It's clearly stated the importance of establishment of Public Sector Transparency Board, which would work alongside government on Open Data activities including:

• Creation of publication "Open Data Principles"

• Establishment of gov.uk portal in order to channel and classify open data for ease of usage

• Creation of e-petitions portal - general idea of E-petitions portal is that the Government is responsible for it and if any petition gets at least 100.000 signatures, it will be eligible for debate In House of Commons.

• Independent review of the impact of Transparency on privacy, in a form of review tool

After analyzing available materials these are conclusions:

1. UK Action Plans are focused to specific sets of data (even Ministry specific), mostly Healthcare, Education and Transport. These seems like a good datasets for initial Open Government initiative.

2. UK Open Government initiative strongly focuses on civilian sector (Public Sector Transparency Board) which works alongside government.

3. Most interesting services related to UK use case are: 1) transport services and 2) public petitions portal.

4. There is no clear explanation how Digital Identity is maintained within UK. Seems like they don't address this issue with their Action Plains, meaning that they probably consider this prerequisite.

III. AP2A Methodology

After analysis in previous section, it's clear that methodologies and action plans from advanced governments can't be directly applied to Balkan countries, and that some infrastructure preparations are in order before building stable Open data Applications. Idea is to create methodology which would ensure creation of successful action plan, and implementation of this action plan up to Applications level.

Some of key infrastructure issues that needs to be resolved are:

1. Determining data repositories and owners

2. Defining current state of data repositories

3. Defining set of rules for making data open or not, and defining their current state

4. Creation of Action Plan for Data Repositories preparation and publishing

5. Infinity plan - handling new requests for data repositories and requests by data owners

D. Determining data repositories and owners

In government environment every database is most likely defined by some kind of legislative. If you take example of Republic of Serbia legislative or Bosnia and Herzegovina legislative with entities legislative of Republic of Srpska and Federation of Bosnia and Herzegovina, it's visible that most databases are defined by law, or other sub acts related to specific law.

We can conclude that every database needs to be defined by law. It implies that, if database exist, one or more laws define this database, how it is created, who is data owner, who maintains database and for which purposes. Finding these records is actually Phase 1 of A2AP methodology.

Best way to determine data repositories and owners would be reading trough all legislative for a key phrases such as "data", "repository", "data entry", etc. Also, we need to keep in mind that each legislative is handled by specific Ministry and that Ministry should be aware of what that database represents and where it is implemented.

For example, within a Law on Electronic Signature of Republic of Srpska, two Regulations are introduces. Firstly, there is Regulations on records of certificate entities and second is Regulations on qualified certification entities. Both of these define databases of these entities, which is handled by Ministry of Science and Technology of Republic of Srpska. So, reading trough these regulations points out to Ministry, and they are able to provide additional information about these databases/registries.

Now, reading trough all legislative could be very challenging job, where simple electronic service can be quite useful. Most of the countries are in process or already digitalized their Official Gazette's, with full text of all active legislative. If we imagine automated software that simply reads through these documents for specific set of keywords. These keywords points out to possible existence of database. As a result, software would provide array of potential databases described with example below (JSON format used in example):

```
{"PotentialDatabases":[
   {"Id":"10045", "Article":"25",  "TriggerList":
"data,database,registry", "Probability":"70"},
   {"Id":"10492", "Article":"1", "TriggerList": "registry,
entry", "Probability":"50"},
   {"Id":"20424", "Article":"80", "TriggerList":"data",
"Probability":"40"}
]}
```

After receiving result from service, administrator/user would manually read through selected articles and create list of databases linked with Regulations/Ministries who are responsible for them. End result of this activity would be

presented as a list of databases sorted by owners. This would enable to proceed to next step of methodology.

Creation of described service is a challenge for itself, because it's idea to hit most accurate results with specific set of rules. This can be implemented by some selective logics or "IF- ELSE IF - ELSE " oriented systems, or even with some neural network. This neurological network would use supervised learning algorithms to "learn" to recognize database from legislative.

IV. CURRENT STATE OF DATA REPOSITORIES

After successful determination of data repositories, good system should find out more about these repositories and their owners. Acquiring set of metadata that describes current state of databases / data repositories is vital step in AP2A methodology.

Approach is quite straightforward, create a set of unified queries that describe technical and non-technical details of data repositories and ask potential owner. Get the answers, archive them, check if these answers generated any new owners and/or data repositories and repeat the process.

Set of questions that should be asked in any iteration could look like this:

• IS DATA REPOSITORY IMPLEMENTED?

○ IF (TRUE) CONTINUE WITH QUESTIONS

§ IS DATA REPOSITORY IN DIGITAL FORM?

□ IF (TRUE) CONTINUE

® ASK TECHNICAL SET OF QUESTIONS

◊ FORM OF DATABASE (FILE SYSTEM, RELATIONAL, OODB, etc.)

◊ DATABASE ACCESS (WEB SERVICE, VPN, RESTRICTED, etc.)

◊ TECHNICAL DOCUMENTATION ON DATABASE

◊ MODULARITY OF DATABASE

◊ other important technical questions

□ ELSE IF (MIXED FORM) CONTINUE

® ASK ABOUT DATES WHEN REPOSITORY WILL BE FULLY DIGITALIZED

® ASK ABOUT METHODOLOGIES THAT WILL DIGITALIZE DATA

® ASK ABOUT DATA OWNER

□ ELSE

® ASK IF REPOSITORY IS PLANNED TO BE DIGITALIZED

○ ELSE END;

Providing answers to presented set of questions (for each database defined in Phase 1) can be viewed as Phase 2 of A2AP methodology. It's important to understand that this phase represents only current state of data repositories, and that this state doesn't recognize time.

So, to make it completely clear, main goal of Phase 2 of A2AP methodology is information gathering. This includes gathering information about databases from its potential owners, through a set of questions unified for all datasets. This paper provides example set of question (in order to present logic of this phase). Creating real time set of question can even be considered as a creation of sub methodology and challenge itself.

E. DTL (Databases Time Lines) and LOI (Level of Implementation)

Upon Phase 2 completion, we have databases, their owners and their descriptions, but we don't have two important information. We don't know databases time line, we're aware that info is related to present database but we don't know database chronology. This is very important and it will be explained in detail.

Also, we're not aware of Level of Implementation of database, thus we have all provided data but we're yet to categorize it. So, we have two challenges, DTL recreation and LOI defining.

Second challenge, LOI defining can easily be solved. As we mentioned in Section 2.1., this issue has been addressed by Tim Berners Lee, the inventor of the Web and Linked Data initiator, who suggested 5-star deployment scheme, as presented within [12]. We can use his scheme to validate inputs from Phase 2 for each database, and simply categorize each database with 5-star deployment scheme. Ofcourse, depending of number of stars some database got, there is new opportunity for this data improvement, but this will be considered in next phase. For now, this resolves LOI challenge.

Let's define DTL challenge with couple of statements:

• Each database is created according to some legislative (law, regulation, etc.) which should clearly describe structure of that database (at least in Use Case form, maybe even in technical).

• Laws and regulations change over time (Use Cases change, requirements change, etc.)

• We can conclude that if laws and regulations change databases change too.

• When database change (updates, gets new version) we still have old data inside (updated in some cases).

• When user asks for data from database he is concerned not only about quantity of data, but about quality too.

• There is issue on definition of quality of data:

○ Different users can have different ideas of what quality of data is.

○ Different versions of law and regulation can define different quality standards.

○ Different Use Cases for data define different quality expectations.

So, we can define that DTL challenge is actually a quality challenge, where the quality of data is challenged by three aspects: Use Case for which is this data required, time when this data is gathered (there can be different timeframes if data is gathered within longer period of time), and compliance of that data with legislative (not with current legislative, but with legislative which was active at the time when data was

gathered). This is very complex issue, and it can be applied to any form of database (Medical Records, Tax Administration, Land Registry, etc.)

Further considerations of DTL challenge are out of scope of this paper. Idea is to point out importance of time frames in databases and its relations to legislative. In that manner, each database should aim to have large set of Meta data (mostly time and owner related), to describe entries, so that these datasets can be of any real use. AP2A methodology is not intended to change database or logic of these databases, but it should try to gather as much as time data as possible for these databases. After gathering all DTL relevant data and after LOI classification, for each database from Phase 1 and Phase 2, this phase (Phase 3) can be considered completed.

F. Action Plan for Data Repositories

Action Plan for data repositories presents Phase 4 of AP2A methodology. This clearly indicates that Action Plan which is initial activity in OGP of developed countries, is actually proposed as Phase 4 of methodology for Balkan countries. This only means that there is, as previously stated, significant need for preparations, described in Phases 1,2 and 3.

Action Plan itself is bureaucratic process and there are already defined mechanisms for creating and accepting documents like this. It's important to point out that proposed Action plan should have two main goals:

1. Preparation (in technical and nontechnical manner) recognized databases and turning them into data sets for Open Data portal.

2. Increase of LOI for specific data sets

As we were able to see in analysis from Sections 2.1, 2.2.and 2.3, all Action Plans for Open Data are "owner driven". This means that future action plan should recognize not only what will new data sets be, but also who is owner and which responsibility is to create these data sets.

In that manner, logical concept of Action Plan should be presentable by TORR (Table of Roles and Responsibilities).

Table 3.4.1. - Table of Roles and Responsibilities for 4th Phase of AP2A

CANDIDATE FOR NEW DATA SET	OWNER OF REQUIRED DATABASE	CURRENT LOI	TARGETED LOI	DATE OF COMPLETION
Candidate 1	Owner(s) 1	LOI 1-5	LOI 1-5	DATE
Candidate 2	Owner(s) 2	LOI 1-5	LOI 1-5	DATE
…	…	…	…	…
Candidate N-1	Owner(s) N-1	LOI 1-5	LOI 1-5	DATE
Candidate N	Owner(s) N	LOI 1-5	LOI 1-5	DATE

G. Infinity plan / listener phase

Infinity plan is Phase 5 of AP2A methodology and it is actually a recursive activity which happens on defined time interval after previous phases are completed. This means that AP2A methodology proposes four previous phases in linear order, and this phase as a recursive one (with specific time intervals to iterate).

This phase is considered as listener phase, where system listens for events/triggers from external sources and if these triggers are important enough AP2A will recognize need for new databases and/or datasets and it will iterate trough several or all phases of AP2A again.

Let's explain this trough example. If new legislation is created (new law or regulation) then this legislative needs to be checked for potential databases (Phase 1), and ofcourse all following phases needs to be completed. So, in case of new legislative AP2A needs to be iterated form the start, from Phase 1.

If some legislative is changed and that changes affect already existing repository than repository owner needs to be asked about current/new state of repository, which is Phase 2 of this methodology. So, changes on current legislative on already existing database will trigger AP2A methodology, but from Phase 2. If external triggers recognize some kind of development or infrastructure project that aims to increase level of some information system, than most likely affected databases will change LOI, which means AP2A should be iterated from Phase 3. If citizens or other parties have specific requests and these reflects to Action Plan, it might cause previous phase to be repeated.

Ofcourse, after each repetition inside AP2A methodology, system returns to state of Phase 5, Infinity plan, a.k.a. Listener phase. As a conclusion to this, image 3.5.1. presents diagram of complete AP2A methodology proposal, with trigger events clearly defined.

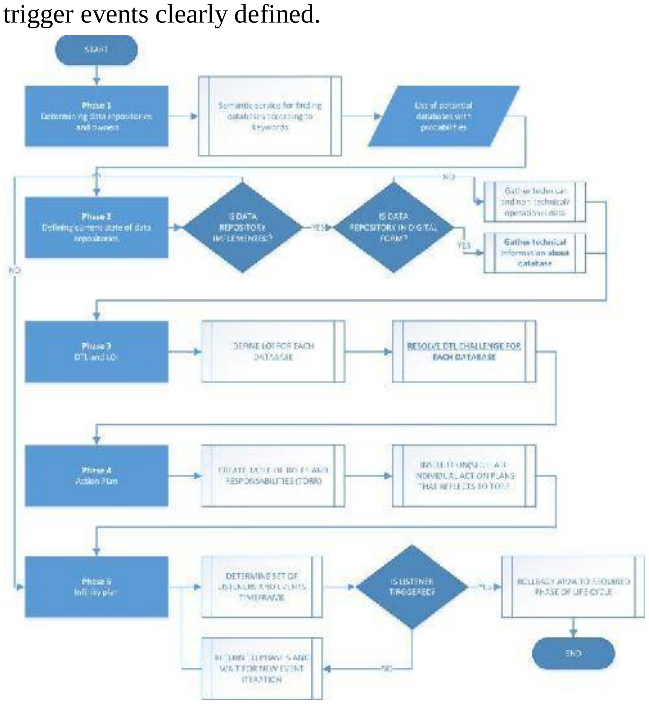


Image 3.5.1. - AP2A methodology, flow chart diagram

V. Conclusion

This paper proposes new and innovative methodology called AP2A, with goal to define roadmap to handle OGD from first phase to action plan, through five proposed phases. This methodology is created after research based on OGD implementations in Netherlands, Estonia and United Kingdom, described in Section 2 of this paper, with resources marked as [1],[2],[3],[4],[5],[6] and [7].

Main goal of AP2A methodology is to create business process for to help decision making in process of defining databases, data sets and making them published and publically available, trough couple of phases. Phase 1 of this methodology describes how to define "Register of Registries", reconstructed from current legislative. Also, this phase proposes existence of specific electronic service which is able to read through legislative. Phase 2 proposes set of technical and non-technical questions (future meta data) that should be answered in order to fully describe each existing database. These two phases form initial preparation for further data handling.

Phase 3 of this methodology handles DTL challenge and LOI determination. It's proposed that LOI determination is easy to handle by using same concept as the Netherlands described in Section 2.1. and in [12]. Related to DTL challenge, this paper doesn't resolve it, but it describes it in details and proposes reuse of electronic service from Phase 1 in order to somewhat automate this challenge. Phase 4 is formal creation of action plan, which is final delivery of this methodology.

Once Action plan is fully defined, some kind of maintenance and monitoring of its implementation is in order. Phase 5 of this methodology represent that kind of monitoring tool. Name of this phase is "Infinity phase" since it never actually ends, rather it iterates trough set of listeners and waits for specific set of events to trigger response. After specific event is recognized, this phase shifts AP2A to new iteration, starting from Phase 1,2,3 or 4, depending of severity of event. For more important events AP2A will be shifted to earlier phases of existence and re-iterated all over again.

It's important to notice that described methodology is limited to technology aspects of OGD implementation. In technology aspects, usage of this methodology can help in implementation of OGD, and also it can help better define database structures throughout all government systems. Also, proposed electronic service is reusable with smaller calibrations on keywords and probability matrix.

This research will continue in two paths: first one is defining and prototyping proposed electronic service, and the second one is resolving issues of DTL trough new concepts and possible automation of certain parts of process.

References

[1] Ubaldi, B. (2013), "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives", OECD Working Papers on Public Governance, No. 22, OECD Publishing.

[2] N. Anderson, G. Hodge, (2014), "Repository Registries: Characteristics, Issues and Futures", CENDI Secretariat, c/o Information International Associates, Inc. Oak Ridge, TN

[3] Open Data Consultancy study (November 2013), The Scottish Government, Swirrl IT Limited, ISBN: 978-1-78412-190-7 (web only)

[4] M. Fioretti, (2010), "Open Data, Open Society a research project about openness of public data in EU local administration", DIME network (Dynamics of Institutions and Markets in Europe, www.dime-eu.org) as part of DIME Work Package 6.8

[5] Open Definition Portal, Licence Definitions, [link: http://opendefinition.org/licenses/ visited: 2015/11]

[6] Open Definition Portal, v.2.1, [link: http://opendefinition.org/od/2.1/en/ visited: 2015/11]

[7] RFC 2119, Key words for use in RFCs to Indicate Requirement Levels, [link: https://tools.ietf.org/html/rfc2119 visited: 2015/12]

[8] Ministry of Interior and Kingdom Relations, Open Government Partnership, Self-Assessment Report, The Netherlands, 2014

[9] European Commission, Report on eGovernment in Estonia, 2015

[10] Open Government UK portal [link: http://www.opengovernment.org.uk/about/ogp-action-plans/ visited: 2015/12]

[11] Open Government Partnership portal [link: http://www.opengovpartnership.org/about visited: 2015/12]

[12] LOI concept reference with 5 star data concept [link: http://5stardata.info/en/ visited: 2015/12]

[13] Open State EU portal [link: http://www.openstate.eu/en/2015/11/first-five-dutch-municipalities-release-meetings-agendas-as-open-data/ visited: 2015/12]

Open Government Data in Western Balkans: Assessment and Challenges

Milan Stojkov*, Stevan Gostojić*, Goran Sladić*, Marko Marković**, Branko Milosavljević*
* University of Novi Sad, Faculty of Tehnical Sciences, Novi Sad, Serbia
** Appellate court of Novi Sad, Novi Sad, Serbia
*{stojkovm, gostojic, sladicg, mbranko}@uns.ac.rs, ** marko.markovic@ns.ap.sud.rs

***Abstract*— In order to improve availability and usage of public data, national, regional and local governmental bodies have to accept new ways to open up their data for everyone to use. In that sense, the idea of open government data has become more common in a large number of governmental bodies in countries across the world in the past years. This study gives an overview of open government data that are available on the Internet for Serbia, Montenegro, Bosnia and Herzegovina and Croatia. Three most common methodologies for open data assessment are described and one of them is used to indicate advantages and disadvantages of available data. The detailed research provided enough information to make proposals for eliminating open government data shortcomings in these countries.**

I. Introduction

Public data represents all the information that public bodies in one government produce, collect or pay for. One part of the public data is presented in the form of open data which is defined as data in a machine-readable format that is publicly available under an “open” license that ensures it can be freely used, reused, redistributed by anyone for any legal purpose [1]. Government data is a subset of open data. It is important to consider the distinctions between “open data” and “open government”. Opening up existing datasets is just the first step and does not automatically lead to a democratic government [22]. According to Jonathan Gray [22], director of policy and research at Open Knowledge, opening up is just one step and no replacement for other vital elements of democratic societies, like robust access to information laws, whistleblower protections, and rules to protect freedom of expression, freedom of the press and freedom of assembly. National, regional and local governments have to find appropriate strategies to deliver large amounts of data that is made for public use.

The main reason for opening the data is to increase transparency, the participation of other institutions and citizens, government efficiency and to create a new job and business opportunities [16]. For example, UK Government saved £4 million in 15 minutes with open data [2] and overall economic gains from opening up public data could amount to €40 billion a year in the EU [3]. The European Commission is investing large amounts of finances in finding adequate strategies to use open data which additionally indicates how open data is significant [15]. However, open data strategies are relatively new, so evidence of this expected impact is still limited. One big challenge is the exploitation of the Web as a platform for data and information integration as well as searching and querying the Web. If we combine the Web with sensitive information that government possesses, there we can find answers why some public data is not yet available. Some of the excuses which representatives of different governmental bodies give are that publishing data is technically impossible, data is just too large to be published and used, data is held separately by a lot of different organizations and cannot be joined up or IT suppliers will charge them a fortune to do that [4]. To try to overcome that, it is important that governmental bodies, as well as civil society, are willing to accept the concept of open data. Also, it is very important that data does not collide with existing laws of one country, e.g. data protection law, copyright law, etc.

In this paper, we present representative methodologies for assessing the openness of open government data, as well as their advantages and disadvantages. Further, we pick one of the presented methodologies which we feel that contains principles which open government data should fulfill. After that, we explore available open government data for some Balkan countries to see how the data fits in listed principles. In the end, we propose some solutions for eliminating observed shortcomings of presented methodologies and open government data that is available.

The following text has been organized as follows. The next section describes related work that helped us with our research. In section three, representative methodologies for assessing the openness of data were presented. Section 4 describes the current state of open data in Serbia, Montenegro, Bosnia and Herzegovina and Croatia based on one methodology proposed in section 3. Results of the research are presented in section 5. Proposals to overcome observed shortcomings are presented in section 6. Finally, the last section concludes the paper giving the future directions of this research.

II. Related Work

The main implementations of open government data initiatives are data portals in a number of different ways [21]. Those are catalogs that contain a collection of metadata records which describe open government datasets and have links to online resources [18]. The implementation of a catalog raises an important question - what metadata should be stored and how should it be represented? This question is especially significant when automatic importing of metadata records (also known as harvesting) is performed, as metadata structure and meaning are not usually consistent or self-explanatory. Open data portal software such as CKAN

(Comprehensive Knowledge Archive Network) [11] or vocabularies such as DCAT (Data Catalog Vocabulary) [19] provide solutions for this problem.

The government in the United Kingdom has a site *data.gov.uk w*hich brings all data together in one searchable website [10]. If the data is easily available, people will be easier to make decisions about government policies based on provided information. Website *data.gov.uk* is built using CKAN to catalog, search and display data. CKAN is a data catalog system used by various institutions and communities to manage open data. The UK government continues to use and develop this website and the site has a global reputation as a leading exemplar of a government data portal. Besides the UK portal, there are three more major sites – *data.gov* (the US), *data.gouv.fr* (France) and *data.gov.sg* (Singapore).

In the last few years, the Linked Data paradigm [23] has evolved as a powerful enabler for the transition of the current document-oriented Web into a Web of interlinked data and, ultimately, into the Semantic Web. Aimed at speeding up this process, the LOD2 project [12] (“Creating knowledge out of interlinked data”) partners have delivered the LOD2 Stack, “an integrated collection of aligned state of the art software components that enable corporations, organizations and individuals to employ Linked Data technologies with minimal investments” [13]. As partners of the LOD2 project, the Mihailo Pupin Institute established something similar to *data.gov.uk* website - the Serbian CKAN [14]. This is the first catalog of this kind in the West Balkan countries, with a goal of becoming an essential tool for enforcing business ventures based on open data in this region [15].

There are several studies that contain valuable information that helped us notice what are the challenges every country faces implementing the idea of “open data”. In an inquiry for the Dutch Ministry of the Interior and Kingdom Relations, TNO (the Netherlands Organization for Applied Scientific Research) examined the open data strategies in five countries (Australia, Denmark, Spain, the United Kingdom and the United States) and gathered anecdotal evidence of its key features, barriers and drivers for progress and effects, which is described in [16].

Serbian government hired open data assessment expert Ton Zijlstra to make Open Data Readiness Assessment (ODRA) [9] for Serbia [17] which helped us understand current situation in one of the countries of Western Balkans.

The paper [21] presents an overview of the open government data initiatives. The aim of this research was to answer a set of questions, mainly concerning open government data initiatives and their impact on stakeholders, existing approaches for publishing and consuming open government data, existing guidelines and challenges.

There are some requirements which make government data open government data. Research presented in [8] proposes 14 principles which describe open government data. The number of principles is still expanding since every new principle opens new questions. For example, how can governmental body guaranty that the public data presented is primary? Or what is considered to be safe file format and what is not? Also, what are the ways citizens can review the data? How can an open license be presented in machine-readable form? These are only some questions to bear in mind.

III. Methodologies for Assessing The Openness of Data

The following two methodologies could fall into evaluating implementation category. Sir Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested first presented methodology, a 5-star deployment scheme for open data. The five star Linked Data system is cumulative. Each additional star presumes the data meets the criteria of the previous step(s) [20].

1 Star – Data is available on the Web, in whatever format, under an open license

2 Stars – Available as machine-readable structured data (i.e., not a scanned image)

3 Stars – Available in a non-proprietary format (i.e., CSV, not Microsoft Excel)

4 Stars – Published using open standards from W3C (RDF and SPARQL)

5 Stars – All of the above and links to other Linked Open Data

These steps seem very loose, but exactly that gives them required simplicity. Of course, achieving 4 and 5 stars are not easy and Linked data has its own problems such as the way of publishing and consuming data, etc. For example, although the UK open government program is doing remarkable stuff, only a small percent of all datasets released so far could score 5 stars. It seems that this 5-star scheme mainly targets the technical aspects, but there are more aspects it needs to be considered such as political, social and economic.

Each year, governments are making more data available in an open format. The Global Open Data Index (GODI) tracks whether this data is actually released in a way that is accessible to citizens, media, and civil society and is unique in crowd-sourcing its survey of open data releases around the world [5]. The Index measures and benchmarks the openness of data around the world, and then presents this information in a way that is easy to understand and use. Each year annual ranking of countries is produced and peer reviewed by their network of local open data experts [5]. The Index is not a representation of the official government open data offering in each country, but an independent assessment from a citizen perspective which benchmarks open data by looking at fifteen key datasets in each place (including those essential for transparency and accountability such as election results and governments spending data, and those vital for providing critical services to citizens such as maps and transport timetables). These datasets were chosen based on the G8 key datasets definition [6]. Fifteen key datasets are [7]: election results, company register, national map, government spending, government budget, legislation, national statistical office data, location datasets, public transport timetables, pollutant emissions, government procurement data, water quality, weather forecast, land ownership and health performance data.

Each dataset in each place is evaluated using nine questions that examine the technical and the legal openness of the dataset. In order to balance between the two aspects, each question is weighted differently and worth a different score. Together the technical questions are worth 50 points, the three legal questions are also worth 50 points [7]. Questions that examine technical openness with corresponding weights in parentheses are: does the data exist? (5), is the data in digital form? (5), is the data available online? (5), is the data machine-readable? (15), is the data available in bulk? (10), is the data provided on a timely and up to date basis? (10).

Questions that examine the legal status of openness are: is the data publicly available? (5), is the data available for free? (15), is the data openly licensed? (30).

Contributors to the Index are people who are interested in open government data activity and who can assess the availability and quality of open datasets in their respective locations. The assessment takes place in two steps. The first step is collecting the evaluation of datasets through volunteer contributors, and the second step is verifying the results through volunteer expert reviewers. The reason why this methodology focuses only on fifteen key datasets is because the Global Open Data Index wants to maximize the amount of people who contribute to the Index, across local administrations, countries, regions and languages [7].

The good thing about this methodology is that the Index tracks whether the data is actually released in a way that is accessible to citizens, media, and civil society, and is unique because results are delivered by volunteer contributors [7]. The Index plays a big role in sustaining the open government data community around the world. So, if the government of a country does publish a dataset, but this is not clear to the public and it cannot be found through a simple search, then the data can easily be overlooked [7]. In that case, everyone who is interested to find this particular data can review the Index results to locate it and see how accessible the data appears to citizens [7].

The current problem when looking at national datasets is that there is generally no standardization of datasets between countries. Datasets differ between governments in aggregation levels, metadata, and responsible agency. The Index does not define what level of details datasets have to meet, so we have examples where data about spending is very detailed and data about national maps is very vague.

Another downside of GODI methodology is the number of datasets being assessed. The Index wants to gather as many contributors as possible for a big number of countries, but it seems that 15 datasets on a country level are not enough. Although we can record progress compared to 2014 when there were only 10 datasets, it seems that has room for a few more – for example, datasets referring to public safety (e.g. crime data, food inspection, car crashes, etc), available medications, educational institutions, public works (e.g. road work, infrastructure), transportation (e.g. parking, transit, traffic) and utilities (e.g. water, gas, electrical consumption and prices). Of course, the problem with these datasets can be that they are owned and produced by a company and not the state because some of the government services might be privatized. It is not a bad idea to consider evaluating municipal datasets. With that, public services can gain in efficiency and users in satisfaction by meeting the expectations of users better and being designed around their needs and in collaboration with them whenever possible. Also, another indicator in addition to 9 questions should be the one concerning provenance and thrust of the data. Public data should have some kind of digital license which provides authenticity and integrity of the data. Also, datasets should be multi-lingual because of national minorities.

The next methodology mentioned is called Open Data Readiness Assessment (ODRA) [1]. This methodology could fall into readiness assessment category. The World Bank’s Open Government Data Working Group developed ODRA which is a methodological tool that can be used to conduct an action-oriented assessment of the readiness of a government or individual agency to evaluate, design and implement an Open Data initiative [1]. This tool is freely available for others to adapt and use. The purpose of this assessment is to assist the government in diagnosing what actions the government could consider in order to establish an Open Data initiative [1]. This means more than just launching an Open Data portal for publishing data in one place or issuing a policy. An Open Data initiative involves addressing both the supply and the reuse of Open data, as well as other aspects such as skills development, financing for the government’s Open Data agenda and targeted innovation financing linked to Open Data [1].

The ODRA uses an “ecosystem” approach to Open Data, meaning it is designed to look at the larger environment for Open Data – “supply” side issues like the policy/legal framework, data existing within government and infrastructure (including standards) as well as “demand” side issues like citizen engagement mechanisms and existing demand for government data among user communities (such as developers, the media and government agencies) [1]. The assessment evaluates readiness based on eight dimensions considered essential for an Open Data initiative that builds a sustainable Open Data ecosystem. The readiness assessment is intended to be action-oriented. For each dimension, it proposes a set of actions that can form the basis of an Open Data Action Plan. Eight dimensions are [1][17]: senior leadership, policy/legal framework, institutional structures responsibilities and capabilities within government, government data management policies and procedures, demand for open data, civic engagement and capabilities for open data, funding and open data program, a national technology and skills infrastructure.

In order to make a better assessment, significant numbers of governmental body representatives have to be interviewed. That takes time, and it is questioned if everybody from selected government sections is willing to participate. ODRA is free to use but it can be a big problem for someone to use it to make an assessment on their own. Usually, open data experts are hired by the government to make an assessment for their internal reasons. The process of making the assessment is long, expensive and very detailed. After making an assessment based on ODRA, action plan applies. The suggested actions are provided to be taken into consideration by some kind of Open Data Working Group, and it is suggested to consider them in the context of existing policies and plans to determine priorities and order of

execution in detail [17]. Readiness assessments tend to operate at the country level, although the World Bank suggests their ODRA can also be applied at sub-national levels [1].

We described the GODI methodology most detailed among three widely used methodologies as we feel it is the most accessible way for every citizen to explore open government data and make an assessment. Another reason for choosing this methodology is because Serbia, Bosnia and Herzegovina, Croatia and Montenegro have not been scored in official assessment since they did not submit all datasets to 2015 year's Index. In this way, we can see the true state of open government data in these countries.

IV. Western Balkans Research

This section describes how open data collected from different governmental bodies in Serbia, Montenegro, Bosnia and Herzegovina and Croatia fits in the GODI methodology described in section 3. The first step was to visit open data portals that have specialized data concerning these countries. If the data were not found there, specialized government websites were visited for desired information. Scores were given based on survey score [7]. Summarized results for four countries were given in Tables I-IV. After the research total score for Serbia is 520/1300, for Bosnia and Herzegovina 375/1300, for Croatia 510/1300 and for Montenegro 390/1300. Kosovo[1] is ranked 40th in the 2015 Index with a total score of 555/1300.

V. Results

Considering that Taiwan is 1st on the list with the score of 1010/1300 and based on provided information, it can be concluded that the openness of the data in given countries is not on a high level. Datasets that were observed fulfill minor part of 14 principles defined in [8]. Data is online and free, primary, timely and partly accessible. Further, data is non-discriminatory, permanent and considered in safe file formats. Shortcomings are more visible. There is a lack of information about licenses. Data is not digitally signed or provided with some kind of authenticity and integrity. It has big problem when it comes to machine readability. Datasets are predominantly available in PDF and Microsoft Word files which are not preferred formats for computer processing. Also, data is partly proprietary which refers to datasets available in Microsoft Word file.

TABLE I.
OPEN DATA ASSESSMENT FOR SERBIA

Serbia	Data exists	Digital form	Publicly available	For free	Online	Machine-readable	In bulk	Open license	Timely & up-to-date	Score
Election Results	5	5	5	15	5	15	10	-	10	70
Company Register	5	5	5	15	5	-	-	-	10	45
National Map	5	5	5	15	5	-	-	-	10	45
Government Spending	5	5	-	-	-	-	-	-	-	10
Legislation	5	5	5	15	5	-	-	-	10	45
National Statistical Office Data	5	5	5	15	5	15	-	-	10	60
Location	5	5	-	-	-	-	-	-	-	10
Government budget	5	5	5	15	5	-	10	-	10	55
Pollutant Emissions	5	5	5	15	5	-	-	-	10	45
Gov. procurement data	5	5	5	15	5	-	-	-	10	45
Water quality	-	-	-	-	-	-	-	-	-	0
Weather forecast	5	5	5	15	5	-	-	-	10	45
Land ownership	5	5	5	15	5	-	-	-	10	45

TABLE II.
OPEN DATA ASSESSMENT FOR BOSNIA AND HERZEGOVINA

Bosnia and Herzegovina	Data exists	Digital form	Publicly available	For free	Online	Machine-readable	In bulk	Open license	Timely & up-to-date	Score
Election Results	5	5	5	15	5	-	-	-	10	45
Company Register	5	5	5	15	5	-	-	-	10	45
National Map	5	5	5	15	5	-	-	-	10	45
Government Spending	-	-	-	-	-	-	-	-	-	0
Legislation	5	5	5	15	5	-	-	-	10	45
National Statistical Office Data	5	5	5	15	5	-	-	-	10	45
Location	-	-	-	-	-	-	-	-	-	0
Government budget	-	-	-	-	-	-	-	-	-	0
Pollutant Emissions	-	-	-	-	-	-	-	-	-	0
Gov. procurement data	5	5	5	15	5	-	-	-	10	45
Water quality	-	-	-	-	-	-	-	-	-	0
Weather forecast	5	5	5	15	5	15	-	-	10	60
Land ownership	5	5	5	15	5	-	-	-	10	45

flow provided by the GODI. Scores for transport timetables and health performance are omitted from final

[1] References to Kosovo shall be understood to be in the context of Security Council Resolution 1244 (1999).

TABLE III.
OPEN DATA ASSESSMENT FOR CROATIA

Croatia	Data exists	Digital form	Publicly available	For free	Online	Machine-readable	In bulk	Open license	Timely & up-to-date	Score
Election Results	5	5	5	15	5	15	10	-	10	70
Company Register	5	5	5	15	5	-	-	-	10	45
National Map	5	5	5	15	5	-	-	-	10	45
Government Spending	5	5	-	-	-	-	-	-	-	10
Legislation	5	5	5	15	5	-	-	-	10	45
National Statistical Office Data	5	5	5	15	5	-	-	-	10	45
Location	-	-	-	-	-	-	-	-	-	0
Government budget	5	5	5	15	5	15	10	-	10	70
Pollutant Emissions	5	5	5	15	5	-	-	-	10	45
Gov. procurement data		5	5	15	5	-	-	-	10	45
Water quality	-	-	-	-	-	-	-	-	-	0
Weather forecast	5	5	5	15	5	-	-	-	10	45
Land ownership	5	5	5	15	5	-	-	-	10	45

TABLE IV.
OPEN DATA ASSESSMENT FOR MONTENEGRO

Montenegro	Data exists	Digital form	Publicly available	For free	Online	Machine-readable	In bulk	Open license	Timely & up-to-date	Score
Election Results	5	5	5	15	5	-	-	-	10	45
Company Register	5	5	5	15	5	-	-	-	10	45
National Map	5	5	5	15	5	-	-	-	10	45
Government Spending	5	5	-	-	-	-	-	-	-	10
Legislation	5	5	5	15	5	-	-	-	10	45
National Statistical Office Data	5	5	-	-	-	-	-	-	-	10
Location	-	-	-	-	-	-	-	-	-	0
Government budget	5	5	-	-	-	-	-	-	-	10
Pollutant Emissions	5	5	5	15	5	-	-	-	10	45
Gov. procurement data	5	5	5	15	5	-	-	-	10	45
Water quality	-	-	-	-	-	-	-	-	-	0
Weather forecast	5	5	5	15	5	-	-	-	10	45
Land ownership	5	5	5	15	5	-	-	-	10	45

All data is predominantly found on specialized websites of corresponding governmental bodies and not on existing open data portals. It is unknown if all of the observed countries have appropriate governmental bodies which control open data. Definitely, the big problem is interoperability between different governmental bodies, i.e. lack of it. Besides that, public input is crucial to disseminating information in such a way that it has value, and lack of different datasets prove that this principle is hard to please.

VI. MEASURES FOR SHORTCOMINGS REMOVAL

First two things that should be done are creating a marketing campaign in which relevant political bodies will be familiarized with strategies for open data and creating a specific governmental body which will ensure interoperability between existing governmental bodies. Also, government agencies need to know what are the expenses of collecting and exchanging data as well as which are the ways of income generation. A wider range of datasets can be used if data is anonymous and in accordance with existing data protection laws of one country. In order to include more different government bodies, creating open data pilot projects with participation of different ministries and agencies is advised. Also, publishing government data that is regularly requested as open data is a good way to reduce the workload of government officials. During the research, it was observed that different governmental bodies have their own websites, but it is a bit difficult to find appropriate data. The solution can be found in the United Kingdom and their centralized website. If the data is easily available, people will be easier to make decisions about government policies based on provided information. If making a new website with all the data that exists on known websites is too much work, then this website can contain only links to other websites with adequate information. Different ways for providing access to files would be via File Transfer Protocol (FTP), via torrents or via Application Programming Interface (API). The solution for certifying data to be primary can be found in existing or a new law of a country. Question about safe file format will always be open for debate. The fact is that most of the data are presented in PDF, Microsoft Word, and OpenOffice's OpenDocument files. In most cases later two formats are better for open government data than PDF as they are print-ready like the PDF but also allow for reliable text extraction. The second condition for making file format appropriate for documents would be machine readability. That feature none of the above file formats can satisfy. That is why the data should be available in formats such as XHTML, RDF/XML or CSV, too. As the best solution for open data machine-readability problem, for now, we suggest using linked data paradigm which gives benefits to users like discovering more related data while

consuming the data, making data discoverable and more valuable.

Another shortcoming discovered during our research is related to license under which the data is published. In most jurisdictions there are intellectual property rights in data that prevent third-parties from using, reusing and redistributing data without explicit permission. Licenses conformant with the Open Definition which can be found at *http://opendefinition.org/licenses* are recommended for open data. More precise, using Creative Commons CC0 (public domain dedication) or CC-BY (requiring attribution of source) is suggested. Another flaw is that some government sites have the required data but it is not downloadable in a bulk. Earlier in this section, as one way of providing access to data, API was proposed. It is important to understand when bulk data or an API is the right technology for a particular database or service. Bulk data provides a complete database, but data APIs provide only a small window into the data. Bulk data is static, but data APIs are dynamic. A good data API requires that the agency does everything that good bulk data requires (because ultimately it delivers the same data), plus much more. Therefore, governmental bodies should build good bulk open data first, validate that it meets the needs of users, and after validation, invest in building a data API to address additional use cases. Of course, both bulk and API as possible ways of reading the data are desired.

VII. Conclusion

In this paper were presented some methodologies for assessing the openness of the data. Research concerning Serbia, Montenegro, Bosnia and Herzegovina and Croatia is carried out and observations were expressed. Further, some proposals how to overcome observed shortcomings were listed. During our research, "open data" have been placed on the political and administrative agenda in the Republic of Serbia. Ton Zijlstra made the ODRA for Serbia [17] and different governmental bodies were engaged in improving open government data. There is action plan prepared in Montenegro in accordance with the principles of Open Government Partnership committed to making some difference. In Bosnia and Herzegovina awareness about open data was raised through an EU-funded PASOS project. Government officials in Croatia launched open data portal which offers in a single place different kind of data related to public administration and is an integral part of the e-citizens project. Many hackathons with open government data as a theme were organized for young people, to include them in open data popularization process. The research also demonstrated that opening the data is hard because there is a kind of closed culture within government which is caused by fear of the disclosure of government failures and even escalating political scandals. Also, databases which contain significant data are not well organized and there are not sufficient human and financial resources to collect a big amount of data. Although there is a willingness to apply strategies for open data, governmental bodies still hesitate to actually do this because they do not understand true effects of those strategies.

In the future work, we will try to contribute more to governmental bodies' open data actions like hackathons and open data pilot projects and to provide latest data to the GODI. Also, we will try to expand our research on open data concerning judicial systems and parliaments which would give us a complete picture of open data which one country should provide. The other direction of future work would be developing special software tools for consuming open data by people who are not so technically skilled.

References

[1] World Bank Group, Open Government Data Working Group, Open Data Readiness Assessment (ODRA) User Guide, http://opendatatoolkit.worldbank.org/docs/odra/odra_v3.1_userguide-en.pdf

[2] The Open Data Handbook website, http://opendatahandbook.org/value-stories/en/saving-4-million-pounds-in-15-minutes

[3] Communication from the commissions to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, "Open data – An engine for innovation, growth and transparent governance", Brussels, 12.12.2011.

[4] Andrew Stott, "Implementing an Open Data program within government", OKCON 2011, Retrieved from http://www.dr0i.de/lib/2011/07/04/a_sample_of_data_hugging_excuses

[5] Global Open Data Index, http://index.okfn.org/about/

[6] G8 Open Data Charter and Technical Annex, https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex#technical-annex

[7] Global Open Data Index Methodology, http://index.okfn.org/methodology/

[8] Joshua Tauberer, "Open Government Data: The Book", Second Edition 2014.

[9] The World Bank, Readiness Assessment Tool ODRA, http://opendatatoolkit.worldbank.org/en/odra.html

[10] Opening up Government, https://data.gov.uk

[11] CKAN, The open source data portal software, http://ckan.org

[12] LOD2 Project, http://lod2.eu/Welcome.html

[13] S. Auer, M. Martin, P. Frischmuth, B. Deblieck, "Facilitation the publication of Open Governmental Data with the LOD2 Stack", Share-PSI workshop, Brussels, Retrieved from http://share-psi.eu/papers/LOD2.pdf (2011)

[14] CKAN Serbia, http://rs.ckan.net

[15] V. Janev, U. Milošević, M. Spasić, J. Milojković, S. Vraneš, "Linked Open Data Infrastructure for Public Sector Information: Example from Serbia", Retrieved from http://ceur-ws.org/Vol-932/paper6.pdf

[16] Noor Huijboom, This Van den Broek, "Open data: an international comparison of strategies"

[17] Ton Zijlstra, "Open Data Readiness Assessment – Republic of Serbia", http://www.rs.undp.org/content/dam/serbia/Publications%20and%20reports/English/UNDP_SRB_ODRA%20ENG%20web.pdf

[18] Kučera, J, Chlapek, D, Nečaský, M, "Open government data catalogs: Current approaches and quality perspective. In: Technology-Enabled Innovation for Democracy, Government and Governance", Lecture Notes in Computer Science, vol. 8061, pp. 152–166. Springer Berlin Heidelberg (2013), http://dx.doi.org/10.1007/978-3-642-40160- 2_13

[19] W3C, Data Catalog Vocabulary (DCAT), http://www.w3.org/TR/vocab-dcat/

[20] Tim Berners-Lee, 5-star Linked Data, https://www.w3.org/2011/gld/wiki/5_Star_Linked_Data

[21] Attard J, Orlandi F, Scerri S, Auer S, "A Systematic Review of Open Government Data Initiatives", Article in Government Information Quarterly, August 2015

[22] Jonathan Gray interview, http://www.theguardian.com/media-network/2015/dec/02/china-russia-open-data-open-government

[23] Linked Data, https://www.w3.org/DesignIssues/LinkedData.htm

Survey of Open Data in Judicial Systems

Marko Marković*,**, Stevan Gostojić*, Goran Sladić*, Milan Stojkov*, Branko Milosavljević*
* University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
** Appellate court of Novi Sad, Novi Sad, Serbia
*{markic, gostojic, sladicg, stojkovm, mbranko}@uns.ac.rs, **marko.markovic@ns.ap.sud.rs

***Abstract* — Judicial data is often poorly published or not published at all. It is also missing from datasets considered for evaluation by open data evaluation methods. Nevertheless, data about courts and judges is also data of public interest since it can reveal the quality of their work. Transparency of judicial data has an important role in increasing public trust in the judiciary and in the fight against corruption. However, it also carries some risks, such as publication of sensitive personal data, which need to be addressed.**

***Keywords* — open data, judiciary**

I. Introduction

Transparency of government data gives citizens an insight into how government works. Access to government data is a subject of public interest because its actions affect public in many ways. It is facilitated by widespread use of Internet and rapid development of information technologies. On the other hand, personal data is sometimes part of government data and it is in citizens' best interest to protect their privacy. These opposing expectations make the publishing of such data difficult, especially in judiciary where a considerable amount of personal data is present. In this paper, we will present different aspects of open data in judiciary: from definitions of basic terms, through licenses of published open government data, to specifying typical judicial datasets. Also, some general approaches for opening government data will be presented in the context of judicial data while some current achievements in this field will be discussed.

In [1] definitions of some elementary terms relevant to open government data are given. The term *data* denotes unprocessed atomic statements of facts. Data becomes information when it is structured and presented as useful and relevant for a particular purpose. The term *open data* represents data that can be freely accessed, used, modified and shared by anyone for any purpose with the requirement to provide attribution and share-alike. Open data defined by [2] has two requirements, to be legally open and technically open. Legally open data is available if appropriate license permits anyone to freely access, reuse and redistribute it. Technically open data is available in a machine-readable and bulk form for no more than reproduction cost. The machine-readable form is structured and assumes automatic reading and processing of data by computer. Data is available in bulk when complete dataset can be downloaded by the user. The term open government data is then defined as data produced or commissioned by government bodies (or entities controlled by the government) that anyone can freely use, reuse and redistribute. [3]

In December 2007, a working group consisting of 30 experts interested in open government, proposed the set of eight principles required for government data to be considered open [4]: complete (all public data need to be available), primary (collecting data at its source, unmodified and with the highest level of granularity), timely (to preserve the value of the data), accessible (available to the widest range of users and for the widest range of purposes), machine processable (data structure allows automated processing), non-discriminatory (available to anyone with no need for registration), non-proprietary (format of data not dependable on any entity), and license-free (availability of data is not licensed by any copyright, patent, trademark or trade secret regulation except when reasonable privacy, security and privilege restrictions are needed).

Besides these eight principles, seven additional principles are given: online and free, permanent, trusted, a presumption of openness, documented, safe to open, and designed with public input.

Considering legal openness, besides availability of government data (in the sense of technical openness), it is necessary to specify license under which the data are published. Unfortunately, at government websites, the information about the license is often omitted. In [5], three licensing types of published government data are recognized: case-by-case (licensing is present when published data are subject of copyright and other rights, but permission to reuse these data is given on a case-by-case basis), re-use permitted / automatic licenses (corresponds to cases when copyright and other rights are given by license terms and conditions or another legal statement, while re-use by the public is permitted), and public domain (licensing exempts documents and datasets from copyright or dedicates them to the public domain with no restrictions on public reuse).

All Creative Commons licenses [6] share some base features on the top of which additional permissions could be granted. Among these baseline characteristics are: non-commercial copying and distribution are allowed while copyright is retained; ensures creators (licensors) getting deserved credits for their work; the license is applicable worldwide. Licensors may then choose to give some additional rights: attribution (copying, distribution, and derivation are allowed only if credits are given to the licensor), share-alike (same license terms apply to distribution of derivative work as for the original work), non-commercial (copying, distribution, and derivation are allowed only for non-commercial purposes), and no derivative (only original unchanged work, in whole, may be copied and distributed).

Creative Commons licenses consist of three layers: legal code layer (written in the language of lawyers), commons deed (the most important elements of license written in language non-lawyers could understand), and machine-readable version (license described in CC Rights Expression Language [7] enabling software to understand license terms).

To place their work in public domain, Creative Commons gives creators solution known as CC0. Nevertheless, many legal systems do not allow the creator to transfer some rights (e.g. moral rights). Therefore, CC0 allows creators to contribute their work to the public domain as much as possible by law in their jurisdiction. In [8] it is argued that according to copyright protection regulations, neither databases nor any non-creative part of content cannot be assumed as a creative work.

To provide a legal solution for opening data, the project Open Data Commons launched the open data license called Public Domain Dedication and License (PDDL) [9] in 2008. In 2009, the project was transferred to the Open Knowledge Foundation. PDDL allows anyone to freely share, modify and use work for any purpose.

In [10] is emphasized the importance of opening judicial data in preventing corruption and increasing trust in the judiciary. To achieve this, publishing of data about judges (e.g. first name, last name, biographical data, court affiliation, dates of service, history of cases, statistical data about workload and average time period necessary to make a decision, etc.) and courts (e.g. name, contact data, case schedules, court decisions, statistical data, etc.) is proposed. As an example of open data benefits in the judicial branch, Slovakian OpenCourts portal [11] is given and will be described in the rest of this paper.

In [12] is discussed reidentification as an important issue related to the opening of judicial data. It is a risk of revealing identity for an individual from disclosed information when it is combined with other available information.

In [13] it is emphasized the role of controlled vocabularies in order to achieve semantic interoperability of e-government data. Controlled vocabularies are valuable resource for avoidance of e.g. ambiguities, wrong values and typing errors. Its representation is usually in form of glossaries, code lists, thesauri, ontologies, etc. Some examples of legal thesauri are Wolters Kluwer [14] thesauri for courts and thesauri for German labor law. Also, some examples of ontologies for legal domain are LKIF-Core Ontology [15], Legal Case Ontology [16] and Judicial Ontology Library (JudO) [17].

The rest of this paper is organized as follows. First, available methods for evaluation of open government data will be reviewed. Then, several case studies of open judicial data are discussed. After, some directions for opening judicial data will be proposed. At the end, concluding remarks will be given and directions for future research.

II. Open Data Evaluation Methods

In this section several methods for evaluation of open government data will be reviewed: Global Open Data Index [18], 14 principles of open government data defined in [19] and Open Data Barometer [20].

Global Open Data Index tracks the state of open government data (currently in 122 countries) and measures it on an annual basis. It relies on Open Definition [2] saying that “Open data and content can be freely used, modified, and shared by anyone for any purpose”. The Global Open Data Index gives to the civil society actual openness levels of data published by governments based on feedback given by citizens and organizations worldwide. Some benefits of using Global Open Data Index are: it gives citizen’s perspective on data openness instead of government claims; comparison of dataset groups across the countries; helps citizens to learn about open data and available datasets in their countries; tracks changes in open data over time. During collection and assessment of the data some assumptions were taken into consideration: open data is defined by the Open Definition (while, as an exception, non-open machine-readable formats such as XLS were assumed open); governments are responsible for data publishing (even if some field is privatized by third-party companies); government, as a data aggregator, is responsible for publishing open data by all its sub-governments. Datasets considered by Global Open Data Index are national statistics, government budget, government spending, legislation, election results, national map, pollutant emissions, company register, location datasets, government procurement tenders, water quality, weather forecast, and land ownership. Scoring for each dataset is based on evaluation consisted of nine questions. Questions and its scoring weights (in brackets) are as follows: “Does the data exists?” (5); Is data in digital form?” (5); “Publicly available?” (5); “Is data available for free?” (15); “Is data available online?” (5); “Is the data machine-readable?” (15); “Available in bulk?” (10); “Openly licensed?” (30); and “Is the data provided on a timely and up to date basis?” (10).

Since there are 13 datasets, each with a maximum possible score of 100, the percentage of openness is calculated as a sum of scores for all datasets divided by 1300.

Although Global Open Data Index considers a wide range of government data, only legislation data are tracked in the legal domain. Same evaluation method, applied to supported datasets, could also be applied to judiciary datasets.

In [19] are given essential qualities for open government data, subsumed in four “A”s: accessible, accurate, analyzable and authentic. In detail, these qualities are defined by 14 principles: online and free, primary, timely, accessible, analyzable, non-proprietary, non-discriminatory, license-free, permanent, safe file formats, provenance and trust, public input, public review, and interagency coordination.

Open Data Barometer analyzes open data readiness, implementation, and impact. It is a part of World Wide Web Foundation’s work on common assessment methods for open data. Currently, results in 2014 are available for 86 countries. Open Data Barometer based its ranking on three types of data: peer-reviewed expert survey responses (country experts answer questions related to open data in their countries), detailed dataset assessments (a group of technical experts gives an assessment based on the results of a survey answered by country experts), and secondary data (data based on expert surveys answered by World Economic Forum, Freedom House, United Nations Department of Economic and Social Affairs, and World Bank).

For the ranking purposes, three sub-indexes are considered: readiness, implementation, and impacts. Readiness sub-index measures readiness to enable successful open data practices. Implementation sub-index is based on 10 questions for every 15 categories of data. Categories are as follows: mapping data, land ownership

data, national statistics, detailed budget data, government spend data, company registration data, legislation data, public transport timetable data, international trade data, health sector performance data, primary and secondary education performance data, crime statistics data, national environmental statistics data, national election results data, and public contracting data. Impacts sub-index reflects the impact of open data on different categories such as political, social and economic spheres of life. In the calculation of final ranking, implementation participates with 50% while readiness and impacts are weighted with 25% each.

Among datasets assessed by Open Data Barometer, there are no judiciary data, which in addition to legislative and crime datasets, could improve assessment of public data in the legal domain.

III. Open Judicial Datasets

This section gives an overview of currently available judicial open datasets (or open data portals) for Slovakia, Croatia, Slovenia, Bosnia and Herzegovina, the Republic of Macedonia, Serbia, UK, and the US. These countries were chosen as samples of legal systems in both Anglo-Saxon and continental European countries. Adopting best practices might be helpful for opening judicial data in developing countries such as Serbia.

Besides legislation as the most common dataset in the legal domain, there are many types of judicial data which could be considered for the opening. Most of them are defined by regulations on court proceedings (e.g. [21]). A list of judicial dataset which could be proposed for opening might be summarized as follows: receipted documents records data (e.g. date and time, number of copies, whether the fee is paid or not, etc.), case register data (e.g. case number, date of receipt, date of receipt of the initial document, judge name, date of decision, hearings information, performed procedural actions, etc.), and delivered decisions.

Also, some derived statistical datasets could be the subject of public interest. Such data could be the first step until full opening of judicial datasets occurs. For example, these statistical datasets could be: statistical report for a judge (e.g. number of unresolved cases, received cases and solved cases for some time period, number of relevant solved cases and number of cases solved by other manners, number of confirmed, repealed, partially repealed, commuted and partially commuted appealed judgments, etc.) and statistical report for a court (e.g. number of judges, number of unresolved cases, received cases and solved cases for some time period, number of relevant solved cases and the number of cases solved by other manners, number of confirmed, repealed, partially repealed, commuted and partially commuted appealed judgments, etc.).

A. Slovakia

In [10], OpenCourts portal www.otvorenesudy.sk is given as an example for re-use of open data published by the judiciary. The portal is initiated by Transparency International Slovakia [22] and its purpose is more transparent and more accountable judiciary. The portal is based on data already publicly available but placed at different government websites and sometimes not easily searchable. OpenCourts portal collect these data and provides them in a user-friendly form for free. Court decisions are published in PDF format while other data (e.g. about courts, judges, proceedings, hearings, etc.) are available in HTML format. Notifications about the presence of new data matching search criteria given by the user are also provided. Therefore, registration is required for the user to receive such notifications by e-mail. In [23], judge rankings are emphasized as a purpose of OpenCourts portal to give public and advocates insight into scores of individual judges. No open license is provided for published data.

B. Croatia

On March 19, 2015. Croatian government launched Open Data Portal [24] for collection, classification and distribution of open data from the public sector. It is a catalog of metadata enabling users to perform a search of public data of interest. It is developed on the basis of open source software, Drupal [25] and CKAN [26], just like UK open data portal [27]. Among published datasets, only a few are available in the legal domain (registers of organizations providing free legal aid, mediators, interpreters, and expert witnesses) mostly in CSV format and some in XML format. The work is licensed under Creative Commons CC BY license [6].

Portal e-Predmet [28] provides public access to court case data of municipal, district and commercial courts in Croatia. Updates of published data are performed on a daily basis and retrieval of case data is based on the court name and the case number. Names of the parties are anonymized while juvenile court cases, investigation cases, war crime cases and the cases under the jurisdiction of The Office for the Suppression of Corruption and Organized Crime are not published at all. Case data are presented in HTML format.

Electronic bulletin board e-Oglasna [29] publishes delivered judgments and other documents from municipal, district, commercial, minor offenses, administrative courts in the Republic of Croatia, Financial Agency enforcement proceedings, and public notaries. Published data are in DOCX or PDF format.

Another open data project in Croatia is Judges Web [30]. It is started by a non-government and non-profit organization consisting of judges and legal experts. Judges Web portal publishes case-law as a collection of selected decisions in HTML format rendered by Croatian municipal and district courts, High Commercial Court of the Republic of Croatia and European Court of Justice. Free of charge user registration is required to access court decisions.

C. Slovenia

The open data portal [31] provides links to available open data in Slovenia and to projects developed on the basis of open data. Judicial data are not currently included.

The case law portal Sodna Praksa [32] publishes selected court decisions delivered by Slovenian courts. Decisions are anonymized and available in HTML format. The portal provides free public access for both commercial and non-commercial purposes while reusing of data is permitted if credits to the Supreme Court of Slovenia are given.

D. Bosnia and Herzegovina

Open data portal [33] publishes government data in Bosnia and Herzegovina. The data in available datasets is mostly data about public finances and, therefore, neither legislation data nor judicial data are available. There is no license information provided on the website.

Judicial Documentation Centre [34] publishes selected decisions from the courts of Bosnia and Herzegovina while access to decision database is charged for public. A special commission of Judicial Documentation Centre performs both selections of decisions for publishing and anonymization of personal data. Documents are available in HTML, DOC, and PDF format. Open license is not provided.

E. Republic of Macedonia

Open data portal of the Republic of Macedonia [35] currently publishes 154 datasets. Portal distinguishes three types of datasets: link (URL to an external web page), file (e.g. DOC, ODS) and database (data downloadable in CSV, Excel and XML format). Datasets published by Ministry of Justice are given as links to web pages related to proposed and adopted laws, bailiffs, mediators, notaries, lawyers who provide free legal aid, interpreters, and expert witnesses. License information is not available on the portal website.

The Supreme Court of Macedonia [36] provides case law database of selected decisions delivered by Macedonian courts. Decisions are anonymized and can be retrieved either in HTML or PDF format. The website does not contain license information.

F. Serbia

The website Portal of Serbian Courts [37] provides public information about court cases. It is adapted version of portal developed for commercial courts during 2007. and 2008. Portal of Serbian Courts started operation on December 17, 2010. and published data about cases of basic, higher and commercial courts. The portal became inactive since December 12, 2013. due to ban pronounced by The Commissioner for Information of Public Importance and Personal Data Protection [38]. The ban was pronounced because Portal was publishing personal data (such as full names and addresses of the parties) without legal grounds. Portal continued with work on February 24, 2014. without personal data included. Since October 9, 2015. data about cases of The Supreme Court of Cassation, The Administrative Court, and appellate courts are also published on the portal. However, data about filings received by the basic, higher and commercial courts still contains names of the parties. Published data are in HTML format. Regarding license information, the Portal of Serbian Courts has “all right reserved” notice.

Legal Information System [39] provides free access to regulations currently in force. Case law database of selected decisions is also available but access is charged for public. Both regulations and court decisions are published in HTML format. Open license is not provided.

G. United Kingdom

The website data.gov.uk [27] helps people to search government data and to understand the working of UK government. Dataset openness is rated by stars: one star for unstructured data (e.g. PDF), two stars for structured data in proprietary format (e.g. Excel), three stars for structured data in open format (e.g. CSV), four stars for linkable data served at URIs (e.g. RDF) and five stars for linked data with links to other data. Considering legal domain, UK legislation is marked as unpublished while referencing to the website [40] is given. The license information is available for every dataset and most of them are available under Open Government License (OGL) [41]. This license allows copying, publishing, distribution and adapting of information for commercial and non-commercial purposes only if attribution statement is specified.

The official website of UK legislation [40] publishes original (as enacted) and revised versions of legislation. Public access to legislation is free of charge while legislation is available in HTML, PDF, XML and RDF formats. Bulk download of legislation is also provided. All legislation is published under Open Government License (OGL) except if stated otherwise.

The website of British and Irish Legal Information Institute (BAILII) [42] provides access to the database of British and Irish case law and legislation. Anonymization of personal data found in court decisions is performed by the court of its origin. Documents are available in HTML format while some of them also have RTF or PDF version. Access to the website is public and free of charge. It is allowed to copy, print and distribute published material if BAILII is identified as a document source.

H. United States

The website CourtListener [43] provides free access to legal opinions from federal and state courts. Containing millions of legal opinions, it is a valuable source for academic research. After specifying queries of interest, CourtListener provides e-mail alerts which notify users if new opinions matching given query appear. Besides legal opinions, CourtListener also collects other data: oral arguments (as audio data), dockets and jurisdictions. All of these data are available for download as bulk data files. All data are serialized in JSON format (for oral arguments referencing to audio files is performed). Citations between opinions are also provided for bulk download as pairs of document identifiers in CSV format. Data are in public domain and free of copyright restrictions as indicated by Public Domain Mark [44].

Using Global Open Data Index methodology, summary assessment of judicial data openness for selected countries is given in Table I.

Most judicial portals lack data in machine-readable formats. Bulk data might not be practical in the case of court decisions because it results in enormous data sizes. Another issue is publishing on an up-to-date basis. Manually performed time-consuming activities, such as anonymization of personal data, may prevent publishing on a daily basis. Additionally, the practice of publishing only selection of court decisions should also be considered when questioning data existence. Analyzing case studies given in this paper, some guidelines could be proposed. Anonymization is recognized as the most common solution for personal data protection. Instead of publishing court decisions in either HTML or PDF format only, some machine-readable XML format should be adopted (e.g. Akoma Ntoso [45], OASIS LegalDocML [46], CEN Metalex [47], etc.). Also, along with simple CSV format,

TABLE I.
SUMMARIZED OPEN JUDICIAL DATA ASSESSMENT FOR SELECTED COUNTRIES

Country	Dataset	Data exists	Digital form	Publicly available	For free	Online	Machine readable	In bulk	Open license	Timely & up-to-date	Score
UK	Receipted documents data	-	-	-	-	-	-	-	-	-	0
	Case register data	5	5	5	15	5	-	-	30	10	75
	Delivered decisions	5	5	5	15	5	-	-	30	10	75
Slovakia	Receipted documents data	-	-	-	-	-	-	-	-	-	0
	Case register data	5	5	5	15	5	-	-	-	10	45
	Delivered decisions	5	5	5	15	5	-	-	-	10	45
Croatia	Receipted documents data	-	-	-	-	-	-	-	-	-	0
	Case register data	5	5	5	15	5	-	-	-	10	45
	Delivered decisions	5	5	5	15	5	-	-	-	-	35
US	Receipted documents data	-	-	-	-	-	-	-	-	-	0
	Case register data	5	5	5	-	5	15	-	-	10	45
	Delivered decisions	5	5	5	-	5	-	-	-	10	30
Serbia	Receipted documents data	-	-	-	-	-	-	-	-	-	0
	Case register data	5	5	5	15	5	-	-	-	10	45
	Delivered decisions	5	5	5	-	5	-	-	-	-	20
Slovenia	Receipted documents data	-	-	-	-	-	-	-	-	-	0
	Case register data	-	-	-	-	-	-	-	-	-	0
	Delivered decisions	5	5	5	15	5	-	-	30	-	65
Bosnia and Herzegovina	Receipted documents data	-	-	-	-	-	-	-	-	-	0
	Case register data	5	5	-	15	5	-	-	-	10	40
	Delivered decisions	5	5	5	-	5	-	-	-	-	20
Macedonia	Receipted documents data	-	-	-	-	-	-	-	-	-	0
	Case register data	-	-	-	-	-	-	-	-	-	0
	Delivered decisions	5	5	5	15	5	-	-	-	-	35

suitable XML format for court case records could be proposed (e.g. OASIS LegalXML Electronic Court Filing [48], NIEM – Justice domain [49], etc.).

In [50] are given guidelines for opening sensitive data such as data in the judiciary. First, some issues are identified that should be considered before opening data, also some alternatives to completely opening data are suggested and solutions to some issues are proposed. These guidelines are based on analysis of datasets used by Research and Documentation Center (WODC [51]) in the Netherlands. Since these datasets contains crime-related data some directions are established in order to decrease the risk of privacy violation. Therefore, three types of access (open access, restricted access and combined open and restricted access) are suggested. Open access may involve anonymization of personal data because revealing identities through a combination of several datasets should be avoided. Restricted access is an option if data producers want to provide access to data depending on its type, type of user and the purpose of use. The combination of open access and restricted access is suitable when datasets contain both privacy-sensitive and non-privacy-sensitive data. Instead of rigidly closing data, proposed directions gives an alternative and represents general principles since various people in various institutions may interpret it differently.

IV. CONCLUSIONS

In this paper, judicial data, as a special case of open government data is analyzed. First, definitions of some elementary terms related to open government data were given. Then, several methods for evaluation of open government data are reviewed and open judicial data from different countries along with their publishing policies are presented and discussed. At the end, some issues were identified and their solutions are proposed.

Since judiciary is one of three government branches, it is very important to adequately open those datasets. On the other hand, opening judicial data is a challenge with respect to personal data protection acts. There is no universal recipe for opening judicial data because different governments have different approaches to privacy protection.

Considering open judicial datasets discussed in this paper, CourtListener stands out by going further than other judicial portals and offers even court decisions in bulk. Although its size causes some problems, it represents a valuable source for researchers.

Along with publishing open dataset, data mining, and reporting projects would help people understand benefits of open government data. Good opportunities for such promotion of open data are hackathons (e.g. International Open Data Hackathon [52]), where participants interested in open data brainstorm project ideas, share suggestions or creative solutions. For government institutions, it is also a communication channel with data users and a way to get feedback on published datasets.

Developing standardized data structures suitable for judicial data is one direction of future work. It should be performed in order to achieve interoperability with existing software solutions and proposed software tools for judicial data processing. Such tools would be particularly useful to people who are not technically skilled but are interested in using open data.

On the top of open judicial data, development of various services could be achieved and therefore features such as transparency in court proceedings, fight against corruption and protection of the right to trial within a reasonable time would be enabled.

REFERENCES

[1] Open Knowledge Foundation, “The Open Data Handbook”, http://opendatahandbook.org/ (last accessed on January 14, 2016)

[2] Open Knowledge Foundation, “The Open Definition”, http://opendefinition.org/ (last accessed on January 14, 2016)

[3] The Open Knowledge foundation, “What is Open Government Data”, http://opengovernmentdata.org/ (last accessed on January 14, 2016)

[4] J. Tauberer, and L. Lessig, “The 8 principles of Open Data Government”, Open Government Working Group, 2007, http://opengovdata.org/ (last accessed on January 14, 2016)

[5] J. Gray, and H. Darbishire, “Beyond Access: Open Government Data & the Right to (Re) Use Public Information”, Access Info Europe and Open Knowledge, 2011

[6] Creative Commons, “About the Licenses”, http://creativecommons.org/licenses/ (last accessed on January 14, 2016)

[7] H. Abelson, B. Adida, M. Linksvayer, and N. Yergler, “ccREL: The creative commons rights expression language, Version 1.0”, Creative Commons, 2008, https://wiki.creativecommons.org/ images/d/d6/Ccrel-1.0.pdf (last accessed on January 14, 2016)

[8] P. Miller, R. Styles, T. Heath, “Open Data Commons, a License for Open Data”, In: Proceedings of the 1st International Workshop on Linked Data on the Web (LDOW), 17th International World Wide Web Conference, Beijing, China, 2008

[9] Open Data Commons: “Public Domain Dedication and License (PDDL)”, http://opendatacommons.org/licenses/pddl/1.0/ (last accessed on January 14, 2016)

[10] K. Granickas, “Topic Report: Open data as a tool to fight corruption”, 2014, https://ofti.org/wp-content/uploads/2014/05/ 221171136-Open-Data-as-a-Tool-to-Fight-Corruption.pdf (last accessed on January 14, 2016)

[11] OpenCourts portal, http://www.otvorenesudy.sk/ (last accessed on January 14, 2016)

[12] A. Conroy, and T. Scassa, “Promoting Transparency While Protecting Privacy in Open Government in Canada”, Alberta Law Review, 53(1), 175, 2015

[13] A. Laudi, “The semantic interoperability centre europe: Reuse and the negotiation of meaning”, Interoperability in Digital Public Services and Administration: Bridging E-Government and E-Business, 2010, pp. 144-161

[14] Legal thesauri, Wolters Kluwer Deutschland, http://vocabulary.wolterskluwer.de/ (last accessed on March 20, 2016)

[15] T. F. Gordon, “The legal knowledge interchange format (LKIF)”, Estrella deliverable d4.1, Fraunhofer FOKUS Germany, 2008

[16] A. Wyner, and R. Hoekstra, “A legal case OWL ontology with an instantiation of Popov v. Hayashi”, Artificial Intelligence and Law, vol. 20.1, 2012, pp. 83-107

[17] M. Ceci, A. Gangemi, “An OWL ontology library representing judicial interpretations”, Semantic Web, Preprint, 2014, pp. 1-25

[18] Open Knowledge Foundation, “Global Open Data Index”, http://index.okfn.org/ (last accessed on January 14, 2016)

[19] J. Tauberer, “Open Government Data: The Book, 2nd edition”, 2014, https://opengovdata.io/ (last accessed on January 14, 2016)

[20] World Wide Web Foundation, “Open Data Barometer, 2nd edition”, http://www.opendatabarometer.org/ (last accessed on January 14, 2016)

[21] “The Court Rules of Procedure”, Official Gazette of the Republic of Serbia, no. 104/2015

[22] Transparency International Slovakia, http://www.transparency.sk/ (last accessed on January 14, 2016)

[23] S. Spáč, “Identifying Good and Bad Judges: Measuring Quality and Efficiency of District Court Judges in Slovakia”, ECPR General Conference, Montreal, 2015

[24] Open Data Portal, The Government of the Republic of Croatia, http://data.gov.hr/ (last accessed on January 14, 2016)

[25] Drupal - Open Source CMS, https://www.drupal.org/ (last accessed on January 14, 2016)

[26] CKAN - The Open Source Data Portal Software, http://ckan.org/ (last accessed on January 14, 2016)

[27] UK Open Data Portal, https://data.gov.uk/ (last accessed on January 14, 2016)

[28] e-Predmet, Ministry of Justice of the Republic of Croatia, http://e-predmet.pravosudje.hr/ (last accessed on January 14, 2016)

[29] e-Oglasna, Ministry of Justice of the Republic of Croatia, https://e-oglasna.pravosudje.hr/ (last accessed on January 14, 2016)

[30] Judges Web, http://www.sudacka-mreza.hr/ (last accessed on January 14, 2016)

[31] Open Data Portal, http://opendata.si/ (last accessed on January 14, 2016)

[32] The Case-law Search Engine, the Supreme Court of Slovenia, http://sodnapraksa.si/ (last accessed on January 14, 2016)

[33] Open Data Portal of Bosnia and Herzegovina, http://opendata.ba/ (last accessed on January 14, 2016)

[34] Judicial Documentation Centre, High Judicial and Prosecutorial Council of Bosnia and Herzegovina, http://csd.pravosudje.ba/ (last accessed on January 14, 2016)

[35] Open Data Portal, Ministry of Information Society and Administration of the Republic of Macedonia, http://www.otvorenipodatoci.gov.mk/ (last accessed on January 14, 2016)

[36] Supreme Court of Republic of Macedonia, http://www.vsrm.mk/ (last accessed on January 14, 2016)

[37] Portal of Serbian Courts, http://www.portal.sud.rs/ (last accessed on January 14, 2016)

[38] Commissioner for Information of Public Importance and Personal Data Protection, "Always, on the Portal of Courts as Well, Data Processing Requires Valid Legal Basis", 2014, http://www.poverenik.rs/en/press-releases-and-publications/1730-uvek-pa-i-na-portalu-sudova-za-obradu-podataka-o-licnosti-nuzan-je-valjan-pravni-osnov.html (last accessed on January 14, 2016)

[39] Legal Information System, http://www.pravno-informacioni-sistem.rs/ (last accessed on January 14, 2016)

[40] UK legislation, http://www.legislation.gov.uk/ (last accessed on January 14, 2016)

[41] Open Government License (OGL), http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ (last accessed on January 14, 2016)

[42] British and Irish Legal Information Institute (BAILII), http://www.bailii.org/ (last accessed on January 14, 2016)

[43] Non-Profit Free Legal Search Engine and Alert System, Free Law Project, https://www.courtlistener.com/ (last accessed on January 14, 2016)

[44] Creative Commons, “Public Domain Mark 1.0”, http://creativecommons.org/publicdomain/mark/1.0/ (last accessed on January 14, 2016)

[45] Akoma Ntoso - XML for Parliamentary, Legislative and Judiciary Documents, http://www.akomantoso.org/ (last accessed on January 14, 2016)

[46] OASIS LegalDocML, https://www.oasis-open.org/committees/ legaldocml/ (last accessed on January 14, 2016)

[47] CEN MetaLex - Open XML Interchange Format for Legal and Legislative Resources http://www.metalex.eu/ (last accessed on January 14, 2016)

[48] OASIS LegalXML Electronic Court Filing, https://www.oasis-open.org/committees/legalxml-courtfiling/ (last accessed on January 14, 2016)

[49] National Information Exchange Model - Justice Domain, https://www.niem.gov/communities/justice (last accessed on January 14, 2016)

[50] A. Zuiderwijk, M. Janssen, R. Meijer, S. Choenni, Y. Charalabidis, and K. Jeffery, “Issues and guiding principles for opening governmental judicial research data”, In Electronic Government, Springer Berlin Heidelberg, 2012, pp. 90-101

[51] Research and Documentation Centre of the Dutch Ministry of Justice, https://www.wodc.nl/ (last accessed on January 14, 2016)

[52] International Open Data Hackathon, http://opendataday.org/ (last accessed on January 14, 2016)

Clover: Property Graph based metadata management service

Miloš Simić
Faculty of Technical Sciences, Novi Sad, Serbia
milos.simic@uns.ac.rs

Abstract—**As the file systems continue to grow, metadata search is becoming increasingly important way to access and manage files. Applications are capable to generate huge amount of files and metadata about various things. *Simple metadata* (e.g., file size, name, permission mode), has been well recorded and used in current systems. However, only limited amount of *metadata*, which not record only attributes of entities but also relationships between them, are captured in current systems. Collecting, processing and querying such large amount of files and metadata is challenge in current systems. This paper present Clover, a metadata management service that unifies files/folders, tags, relationships between them and metadata into generic property graph. Service can also be extended with new entities and metadata, by allowing users to add their own of nodes, properties and relationships. This approach allow not only simple operations such as directory traversal and permission validation, but also fast querying large amount of files and metadata by name, size, date created, tags etc. or any other metadata provided by users.**

I. INTRODUCTION

The continuous increase of data stored in the cloud, storage systems, enterprise systems etc. is changing the way we search and access data. Compared to the various database solutions, including the traditional SQL databases [1] and the NoSQL databases [2-4], file systems usually shine in providing better scalability(i.e. larger volume and higher parallel I/O performance). They also provides better flexibility (i.e. supporting both structured and unstructured, as well as non-fixed data schemas). Therefore, a large fraction of existing applications are still using file systems to access raw data. However, with large volumes of complex datasets, the decades-old hierarchical file system namespace concept [5] is starting to show the impact of aging, falling short of managing such complex datasets in an efficient manner, especially when these data comes with some simple metadata. In other words, organizing files (data) in the directory hierarchy can only be effective and efficient for the file lookup requests that are well aligned with the existing hierarchical structures. For today's highly variable data a pre-defined directory structure can hardly foresee, let alone satisfy the ad-hoc queries that are likely to emerge [17]. Metadata can contain user-defined attributes and flexible relationships. Metadata describes detailed information's about different entities like files, folders, users, tags etc. and relationships between them. These information's extend simple metadata which contains attributes from individual entity and basic relationships, to more detail level. Current file systems are not well-suited for search because today's metadata resemble those designed over four decades ago, when file systems contained orders of magnitude fewer files and basic navigation was enough [5]. Metadata searches can require brute-force traversal, which is not practical at large scale. To fix this problem, metadata search is implemented with separate search application, with separate database for metadata as in Linux (locate utility), Apple's Spotlight [6], and appliances like Google [7] or Kazeon [8] enterprise search. This approach have been effective for personal use or small servers, but they face problems in larger scale. These applications often require significant disk, memory and CPU resources to manage larger systems using same techniques. Also these applications must track metadata changes in file system, which is not easy task. Existing storage systems capture simple metadata to organize files and control file access. Systems like Spyglass [10] and Magellan [11] have also been proposed as tools to store and manage these kinds of metadata. While collecting metadata current systems still lack a mechanism to store, process and query such metadata fast. At least some challenges like Storage System Pressure, Effective Processing/Querying, and Metadata Integration should be addressed [12]. The problem with approaches done in the past is that they relied on relational [1] or key-value [14] databases to store and unify metadata. There have been studies that try to fix inefficiency of retrieving and/or managing files, by offering search functionalities from desktop and enterprise systems. For these environments, returning consistent file search results in real-time or near real-time becomes a necessity, which in and by itself is a challenging goal. This paper propose unifying all metadata into one property graph while files remain on file system. All applications and services can store and access metadata by using graph storage and graph query APIs. With this in mind, all applications and services store data on the file system in the same way, and we can further improve performance using optimization techniques for storing data. Complex queries can express easier as graph traversal instead of a join operation in relation databases. Using graph to represent metadata we also gain rapidly-evolving graph techniques to provide better access, speed, and distributed processing.

This paper is organized as follows. Section II present graph model for metadata. Section III present related work. Section IV present system design and implementation, also show used tools. Section V show experimental results. Section VI summarize conclusions and briefly propose ideas for future work.

II. GRAPH-BASED MODEL

Researches already consider metadata as a graph. The traditional directory-based file management generate a

tree structure with metadata stored in *inodes* [15]. Also file system is designed as tree structure. These trees are graphs enriched with annotations that provide more information's. The metadata graph is derived from the *property graph model* [16] (Figure 1), which have vertices (nodes) that represent entities in the system, edges that show their relationships, and properties that annotate both edges and vertices that can store arbitrary information's that user what. These information's are usually stored as properties of vertices or edge in form of *key-value* pair that are usually separated with ':' for example `name: clover, size: 12kb` and so on. **Usually it's not necessary that all vertices or edges contains same set of properties or key-value pairs.**

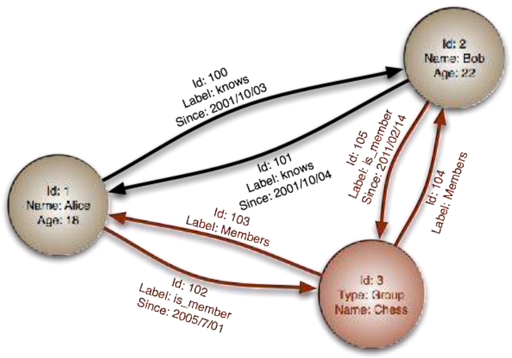


Figure 1. Property graph [28]

A. *Vertices to edges*

Clover define three basic types of vertices, as follow:

Files: represents file on file system, does not contain file content

Folders: represents folder on file system that can contain other folders or files

Tags: represent small metadata information that group other files and/or folders by some (user defined) text. Tags makes filtering easier.

Also users can define their own entities trough APIs for example users or administrators and later on can know which user created some file or folder.

B. *Relationships to edges*

Relationships between entities represent same relationships in file system, and carrying the same semantic. Every file can be *child* of every folder, also every folder can be *child* of every folder in file system structure. Every edge is directed relationship *from child to parent.* Also, relationship between files/folders and tags exist on logic level, and it is not necessarily stored in file system data.

Users are free to add their own relationships and enrich the semantics between data. For example *create* relationship can be added and we can know which user create some edge.

C. *Properties*

Vertices and their relationships have annotations on them. In graph model these annotations are stored as *properties.* These properties are attached to vertices and/or relationships in key-value pars. There is none predefined properties for vertices or relationships, and user add them. Limitation is, that key of every property must be unique in every node/relationship. It can be added more restrictive rule that values for every relationship/edge must be unique like in relation database. Users can always extend model with new properties and enrich semantic of model.

Properties are usually used to select or query specific edges and relationships from others. Examples are name: clover, type: python and so on.

III. Related Work

There is dozens of solutions that have been proposed to fix the inadequacy of file systems in fast file retrieval and filtering, to some extent. These solutions can be broadly divided into the three categories [17]:

File search engines, which rely on the crawling process to catch up with new updates periodically, are unlikely to keep the file index always up-to-date [10-12]. Because of its nature of periodically updates these kind of systems can lead to inaccurate retrieval results. None of the existing file-search engines is designed for large-scale systems. Some of these engines are Apple Spotlight [6], Google Desktop search [7], Microsoft Search [8].

Database-based metadata services use databases as a additional metadata service running on top of file systems. These database-based metadata service have the same limitations like every database-based [2, 3] storage solution. Their performance could not match the I/O workloads on file systems [10, 11]. Also, SQL schema is static and it is not suitable for the exploratory and ad-hoc nature of many big data and HPC analytics activities [18, 19].

Searchable file system interfaces provide file search functions directly through the file systems. Research prototypes that attempt to provide such interfaces include HPC rich metadata management systems [13], Semantic File System [20], HAC [21] and WinFS [22], VSFS [17]. All of these systems serve end-user's needs for retrieving files which means that they will try to find the files based on the keywords provided by users, and have very limited support for the metadata query [10, 23]. These queries might not be useful for analytics applications that rely on range queries or multidimensional queries to fetch the desired data. Furthermore, similar to the file-search engines, these systems perform parsing within the systems, which limits the flexibility in handling the high variety of the datasets.

IV. Design and Implementation

Clover is composed of few parts. Main part is *Cover service* which handles all HTTP requests from clients, and response to them. Also, this service do all communication to storage infrastructure and metadata service.

Clover service understand all basic commands on files/folders that are common on every file system and operating system. Supported commands are: create (folders only), rename, copy, move, remove, list. With this

approach, clover is released from any scheduled tasks to update metadata, and potentially show not consistent state. On every command intended to the storage infrastructure, metadata is updated. With this in mind, clover can use some current and/or future algorithms to improve speed of these operations. All of these operations are done through Python modules.

In additions to these, Clover provides tagging, search, filter operations on metadata storage. When files/folders are opened, their rate is updated. If search provide more results than single item, the higher rated files/folders are on top of the list as top hit. Figure 2 show Clover architecture.

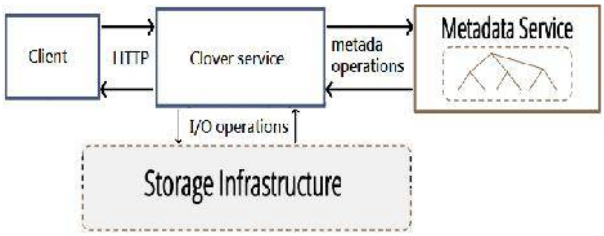


Figure 2. Clover architecture

Metadata service is implemented using graph database [24]. This database store all metadata in form of edges, relationships and properties for every item in storage. Service is also able store information's that exists only on logical level, like who created element, who send item, where is download from etc. and enrich metadata and provide more semantics to it. With this in mind much powerful search can be provided.

Vertices contains at least: name and path to files/folders on storage infrastructure. **Path must be unique**, so *unique* constraint is added to every file/folder vertices. It is recommended, that vertices and/or relationships contains also date created, date modified, last accessed date for better querying, but it is not necessary.

Users are free to extend these properties. Every vertices and relationship, or group of them, can be *labeled* with some free text and make search even sassier and simpler. Metadata service labeled every file with *FILE*, folder with *FOLDER* and tag with *TAG* label to logically distinguish these items, and make querying a lot easier and faster. When file/folder is child of some other folder, that relationship is created and labeled with *CHILD* label. This relationship should have at least *since* property to describe since when files/folders are children of that specific folder. Files and/or folders that are that are tagged are connected to tag vertices over *TAGGED* labeled relationship. Recommendation for *since* property is applied here as well.

Metadata service must provide fast search mechanism, and indexing. Labels are mechanism to relatively fast filter items. This might be acceptable in some cases but if we're going to be looking up some fields frequently, then we'll see better performance if we create an index on that property for label that contains that property. Users can add their own indexes trough clover service APIs. Metadata service provide indexes on *name* property, assuming that file name is used mostly in searching.

Why graph database and not relational database? A graph database... is an online database management system with CRUD methods that expose a graph data model [18]. Two important properties:

- Native graph storage engine: written from the ground up to manage graph data
- Native graph processing, including index-free adjacency to facilitate traversals

The problem with relational approach are joins. All joins are executed **every time** when query is executed, and executing a join means to search for key in another table. With indices executing a join means to lookup a key, B-Tree index speed is O (log (n)).

Graph databases are designed to: store inter-connected data, make it easy to evolve database and to make sense of that data. Enable extreme-performance operations for discovery of connected data patterns, relatedness queries greater than depth 1 and relatedness queries of arbitrary length. People usually use them when have problems with join performance, continuously evolving data set (often involves wide and sparse tables) or the shape of the domain is naturally a graph (like file system).

Early adopters of these databases were Google: Knowledge graph [25], Facebook: Graph search [26]. It show's it is easy to use, it is really fast, and users can query almost on their natural language.

A. Neo4j

Neo4j [27] is used as the database for storing metadata. Neo4j is open source, it has largest ecosystem of graph enthusiast, community is large 1000000+ downloads 150+ enterprise subscription customers including 50+ global 2000 companies (January 2016). Most mature product is in development since 2000, in 24/7 production since 2003.

This database show good connected query performance. Query response time (QRT) [28] is given by formula (1)

```
QRT=f(graph density, graph size, query
            degree)
```
(1)

Graph density is average number of relationships per node

Graph size is total number of nodes in the graph

Query degree is number of hops in ones query

RDBMS has exponential slowdown as each factor increases, and Neo4j performances remains constant as graph size increases. Performance slowdown is linear or batter as density and degree increase.

Neo4j using pointers instead of lookups and doing all joining on creation of vertices and relationships. Also contains *profiler* embedded, and we can detect bottle necks and fix them. Figure 3 show comparison of RDBMS and Neo4j.

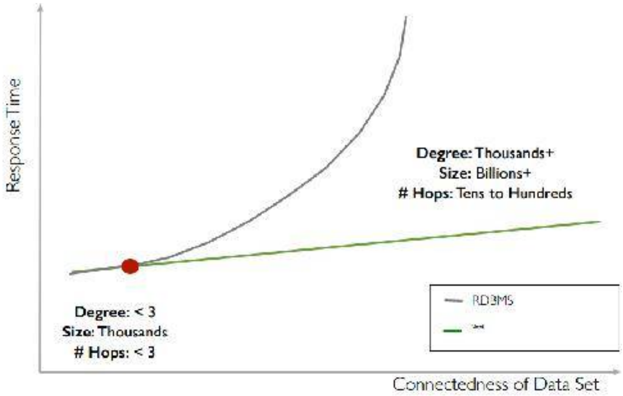


Figure 3. Comparison RDBMS vs Neo4j [28]

Neo4j uses *Cypher* [29] query language. This language can easily mapped graph labels on natural language and make querying a lot easier. For example, if we have two nodes *A*, *B* and both of them contains name property, and one relationship between them labeled with *LOVES*. Simple query to figure out *friends who likes pie*

```
START me = (p:PEOPLE{name:'me'}),
        pie = (t:THING{name:'pie'})
MATCH me-[:FRIEND]-> (friends:PEOPLE),
        friends -[:LIKES]->pie
RETURN people
```

V. Experimental Results

In order to evaluate the performance system presented in Sections 4, search engine was implemented in Python and Neo4j version 2.3.2 for windows. All experiments were made using a 2 GHz Pentium 4 workstation with 4 GB of memory running Windows 7 and Linux. For dataset, Python27 folder is crawled containing 15084 nodes, 15083 relationships and deep recursive structure of sub files and/or folders. No attempts have been made to optimize Java VM (java version "1.8.0_71", SE build 1.8.0_71-b15,), the queries etc.
Experiments were run on Neo4j and MySQL out of the box with natural syntax for queries. The graph data set was loaded both into MySQL and Neo4j. In MySQL a single table.
Figure 4 show comparison results on MySQL, Neo4J, locate and Windows search when searching folder by given name stoneware in *Python27* directory.

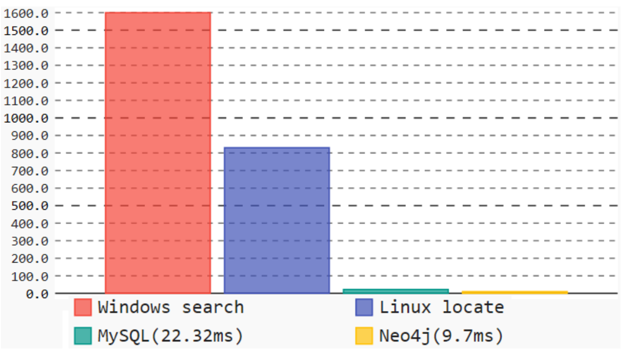


Figure 4. locate, Windows search, MySQL, Neo4J comparison searching folder by given name

Figure 5 show comparison results on MySQL, Neo4j and locate command when searching for child nodes that have **.py* extension of folder by given name stoneware in *Python27* directory.

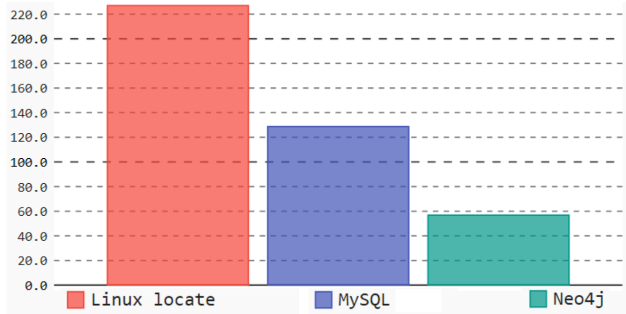


Figure 5. locate, MySQL, Neo4J comparison on searching child nodes that contains *.py as extension of given folder

Figure 6 show comparison results on MySQL, Neo4j and locate when retrieving file/folder attributes for given exact file location.

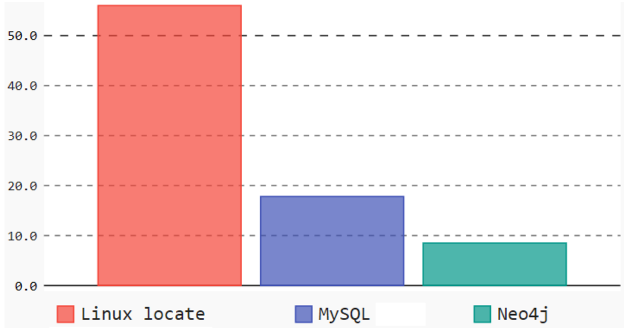


Figure 6. locate, MySQL, Neo4J comparison on retrieving file/folders attributes

Results include time on sending and receiving HTTP requests.

VI. Conclusion

As amount of data and files now days become larger and larger, current systems lack to do fast metadata search. This paper present Clover, a graph-based mechanism to store metadata, and search large-scale systems. Clover model data is in form of property graph, where vertices are presented as edges of graph, and they are connected over relationships. Both vertices and relationships contains properties to more describe them, and give them more semantics to them. These properties are stored in key-value form. Inspiration comes from Facebook and Google which use this approach to enable fast search.

There are numerous ways to improve Clover in future. First to add role-based access control (RBAC) to separate which users can access which files. Second, to improve search by adding *Domain Specific Language* specifically designed for natural language. This will make search even easier. Third, content of text files can be stored in some document database so Clover can search inside content of files. Forth, Clover can be extended with framework to support big-data. Fifth, all operations that

affect storage are currently synchronous. Future work should enable asynchronous operations for every function on file system. This can be handy especially with bigger files and operations that takes a lot of time to be executed (copying or moving big amount of files etc.). Also system should be tested on server configuration with larger amount of files/folders and different kind of not just simple, but also rich metadata by giving more semantic relationships.

REFERENCES

[1] Oracle database. http://www.oracle.com/us/products/database/overview/index.html, accessed 2016.

[2] K. Banker. MongoDB in Action. Manning Publications Co., Greenwich,CT, USA, 2011.

[3] D. Borthakur and et al. Apache hadoop goes realtime at facebook.SIGMOD '11. ACM

[4] G. DeCandia and et al. Dynamo: amazon's highly available key-value store. SOSP '07

[5] Daley R., Neumann P., A general-purpose file system for secondary storage. In *Proceedings of the Fall Joint Computer Conference, Part I* (1965), pp 213-229

[6] Apple Spotlight, https://support.apple.com/en-us/HT204014, accessed 2016

[7] Google inc. Google enterprise, https://www.google.com/work/, accessed 2016

[8] Kazeon, Kazeon search enterprise, http://www.emc.com/domains/kazeon/index.htm, accessed 2016

[9] Microsoft. Windows Search. https://support.microsoft.com/en-us/kb/940157, accessed 2016

[10] A. W. Leung, M. Shao, T. Bisson, S.Pasupathy, E. L. Miller, Spyglass: Fast, Scalable Metadata Search for Large-Scale Storage Systems, in *FAST*, vol 9, 2009, pp. 153-166

[11] A. Leung, I. Adams, E. L. Miller, Magellan: A Searchable Metadata Architecture for Large-Scale File Systems, University of California, Santa Cruz, Tech. Rep. UCSC-SSRC-09-07, 2009.

[12] L. Xu, H. Jiang, L. Tian, and Z. Huang. Propeller: A scalable real-time file-search service in distributed systems. ICDCS '14

[13] D. Dai., R B. Ross, P Carns, D. Kimpe, Y. Chen, Using Property Graphs for Rich Metadata Management in HPC Systems

[14] S. C. Jones, C. R. Strong, A. Parker-Wood, A. Holloway, D. D. Long, Easing the Burdens of HPC File Management, in Proceedings of the sixth workshop on Parallel Data Storage. ACM, 2011, pp. 25-30

[15] J. C. Mogul, Representing Informatinos about Files, Ph.D dissertation, Citeseer, 1986, Texas Tech University

[16] Property Graph, https://www.w3.org/community/propertygraphs/, 2016

[17] L. Xu, Z. Huang, H. Jiang, L. Tian, D. Swanson, VSFS: A Searchable Distributed File System, parallel data storage workshop, 2014

[18] J. Lin and D. Ryaboy. Scaling big data mining infrastructure: the twitter experience. SIGKDD Explor. Newsl., 14(2):6–19, Apr. 2013

[19] A. Pavlo, E. Paulson, A. Rasin, D. J. Abadi, D. J. DeWitt, S. Madden, and M. Stonebraker. A comparison of approaches to large-scale data analysis. SIGMOD '09

[20] D. K. Gifford, P. Jouvelot, M. A. Sheldon, and J. W. O'Toole, Jr. Semantic file systems. In SOSP '91, 1991

[21] B. Gopal and U. Manber. Integrating content-based access mechanisms with hierarchical file systems. In OSDI '99.

[22] Microsoft. WinFS: Windows Future Storage. http://en.wikipedia.org/ wiki/WinFS, accesed 2016

[23] Y. Hua and et al. Smartstore: a new metadata organization paradigm with semantic-awareness for next-generation file systems. In SC '09.

[24] I. Robinson, J. Webber, E. Eifrem, Graph databases, O'Reilly, 2015, ISBN: 9781491930892

[25] Google Knowledge graph, http://www.google.com/intl/es419/insidesearch/features/search/knowledge.html, accessed 2016

[26] Facebook graph search, https://www.facebook.com/graphsearcher/, accessed 2016

[27] Neo4j, http://neo4j.com/, accessed 2016.

[28] Goto conference 2014, http://gotocon.com/dl/goto-chicago-2014/slides/MaxDeMarzi_AddressingTheBigDataChallengeWithAGraph.pdf, accessed 2016

[29] Cypher, http://neo4j.com/docs/stable/cypher-introduction.html, accessed 2016

Optimized CT Skull Slices Retrieval based on Cubic Bezier Curves Descriptors

Marcelo Rudek, Yohan B. Gumiel, Osiris Canciglieri Jr., Gerson L. Bichinho
Pontifical Catholic University of Parana – PUCPR / PPGEPS, Curitiba, Brazil
marcelo.rudek@pucpr.br, yohan_bg@hotmail.com, osiris.canciglieri@pucpr.br, gerson.bichinho@pucpr.br

***Abstract*— The paper presents a method applied to the geometric modelling of skull prosthesis. The proposal deals over definition of the representative descriptors of skull bone curvature based on Cubic Bezier Curves. A sectioning of the bone edge in CT image is performed to create small sections in order to optimize the accuracy of fitness of the Bezier curve. We show that is possible to reduce around 15 times the amount of points from original edge to an equivalent Bezier curve defined by a minimum set of descriptors. The final objective is to apply the descriptors to find similar images from CT databases in order to modelling customized skull prosthesis. A study case shows the feasibility of method.**

I. Introduction

Among several applications involving image processing and automated manufacture, the medical context represents new challenges in engineering area. In the context of machining process, the 3D printers are capable to build complex structures in different materials geometrically compatible with tissues of human body.

The congenital failure or trauma in skull bone require surgical procedures to prosthesis implant as functional or esthetical repairing. In this process, a customized piece built according individual morphology is an essential requirement. Normally in bone repairing, the geometric structure is unrepeatable due its “free form” [1]. Due the complexity in geometry, the free form objects does not have a mathematical expression in a close form to define its structure. However, numerical approximations are feasible way to the geometric representation. The link between the medical problem and the respective manufactured product (i.e. prosthesis) is the geometric modelling and the different approaches in bone modelling have opened new researches interest as in [2],[3] and [8].

In prosthesis modelling, we face different levels of information handling, from the low level of pixel analysis in image to the automated production procedure. In general, there are the following levels: a “preparation level” (containing CT scanning, segmentation, feature extraction, i.e. entire image processing) and a “geometric modelling level” (containing the polygonal model, curve model, extraction of anatomic features, i.e. entire CAD based operations) [2].

In our strategy, we need to generate the geometric representation of bone without enough information (e.g. no mirroring or symmetry applicable). A common image segmentation procedure is executed as pre-processing in the preparation level. Moreover, from segmented images we defined a set of descriptors based on Bezier Curves [7] in order to describe the geometry of skull edge on a CT image. This approach is applied with objective to reduce the amount of points capable to represent de skull bone curvature. It was adapted from method of [8] now using the *de Casteljau* algorithm to define the Bezier parameters. The paper explores the accuracy of prosthesis modelled through the balanced relationship between curve fitness versus number of descriptors.

II. Proposed Method

A. The conceptual proposal

In our study, the main question is related to the information recovering for automation of the prosthesis modelling process. Sometimes it is possible to reconstruct a fragmented image by using information of same bone structure, e. g. by mirroring using body symmetry from same individual. However, in many cases, there are not enough information to be mirrored. A handmade procedure can be performed by a specialized doctor by using a CAD system [4], [5].

In order to circumvent mirroring limitations and user intervention, we are looking for by an autonomous process to geometric modelling of skull prosthesis. Thus, the basis of our hypothesis is to find compatible information from different healthy individuals from image database. The problem addressed here is the method to find a compatible intact CT slice to replace the respective defective CT slice.

When working with medical images, it exists a lot of information to processing [6]. After image segmentation and edge detection, the total of pixels in edge are still to much information for processing. Our approach is a content-based retrieval procedure and a pixel-by-pixel comparing it is a hard processing task we need avoid.

In order to optimize the search by similarity, we propose to define shape descriptors by Cubic Bezier Curves. In this way it is possible to reduce the amount of data to processing to a few parameters. Then, the important question is to find these descriptors capable to describe the edge shape as better as possible. Thus, we also look for by a balance between accuracy and the minimum quantity of information. The next section will explain about our approach in curve modelling.

B. The Curve Modelling

The curve modelling adopted in this research is based on *de Casteljau* algorithm applied in calculation of the points of a *Bézier Curve* [7], [8]. The *de Casteljou* method [9],[10] operates by the definition of a “control polygon” whose vertices are the respective “control points” (or “anchor points”) used to define the shape of the Bezier Curve. A Bezier curve of degree n is built by $n+1$ control points. The Cubic Bezier Curve have two endpoints (fixed points) and two variable points. They define the shape (flatness) of curve. The Figure 1 show an example of a *Cubic Bezier*

Curve where P_0, P_1, P_2, P_3 are the vertices of the control polygon. The points P_0, P_3 are the fixed points and they are respectively the begin and end of the curve; - these points belong to the curve. The P_1, P_2 are the variable points and they occupied any random position in $\mathbb{R}^2$.

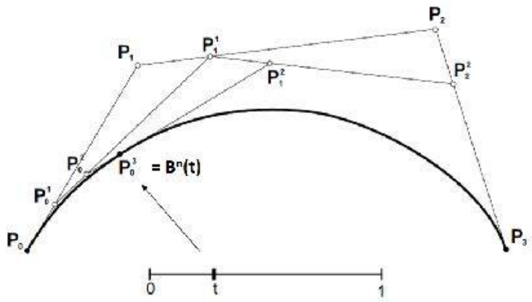


Figure 1. Graphical Representation of *de Casteljau* method. Adapted from [10].

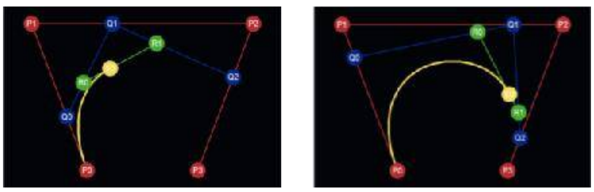
According (1), for all points $P_i^r(t) = P_i$, we have a $P_0^n(t)$ as a point on the Bezier curve. The Bezier curve $B^n(t)$ with degree 'n' is a set of points $P_0^n(t)$, $t \in [0,1]$, i.e. $B^n(t) = \{P_0^n(t); t \in [0,1]\}$. Then the polygon formed by the *n* vertices $P_o, P_1, \dots, P_n$ is so called "control polygon" (or Bezier polygon) [10].

Through the *de Casteljau* algorithm each line's segment results in $(n-1)$ baselines as $\overline{P_0P_1}$, $\overline{P_1P_2}$, $\overline{P_2P_3}$ which are recursively divided to define a new set of control points. By changing the 't' value as defined in (2) we obtain the position of the point in the curve.

$$P_i^r(t) = (1-t)P_i^{r-1}(t) + tP_{i+1}^{r-1}(t), \quad \begin{cases} r = 1,2,\dots,n \\ i = 0,1,\dots,n-r \end{cases} \tag{1}$$

$$t = \frac{t_1 - t_0}{t_2 - t_0} \tag{2}$$

The control points for $P[t0\ t1]\ (t)$, are $P_0^0, P_0^1, P_0^2, \dots, P_0^n$, and the control points for $P[t1\ t2]\ (t)$ are $P_0^n, P_1^{n-1}, P_2^{n-2}, \dots, P_n^0$. In order to avoid misunderstanding in representation, the figure 2 shows the control points, and recursive subdivision of *de Casteljau* algorithm labelled as P, Q, R and S, where S is the final position of a point in the curve for different values of *t*. In figure 2.a the t = 0.360 and in figure 2.b the value of t = 0.770.

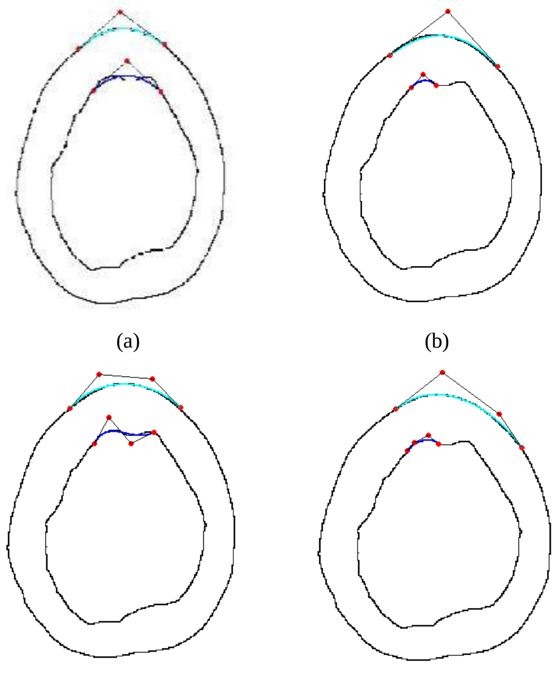


(a) (b)

Figure 2. Position of the control points and its respective Bezier Curve adapted from [9][1].

As presented in literature, for practical applications, the most common is to apply the Cubic Bezier Curves (*n*=3) due large possibilities in adapting shape (flatness) according our necessities. Also in our proposal, the Bezier with *n*=3 is more suitable to fit the skull contour in tomographic cuts. A graphical example of a segmented CT slice is shown in figure 3. In figure 3.a is presented a Quadratic Bezier Curve ($B^n(t)$ with n = 2) adjusted on skull edge. In this case we have three control points and only two baselines. Note that the adjustment in outer edge seems satisfactory, but in inner edge, the result is weak. By the same way, in figure 3.b was applied the *de Casteljau* algorithm to a smallest segment of inner edge; then in this case the curve representation was improved. In figure 3.c is presented a *Cubic Bezier Curve*. Now it exists more control points and as result the adjust looks like very good for the both outer and inner edge. Also, in figure 3.d the method applied in a small section (the inner edge) is more accurate.

The question that we intend to discuss in next section is about the similarity measurement. In other words, how good is the quality of a Bezier curve that represents a CT skull edge? This is essential in our approach, because we need to define the best curve based descriptor. The good descriptors will permit us to retrieve compatible CT images to produce the skull prosthesis.



(a) (b) (c) (d)

Figure 3. A CT slice sample and respective Bezier representation. (a) The Quadratic Bezier Curve (n=2). (b) The Quadratic Bezier Curve on small region. (c) The Cubic Bezier Curve (n=3). (d) The Cubic Bezier Curve on small region.

III. Application of the Method

The aim of this research is to define a small set of descriptors to represent the bone curvature. The strategy is

---

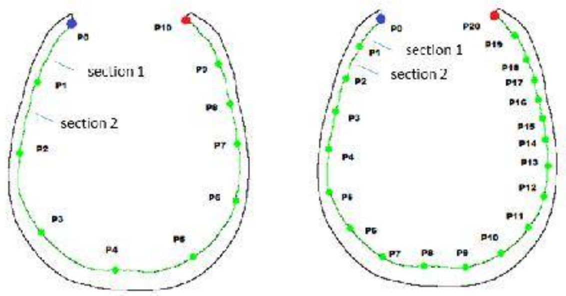
[1] *under an Attribution-NonCommercial-ShareAlike CC license (CC BY-NC-SA 3.0)*

to use the Cubic Bezier Curve method calculated through *de Casteljou* algorithm.

In our previous section we presented that the accuracy of curve fitting in our approach by Bézier depends of its degree *n* and the length of edge section to be reproduced.

A. *The Secctioning of Edge*

As before presented in figure 3, the curve generated on the edge seems to fit better to smallest length region (i.e., the shape of curve looks like similar with original edge shape). The first question is about the best number of sections to produce the best-fitted curve. As an example, the edges of a CT image can be sectioned as figure 4.



(a) (b)

Figure 4. A CT slice sample with (a) k=10 sections; (b) k=20 sections.

Figure 4.a show the total of *k*=10 cuts (with P_0 to P_{10} fixed points) whose section edges lengths are bigger than sections of figure 4.b with *k*=20 cuts (with P_0 to P_{20} fixed points). For each section a fitness value *F* is calculated trough (3), defined in [8] as:

$$F(k) = \sum_{i=1}^{n} \sqrt{\left(x_B(i) - x_O(i)\right) + \left(y_B(i) - y_O(i)\right)} \qquad (3)$$

Where *F*(*k*) is the fitness value to each section *k*. The fitness *F* calculates the error between Bezier coordinates (x_B, y_B) and original edge coordinates (x_O, y_O) for each pixel '*i*' in the edge. The sectioning procedure and control points calculation are fully covered in [8]. The table 1 shows the average of fitness (error) to the respective 5, 10, 15 and 20 sections cuts.

Table 1 shows the cumulative error evaluated by (3) and the average of fitness for different values of sectioning. As expected, the error is minimized with larger values of *k*. The graph in figure 5 presents the relationship between number of sections and calculated error (difference between original edge and calculated Bézier).

As presented in figure 5, the average of error calculated from fitness equation goes down according the number of sections are increased. Then, in this condition, maybe we could define the *k* value as the maximum possible, i.e. the length of total of pixels of edge. However, the computational cost to Cubic Bézier Curve calculation for hundreds of sections is also increased. The same proportion of error occurs for all CT slices from different images. From the graph, selecting the value of *k*=20 is enough to match a relative good fitness with small error and give us an adequate balance between precision and computational cost. Thus, for *k*=20 we have in *de Casteljau* algorithm, 20 "fixed points" and another 20 "variable points", (i.e. 4 points per section) calculated as in [8]. Now, it is possible represent the total length of each edge (inner and outer) in a CT slice with 80 points descriptors each instead ≈ 1250 in original edge (around 15 times information reduced).

Table 1. Relationship between number of cuts and respective fitness (error).

# of Section	*F*(5)	*F*(10)	*F*(15)	*F*(20)
1	61.81640	26.03820	8.96430	6.22810
2	87.16330	22.63830	17.94760	9.27160
3	99.14170	21.17040	20.07230	9.34800
4	78.18280	25.94340	16.91680	10.25800
5	68.10270	26.46990	18.00790	10.09840
6	-	30.86040	17.19840	9.31330
7	-	28.55470	19.37670	14.06260
8	-	24.36930	22.17160	9.27060
9	-	36.52540	17.10820	9.50420
10	-	48.82190	19.46810	11.26640
11	-	-	19.99220	12.80600
12	-	-	21.37690	7.48890
13	-	-	18.15550	10.65490
14	-	-	27.96400	12.36890
15	-	-	40.85040	8.37890
16		-	-	9.40440
17	-	-	-	9.74180
18	-	-	-	17.47880
19	-	-	-	15.90830
20	-	-	-	25.4706
Σ (error)	394.40690	291.39190	305.57090	228.32270
Fitness (avg.)	78.88138	29.13919	20.37139333	11.416135

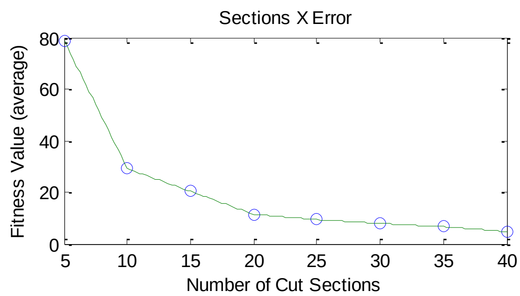


Figure 5. The relationship between the number of cuts in the edge and respective error from fitted Bézier curve in each section.

B. *Compatible CT slice Recovering*

The curve fitting procedure is applied on each CT slice of defective skull. The same procedure is also applied on each searching image on database. The compatible answer image is retrieval as example presented in figure 6.

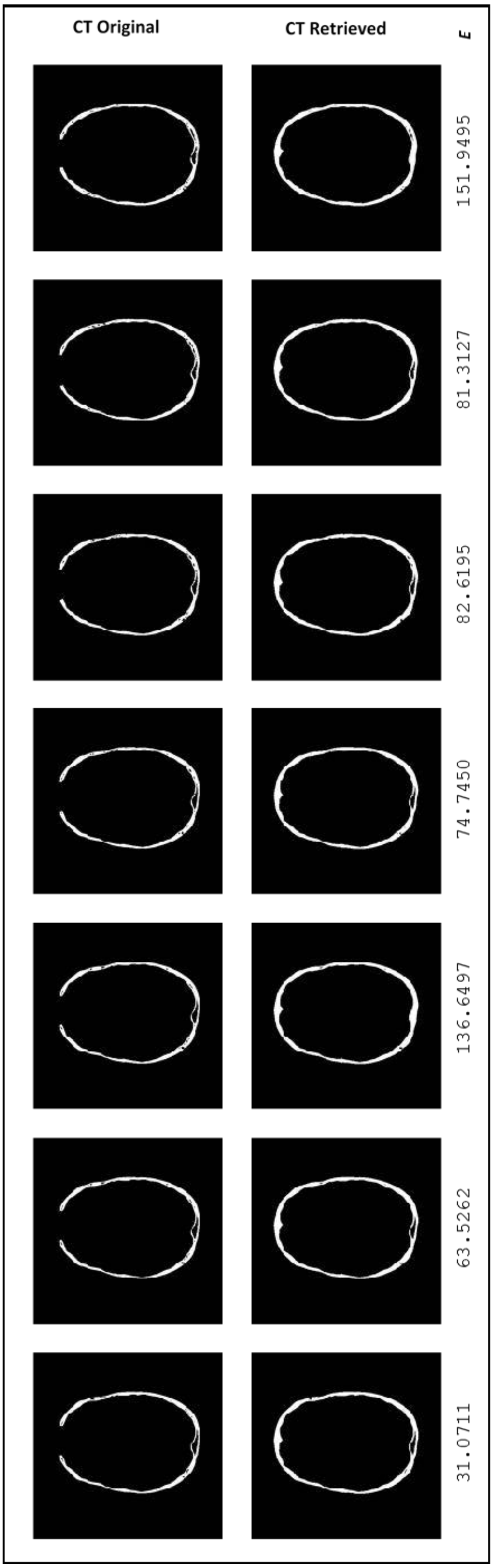
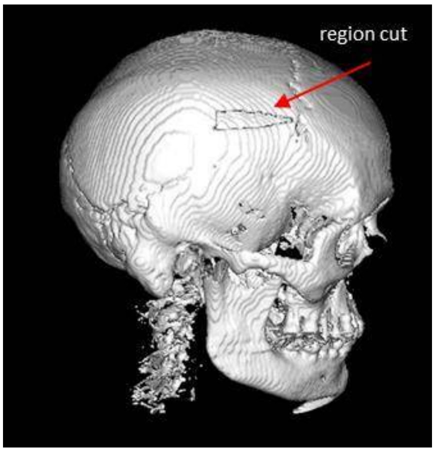


Figure 6. The defective set of slices and respective compatible CT recovered from medical image database.

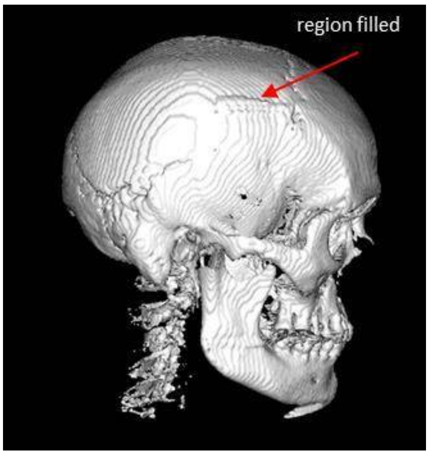
In figure 6 it is shown some samples of defective CT from the original dataset and respective retrieved CT with compatible descriptors (*i. e.*, minimum error in descriptors). The figure 6 also shows the error value (E) for each images pairs. The error is the cumulative difference between original bone and calculated Bezier curve by applying (3).

IV. 3D Example from Recovered Data

A handmade testing failure was built in a skull trough the FIJI software [11]. It is an open source Java suite to medical image analysis. A set of toolboxes permits to handle the CT slices from DICOM file [6]. The edge from individual slices can be cut in sequence to build a failure on a region. Thus, after 3D reconstruction it is obtained a synthetically built failure in skull as the example in figure 7. a.



(a)



(b)

Figure 7. Testing image. (a) A handmade testing failure built on original image. (b) Region filled with prosthesis modeled.

In figure 7.b is presented the failure region filled with compatible CT slice from medical database. The piece was cut from different slices where Bezier descriptors were compatible, i.e. those all CT retrieved with minimum error. The retrieved slices numbers and respective patient are shown in table 2. Note that the slices retrieved is not ever from same patient. The set of retrieved slices (good slices) were recovered from healthy individuals (intact skull) whose descriptors matches with original image in each CT slice (defective slice). From table 2 it is possible to see that many CTs are coming from individual #6.

In fact, the individual #6 have similar morphological characteristic with testing patient, as similar age and gender.

Table 2. Retrieved set of CT slices.

Original Slice #	Compatible individual #	Compatible CT image #	Fitness
279	6	280	128.0724
280	7	278	130.2566
281	5	279	118.3946
282	6	285	143.4175
283	5	287	176.9402
284	6	282	128.9885
285	7	281	179.4653
286	6	285	139.1415
287	6	288	127.1768
288	5	284	189.7201
289	6	287	132.4534
290	7	286	226.1416
291	6	289	137.2678
292	6	290	140.9704
293	5	290	113.8132

The filled region was evaluated through Geomagic® software [12]. The differences between original bone and prosthesis piece are presented in figure 8. The software permits to overlap 3D structures and provide a colored scale to show spatial the difference, whose the lower error values are represented in green and the higher error tends to red. As shown in figure 8 the region A008 has error closest to zero because is the original skull (skull of patient). The other colored identifications are into prosthesis area like A001, A003, A005 and A006; they are below of 1mm difference and the maximum error is in region A004 with value about 1.7087 mm.

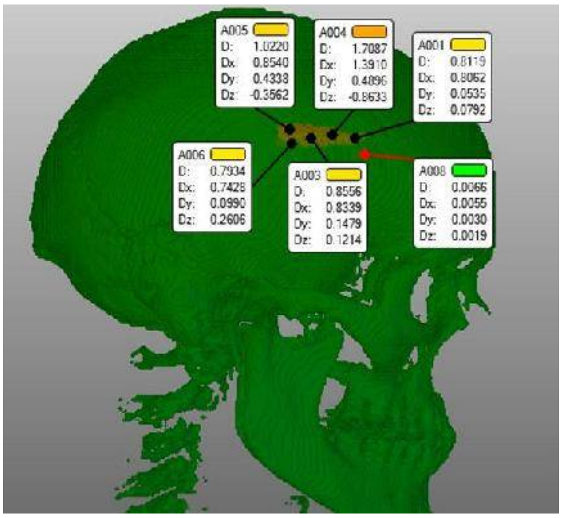


Figure 8. 3D evaluation of filled region.

V. Conclusions

The paper presented a method to generate skull shape descriptors based on Bezier Curves whose parameters were generated by *de Casteljou* algorithm. A sectioning of edge in k=20 sections with same length permits define two markers as the respective "fixed points" in Bezier curve generator. More two "variable points" calculated by *de Casteljau* defines the total of 4 descriptors for each section. Thus, it is possible to reduce all edge size in CT to be represented by a set of 80 descriptors. The descriptors are used to look for compatible CT images whose bone edge shape versus Bezier curve calculated by their descriptors have a minimum error. The example shows the result with maximum error in image around 1.7mm. We show it is possible represent a missing region of a patient's skull by a set of similar CT from healthy individuals selected by a reduced descriptors group.

In addition, the example shows that retrieved slices are from individuals with similar characteristic as age and gender. In a future work, the database searching engine can group individuals with these characteristics before proceed with descriptors calculation.

Acknowledgment

The author would like to thanks the Pontifical Catholic University of Parana – PUCPR trough its respective Graduate Programs PPGEPS and PPGTS by collaboration in providing all master student support and CT image data.

References

[1] M. Trifunovic, M. Stojkovic, M. Trajanovic, M. Manic, D. Misic, N. Vitkovic, "Analysis of semantic features in free-form object reconstruction", *Artificial Intelligence for Engineering Design, Analysis and manufacturing*, 2015, pp.1-20.

[2] V. Majstorovic, M. Trajanovic, N. Vitkovic, M. Stojkovic, "Reverse engineering of human bones by using method of anatomical features", *CIRP Annals – Manufacturing Technology*, vol. 62, 2013, pp.167-170.

[3] M. Rudek, Y. B. Gumiel, O. Canciglieri Jr., G. L. Bichinho and M. da Silveira, "Anatomic Prosthesis Modelling Based on Descriptors by Cubic Bézier Curves", *CIE45 Proceedings*, Metz / France, 2015, pp. 1-8.

[4] Mimics, "Medical Image Segmentation for Engineering on Anatomy", available in http://biomedical.materialise.com/mimics, accessed in JAN. 2016.

[5] Osirix, "Osirix Imaging Software – Advanced Open-Source Pacs Workstation DICOM Viewer", available in: http://www.osirix-viewer.com/, accessed in JAN. 2016.

[6] Dicom, "Digital Imaging and Communications in Medicine Part 5: Data Structures and Encoding", Available online at < www.medical.nema.org/dicom>, Accessed DEC. 2015.

[7] L. J. Shao, H. Zhow, "Curve Fitting with Bezier Cubics". *Graphical Models and Image Processing*, vol. 58(3), 1996, pp.223-232.

[8] M. Rudek, Y. B. Gumiel, O. Canciglieri Jr., "Autonomous CT Replacement Method For The Skull Prosthesis Modelling", *Facta Universitatis, series: Mechanical Engineering*, vol. 13, n. 3, 2015, pp. 283-294.

[9] M. Christersson, "De Casteljau's Algorithm and Bézier Curves", in: http://www.malinc.se/m/DeCasteljauAndBezier.php, 2014.

[10] R. Simoni, "Teoria Local das Curvas", monografia de conclusão de curso, UFSC, 2005, (*in portuguese*).

[11] J. Schindelin et al, "Fiji: an open-source platform for biological-image analysis", *Nature Methods* vol. 9, n.7, 2013, pp. 676-682.

[12] Geomagic, "Modeling Easter Island's Moai with Geomagic 3D Scan Software", available in http://www.geomagic.com/en/. Accessed DEC. 2015.

Enhancing Semantic Interoperability in Healthcare using Semantic Process Mining

Silvana Pereira Detro[*,**], Dmitry Morozov[*], Mario Lezoche[*], Hervé Panetto[*], Eduardo Portela Santos[**], Milan Zdravkovic[***]

[*] Research Center for Automatic Control of Nancy (CRAN), Université de Lorraine, UMR 7039, Boulevard des Aiguillettes B.P.70239, 54506 Vandoeuvre-lès-Nancy, France.

[**] Graduate Program in Production Engineering and Systems (PPGEPS), Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, Paraná, Brasil

[***] University of Nis, Mechanical Engineering department

{silvana.pereira-detro, dimitrii.morozov, mario.lezoche, herve.panetto}@univ-lorraine.fr, eduardo.portela@pucpr.br, milan.zdravkovic@gmail.com

Abstract—**Semantic interoperability plays an important role in healthcare domain, essentially it concerns the action of sharing the meaning between the involved entities. The enterprises store all the execution processes data as event log files. The process mining method is one among the possible methods that enable the processes analysis behavior in order to understand, optimize and improve them. However, the standard process mining approaches analyze the process based only on the event log label strings, without consider the semantics behind this label. A semantic approach on the event logs might overcome this problem and could enable the use, reuse and sharing of the embedded knowledge. Most of the research developed in this area focuses on the process dynamic behavior or in clarifying the meaning of the event log label. Therefore, less attention has been paid in the knowledge injection perspective. In this context, the objective of this paper is to show a procedure, in its preliminary state, to enhance the semantic interoperability through the semantic enrichment of event logs with domain ontologies and the application of a formal approach, named Formal Concept Analysis.**

I. Introduction

Healthcare organizations are under constant pressure to reduce costs while delivering quality care to their patients. However, this is a challenge task due the characteristics of this environment.

Healthcare practices are characterized by complex, non-trivial, lengthy in duration, diverse and flexible clinical processes in which high risk and high cost activities take place and by the fact that several organizational units can be involved in the treatment process of patients [1], [2].

In this environment organizational knowledge is necessary to coordinate the collaboration between health care professionals and organizational units.

The knowledge is the most important asset to maintain competitiveness. Defined from different points of view, in this research we consider that the knowledge is composed by data and/or information that have been organized and processed to convey the understanding, the experience, the accumulated learning, and the expertise as they are applied to a current problem or activity [3].

The knowledge representation is the result of embodying the knowledge from its owner's mind into some explicit form. It enables external agents to perform some specific operations for achieving their particular needs. The knowledge representations act as the carriers of knowledge to assist collaboration activities [4].

Interoperability is the ability of two or more systems to exchange information and to use the information that have been exchanged [5], thus supporting collaboration. This research has as focus the semantic interoperability, which is concerned with the meaning of the elements. In healthcare, achieving semantic interoperability is a challenge due to many factors as: the ever-rising quantity of data spread in many systems, the existed ambiguity between the different terms, the fact that data are related to organizational and medical processes, to cite only the most known problems [6], [7], [8].

In healthcare, collaboration between processes is of the utmost importance to deliver a quality service care. To improve the collaboration between processes is necessary to understand how the processes collaborate. Many authors claim the existence of a gap between what happens and what is supposed to happen. The process mining approach extracts information from the event log, providing a real image of what is happening, showing the gap, if it exists, between the planned and the executed process [9].

However, there is a lack of automation between business world and IT world. Thus, the translation between both worlds is challenging and requires a huge human effort. Besides, the analysis provided by process mining technology are purely syntactic, i.e. based in the string of the label. These drawbacks leads to the development of Semantic Business Process Management.

The use of semantics in combination with event logs analysis is a bridge between the IT world and business world. It brings advantages to both worlds, as less human effort in the translation between them, the possibility to reason on processes, the possible analyses to complex processes, etc.

In this context, this paper proposes a formal approach to enhance the semantic interoperability in healthcare through the semantic enrichment of the event log. We highlight that this is a preliminary work and not yet validated.

The article is organized as follows: In section II, the research problem is presented. The section III introduces the proposed approach to the enrichment of the event log. The section IV provides the required background knowledge. In Section V, the conclusions and the future works are discussed.

II. Overview of the Proposed Approach

Nowadays the enterprises have been extremely efficient in collecting, organizing, and storing a large amount of data obtained in their daily operations. Healthcare is an information rich environment and even the simplest healthcare decisions require many pieces of information. But, the healthcare enterprises are also 'knowledge poor' because the healthcare data is rarely transformed into a strategic decision-support resource [10].

In this environment, the success in the activities depends of different factors such as the physician's knowledge and experience, the availability of resources and data about patient's condition, and the access to the domain models (which formalize the knowledge needed for taking decisions about the therapeutic actions). All this information must be uniquely accessed and processed in order to make relevant decisions [11], [12], [13].

However, the wide variety of clinical data formats, the ambiguity of the concepts used, the inherent uncertainty in medical diagnosis, the large structural variability of medical records, the variability of organizational and clinical practice cultures of different institutions makes the semantic interoperability a hard task [6], [7], [8].

The semantic interoperability between processes in the healthcare environment is mandatory when the processes need to collaborate during the patient treatment. The analysis of the event log provide knowledge about how processes collaborate and how improve it.

However, the event log may contain implicit relations. The semantic enrichment of the event log enables the discovery of unknown dependencies which can improve the semantic interoperability. Formal Concept Analysis is applied in our approach to discover these unknown dependencies, enabling an improvement in the semantic interoperability.

Ensuring the semantic interoperability between medical and organizational processes is of the utmost importance to improve patient care, reducing costs by avoiding unnecessary duplicate procedures, thus reducing time of the treatment, errors, etc.

III. Literature Review

A. Process Mining

The Process mining technique aims to enable the understanding of process behavior and in this way to facilitate decision making to control and improve that behavior.

However, process mining can have different types of results and is not reduced only to the discovery of process models [14]. In the last decade, process mining techniques were implemented under different perspectives and hierarchy levels: either for the identification of the business process workflow, for the verification of conformance and machine optimization, for the monitoring of the system performance, among others [15].

The application of process mining in healthcare allows one to discover evidence-based process models, or maps, from time-stamped user and patient behavior, to detect deviations from intended process models relevant to minimizing medical error and maximizing patient safety and to suggest ways to enhance healthcare process effectiveness, efficiency, and user and patient satisfaction [16].

The base of process mining are the event logs (also known as 'history', 'audit trail' and 'transaction log') that contain information about the instances (also called cases) processed in systems, the activities (also named task, operation, action or work-item) executed for each instance, at what time the activities were executed and by whom, named respectively as timestamp and performer or resource. The event logs may store additional information about events as costs, age, gender etc. [17], [18].

Fig. 1 shows that the three basic types of process mining are discovery, conformance, and extension.

The discovery is the most prominent type. It takes an event log and produces a model without using any a-priori information.

The second type is conformance checking which compares an a-priori or reference model with the observed behavior as recorded.

The extension is the third type, the idea is to extend or improve an existing process model using information

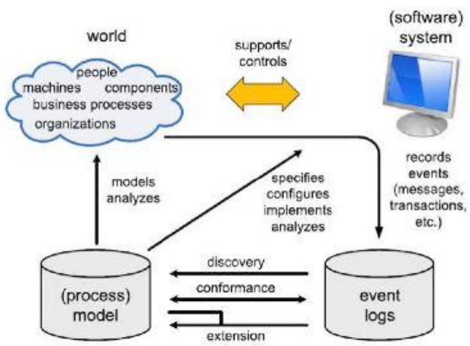


Figure 1. Three main types of process mining

about the actual process recorded in some event log. The mining techniques are aimed at discovering different kinds of models for different perspectives of the process, namely: the control-flow or process perspective, organizational perspective and the data or case perspective. The format of the output model will depend on the technique used [9], [14], [16], [17] [19], [20], [21], [22], [23].

However, despite the benefits of process mining technique there are still some issues to be overcome. One problem is related to the inconsistency of the activity labels. Naming the activity is realized freely by the designer, this action may create complex and inconsistent models generating difficulties in the model analysis. The result of this situation is that the mining techniques are unable to reason over the concepts behind the labels in the log [24]. It is very common the situations where different activities are represented by the same label or different labels are described by the same activity.

Besides, Business Process Management suffers from a lack of automation that would support a smooth transition between the business world and the IT world. This gap is due to the lack of understanding of the business needs by IT experts and of technical details by business experts. One of the major problems is the translation of the high-level business process models to workflow models, resulting in time delays between design and execution phases of the process [25], [26].

In this way, moving between business and IT world requires huge human effort which is expensive and prone to errors [27].

To overcome these issues, the semantic technologies were combined with BPM, enabling the development of the Semantic Business Process Mining approach, which aims to access the process space (as registered in event logs) of an enterprise at the knowledge level so as to support reasoning about business processes, process composition, process execution, etc. [25], [28].

B. Semantic Business Process Mining

The basic idea of semantic process mining is to annotate the log with the concept in an ontology, this action will let the inference engine to derive new knowledge.

The combination of the semantics and the processes can help to exchange process information between the applications in the most correct and complete manner, and/or to restructure business processes by providing a tool for examining the matching of process ontologies [29].

The ontologies are used to capture, represent, (re) use, share and exchange knowledge. There is no official definition about ontology but the most accepted one is from [30] that states that the ontology is an explicit specification of a conceptualization, meaning that the ontology is a description of the concepts, relationships and axioms that exist in a domain.

The ontology is built, mostly, to share common understanding of the information structure among people or software agents. The ontology is also used to separate domain knowledge from the operational, to analyze and to reuse domain knowledge and to make assumptions about a domain explicit [31], [32], [33], [34], [35].

The ontology describes the domain of interest, but for knowledge sharing and reuse among applications and agents, the documents must contain formally encoded information, called semantic annotation.

The annotation process enables the reasoning over the ontology, so to derive new knowledge. Annotation is defined by [36] as "a note by way of explanation or comment added to a text or diagram". An annotation can be a text, a comment, a highlighting, a link, etc. According [37], semantic annotation is the process of annotating resources with semantic metadata. In this way, semantic annotation is machine readable and processable; and it contains a set of formal and shared terms in the specific context [4].

There are three options to annotate the event log. The first one is to create all the necessary ontologies, or to use the existing ones, about the chosen domain and to annotate the elements. The second option is to use tools to (semi-) automatically discover ontologies based on the elements in event logs. In this case, these mined ontologies can be manually improved. The third option is a combination of the previous two in which models/logs are partially annotated by a person and mining tools are used to discover the other missing annotations for the remaining elements in logs/models. The discovery and extension process mining techniques can play a role in the last two options [25].

The idea of adding semantic information to business processes was initially proposed by [38], which aimed to improve the degree of mechanization on processes by combining Semantic Web Services and BPM.

A similar idea was proposed in SUPER (Semantic Utilized for Process Management within and between Enterprises), an European project, which fundamental approach is to represent both the business perspective and the systems perspective of enterprises using a set of ontologies, and to use machine reasoning for carrying out or supporting the translation tasks between both worlds [28].

Reference [39] addressed the problem of inconsistency in the labeling of the elements of an organizational model through the use of semantic annotation and ontologies. The proposed approach uses the i* framework, one of the most widespread goal-oriented modeling languages, and the two i* variants Tropos and service-oriented i*. However, the proposed approach can be applied to other business modeling techniques.

In [40] semantic annotation was use to unifying labels on process models that represent the same concept and abstracting them into meaningful generalizations. The business processes are semantically annotate with concepts taken from a domain ontology by means of standard BPMN textual annotations, with the semantic concept prefixed by an '@'.

Reference [41] proposes an approach for (semi-) automatic detection of synonyms and homonyms of process element names by measuring the similarity between business processes models semantically modeled with the Web Ontology Language (OWL).

An ontological framework was introduced by [42] for the representation of business process semantics, in order to provide a formal semantics to Business Process Management Notation (BPMN). Reference [43] introduces a methodology that combines domain and company-specific ontologies and databases to obtain multiple levels of abstraction for process mining and analysis.

Reference [44] proposes an approach to semantically annotated activity logs in order to discover learning patterns automatically by means of semantic reasoning. The goal is automated learning that is capable of detecting changing trends in learning behaviors and abilities through the use of process mining techniques.

The most of the studies developed in this area focuses on process behavior analysis and in clarifying the meaning of the event log label. Thus, less attention has been paid in the knowledge injection perspective and the semantic discovery perspective [40], [45].

C. Formal Concept Analysis

The Formal Concept Analysis (FCA) is a mathematical formalism based on the lattice theory whose main purpose is structuring information given by sets of objects and their descriptions. It brings knowledge

representation framework that allows discovery of dependencies in the data as well as identification of its intrinsic logical concepts [55].

The FCA theory was introduced in the early 1980s by Rudolf Wille, as a mathematical theory modeling the concept of 'concepts' in terms of lattice theory. The FCA is based on the works of Barbut and Monjardet (1970), Birkhoff (1973) and others for the formalization of concepts and conceptual thinking [46], [47], [48].

During recent years the FCA was widely applied in research studies and practical applications in many different fields including text mining and linguistics, web mining (processing and analysis of data from internet documents), software mining (studying and reasoning over source code of computer programs), ontology engineering and others.

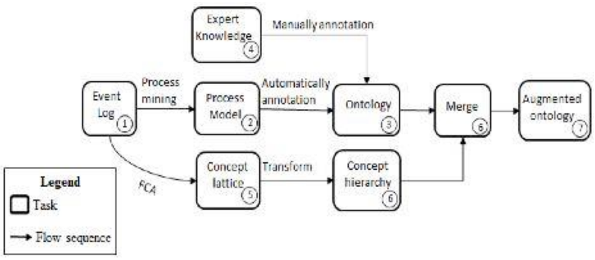


Figure 2. Procedural model of the research

In ontology engineering, the FCA is mainly used for construction of a conceptual hierarchy. The resulting taxonomy of concepts with "*is-a*" relation serves as a basis for successive ontology development. An FCA diagram of the concepts visualizes the structure and, therefore, is a useful tool to support navigation and analytics [49]. Another application of FCA is merging of ontologies, where its power in discovering relations is exploited in order to combine several independently developed ontologies from the same domain [50][26] [46] [51] [52].

IV. Proposed Approach

The proposed method, yet to be validated, for semantically enrich the event log on domain ontologies using Formal Concept Analysis is presented in Figure 2.

The "Step 1" is related to the capture of the event log, which must contain the information about the process executions. The process mining techniques will be used to obtain the process model in the "step 2". The process model provides knowledge about how the activities are connected, who performed the activities, the social network, the time of execution, and others. This acquired knowledge can also be helpful in the annotation process.

The Process Mining Framework (ProM) [53] was the first software developed to support process mining techniques. Initially, ProM accepted as input process execution data in the form of MXML log files, which has been extended to SA-MXML to support semantic annotation. The advance of process mining leveraged the development of another tools as Disco, Interstage Business Process Manager and Perceptive Process Mining [53], [54].

Thus, ProM will be used to discover automatically the ontologies based on the elements of the event log in the "step 3". The resulting ontology will be improved with the expert knowledge (step 4).

The method suggested in our research to enhance the standard event log is the application of Formal Concept Analysis.

The application of FCA (step 5) produces a conceptual structure organizing the domain knowledge. It gives a better understanding about the interoperability between processes and also can be helpful in the discovery of knowledge gaps or anomalies.

In order to establish the correspondence between the concepts in the ontology with the concepts suggested by the FCA knowledge discovery procedure we propose to apply following methods. Firstly, we can identify the ontology concepts within the formal concepts of the FCA. We will consider attributes as concepts. The goal is to build a concept network to express in the best way possible the knowledge [50], [49].

The lattice produced by FCA can be transformed into a type of concept hierarchy (step 6) by removing the bottom element, introducing an ontological concept for each formal concept (intent) and introducing a sub-concept for each element in the extent of the formal concept in question [49].

In our approach, the patients are used as objects and the processes activities (events) are used as the attributes. For the transformation of the lattice in the concept hierarchy we can consider just the attributes. Thus, as proposed by [56], before the transformation we can eliminate lattice of extents (objects) and get as result a reduced lattice of intent (attributes) of formal concepts.

In step 5, it is necessary to incorporate the new data into the ontology. This can be done manually or we can apply a method for ontology merging. Some methods to merge (semi) automatically ontologies have been developed as Prompt, OM algorithm, Chiamera, OntoMerge, FCA-Merge, IF-Map and ISI [57].

The resulting ontology has an augmented knowledge (step 7), thus improving the semantic interoperability.

The proposed approach is under validation procedure. The Nancy University Hospital applications will represent the first case study. The goal of the hospital direction is to optimize the processes interoperability to reduce the costs and increment the quality.

V. Example of Application

An hospital stores the data of the patients, the associated medical data set, the department organizational data, and laboratory data. It stores also the data related to the costs of all events (appointments, treatments, surgeries, exams, materials, and medicines).

The recovered data are stored and related to the patient, doctor, department, and laboratory ID, the events, the date of the event and requests.

Initially one process, for example, the breast cancer treatment is chosen to be analyzed. Following our approach, processes mining techniques can be applied to provide process behavior. The ontology related to the processes is built and annotate. The new concepts will be added in the ontology after the application of the FCA

approach that will semantically enrich the event log showing the implicit relations.

Through process mining techniques is possible to analyze the length of stay, treatment time, pathway followed by the patients, if guidelines or protocols are been followed, etc.

However, normally process model resulted from this kind of data are complex and difficult to analyze. The proposed approach enable the analysis of these complex processes showing the roots of the problems, for example, the causes of the increased length of stay, the lack of some essential care interventions in the treatment, the problems in following clinical guidelines, the discovery new care pathways, the discovery of best practices, the anomalies and the exceptions which may exist in the process providing a better understood where to take action to improve the healthcare processes.

VI. Conclusion

In healthcare domain the access to the information at the right place and at the right time is crucial to provide quality services. In this environment, organizational and medical processes are constantly exchanging information.

The processes analysis shows what are really happening, thus it is providing knowledge about possible improvements. Besides, the data related to the traces of the processes may show problems related to the interoperability, and also ways to improve it.

The process mining techniques enables this kind of analysis. In healthcare, this method is normally used to discover clinical pathways, to discover best practices, adverse events, conformance checking between medical recommendations and guidelines, etc.

Due to the limitations of the process mining techniques, the semantics was combined with the event logs. This combination brings many benefits to process improvement and for knowledge management.

There is a lack of studies about knowledge injection perspective. This research aims to fulfill this gap. Our objective is the enhancement of the semantic interoperability in the healthcare domain using semantic process mining.

Our approach proposes to apply the formal concept analysis method to capture knowledge from the event log, which is not implicit in the ontology, thus improving the semantic interoperability. The semantic enrichment of the event log may also provide knowledge about processes improvement.

The next step is related to the operational development of the proposed approach and the following validation.

References

[1] R.S., Mans, M.H., Schonenberg, G., Leonardi, S., Panzarasa, A., Cavallini, S., Quaglini, W.M.P., van der AALST, "Process mining techniques: an application to stroke care," in *Studies in health technology and informatics*, 136, 2008, pp. 573-578.

[2] A., Rebuge, D.R., Ferreira, "Business process analysis in healthcare environments: A methodology based on process mining", in *Information Systems,* vol. 37, 2012, pp. 99-116.

[3] E., Turban, R.K., Rainer, R.E., Potter, "Introduction to information technology", in *Chapter 2nd*, Information technologies: Concepts and management, 2005.

[4] Y.X., Liao. "Semantic Annotations for Systems Interoperability in a PLM Environment", PhD diss., Université de Lorraine, 2013.

[5] IEEE, Standard Computer Dictionary - A compilation of IEEE Standard Computer Glossaries, 1990.

[6] J., Ralyté, M.A., Jeusfeld, P., Backlund, H., Kühn, N., Arni-Bloch, "A knowledge-based approach to manage information systems interoperability", in *Information Systems*, vol. 33, no. 7, pp. 754-784, 2008.

[7] S.V.B., Jardim, "The electronic health record and its contribution to healthcare information systems interoperability," in *Procedia Technology*, vol. 9, pp. 940-948, 2013.

[8] B., Orgun, J., Vu, "HL7 ontology and mobile agents for interoperability in heterogeneous medical information systems", in *Computers in biology and medicine* vol. 36, no. 7, pp. 817-836, 2006.

[9] U., Kaymak, R., Mans, T., van de Steeg, M., Dierks, "On process mining in healt care", in Systems, Man, and Cybernetics (SMC), *IEEE International Conference on*, pp. 1859-1864, 2012.

[10] S.S.R., Abidi, "Knowledge management in healthcare: towards 'knowledge-driven' decision-support services", in *International Journal of Medical Informatics*, vol. 63, no. 1 pp. 5-18, 2001.

[11] R., Lenz, M., Reichert, "IT support for healthcare processes – premises, challenges, perspectives", in *Data and Knowledge Engineering*, vol. 61, pp. 39-58, 2007.

[12] M., Zdravković, M., Trajanović, M., Stojković, D., Mišić, N. Vitković, "A case of using the Semantic Interoperability Framework for custom orthopedic implants manufacturing". *Annual Reviews in Control*, vol. *36, no.*2, pp.318-326, 2012.

[13] D.C., Kaelber, D.W., Bates, "Health information exchange and patient safety", in *Journal of biomedical informatics,* vol. 40, no. 6, pp. S40-S45, 2007.

[14] P., Homayounfar, "Process mining challenges in hospital information systems", *Federated Conference on Computer Science & Information Systems (FedCSIS)*, pp. 1135-1140, 2012.

[15] W.M.P., van Der Aalst, "Process mining: discovery, conformance and enhancement of business processes", Springer Science & Business Media, 2011.

[16] C.J., Turner, A., Tiwari, R., Olaiya, Y., Xu, "Process mining: from theory to practice", in Business Process Management Journal, vol. 18, no. 3, pp. 493-512, 2012.

[17] M., Jans, J., van der Werf, N., Lybaert, K., Vanhoof, "A business process mining application for internal transaction fraud mitigation", in *Expert Systems with applications*, vol. 38, pp. 13351-13359, 2011.

[18] W.M.P., van der Aalst, "Service Mining: Using Process Mining to Discover, Check, and Improve Service Behavior", in IEEE Transactions on Service Computing, vol. 6, no. 4, 2012.

[19] R.S., Mans, M.H., Schonenberg, M., Song, W.M.P., van der Aalst, P.J.M., Bakker, "Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital", Chapter, Biomedical Engineering Systems and Technologies, Volume 25 of the series Communications in Computer and Information Science, pp 425-438, 2009.

[20] A.J.S., Rebuge, "*Business process analysis in healthcare environments",* (Doctoral dissertation, Master, dissertation, The Technical University of Lisboa), 2012.

[21] C., Webster, "EHR Business Process Management: from process mining to process improvement to process usability", Health care systems process improvement, Conf. 2012.

[22] P., Weber, B., Bordbar, P., Tino, "A framework for the analysis of process mining algorithms", *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, vol. *43, no.* 2, pp.303-317, 2013.

[23] M., Song, W.M.P., van der Aalst, "Towards comprehensive support for organizational mining", in Decision Support Systems, vol. 46, pp. 300-317, 2008.

[24] A.K.A., de Medeiros, W.M.P., van der Aalst, C., Pedrinaci, "Semantic process mining tools: core building blocks", 2008.

[25] A.K.A, De Medeiros, C., Pedrinaci, W.M.P., van der Aalst, J., Domingue, M., Song, A., Rozinat, B., Norton, L., Cabral, "An outlook on semantic business process mining and monitoring", in *On the Move to Meaningful Internet Systems* 2007: OTM 2007 Workshops, pp. 1244-1255. Springer Berlin Heidelberg, 2007.

[26] B., Wetzstein, Z., Ma, A., Filipowska, M., Kaczmarek, S., Bhiri, S., Losada, J.M., Lopez-Cob, L., Cicurel, "Semantic Business

Process Management: A Lifecycle Based Requirements Analysis", in SBPM, 2007.

[27] C., Pedrinaci, J.,Domingue, C., Brelage, T., van Lessen, D., Karastoyanova, F., Leymann, "Semantic business process management: Scaling up the management of business processes", in Semantic Computing, 2008 IEEE International Conference on, pp. 546-553, 2008.

[28] C., Pedrinaci, J., Domingue, "Towards an ontology for process monitoring and mining", in CEUR Workshop Proceedings, vol. 251, pp. 76-87, 2007.

[29] I., Szabó, K., Varga, "Knowledge-Based Compliance Checking of Business Processes", in On the Move to Meaningful Internet Systems: OTM 2014 Conferences, pp. 597-611, Springer Berlin Heidelberg, 2014.

[30] T.R., Gruber, Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, *43*(5), pp.907-928, 1995.

[31] M.A., Musen, "Dimensions of knowledge sharing and reuse", in *Computers and Biomedical Research*, vol. 25, pp. 435-467, 1992.

[32] T.R., Gruber, "A translation approach to portable ontology specifications", in *Knowledge acquisition*, vol. 5, no. 2, pp. 199-220, 1993.

[33] M., Obitko, V., Snášel, J., Smid, "Ontology design with formal concept analysis", in CLA, vol. 110, 2004.

[34] A., Tagaris, G., Konnis, X., Benetou, T., Dimakopoulos, K., Kassis, N., Athanasiadis, S., Rüping, H., Grosskreutz, D., Koutsouris, "Integrated Web Services Platform for the facilitation of fraud detection in health care e-government services", in *Information Technology and Applications in Biomedicine*, ITAB 2009. 9th International Conference on, pp. 1-4, 2009.

[35] C.S., Lee, M.H., Wang, "Ontology-based intelligent healthcare agent and its application to respiratory waveform recognition", in *Expert Systems with Applications*, vol. 33, no. 3, pp. 606-619, 2007.

[36] Oxford Dictionaries, Language matters. Available in http://www.oxforddictionaries.com.

[37] Nagarajan, M., Semantic annotations in web services, In Semantic Web Services, Processes and Applications, pp. 35-61. Springer US, 2006.

[38] M., Hepp, F., Leymann, J., Domingue, A., Wahler, D., Fensel, "Semantic business process management: A vision towards using semantic web services for business process management", in e-Business Engineering, 2005. ICEBE 2005. IEEE International Conference on, pp. 535-540, 2005.

[39] B., Vazquez, A., Martinez, A., Perini, H., Estrada, M., Morandini, "Enriching organizational models through semantic annotation", in *Procedia Technology*, vol. 7, pp. 297-304, 2013.

[40] C., Di Francescomarino, P., Tonella, "Supporting ontology-based semantic annotation of business processes with automated suggestions", in Enterprise, Business-Process and Information Systems Modeling, pp. 211-223. Springer Berlin Heidelberg, 2009.

[41] M., Ehrig, A., Koschmider, A., Oberweis, "Measuring similarity between semantic business process models", in Proceedings of the fourth Asia-Pacific conference on Conceptual modelling, vol. 67, pp. 71-80. Australian Computer Society, Inc., 2007.

[42] A., De Nicola, M., Lezoche, M., Missikoff, An ontological approach to business process modeling, in 3th Indian International Conference on Artificial Intelligence, pp. ISBN-978, 2007.

[43] W., Jareevongpiboon, P., Janecek, "Ontological approach to enhance results of business process mining and analysis", in *Business Process Management Journal*, vol. 19, no. 3, pp. 459-476, 2013.

[44] K., Okoye, A.R.H., Tawil, U., Naeem, R., Bashroush, E., Lamine, "A Semantic Rule-based Approach Supported by Process Mining for Personalised Adaptive Learning", in *Procedia Computer Science*, vol. 37, pp. 203-210, 2014.

[45] Y.X., Liao, E.R., Loures, E.A., Portela Santos, O., Canciglieri, "The Proposition of a Framework for Semantic Process Mining", in *Advanced Materials Research,* vol. 1051, pp. 995-999, Trans Tech Publications, 2014.

[46] J., Poelmans, D.I., Ignatov, S.O., Kuznetsov, G., Dedene, "Formal concept analysis in knowledge processing: A survey on applications", *Expert systems with applications,* vol. 40, no. 16, pp. 6538-6560, 2013.

[47] B., Ganter, G., Stumme, R., Wille, "Formal concept analysis: methods and application in computer science, in TU Dresden, http://www. aifb. uni-karlsruhe. de/WBS/gst/FBA03. shtml 2002.

[48] F., Škopljanac-Mačina, B., Blašković, B., "Formal Concept Analysis–Overview and Applications", in. *Procedia Engineering*, *69*, pp.1258-1267, 2014.

[49] P., Cimiano, A., Hotho, G., Stumme; J., Tane, "Conceptual Knowledge Processing with Formal Concept Analysis and Ontologies in Concept Lattices", in Springer Berlin Heidelberg, pp. 189-207, 2004.

[50] S.A., Formica, "Ontology-based concept similarity in formal concept analysis", in *Information Sciences*, vol. 176, no. 18, pp. 2624-2641, 2006.

[51] G., Stumme, A., Maedche, "FCA-Merge: Bottom-up merging of ontologies". In *IJCAI,* vol. 1, pp. 225-230, 2001.

[52] R.C., Chen, C.T., Bau, C.J., Yeh, "Merging domain ontologies based on the WordNet system and Fuzzy Formal Concept Analysis techniques", in Applied Soft Computing, vol. 11, no. 2, pp. 1908-1923, 2011.

[53] PROM, Available in http://www.promtools.org/doku.php

[54] R., Mans, H., Reijers, D., Wismeijer, M., van Genuchten, "A process-oriented methodology for evaluating the impact of IT: A proposal and an application in healthcare", in *Information Systems*, vol. 38, no. 8, pp. 1097-1115, November 2013.

[55] R., Wille, Restructuring lattice theory: an approach based on hierarchies of concepts (pp. 445-470). Springer Netherlands, 1982.

[56] H.M., Haav, "A Semi-automatic Method to Ontology Design by Using FCA", in *CLA*, 2004.

[57] A.D.C., Rasgado, A.G., Arenas, "A language and Algorithm for Automatic Merging of Ontologies", in *Computing, 2006, CIC'06, 15th International Conference on*, pp. 180-185, IEEE, 2006.

Expert System for Implant Material Selection

Miloš Ristić*, Miodrag Manić**, Dragan Mišić**, Miloš Kosanović*
* College of Applied Technical Sciences Niš, Niš, Serbia
** University of Niš / Faculty of Mechanical Engineering, Niš, Serbia
milos.ristic@vtsnis.edu.rs, miodrag.manic@masfak.ni.ac.rs, misicdr@gmail.com, milos.kosanovic@vtsnis.edu.rs,

***Abstract*— Workflow Management System (WfMS) is a software that enables collaboration of people, processes and monitoring of a defined sequence of tasks, involved in the same business enterprise. There are some WfMSs which enable integration of project activities realized among different institutions. They enable that comprehensive activities carried out at various locations are more easily monitored, with improved internal communication. This paper presents an example of decision support system in Workflow Management System for design, manufacturing and application of customized implants. This support system is based on expert system. Its task is to carry out a selection of biomaterial (or class of material) for a customized implant and then to propose a technological process for implant manufacturing. This model significantly improves the efficiency of WfMS for preoperative planning in medicine.**

***Key words:* Customized implant; Workflow Management System, Expert system.**

I. Introduction

Technological development has influence to all spheres of the society, especially in the field of information technologies, economy, and the needs of a user. With constant innovation and invention, there are thousands of computer application created every day, on various topics, available worldwide, whose functionality meets the customer needs and market demands. The requirement that the quality low-cost product should be first on the market has led to the formation of multidisciplinary teams of different area experts [1]. On the other hand, Knowledge based technologies have provided the integration of different area knowledge into a single software environment. Such systems are usually based on the application of certain methodologies from the domain of artificial intelligence [2]. The most commonly used are expert systems, genetic algorithms and neural networks. Their application in biomedicine is significant, both in the data monitoring systems, and in advanced decision-making systems.

In comparison to the personalization in industry, personalization in medicine has just recently begun to gain importance. Personalized medicine derives from the belief that same illnesses that afflict different patients cannot be treated in the same manner [3].

An implant is a medical device manufactured to replace a missing biological structure, support a damaged biological structure, or fix an existing biological structure [4]. Implants must respond to specific demands in patient treatment. As such, they are used in almost all the areas and fields of medicine.

Unlike standard implants, which have predetermined geometry and topology, customized implants are completely adjusted to match anatomy and morphology of the selected bone of the specific patient [5]. In this way they fully meet the needs of the patient, thus shortening post-operative treatment period and significantly reducing adverse reactions to the acceptance of implants or possible pain. The patient-specific implant concept is evidenced since 1996, research on implants for hip replacement that is manifested in the need for adaptation and customizing implants [6]; then, since 1998 the first cases about patient-specific implants for skull were developed [7]. These kind of implants are custom devices based on patient-specific requirements [8].

This paper focuses on the presentation of the concept of support expert systems for manufacturing personalized orthopedic implants, more precisely for the selection of implant materials and manufacturing method. Implant material selection using expert system can be made by rankings properties such as strength, formability, corrosion resistance, biocompatibility and small implant price [9]. The application of quantitative decision-making methods for the purpose of biomaterial selection in orthopedic surgery is presented in the paper [10]. A decision support system based on the use of multi-criteria decision making (MCDM) methods, named MCDM Solver, was developed in order to facilitate the selection process of biomedical materials selection and increase confidence and objectivity [11]. Based on these researches, we propose a decision support system in a WfMS.

Bearing in mind that the implants are complex geometric forms, most commonly used method for their design is reverse engineering [12].

In order to manufacture adequate customized implants it is necessary, beside the geometry and topology, to select the appropriate material and manufacturing technology. This process physically takes place partly at clinics (where the process of diagnosing and identifying the problem is performed), then at the implant manufacturing facility the implant is designed and configured (and if possible- manufactured) and finally again at clinics where the implant is embedded. This process requires the knowledge of experts from various fields (doctors, biologists, engineers, etc.). The separation of the processes emphasises the importance of the need for an information system for monitoring action flows within the institutions and mutual communication between them. Such a model is made possible by using the Workflow Management System (WfMS). WfMS is a system that completely defines, manages and executes workflows through the execution of software whose order of execution is driven by a computer representation of the workflow logic [13].

A model of integration technologies for ideation of customized implants [14] work with several interoperability flows, between Bio-CAD, CAD and CAE software, based on requirements. Yang y Cols. [15]

propose integration of technologies as mechanisms to attaching or exchange independent systems that interoperate, and promote results optimization, automation and reduce of process time. Newer researches are based on semantic interoperability for custom orthopedic implants manufacturing [16].

This paper gives the concept of an expert system which is a decision support system to WfMS. The purpose of this system is the selection of materials and the selection of customized implant manufacturing process. Therefore, in the definition of the implant model, the implant knowledge is additionally inserted in the form of facts, which actually forms a knowledge model about the implant. This knowledge, connected by appropriate relations to rule databases for the material selection (or the material class selection), provides the prerequisites for the start of the customized implant production technology selection process.

II. EXPERT SYSTEM FOR IMPLANT MATERIAL SELECTION

The connection between activities which are carried out in the clinics and those which take place in the manufacturing facility is achieved by the workflow management system. The concept of Workflow Management System was successfully tested on WfMS-MD system [17]. In [18], a concept is proposed that integrates all activities for the design and manufacturing of customized implants. This concept implies the integration of the processes in orthopaedic clinics and manufacturing facilities by using the information system which would manage the whole process, starting from the doctor's diagnosis and ending with the manufacturing of the implant which is adjusted to a specific patient. . This system provides flexibility of the process of selecting and embedding implants and is realized through cooperation with the manufacturing of customized implants, on one side, and through the possibility to respond to exceptions which may occur during the process, on the other.

The process of designing and manufacturing customized implants [19] may be physically realized by defining activities which would be monitored by WfMS. The process diagram realized [20] is shown in Fig. 1.

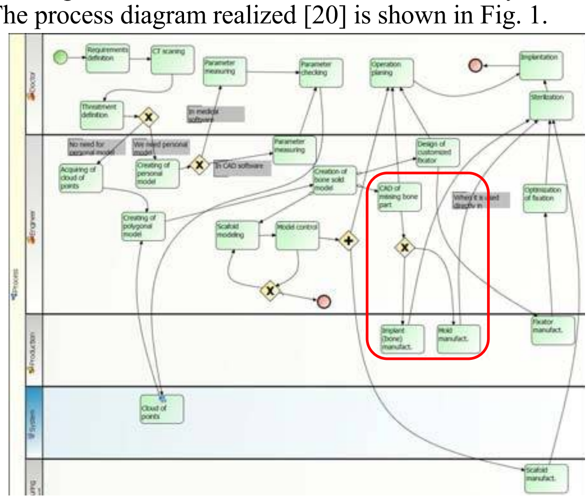


Figure 1. The process of creating customized osteofixation material.

The presented system is extensive and monitors the complete integration process, from defining requirements, CT / MRI scans (doctor's task), then the process of reverse engineering to obtain a complete 3D CAD model of bone missing part (engineer's task) in order for an implant to be manufactured, and then sterilized immediately prior to embedding. The system is connected to decision support system based on the use of expert systems, which should help in the decision making process necessary to produce a customized implant. It incorporates the knowledge (in the form of rules) of an expert who is not even a part of the team producing implants.

Expert systems are meant to solve real complex problems by reasoning about knowledge which normally would require a specialized human expert (such as a doctor). The typical structure of an expert system is shown in Figure 2 and is made of: a knowledge base, an inference engine, and an interface.

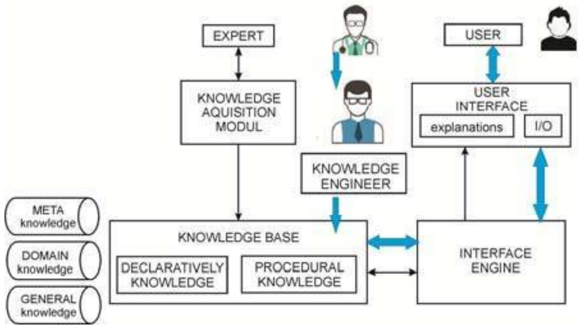


Figure 2. Architecture of expert system.

Since in the expert system the decision making process and knowledge base are separated, parts of knowledge within the knowledge base can be easily supplemented or modified. Knowledge base contains certain rules, which describe the knowledge and work logic of a particular field expert, in this case of a technologist.

The task of the expert system designed in this paper is to recommend a suitable material to meet the requirements of a customized implant, and then to decide on the selection of the manufacturing technological process.

Integration between expert systems and WfMS is performed by means of rule-based Web application. This Web application receives input parameters from the user through the user interface. Input parameters are, in our case, represented in the customized implant knowledge model. In this way the values of characteristics that describe the implant are defined.

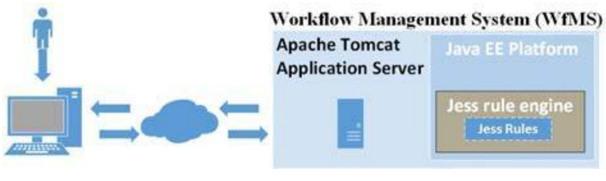


Figure 3. Architecture of Web rule-based application.

The role of a Web browser is to process and forward the parameters to the application on a server by using the appropriate web application. Web application itself is implemented using JavaEE technologies and represents the part of the WfMS system as shown in Figure 3. WfMS system receives parameters, processes them further, and then forwards them to the expert system comprising a knowledge base, i.e. rules. This expert system is actually a rule-based application implemented by Jess rule engine [21].

III. IMPLANT KNOWLEDGE MODEL

The basic building block of every expert system is knowledge. Knowledge in expert system consists of facts and heuristics. While heuristics is made of rules of judgment based on experience or intuition (tacit knowledge domain), the facts are widely distributed, publicly available information that were agreed on at the expert level in subject areas (explicit knowledge domain). For successful work of our expert system it is necessary that there is an adequate knowledge transfer (Figure 4) from the certain field expert to knowledge engineer, in order for an engineer to adequately present accumulated knowledge in the knowledge base.

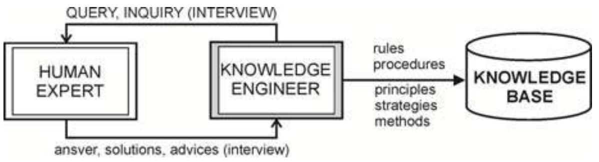


Figure 4. Knowledge transfer from an expert to an expert system knowledge base.

In order for a resulting database of expert knowledge to have its function, it needs to connect on one side to the specific problem database (in our case it is the knowledge model about customized implant), and on the other, with reasoning mechanisms (which is part of the expert shell). The following table gives a part of the knowledge base about customized implants. This knowledge base is adequately filled by orthopaedist and engineers who designed and manufactured the implant. Since these parameters are essential, it is important to present a knowledge model about customized implant with the facts, characteristics, as well as with the description of the facts or by defining certain parameters values.

TABLE I.
CUSTOMIZED IMPLANT MODEL KNOWLEDGE

Parameter	Fact
Gender	Male
Age	30 – 49
Diagnosis	Fracture
Bone	Tibia
Location	Diaphyseal
AO/OTA Classification of fracture	42-B2
Implant type	Internal
Implant kind	Plate
Implant volume	10-15 cm^3
Weight	Low – Medium
Number of necessary joints – screw	8
Number of necessary joints – K wire	2
Biocompatibility	Very high
Sterilizability	Very high
Endurance Limit	High
Lifetime	Max

Since the expert system is connected to WfMS, it is important that they are designed with the same technology and programming languages in order to ensure compatibility. WfMS MD uses a modified open source system Enhydra Shark as workflow engine. Enhydra Shark is a flexible and extendable workflow management facility compliant, embeddable or standalone Java workflow engine. For execution of the rules, we use the expert shell JESS, a rule engine and scripting environment written entirely in the Java language.

IV. REVIEW OF A PART OF BIOMATERIAL CLASS KNOWLEDGE BASE AND AN EXAMPLE OF DECISION-MAKING PROCESS

As there is no universal or optimal material, whose characteristics fit each implant model, it is necessary from a large number of available biomaterials to choose the one that, according to certain specific requirements, fully corresponds to the model.

On the other hand, a wide range of materials ensures that several types of materials belonging to different classes of biomaterials will have certain properties. In order to make a decision on the selection of the concrete material, it is often necessary to predict the resolution of a conflict that will clearly define the procedure for determining priorities, thus the process of material selection in fully defined.

The structure of a thus designed expert system consists of 3 modules: a module for material class selection, a module for material type selection, and a module for customized implant manufacturing technology selection. A model of a part of the expert system for biomaterial class selection is shown in Figure 5.

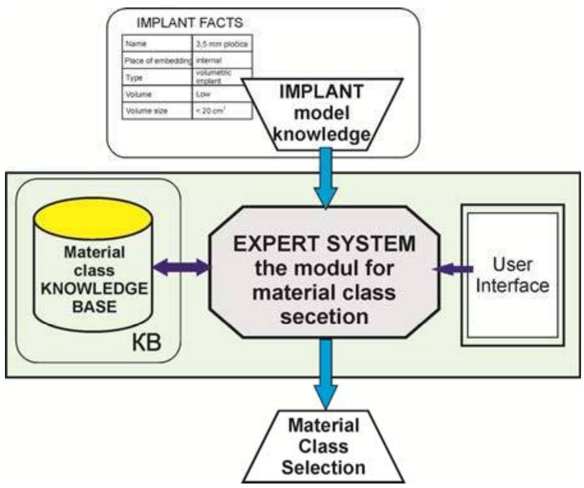


Figure 5. Model of a part of the designed expert system.

Based on the recognized class of materials, by further use the new module of the expert system, we can achieve the selection of specific material for implant manufacturing. Through the latest module of a designed expert system, the customized implant manufacturing technology is determined, according to available resources and applicable technologies.

In Table II there are some of the rules about material classes. For defined parameters in the form of facts, there are three classes of biomaterials presented and their comparison in the certain value range.

After integrating the knowledge about the model, and the biomaterial classes and other necessary knowledge models, in the expert system, an opportunity is created that a user of such a proposed system, e.g. a doctor, can select material (class) for the customized implants.

TABLE II.
RULE BASE ON MATERIAL CLASSES (EXTRACT)

	Tensile modulus	Yield strength	Ultimate strength	Strain to failure	Ductility	Toughness	Resistance to *in vivo* attack	Local host response (bulk)	implant manufacturing location
Metals	M	H	H	M	M	H	L	H	O
Ceramics	H	/	M	L	L	M	H	L	I/O
Polymers	L	L	L	H	H	L	M	M	O
explanations									
L- Lowest; M- Intermediate; H- Highest;					O – Out I – In I/O – In and Out				

By inserting this knowledge in JESS a code is in the following form:

```
;(watch all)
(reset)
(deftemplate feature_has_value (slot feature)
(slot value))

(defrule choose_P
    (and ... )

(defrule choose_C
    (and ... )

(defrule choose_M
    (and
          (or (not (feature_has_value (feature
TM))) (feature_has_value (feature TM) (value
L)))
          (or (not (feature_has_value (feature
YS))) (feature_has_value (feature YS) (value
L)))
          (or (not (feature_has_value (feature
US))) (feature_has_value (feature US) (value
L)))
          (or (not (feature_has_value (feature
SF))) (feature_has_value (feature SF) (value
H)))
          (or (not (feature_has_value (feature
DT))) (feature_has_value (feature DT) (value
H)))
          (or (not (feature_has_value (feature
UT))) (feature_has_value (feature UT) (value
L)))
          (or (not (feature_has_value (feature
HRC))) (feature_has_value (feature HRC) (value
L)))
          (or (not (feature_has_value (feature
D))) (feature_has_value (feature D) (value L)))
          (or (not (feature_has_value (feature
R))) (feature_has_value (feature R) (value M)))
          (or (not (feature_has_value (feature
LHR))) (feature_has_value (feature LHR) (value
M)))
          (or (not (feature_has_value (feature
M))) (feature_has_value (feature M) (value L)))
          (or (not (feature_has_value (feature
PP))) (feature_has_value (feature PP) (value
P)))
          (or (not (feature_has_value (feature
W))) (feature_has_value (feature W) (value L)))
          )
    => (printout t "Choose Metal" crlf))
```

```
(bind ?another y)
(while (= ?another y)
              (printout t "Type the feature?")
              (bind ?f (read))
              (printout t "Type the value?")
              (bind ?v (read))
       (assert (feature_has_value (feature ?f)
(value ?v)))

(bind ?another (read)))

(run)
```

As a result Jess has, based on the criteria given by user and the defined rule base, selected the biomaterial class. In this scenario the suggested solution is the metallic biomaterial (Figure 6).

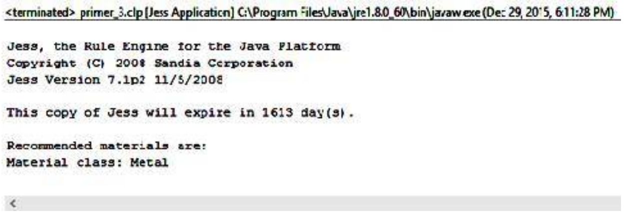


Figure 6. Biomaterial class recommended by Jess

Biomaterial class recommended by Jess (Fig. 6) is further presented to the user through Web interface.

V. CONCLUSION

Integration of business systems that take place in different institutions and in different locations is successfully implemented through the use of information technologies and Workflow Management System. Comprehensiveness and massiveness of WfMS in a certain part requires the use of other technologies, and in this way, the system can be upgraded.

Web based application connects expert system with WfMS. Thus, the business system is upgraded with the appropriate decision support system.

The proposed model concept has been successfully verified by selecting the appropriate class of biomaterials for the purposes of customized implant manufacturing.

Future research will focus on the development of the concept of expert system for customized implant manufacturability analysis, as well as on the development of expert system modules for the selection of materials and manufacturing process.

ACKNOWLEDGMENT

This paper is a result of the project III 41017, supported by the Ministry of Science and Technological Development of the Republic of Serbia.

Authors express their gratitude to Sandia National Laboratories from USA, Albuquerque, New Mexico, for the license to use JESS software for Academic use, Research Agreement for Jess, No. #15N08123.

References

[1] X. Guangleng, Z. Yuyun, “Concurrent Engineering Systematic Approach and Application”, Tsinghua science and technology, Vol. 1, No. 2, Jun 1996, pp. 185-192

[2] Blount G. N., Clarke S., Artificial Intelligence and Design Automation Systems, Journal of engineering Design, Vol. 5, No. 4, pp. 299-314., 1994.

[3] Annas, G. J., 2014, “Personalized medicine or public health? Bioethics, human rights, and choice” Revista Portuguesa de Saúde Pública, Volume 32, Issue 2, Pages 158-163.

[4] M. Manić, Z. Stamenković, M. Mitković, M. Stojković, D. Shepherd, “Design of 3D model of customized anatomically adjusted implants”, Facta Universitatis, Series: Mechanical Engineering, Vol. 13., No: 3, pp. 269-282, 2015.

[5] V. Chulvi. D. Cabrian-Tarrason, A. Sancho, and R. Vidal, “Automated design of customized implants”, Rev. Fac. Ing. Univ. Antioquia N.° 68 pp. 95-103. Septiembre, 2013.

[6] J. M. Bert, “Custom total hip arthroplasty,” J. Arthroplasty, vol. 11, no. 8, pp. 905–915, Dec. 1996.

[7] E. Heissler, F. S. Fischer, S. Bolouri, T. Lehmann, W. Mathar, A. Gebhardt, W. Lanksch, and J. Bier, “Custom-made cast titanium implants produced with CAD/CAM for the reconstruction of cranium defects.,” Int. J. Oral Maxillofac. Surg., vol. 27, no. 5, pp. 334–338, Oct. 1998.

[8] C. Götze, W. Steens, V. Vieth, C. Poremba, L. Claes, and J. Steinbeck, “Primary stability in cementless femoral stems: custom-made versus conventional femoral prosthesis.” Clin. Biomech. (Bristol, Avon), vol. 17, no. 4, pp. 267–273, May 2002.

[9] M. Ipek. I.H. Selvi, F. Findik, O. Torkul, I.H. Cedimoglu, “An expert system based material selection approach to manufacturing”, Materials and Design Vol. 47, 2013. pp. 331–340

[10] B. Ristić, Z. Popović, D. Adamović, G. Devedžić, “Izbor biomaterijala u ortopedskoj hirurgiji”, Vojnosanitetski pregled, Vol. 67, Br. 10, 2010. pp. 847-855

[11] D. Petković, M. Madić, G. Radenković, M. Manić, M. Trajanović, “Decision Support System for Selection of the Most Suitable Biomedical Material”, Proceedings of 5th International Conference on Information Society and Technology, Kopaonik, March 8-11 2015., Society for Information Systems and Computer Networks, Issued in Belgrade, Serbia, 2015. pp. 27 – 31.

[12] V. Majstorovic, M. Trajanovic, N. Vitkovic, M. Stojkovic, “Reverse engineering of human bones by using method of anatomical features”, CIRP Annals - Manufacturing Technology, Vol. 62, pp. 167–170, 2013.

[13] WfMC. “Workflow Management Coalition Terminology & Glossary”, Document Number WFMC-TC-1011, Document Status- Issue 3.0. Technical report, Workflow Management Coalition, Brussels, Belgium (Februrary 1999)

[14] C. I. López, E. Bravo, J. C. Moreno, and L. E. Bautista, “Approach Study to a Model of Integration Technologies for Ideation of Customized Implants”, Journal of Advanced Management Science Vol. 3, No. 4, December 2015., pp. 323-328

[15] L. R. Yang, J. T. O’Connor, and C. C. Wang, “Technology utilization on different sizes of projects and associated impacts on composite project success,” Int. J. Proj. Manag., vol. 24, no. 2, pp. 96–105, Feb. 2006.

[16] M. Zdravković, M. Trajanović, M. Stojković, D. Mišić, and N. Vitković, “A case of using the semantic interoperability framework for custom orthopedic implants manufacturing,” Annu. Rev. Control, vol. 36, no. 2, pp. 318–326, Dec. 2012.

[17] D. Mišić, “Adaptibilni sistemi za upravljanje proizvodnim poslovnim procesima”, Univerzitet u Nišu, Mašinski fakultet, Doktorska disertacija, Niš, 2010.

[18] D. Mišić, M. Manić, N. Vitković, N. Korunović, “Toward integrated information system for design, manufacturing and application of customized implants”, Facta Universitatis, Series: Mechanical Engineering, Vol. 13., No: 3, pp. 307-323, 2015.

[19] M. Ristić, M. Manić, B. Cvetanović, “Framework for early manufacturability and technological process analysis for implants manufacturing”, Proceedings of 5th International Conference on Information Society and Technology, Kopaonik, March 8-11 2015., Society for Information Systems and Computer Networks, Issued in Belgrade, Serbia, 2015. pp. 460 – 463.

[20] JaWE editor, http://www.together.at/prod/workflow/twe (last access: 30.03.2015.)

[21] JESS, the Rule Engine for the JavaTM Platform, Sandia National Laboratories, http://herzberg.ca.sandia.gov/, last access 12.01.2016.

A supervised named entity recognition for information extraction from medical records

Darko Puflović[*], Goran Velinov[**], Tatjana Stanković[*], Dragan Janković[*], Leonid Stoimenov[*]
[*] Faculty of Electronic Engineering, University of Niš, 18000 Niš, Serbia
[**]Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University in Skopje
[*]{darko.puflovic, tatjana.stankovic, dragan.jankovic, leonid.stoimenov}@elfak.ni.ac.rs
[**]goran.velinov@finki.ukim.mk

***Abstract*— Named entity recognition is a widely used task to extract various kinds of information from unstructured text. Medical records, produced by hospitals every day contain huge amount of data about diseases, medications used in treatment and information about treatment success rate. There are a large number of systems used in information retrieval from medical documentation, but they are mostly used on documents written in English language. This paper contains the explanation of our approach to solving the problem of extracting disease and drug names from medical records written in Serbian language. Our approach uses statistical language models and can detect up to 80% of named entities, which is a good result given the very limited resources for Serbian language, which makes the process of detection much more difficult.**

I. INTRODUCTION

Named entity recognition [1, 2, 3, 4] is part of the process called information extraction [5, 6, 7]. It is used to classify parts of text into predefined categories. Categories can vary, depending on task. Usually, text is divided into categories such as names of persons, organizations, locations, numbers, that can represent quantities of money, times, dates, percentages, etc. Entities relevant in this paper are names of diseases and medications and numbers that can represent dates, times and quantities as well as the abbreviations that medical staff often use.

Problem of named entity recognition is often solved using a grammar based or statistical methods [3, 4]. Commonly used statistical methods are supervised, semi-supervised and unsupervised methods.

The systems used for this task are primarily developed for English language and they use a variety of techniques to detect named entities, but most of them are useless for other languages. Statistical language models [15] can be of great help in dealing with languages that have sparse language resources, especially when used in combination with several other available techniques, like stemming [16]. Another advantage of using the statistical language models is the ability of their use on texts written in other languages with minor changes.

Today, hospitals and medical institutions produce huge amount of data about diseases and medications used in treatment and information about treatment success rate. Diagnoses written in text format are usually not structured and do not contain categorized information. This makes process of their understanding much more difficult. Computer systems which should recognize certain entities need to convert text into structured, and then to carry out the identification of specific parts that could be useful in the analysis.

Anamneses are composed of large amounts of useful information. It is possible to find the patient's history of illness, category the patient belongs like habits of smoking, drinking, etc. In addition to the historical background, text contains the tests that were carried out as well as the diagnosis that has been established. After the diagnosis, the patient is receiving a particular treatment that is contained in the document also. In addition, the record can have information about the amount in which the drugs are used, therapy duration and other useful details.

After that period, the patient undergoes examination when a doctor decides whether the treatment was successful or not. A system that allows obtaining such information from medical records can very easily highlight meaningful information in text or store them in structured way which can provide insights for better diagnosis or browsing history of the disease. Linking extracted named entities with the categories to which they belong can also expedite and facilitate the treatment process.

There are several approaches that can achieve these results. In the next section we will discuss the different methods used for solving these problems. In section III we will present our approach to solving this problem in documents written in Serbian language. The system is subject to changes and adding new functionality, which will be discussed in section 4 (Conclusion and future work).

II. RELATED WORK

There are several approaches to data extraction from text documents. Most of them are based on a method described in previous section, named entity recognition. This method can be accomplished in several ways, through supervised or semi-supervised learning algorithms, unsupervised learning algorithms [8, 9, 10].

Named entity recognition has the ability to recognize previously unknown entities using examples and rules. Examples are usually composed of positive and negative

ones. The algorithm uses these examples to create rules, later used to detect entities from new sentences.

Supervised learning [8] uses different techniques for named entity recognition, some of which are support vector machines [11], maximum entropy models [12], hidden Markov models [13], etc. Supervised learning approach uses huge set of training data, manually annotated, from which the system creates rules that are later used to identify entities in new sentences.

Unsupervised approach [9] uses unlabeled data to look for patterns in sentences. This is good approach to look for structure in data and to classify data into different categories.

Semi-supervised [10] learning approach is different because it uses smaller labeled training set and usually larger unlabeled one to create rules. This approach is useful in cases of insufficient data. Labeled data is expensive, but gives good results which makes this approach a good combination of supervised and unsupervised approaches.

A large number of named entity recognition systems for English language use unsupervised learning. This approach gives very good results, but it uses a large number of lexical resources such as WordNet and systems for part of speech tagging [24, 25, 26]. Also, semi-supervised learning [27] is widely used for bio-named entity recognition. Language resources in this approach are used to learn accurate word representations, but to a much smaller extent or even without using them in cases when this task is performed manually. Hidden Markov models [13], support vector machines [11] and conditional random fields [14] are often used as supervised learning [28] techniques. This method is not preferred for named entity recognition, because huge training dataset is needed. But, in some cases, like medical or biological texts, training data is already available.

Named entity recognition is used in the number of different tasks. Results depend on the methods used, as well as on the language over which those methods are applied. Typically, results are between 64% and 90%, but in some specific tasks can be near 100% [29, 30].

Some of the currently available tools for solving problem of named entity recognition in medical records are *Apache cTakes*[1] which is used to extract information from electronic medical records, written as free text, *CliNER*[2] (Clinical Named Entity Recognition system), an open source natural language processing system for named entity recognition in clinical text of electronic health records that uses conditional random fields (CRF) and support vector machines (SVM) and *DNorm*[3].

III. Our Approach

The previous section provides a list of tools used in solving problem of named entity recognition in medical records. Listed tools are designed to work with documents written in English language. Medical records that we use are written in Serbian language. It is impossible to use any of these tools for documents written in any language other than English.

Our approach uses different technics for named entity recognition. Detection of disease and medication names is carried out using character and word n-gram models. Detection of dates, times and time intervals, on the other hand, uses parsed text to detect sequences of numbers and special abbreviations that are used to represent time intervals. Based on format, sequences obtained in this manner can distinguish between dates, times and time intervals. Other abbreviations are detected using dictionary that contains mostly used ones.

The process of named entity recognition that we used is carried out using statistical language models [15, 16]. It is necessary to divide the text first into words or characters and calculate the probability of their occurrence:

$$P(x_1^n) = \prod_{i=1}^{n} P(x_i \mid x_1^{i-1})$$

or

$$P(x_k \mid x_{k-3}, x_{k-2}, x_{k-1}) = \frac{Count(x_{k-3}, x_{k-2}, x_{k-1}, x_k)}{Count(x_{k-3}, x_{k-2}, x_{k-1})}$$

Character models [17] can be useful in recognizing words that do not appear in training data. Combinations of characters that appear in words are specific and can be indication of certain kinds of words.

For example, medications often contains "oxy" and "axy" group of characters. On the other hand, disease names frequently contain trigram "oza". Words with these groups of letters are not often used in medical records, so their presence is usually a sign that they represent names of medications or diseases.

Another way to reduce the number of "false positives" is stop words removal. [18] Stop words do not alter the meaning of the sentence, so that their removal does not affect the system accuracy. Depending on position, words can be lowercase of uppercase. Uppercase letters are not of big importance in this task, so text can be transformed to lowercase.

Another important transformation is lemmatization [19]. Words can be used in various forms which can make it difficult for recognition. Lemmatization transforms every word in text in its lemmatized form, the same form used in dictionaries, documents containing names of drugs and diseases mentioned before. This way, the grammatical rules used to create words in sentence are no longer a factor influencing the results.

Abbreviations in the text should be replaced by the words they represent. In case of codes present in list of diseases and medications, codes can be found in the documents. Other abbreviations are not easy to find. Incomplete list of commonly used English ones can be found online[4].

[1] http://ctakes.apache.org/index.html
[2] http://text-machine.cs.uml.edu/cliner/
[3] http://ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/DNorm
[4] http://studenti.mef.hr/abbrevations.doc

The numbers contained in the text do not necessarily represent categories that are of the interest to the system. Finding measures that are located next to the numbers is very important task. Fortunately, a list that contains measures is short and can be further divided in those that represent weight or time.

The process of creating of the models is applied over all data. The medication names and names of diseases do not include additional text that could interfere with the detection process, so the normalization [20] of the text could be skipped.

Once models are created, it is necessary to carry out a comparison over the parts of the text [21, 22, 23]. Elements of models that were created from medical records are sequentially compared with other models. If the similarity exceeds a certain minimum value, it is likely that part of the sentence is entity to be detected. Which category entity belongs, depends on the similarity with the different models created for those categories. The greatest similarity between model of some category and medical record model is an indication that named entity belongs to that category. However, the similarities between the different categories are rare, so resemblance with one of them usually means the affiliation to that category.

Numbers use a slightly altered approach. Primarily, system should detect numbers. Once this step is completed, detection of units of measure is performed. Units of measure which are located next to the numbers allow the identification of category that numbers belong. If being close to the number, units are an indication that number is an entity and they are remembered as one entity.

Medical document usually ends with a final diagnosis by a doctor established after the treatment. It may indicate that the patient is cured or it is necessary to continue treatment and further analysis. Detection of entity b.o. in the section that describes the condition of the patient after treatment means that he is healed successfully and does not need further actions. However, this is not always the case. On some occasions, the patient is referred for further treatment and tests or receives new therapy. This part of document can be similar to the previous parts, so it is possible to apply the same method for detection entities. This allows some of the entities present in that part of the document to be detected.

Entities that this system should recognize can be divided into the following categories:

- Names of diseases

List of disease names in Serbian language can be found in *International Statistic Classification of Diseases and Related Health Problems (ICD 10)*[5].

This list is divided into categories and contains code that represent every disease, category name which that disease belongs and names in Serbian and Latin.

Also, sometimes, doctors use these codes when writing diagnosis and this list makes it possible to decode them and replace with appropriate names. There are 14405 diseases listed, but using categories we can divide this list into smaller ones and to use them as training data.

- Names of medications

List of all medications used in Serbia can be found inside of *National Register of Medications (NRL)*[6]. Every entry in this register is represented by medication name, the company that produces it, the category to which the drug belongs, the code (*ATC code*), the date since when it is listed, the dosage and a detailed description of the cases where it is applicable, how it is used and what it consist of. This information can be very useful for linking with diseases that can be treated and to check whether the drug is suitable for treating disease.

- Abbreviations

Some of the abbreviations can be found in *NRL* described before, but doctors use other ones in medical records. A list of abbreviations can never be complete and it is hard to train system to recognize ones that are not listed, but they are important part of understanding meaning of medical record because doctors use them as substitute for many key parts of diagnosis.

- Numbers that represent dosage or dates and times

Numbers are often used in medical records as an indicator of the amount of medications prescribed by a doctor. In addition, the numbers may represent a time after which it is necessary to take medication or the dates when the patient needs to see doctor for examination. Amounts are usually accompanied by measures like milligrams (mg) or grams (g). On the other hand, dates and times are typically accompanied by measures of time like hours (h) or days. This can be helpful when determining category into which number belongs.

- Medical treatment success

Medical record usually ends with information about how successful treatment was. In the case of success, record typically ends with abbreviation b.o. (*English, N.A.D. nothing abnormal detected*). In case of unsuccessful treatment, doctor can list everything abnormal detected to that point and recommendations.

A. Overall Results

Records that we used in this paper consist of 42526 medical diagnoses from neurologic clinic. Those records consist of chief complains, description, history of disease and family diseases, psychosomatic and neurological symptoms all written as text in Serbian language. In the end, every record contains information about who prescribed therapy to the patient and was it successful.

Statistical linguistic models have proved to be a good choice because a large number of these diagnoses contain typographical errors. Despite attempts to rectify a number of them, some remained faulty. These, however, did not significantly affect the accuracy of the system, especially when order of some letters is replaced or a letter is misspelled.

[5] http://www.batut.org.rs/download/MKB102010Knjiga1.xls

[6] http://www.alims.gov.rs/ciril/files/2015/04/NRL-2015-alims.pdf

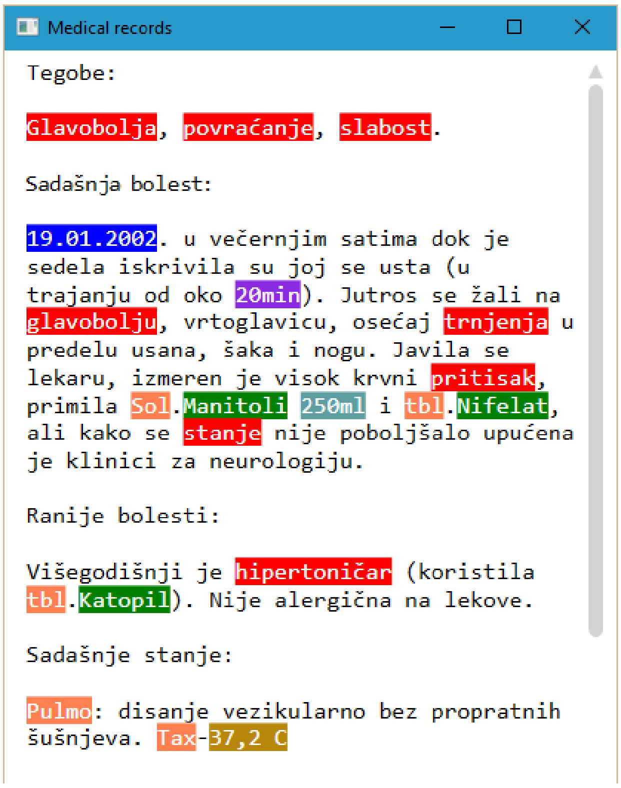


Figure 1. Detected entities shown in different colors

Figure 1 presents different entities shown in different colors. Words with red background color represent disease names and the green color is used for the names of drugs. The dates are marked in blue, time intervals in violet and numbers and abbreviations that represent quantities use cyan color. Words colored orange represent different kinds of abbreviations that doctors use in medical records. Also, dark yellow color is used to indicate other numbers that can provide more information, like temperature shown above.

Adjusting the length of the model gives different results. The best results were obtained using model length from 6 to 8 characters. Model used in this approach transform all disease and drug names in character models, but on the word level.

This approach gives good results in cases of small corpuses of disease and medications names, because of its ability to recognize similar words. This is not beneficial in cases of huge corpuses when there are a lot of false positives.

In order to remove false positives, models are produced differently. Producing models from large list of existing names word by word do not benefit from word relations in those names. Best way to solve this problem is to make character models from entire sentences of names. That raises another problem of variable length of those names. Solution we used is the use larger models and filling of empty spaces in cases of shorter ones with some special character that does not appear in text. This way, the process of the comparison comes down to a comparison of entire disease and medication names that removes possibility to detect similar words. This approaches have their advantages and disadvantages, therefore we use both.

Shorter records with a higher amount of specific data give the best results for models on entire names and sentences of length from 20 to 200 characters. Longer models eliminate the possibility of incorrectly detected named entities and results obtained using them are shown in Table I.

TABLE I.
CHARACTER MODELS MADE FROM SENTENCES AND NAMES

Text kind, length	Correct	Wrong	Not detected
Structured, 20	76%	2%	22%
Structured, 100	73%	3%	24%
Structured, 200	72%	1%	27%
Unstructured, 20	76%	1%	23%
Unstructured, 100	77%	0%	23%
Unstructured, 200	76%	0%	24%

Using 100 manually checked medical records

Results obtained using models of length 6 to 8 characters, created from words detect a large number of words that are not of our interest when used on small, structured text, but were good in case of long unstructured medical records. These results are shown in Table II.

TABLE II.
CHARACTER MODELS MADE FROM WORDS

Text kind, length	Correct	Wrong	Not detected
Structured, 6	35%	65%	0%
Structured, 7	37%	63%	0%
Structured, 8	38%	60%	2%
Unstructured, 6	78%	22%	0%
Unstructured, 7	77%	22%	1%
Unstructured, 8	81%	18%	1%

Using 100 manually checked medical records

It is very difficult to compare obtained results with those from other experiments. The task of finding a named entity depends on the category of entities that should be detected, but also on the language in which text is written. For this particular task the percentage of recognized entities is about 64%-90%. The complexity of the Serbian language makes this task even more difficult but the results obtained in our experiments are satisfactory, with the possibility for improvement.

Abbreviations that are impossible to recognize can be a problem during detection, but it is easy to differentiate them from other types of words used in text, so it is easy to detect and mark them. It is possible to request update of list of abbreviations or to simply use one in its original form.

All diagnoses are related to neurological diseases. It is possible to shorten list of medications and diseases to

speed up the system. This, however, can cause the problem in recognizing entities in areas of record that are containing history of illness or possible complications after received treatment. Yet the division into different categories of diseases and medications can help in classification of entities into groups that belong only to certain types of diseases or medications.

IV. Conclusion and future work

The information obtained by these methods, separated by categories make the overview of the diagnosis much easier. However that is not the only advantage of this system. The entities of all categories are directly related because they are obtained from the same medical record. This allows determination of the correlation between the relevant entities. Names of diseases are linked with medications used to treat them and many other features. As stated in previous sections, the list of drugs includes detailed descriptions of each of them, like substances they contain, but also list of possible replacements. This can help the doctor to select a suitable replacement for the drug if it is necessary, but also to suggest that a drug can create problems to the patient, in case of allergies or in case of medication intolerance.

No less important is the ability to determine in which cases the therapies proven effective and led to the healing of the patient, and which have caused additional complications and required further treatment.

A large amount of information and number of drugs and diseases makes it difficult for doctors to choose the most effective treatment that could be applied. Systems like this could provide a better insight and a lot of helpful information extracted from different sources.

System meets the requirements of detection on the medical records described in this paper, but there are possible improvements to make it better in more complex cases. Lack of the part of speech tagger is the biggest problem and a major handicap, but realization is not an easy task. Its use would facilitate the identification of potential parts of sentences that could be identified as a named entity.

Although the information is very useful, data obtained in this way can be used in many other purposes. The connection of the disease history with a current diagnosis and possible complications is one of the interesting approaches that might give good results in the creation of patterns, both in terms of diseases, and the time when they occurred.

The next step in improvement of this system should be the implementation of other detection methods for named entity recognition. Different techniques should be applied to large amounts of medical records in order to find a suitable method for various types of diagnoses. Medical documentation contains a wide range of diagnoses depending on the area in which they are written, so it is necessary to find a suitable combination of techniques that give the best results.

Huge amount of information requires finding better methods and the ways to minimize time required for their processing.

Acknowledgment

Research presented in this paper was funded by the Ministry of Science of the Republic of Serbia, within the project "Technology Enhanced Learning in Serbia", No. III 47003.

References

[1] Erik F. Tjong Kim Sang, Fien De Meulder, *Introduction to the CoNLL-2003 shared task: language-independent named entity recognition*, CONLL '03 Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003. - Volume 4, pp. 142-147, 2003.

[2] Nadeau, David; Sekine, Satoshi, *A survey of named entity recognition and classification*, Lingvisticae Investigationes, Volume 30, Number 1, pp. 3-26(24), 2007.

[3] GuoDong Zhou, Jian Su, *Named entity recognition using an HMM-based chunk tagger*, ACL '02 Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 473-480, 2002.

[4] Dan Klein, Joseph Smarr, Huy Nguyen, Christopher D. Manning, *Named entity recognition with character-level models*, CONLL '03 Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4, pp. 180-183, 2003.

[5] Stephen Soderland, *Learning Information Extraction Rules for Semi-Structured and Free Text*, Machine Learning, Volume 34, Issue 1, pp. 233-272, 1999.

[6] Ellen Riloff, Rosie Jones, *Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping*, AAAI-99 Proceedings, 1999.

[7] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, *Introduction to Information Retrieveal*, Cambridge University Press, 2008.

[8] Cezary Z. Janikow, *A knowledge-intensive genetic algorithm for supervised learning*, Machine Learning, Volume 13, Issue 2, pp. 189-228, 1993.

[9] Trevor Hastie , Robert Tibshirani, Jerome Friedman, *Unsupervised Learning*, The Elements of Statistical Learning, Part of the series Springer Series in Statistics, pp. 485-585, 2009.

[10] Olivier Chapelle, Bernhard Scholkopf, Alexander Zien, *Semi-Supervised Learning*, The MIT Press, Cambridge, Massachusetts, 2006.

[11] Jun'ichi Kazama, Takaki Makino, Yoshihiro Ohta, Jun'ichi Tsujii, *Tuning support vector machines for biomedical named entity recognition*, BioMed '02 Proceedings of the ACL-02 workshop on Natural language processing in the biomedical domain - Volume 3, pp. 1-8, 2002.

[12] Hai Leong Chieu, Hwee Tou Ng, *Named entity recognition: a maximum entropy approach using global information*, COLING '02 Proceedings of the 19th international conference on Computational linguistics - Volume 1, pp. 1-7, 2002.

[13] Sean R Eddy, *Hidden Markov models*, Current Opinion in Structural Biology, Volume 6, Issue 3, pp. 361–365, Elsevier, 1996.

[14] Burr Settles, *Biomedical named entity recognition using conditional random fields and rich feature sets*, JNLPBA '04 Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications, pp. 104-107, 2004.

[15] ChengXiang Zhai, *Statistical Language Models for Information Retrieval, Synthesis Lectures on Human Language Technologies*, pp. 141, 2008.

[16] Darko Puflović, Leonid Stoimenov, *Plagiarism detection in homework assignments and term papers*, The Sixth International Conference on e-Learning, pp. 204-209., Belgrade, Serbia, 2015.

[17] Praneeth M Shishtla, Prasad Pingali, Vasudeva Varma, *A Character n-gram Based Approach for Improved Recall in Indian Language NER*, IJCNLP-08 Workshop on NER for South and South East Asian Languages, pp. 67-74, 2008.

[18] Akiko Aizawa, *Linguistic Techniques to Improve the Performance of Automatic Text Categorization*, Proceedings of the Sixth

Natural Language Processing Pacific Rim Symposium (NLPRS2001), pp. 307–314. 2001.

[19] Vlado Kešelj, Danko Šipka, *A suffix subsumption-based aproach to building stemmers and lemmatizer for highly inflectional languages with sparse resources*, INFOthecha, 2008.

[20] Andrei Mikheev, *Document centered approach to text normalization*, SIGIR '00 Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 136-143, 2000.

[21] Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn, Supachanun Wanapu, *Using of Jaccard Coefficient for Keywords Similarity*, Proceedings of the International MultiConference of Engineers and Computer Scientists 2013., Vol I, IMECS 2013, Hong Kong, 2013.

[22] Shie-Jue Lee, *A Similarity Measure for Text Classification and Clustering*, IEEE Computer Society, Issue No.07 - July (2014 vol.26), pp. 1, 2014.

[23] Subhashini, R., Kumar, V.J.S., *Evaluating the Performance of Similarity Measures Used in Document Clustering and Information Retrieval*, First International Conference on Integrated Intelligent Computing (ICIIC 2010), pp. 27-31, 2010.

[24] Shaodian Zhang, Noémie Elhadad, *Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts*, Journal of Biomedical Informatiocs, Volume 46, Issue 6, December 2013, pp. 1088-1098, 2013.

[25] George A. Miller , *WordNet: a lexical database for English*, Communications of the ACM, Volume 38 Issue 11, Nov. 1995, pp. 39-41, 1995.

[26] Christopher D. Manning, *Part-of-speech tagging from 97% to 100%: is it time for some linguistics?*, CICLing'11 Proceedings of the 12th international conference on Computational linguistics and intelligent text processing - Volume Part I, pp. 171-189, 2011.

[27] Pavel P. Kuksa, Yanjun Qi, *Semi-supervised Bio-named Entity Recognition with Word-Codebook Learning*, Proceedings of the SIAM International Conference on Data Mining, SDM 2010, pp. 25-36, 2010.

[28] Bodnari, A., Deleger, L., Lavergne, T., Neveol, A., Zweigenbaum, P.: *A supervised named-entity extraction system for medical text*, Online Working Notes of the CLEF 2013 Evaluation Labs and Workshop, September 2013.

[29] Xiao Fu, Sophia Ananiadou, *Improving the Extraction of Clinical Concepts from Clinical Records*, Proceedings of the Fourth Workshop on Building and Evaluation Resources for Biomedical Text Mining (BioTxtM 2014) at Language Resources and Evaluation (LREC) 2014.

[30] Giuseppe Attardi, Vittoria Cozza, Daniele Sartiano, *Annotation and extraction of relations from Italian medical records*, Proceedings of the 6th Italian Information Retrieval Workshop, Cagliari, Italy, May 25 – 26, 2015, CEUR Workshop Proceedings, vol. 1404, 2015.

Software-hardware system for vertigo disorders

Nenad Filipovic[*,**], Zarko Milosevic[*,**], Dalibor Nikolic[*,**], Igor Saveljic[*,**]
Kikiki Idididid[***]and Athanasios Bibas[***]
[*] Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia
[**] BIOIRC Bioengineering Research and Development Center, Kragujevac, Serbia
[***] National & Kapodistrian University of Athens,
1st Department of Otolaryngology – Head & Neck Surgery, Athens, Greece
fica@kg.ac.rs, zarko@kg.ac.rs, markovac85@kg.ac.rs, isaveljic@kg.ac.rs
dimitriskikidis@yahoo.com, thanosbibas67@gmail.com

Abstract—**The (benign paroxysmal positional vertigo) BPPV is the most common type of vertigo, influencing the quality of life to considerable percentage of population after the age of forty (25 out of 100 people are facing this problem after 40). The semicircular canals which are filled with fluid normally act to detect rotation via deflections of the sensory membraneous cupula. We are modeling human semicircular canals (SSC) which considers the morphology of the organs and the composition of the biological tissues and their viscoelastic and mechanical properties. For fluid-structure interaction problem we use loose coupling methodology with ALE (Arbitrary Lagrangian Eulerian) formulation. The tissue of SSC has nonlinear constitutive laws, leading to materially-nonlinear finite element formulation. Our numerical results are compared with nystagmus from real clinical patient. The initial results of 3D Tool software user interface and fluid motion simulation and measurement of the video head impulse test (vHIT) with the Oculus system are presented.**

I. Introduction

Benign paroxysmal positional vertigo (BPPV) is the most commonly diagnosed vertigo syndrome that affects 10% of older persons. BPPV is characterized by sudden attacks of dizziness and nausea triggered by changes in head orientation, and primarily afflicts the posterior canal [1]. The semi-circular canals are interconnected with the main sacs in the human ear: the utricle and the saccule which make up the otolith organs. These organs are responsible for detecting linear movement, such as the sensation when someone goes up or down with an elevator.

We are focused on the semi-circular canals, fluid-filled inner-ear structures designed to detect circular or angular motion. In situations such as rolling at high speed in an airplane, performing ballet spins, or spinning in a circle, our body detects circular motion with these canals. Sometimes this sense of moving in a circle may lead to dizziness or, in extreme cases, even nausea. People who have something wrong with this motion-sensing system often suffer from a condition known as vertigo and feel as if they are spinning even when they are not [2].

Each ear contains three semi-circular canals. Each set of canals is oriented in a different plane that corresponds to a major rotation axis of the head in space.

Firstly, we described numerical procedures fluid flow and fluid-structure interaction with cupula deformation. Some results for fluid velocity and particle tracking are presented. Finally, numerical results which are correlated with experimental measurement with Oculus Rift and conclusions are given

II. Methods

A. Fluid domain

For fluid domain we solved full 3D Navier-Stokes equation and continuity equation. We are using Penalty method to eliminate the pressure calculation in the velocity-pressure formulation. The procedure is as follows. The continuity equation is approximated as

$$v_{i,i} + \frac{p}{\lambda} = 0 \tag{1}$$

where λ is a selected large number, the penalty parameter. Substituting the pressure p from Equation 1 into the Navier-Stokes equations we obtain

$$\rho\left(\frac{\partial v_i}{\partial t} + \partial v_{i,k} v_k\right) - \lambda v_{j,ij} - \mu v_{i,kk} - f_i^V = 0 \tag{2}$$

then the FE equation of balance becomes

$$\mathbf{M}\dot{\mathbf{V}} + \left(\mathbf{K}_{vv} + \mathbf{K}_{vv}^{\lambda}\right)\mathbf{V} = \mathbf{F}_v + \mathbf{F}_{\lambda} \tag{3}$$

where

$$\left[K_{KJ}^{\lambda}\right]_{ik} = \lambda \int_V N_{K,i} N_{J,k} dV, \qquad \left(\mathbf{F}_{\lambda}\right)_{Ki} = \lambda \int_S N_K v_{j,j} n_i dS \tag{4}$$

In examples above we showed a selection of the range of the penalty parameter λ and its effect on the solution.

B. Solid-fluid interaction

There are many conditions in science, engineering and bioengineering where fluid is acting on a solid producing surface loads and deformation of the solid material. The opposite also occurs, i.e. deformation of a solid affects the fluid flow. There are, in principle, two approaches for the FE modeling of solid-fluid interaction problems: a) strong coupling method, and b) loose coupling method. In the first method, the solid and fluid domains are modeled as one mechanical system. In the second approach, the solid and fluid are modeled separately and the solutions are obtained with different FE solvers, but the parameters from one solution which affect the solution for the other medium are transferred successively.

If the loose coupling method is employed, the systems of balance equations for the two domains are formed separately and there are no such computational difficulties. Hence, the loose coupling method is advantageous from the practical point of view and we further describe this method.

As stated above, the loose coupling approach consists of the successive solutions for the solid and fluid domains.

A graphical interpretation of the algorithm for the solid-fluid interaction problem is shown in Fig. 2 [3].

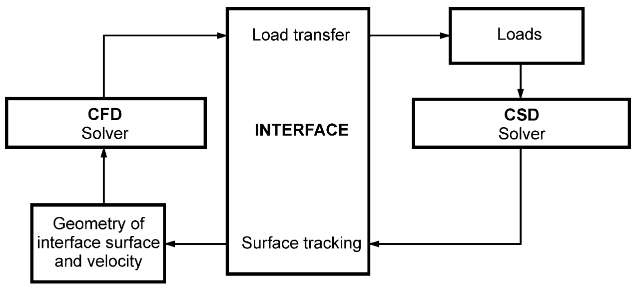


Figure 1. Block-diagram of the solid-fluid interaction algorithm Information and transfer of parameters between the CSD (computational solid dynamics) and CFD (computational fluid dynamics) solvers through the interface block.

Iteration scheme for the solid-fluid interaction, loose coupling approach is presented in Table 1 [3].

Table 1. Iteration scheme for the solid-fluid interaction, loose coupling approach.

1. **Initial conditions for the time step** '*n*'
   Iteration counter I=0:
   configuration of solid ${}^{n+1}\mathcal{B}^{(0)} = {}^{n}\mathcal{B}$; common velocities ${}^{n+1}\mathbf{V}^{(0)} = {}^{n}\mathbf{V}$.
2. **Iterations for both domains**: I=I+1
   a) Calculate fluid flow velocities and pressures ${}^{n+1}\mathbf{V}_f^{(I)}$ and pressures ${}^{n+1}\mathbf{P}^{(I)}$ by an iterative scheme.
   b) Calculate interaction nodal forces from the fluid acting on the solid as

$$ {}^{n+1}\mathbf{F}_S^{(I)} = -\int_S \mathbf{N}^T \, {}^{n+1}\boldsymbol{\sigma}_{Sf}^{(I)} dS $$

   c) Transfer the load from the fluid to solid. Find a new deformed configuration of the solid ${}^{n+1}\mathcal{B}^{(I)}$. Calculate velocities of the common nodes with the fluid ${}^{n+1}\mathbf{V}^{(I)}$ to be used for the fluid domain.
3. **Convergence check**. Check for the convergence on the solid displacement and fluid velocity increments for the loop on I:

$$ \left\| \Delta\mathbf{U}_{solid}^{(I)} \right\| \le \varepsilon_{disp}, \quad \left\| \Delta\mathbf{V}_{fluid}^{(I)} \right\| \le \varepsilon_{velocity} $$

If the convergence criteria are not satisfied, go to the next iteration, step 2. Otherwise, use the solutions from the last iteration as the initial solutions for the next time step and go to step 1.

III. RESULTS

Real patient specific geometry of three SCC is presented in Figure 2. A 3D reconstruction was done from original DICOM images from clinical partners.

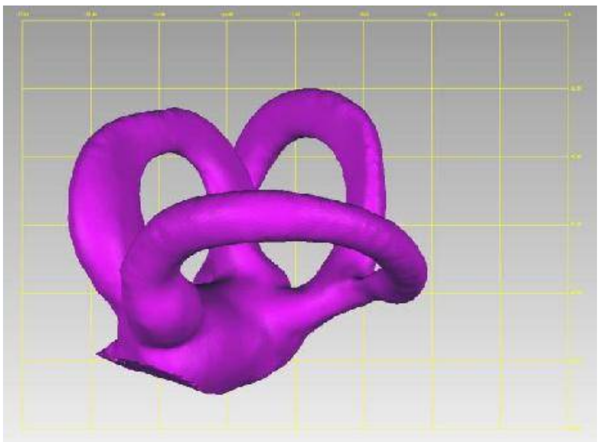


Figure 2. Geometry for three SCC generated from original DICOM images

The video head impulse test (vHIT), which measures the eye movement response to head impulse (brief, unpredictable, passive head rotations), has been used as a simple valid clinical tool for testing the function of the horizontal semicircular canals [4-5].

At the same time, it is possible to use vHIT to identify vertical canal function by measuring vertical eye movement responses to pitch head movements.

The vertical canals lie in planes approximately 45 deg to the sagittal plane of the head and each vertical canal is approximately parallel to the antagonistic canal on the other side of the head [6, 7]. It is possible to test vertical canal function by moving the head of the patient in a diagonal plane, but this turns out to be difficult for the operator and uncomfortable for the patient. A better way of delivering head impulses in the planes of the vertical canals is to use a simple head turned position: the patient

is seated with the head and the body facing the target on the wall at a distance of about 1 meter. The clinician then turns the patient's head on the body about 35 degrees to the left or to the right while the patient's gaze remains on the target and aligned with the patient's sagittal plane (Figure 3). The clinician then pitches the patient's head up or down in the sagittal plane of the body and in this way maximally stimulates the vertical semicircular canals. The angular extent of the head rotations is small (about 10-20 degrees), so the risk of neck injury is very small [8]. We implemented the vHIT method together with the Oculus system (Figure 4).

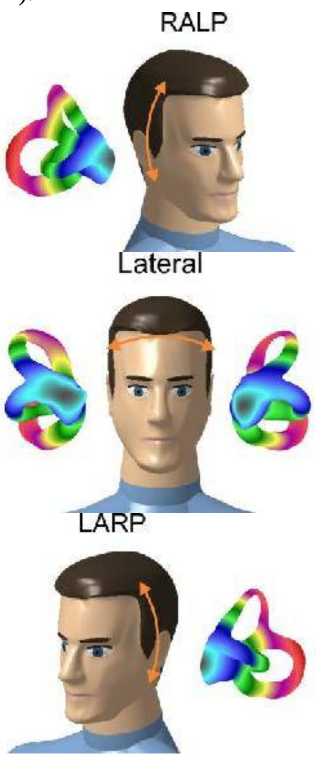


Figure 3. Head rotation during vHIT

Figure 4. Oculus system during vHIT

The Oculus Rift that we are using is a new virtual reality headset that enables us to step into the virtual worlds (Figure 5). This device can provide a custom tracking technology for 360° head tracking with an ultra-low latency.

Each movement of the head is tracked in real time using 4 different sensors: Gyroscope (refresh rate of 1000 Hz), Accelerometer (refresh rate of 1000 Hz), Magnetometer (refresh rate of 1000 Hz), and a Near Infrared CMOS Sensor (refresh rate of 60 Hz), for positional tracking Maps of head movements.

With these sensors we can easily track the head orientation, velocity and acceleration, as well as the eye movement, with respect to a head reference system.

This device creates a stereoscopic 3D view with excellent depth, scale and parallax. Using these features we can easily test the vHIT in the virtual interactive testing room. Applying graphical animations in such virtual interactive testing room, the Oculus device can force the user to move the head or the eyes in the desired position at a certain speed.

In order to test eye movements, we installed a small camera with the IR filter to read the shape and location of a user's eye to determine the direction in which the user is looking. We investigated the correlation with a nystagmus measurement and a fully 3D fluid with particle tracking simulation inside SCC which can be used for self-patient diagnosis as well as therapy.

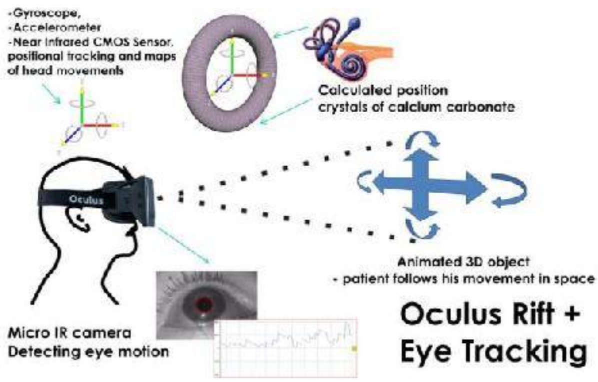


Figure 5. Oculus measurement and computer simulation.

The vHIT test was first described in 1988 by Curthoys and Halmagyi. It is used for the daily work in the set up practice and the clinical application. In many studies, the head impulse test is a standard test in vestibular diagnosis. Our 3D Tool can work separately from standard equipment for vHIT or can be integrated with some existing hardware solutions. We tested our 3D tool system on Oculus. User friendly interface for different head rotations around Z axis is presented in Figure 6 for anti-clock wise direction. For clock wise direction, with head rotation around Z axis, velocity solution for SCC is presented in Figure 7.

Comparison of eye movement, head motion, obtained from measurement, and fluid motion, obtained from computer simulation, has been shown in Figure 8. As it can be seen, there is a small delay in fluid motion response. We think that the reason lies in inertial forces which are incorporated in fluid motion solver based on full 3D Navier-Stokes equation with continuity equation.

User can prescribe boundary conditions through user friendly interface: the axis of rotation, X, Y or Z. Different angles of viewing for velocity, shear stress, pressure and forces on the wall can be defined. The motion of the head is directly connected to the prescribed

wall motion of the SCC. We introduce the assumption that axes of the SCC are the main axes X, Y and Z, which is not totally accurate due to patient specific anatomy. The next software 3D Tool version will incorporate different rotation axis other than main axis X, Y and Z. The current software version is using two kinds of approach for mathematical model. The finite element approach gives very accurate calculation of the fluid motion parameters such as velocity profile, pressure, shear stress etc. Also, boundary condition for head motion can be prescribed very precisely. The only drawback is speed of calculation which cannot be in real time. Another approach for solving fluid motion inside SCC is using LB method and implementation GPU with parallel computing algorithm. We are still testing and incorporating both approaches in the current version of 3D Tool software.

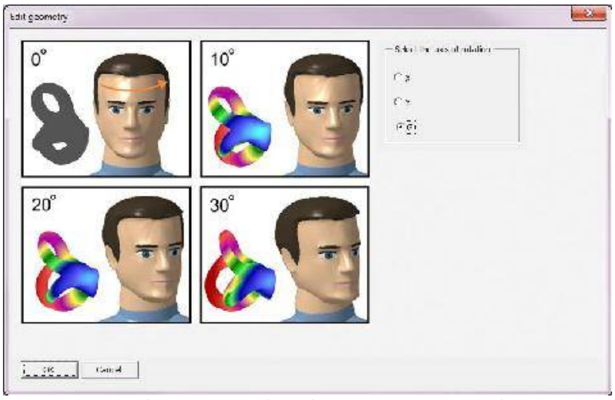


Figure 6. User friendly interface for 3D Tool, results for 10, 20, 30 degrees, head rotation around Z axis, direction anti clock wise.

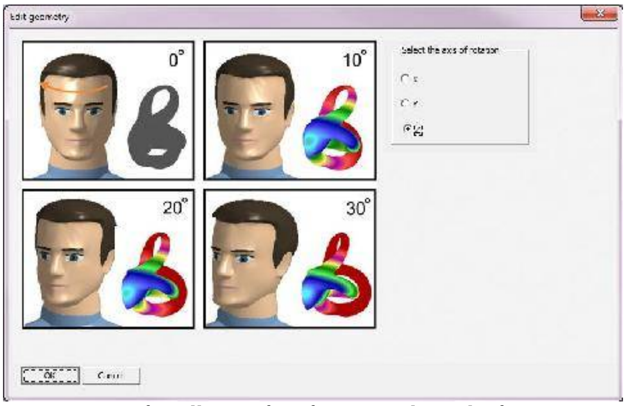


Figure 7. User friendly interface for 3D Tool, results for 10, 20, 30 degrees, head rotation around Z axis, direction clock wise

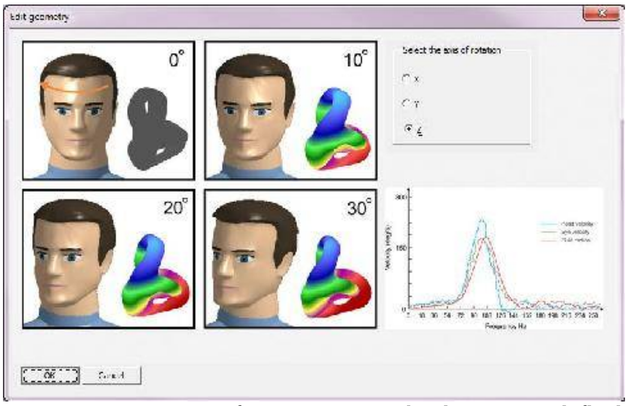


Figure 8. Comparison of eye movement, head motion and fluid motion (computer simulation)

IV. Conclusions

BPPV is the most common type of vertigo, influencing the quality of life to considerable percentage of population after the age of forty. We have developed 3D software tool for specific measuring dimension of the SCC in the axial, coronal, and sagittal planes. Using viscous fluid flow fluid-structure interaction and dynamics finite element analysis for solid domain we can determine velocity distribution, shear stress, forces, deformation of the cupula. We presented the initial results 3D Tool software user interface and fluid motion simulation and measurement of the video head impulse test (vHIT) with the Oculus system. Different methodologies for 3D visualization tool have been developed using C++, OpenGL, and VTK tools.

Acknowledgment

This work was supported by grants from FP7 ICT EMBALANCE 610454 and Serbian Ministry of Education and Science III41007, ON174028.

References

[1] JP Caray, CCD Santin, “Vestibular Disorders: Principles of applied vestibular physiology”, Flint Chapter, 2009.

[2] T. Mergner, Rosemeier T “Interaction of vestibular, somatosensory and visual signals for postural control and motion perception under terrestrial and microgravity conditions: a conceptual model.” *Brain Research Review* 198, 28, pp.118-135.

[3] N. Filipovic, S. Mijailovic, A. Tsuda, M. Kojic, “An Implicit Algorithm Within The Arbitrary Lagrangian-Eulerian Formulation for Solving Incompressible Fluid Flow With Large Boundary Motions”. *Comp. Meth. Appl. Mech. Engrg.* 2006, 195, pp. 6347-6361.

[4] KP Weber, MacDougall HG, Halmagyi GM, Curthoys IS. “Impulsive testing of semicircular canal function using video-oculography”. *Ann N Y Acad Sci* 2009;1164, pp 486-91.

[5] J. Clerk Maxwell, Curthoys IS, MacDougall HG, Manzari L et al. “Klinische Anwendung eines objektiven Tests zur Prüfung der dynamischen Bogengangsfunktion – der Video-Kopfimpuls-Test (vHIT). In: Iro H, Waldfahrer F, eds. Vertigo – Kontroverses und Bewährtes”. 8, *Hennig-Symposium. Vienna: Springer*, 2011; pp 53-62.

[6] RHI Blanks, Curthoys IS, Markham CH. “Planar relationships of semicircular canals in man”. *Acta Otolaryngologica* 1975;80, pp.185-196.

[7] AP Bradshaw, Curthoys IS, Todd MJ et al. “An accurate model of human semicircular canal geometry: a new basis for interpreting vestibular physiology”. *JARO* 2010;11, pp145-159

[8] HG MacDougall, McGarvie LA, Halmagyi GM, Curthoys IS, Weber KP “The Video Head Impulse Test (vHIT) Detects Vertical Semicircular Canal Dysfunction”. *PLoS ONE* 2013, 8(4): e61488. doi:10.1371/journal.pone.0061488

Using of Finite Element Method for Modeling of Mechanical Response of Cochlea and Organ of Corti

Velibor M. Isailovic, Milica M. Nikolic, Dalibor D. Nikolic, Igor B. Saveljic and Nenad D. Filipovic, *Member, IEEE*

***Abstract*—Human hearing system is very interesting for investigation. There are several parts in hearing system, but the most important parts in sense of conversion of audio signal in electrical impulse are cochlea and organ of Corti. The reason why scientists investigate mechanical behavior of human hearing system is hearing loss – a health problem that affects a large part of the world's human population. That problem can be caused by aging, as consequence of mechanical injuries or some disease or even can be congenital. The experimental auditory measurements provide information only about the level of hearing loss, but without information what is happening in the hearing system. Therefore, it is very helpful to develop a numerical model of the parts of hearing system such as cochlea and organ of Corti to demonstrate process of conversion of acoustic signals into signals recognizable by human brain. Two numerical models are developed to investigate hearing problems: tapered three-dimensional cochlea model and two-dimensional cochlea cross-section model with organ of Corti.**

I. Introduction

HUMAN hearing system consists of several parts: external auditory canal, tympanic membrane, three very small ossicles (malleus, incus and stapes) and cochlea [1], [2], [3]. The role of all parts situated before the cochlea is to transmit audio signals from the environment into the most important part of the inner ear – cochlea [4]. The role of cochlea is to transform mechanical vibrations into electrical impulses and send them via the cochlear nerve to the brain. The main mechanisms in the cochlea take place along the basilar membrane and the organ of Corti. Oscillations of the basilar membrane occur due to oscillations in the fluid chambers, which are transmitted through the middle ear. The organ of Corti, which contains an array of cells sensitive to basilar membrane vibration, lies at the surface of the basilar membrane. Those cells are known as outer and inner hair cells. They produce electrical signal under the influence of the basilar membrane.

For modeling the whole process, we have developed two different models. The first one is a three-dimensional tapered cochlea model which contains several parts: basilar membrane, fluid chambers (scala tympani and scala vestibuli), oval window, round window and outer shell. This model is used only to simulate response of basilar membrane. The second model is a two-dimensional slice model which is used to simulate the motion of all parts of the organ of Corti.

II. Methods

The mathematical model for mechanical analysis of the behavior of the cochlea includes acoustic wave equation for fluid in the cochlear chambers and Newtonian dynamics equation for the solid part of the cochlea (vibrations of the basilar membrane) [5].

Acoustic wave equation is defined as:

$$\frac{\partial^2 p}{\partial x_i^2} - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0 \tag{1}$$

where p stands for fluid pressure inside the chambers, x_i are spatial coordinates in Cartesian coordinate system, c is the speed of sound, and t is time.

Matrix form of the acoustic wave equation, obtained by using Galerkin method, can be presented in the following formulation:

$$Q\ddot{p} + Hp = 0 \tag{2}$$

where Q is the acoustic inertia matrix, and H represents the acoustic “stiffness” matrix.

The motion of the solid part of the cochlea was described by Newtonian dynamics equation:

$$M\ddot{U} + B\dot{U} + KU = F^{ext} \tag{3}$$

In equation (3), M, B and K stands for mass, damping and stiffness matrix, respectively.

The real material properties of the basilar membrane are nonlinear and anisotropic. Also, dimensions of the cross-section of the basilar membrane are not constant along the cochlea. In order to match place-frequency mapping, the value of stiffness or geometry of the basilar membrane in finite element model should be variable along the cochlea. In this model tapered geometry of basilar membrane was used to obtain frequency mapping. The basilar membrane width increases and thickness decreases from the beginning to the end of membrane.

In the frequency analysis damping could be included using modal damping [5]. In that case, inside the stiffness matrix there is an imaginary part, so equation (3) could be written in the following form:

$$M\ddot{U} + K(1+i\eta)U = F^{ext} \tag{4}$$

where η is the hysteretic damping ratio.

The fluid-structure interaction with strong coupling was used for solving these equations. Strong coupling means that the solution of solid element in the contact with fluid has impact on the solution of fluid element. The coupling was achieved by the equalization of normal fluid pressure gradient with normal acceleration of solid element in the contact, as shown in the equation (5).

$$n \cdot \nabla p = \rho n \cdot \ddot{u} \tag{5}$$

For the mechanical model of the cochlea we defined a system of coupled equations:

TABLE I
MATERIAL PROPERTIES

Quantity	Value	Unit
Length of cochlea	35	mm
Width of fluid chamber	Tapered from 3 to 1	mm
Width of basilar membrane	Tapered from 6e-5 to 6e-4	mm
Width of basilar membrane	5e-6	mm
Density of fluid	1000	kg m^{-3}
Speed of sound	1500	m/s
Density of solid	1000	kg m^{-3}
BM properties:		
E_x	1e+4	Pa
E_y	1e+8	Pa
E_z	1e+8	Pa
ν_{xy}	0.005	
ν_{yz}	0.3	
ν_{xz}	0.005	
G_{xy}	2e+4	Pa
G_{yz}	1e+5	Pa
G_{xz}	1	Pa
Damping ratio	0.3	

$$\begin{bmatrix} M & 0 \\ -\rho_f R & Q \end{bmatrix} \begin{Bmatrix} \ddot{U} \\ \ddot{p} \end{Bmatrix} + \begin{bmatrix} K(1+i\eta) & -S \\ 0 & H \end{bmatrix} \begin{Bmatrix} U \\ p \end{Bmatrix} = \begin{bmatrix} F \\ q \end{bmatrix} \tag{6}$$

where R and S are coupling matrices.

The solutions for displacement of the basilar membrane and pressure of fluid in the chambers were assumed in the following form:

$$\begin{aligned} U &= A_U \sin(\omega t + \alpha) \\ p &= A_p \sin(\omega t + \alpha) \end{aligned} \tag{7}$$

In equation (7), A_U and A_p represent amplitudes of displacement and pressure, respectively. The circular frequency is ω, t is time and α is phase shift.

When displacement and pressure solution (7) were substituted in the equation (6) we obtain a system of linear equations that can be solved:

$$\begin{bmatrix} K(1+i\eta) - \omega^2 M & -S \\ -\rho_f R & H - \omega^2 Q \end{bmatrix} \begin{Bmatrix} A_U \\ A_p \end{Bmatrix} = \begin{bmatrix} 0 \\ q \end{bmatrix} \tag{8}$$

For solving the system of equations (8), in-house numerical program was developed. The program is part of PAK software package [7], [8].

III. FINITE ELEMENT MODELS

As already mentioned in the introduction, we have developed two different models. The first one is a three-dimensional tapered cochlea model. This model is used only for investigation of mechanical response of the cochlea. The length of the model is 35 mm, which corresponds to the real length of the cochlear chambers. The cross-section is square with 3 mm edge length at the beginning of basilar membrane and 1 mm edge length at the end of basilar membrane. Here is used orthotropic material model for modeling of basilar membrane. The material properties are given in Table 1. 3D model of the cochlea is given in Fig. 1.

The boundary condition in the fluid domain is prescribed acoustic pressure at round window, the beginning of upper fluid chamber. That excitation corresponds to reality when an audio signal comes into the hearing system. This signal is then transmitted through the elements of the middle ear to the cochlea. The unit value is prescribed because the value is not significant. This model is used only to analyze modal shapes.

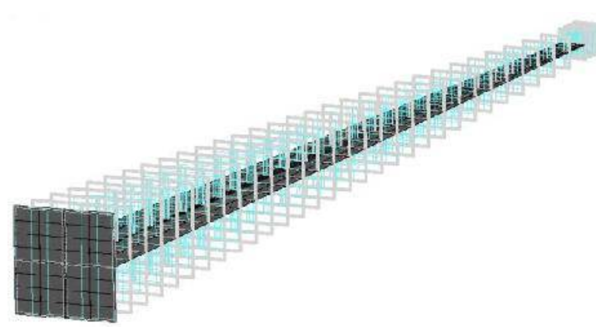


Fig. 1. 3D finite element model of the tapered cochlea

The boundary conditions for the basilar membrane are clamped edges.

The third boundary condition is fluid-structure interface at all surfaces where fluid and solid finite elements are coupled face to face.

The second model that we use to model active cochlea is a 2D slice model (Fig. 2). This model is generated depending on excitation frequency in the 3D model. Several different excitation frequencies were investigated. For each specific frequency we solve the 3D model and determine the peak in basilar membrane response.

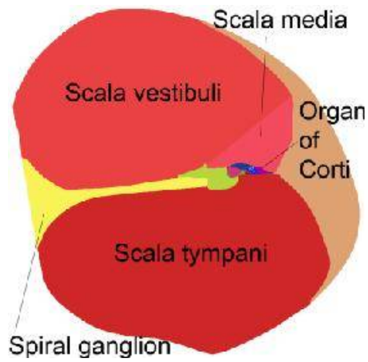


Fig. 2. Two-dimensional slice model.

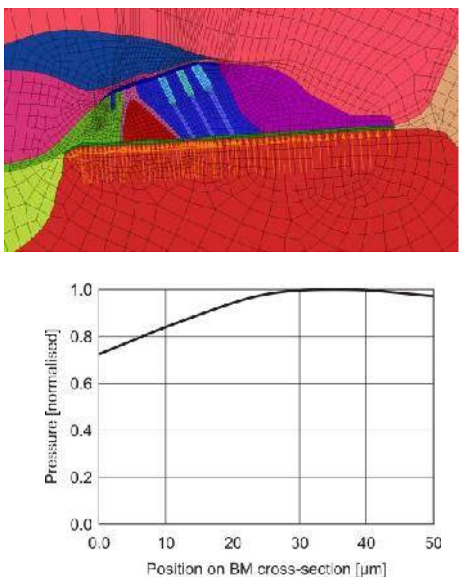


Fig. 3. Organ of Corti with pressure distribution along BM cross-section.

After that, we reconstruct the cross-section of the cochlea at that place and generate 2D slice model. The pressure at basilar membrane calculated by the 3D model is used as boundary condition. (Fig. 3).

IV. RESULTS

The response of the basilar membrane was obtained using the tapered three-dimensional model, Fig. 4. Here are presented the results for only one excitation frequency of 3450 Hz. The pressure distribution is obtained by using tapered cochlea model and after that is prescribed in a proper 2D slice model. In Fig. 5 contour plot for fluid pressure in 3D model and displacement field in 2D slice model are shown.

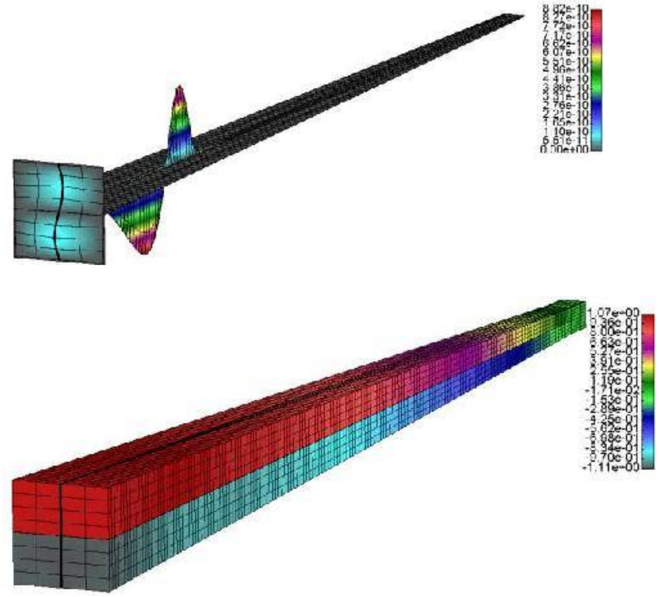


Fig. 4. Tapered cochlea model: basilar membrane displacement and pressure distribution

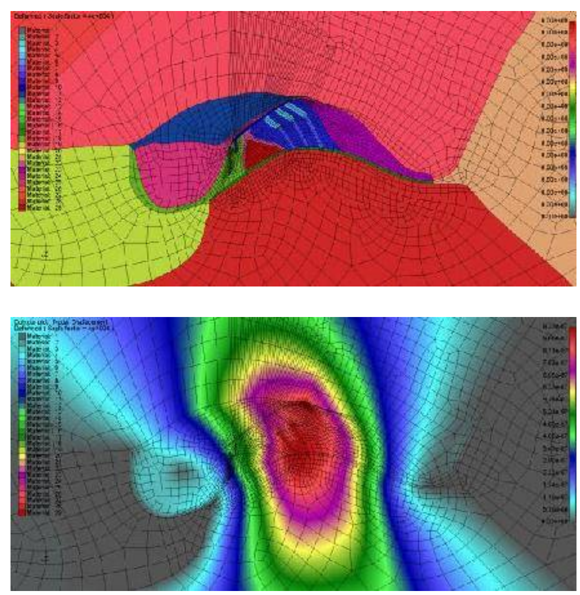


Fig. 5. Two-dimensional slice model

V. CONCLUSION

The processes that appear in human ears are very interesting for research. Medical doctors are mainly engaged in experimental research to measure the level of hearing damage. Their knowledge helps engineers to approximate the auditory system in an appropriate way, to avoid less important parts and to make modeling of the most important

parts in a right way, in order to obtain meaningful results.

ACKNOWLEDGMENT

This work was supported in part by grants from Serbian Ministry of Education and Science III41007, ON174028 and FP7 ICT SIFEM 600933.

REFERENCES

[1] J.O. Pickles, "*An Introduction to the Physiology of Hearing*", 2nd ed. Academic Press, London. 1988.

[2] W.L. Gulick, G.A. Gescheider, and R.D. Fresina, "*Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics*," Oxford University Press, London, 1989.

[3] C.R. Steele (1987), "Cochlear Mechanics". In Handbook of Bioengineering, R. Skalak and S. Chien, Eds.,pp. 30.11–30.22, McGraw-Hill, New York.

[4] R. Nobili,F. Mommano, and J. Ashmore (1998), "How well do we understand the cochlea?". TINS, 21(4), pp.159–166.

[5] Guangjian Ni (2012), "Fluid coupling and waves in the cochlea" – PhD thesis, University of Southampton, Faculty of engineering and the environment, Institute of sound and vibration research.

[6] S.J. Elliott, Ni G, Mace BR, Lineton B. (2013), "A wave finite element analysis of the passive cochlea". The Journal of the Acoustical Society of America, vol. 133, issue 3, p. 1535

[7] N. Filipovic, M. Kojic, R. Slavkovic, N. Grujovic, M. Zivkovic (2009), PAK, Finite element software, BioIRC Kragujevac, University of Kragujevac, 34000 Kragujevac, Serbia.

[8] M. Kojic, N. Filipovic, B. Stojanovic, & N. Kojic, "*Computer Modeling in Bioengineering – Theoretical Background, Examples and Software*". John Wiley and Sons, 978-0-470-06035-3, England, 2008.

INTERFACING WITH SCADA SYSTEM FOR ENERGY MANAGEMENT IN MULTIPLE ENERGY CARRIER INFRASTRUCTURES

Nikola Tomašević, Marko Batić, Sanja Vraneš
University of Belgrade, Institute Mihajlo Pupin

Abstract – *In order to provide more advanced and intelligent energy management systems (EMS), generic and scalable solutions for interfacing with numerous proprietary monitoring equipment and legacy SCADA systems are needed. Such solutions should provide for an easy, plug-and-play coupling with the existing monitoring devices of target infrastructure, thus supporting the integration and interoperability of the overall system. One way of providing such holistic interface that will enable energy data provision to the EMS for further evaluation is presented in this paper. The proposed solution was based on the OpenMUC framework implementing a web-service based technology for data delivery. For testing the system prior to an actual deployment, data simulation and scheduling component was developed to provide a low-level data/signal emulation which would replace signals from the field sensors, monitoring equipment, SCADAs, etc. Furthermore, the proposed interface was implemented in a scalable and adaptable manner so it could be easily adjusted and deployed, practically, within any complex infrastructure of interest. Finally, for demonstration and validation purposes, the campus of the Institute Mihajlo Pupin in Belgrade was taken as a test-bed platform which was equipped with SCADA View6000 system serving as the main building management system.*

1. INTRODUCTION

Contemporary energy management systems (EMS) usually have to supervise and control a number of technical systems and devices, including different energy production, storage and consumption units, within the target building or infrastructure. In such a case, interfacing with the numerous field-level devices, coming from different vendors, using different proprietary communication protocols, is a difficult problem to solve, particularly in case of complex infrastructures with multiple energy carrier supply. Therefore, in order to support more advanced and intelligent EMSs, generic and scalable solutions are needed to be provided in terms of easy, plug-and-play interfacing with different legacy field-level devices, monitoring equipment, etc. Such solutions should provide means for integration with the existing energy related devices and systems, but also to the legacy building management systems (BMS) such as, for instance, widely utilized systems for supervisory control and data acquisition (SCADA). So far, a number of results on interfacing with SCADA and underlying monitoring infrastructure to support energy management were published in the literature. For instance, in [1], interfacing between smart metering devices and SCADA system on the carrier level was investigated for private households and small enterprises. A concept design and communication interfaces with data monitoring system underlying the hybrid renewable energy systems were presented in [2] and [3]. Integration interfaces between new EMS centre and the existing SCADA system respecting the data acquisition and communication aspects were described in [4]. Then, a scalable, distributed SCADA systems for energy management of chain of hydel power stations and power distribution network buildings were proposed in [5] and [6], respectively. Authors of [7] proposed an integrated system for the monitoring and data acquisition related to the energy consumption of equipment or facility. Moreover, the requirements for SCADA systems in terms of standardized protocols for communication and interfaces for large wind farms were discussed in [8]. Finally, an overview of the open architecture for the EMS control centre modernization and consolidation providing modular interfacing with the new applications was provided in [9]. All these results are to the significant extent tailored to the specific scenarios of interest and do not provide a generic and scalable solution.

The objective of this paper was to present one way of providing the interface and the communication means between the energy data acquisition system (such as SCADA system) and EMS of the target infrastructure/building. Through the developed interface, the energy production and consumption data (such as electrical and thermal (hot water) energy), were provided to the EMS for further evaluation in order to optimize the energy flows (both production and consumption) within the target infrastructure [10],[11]. The proposed solution for interfacing with SCADA system was based on the OpenMUC framework [12] which was leveraged upon OSGi specifications and web-service based technology for data delivery. Supported by OpenMUC framework, the interfacing module provided the energy production and consumption data to the rest of EMS components (such as energy optimization module [11]) in a web-service based manner, on demand by querying SCADA database, in flexible and secured manner. Furthermore, different communication protocols and flexible time resolution of acquired data are supported. Apart from providing the acquired data to the EMS, the interface module also provides means for execution of the field device controls (for controlling the actuators, definition of target set-points, etc.). For validation and demonstration purposes of the proposed interface, the campus of the Institute Mihajlo Pupin (PUPIN) in Belgrade was taken as a test-bed platform with multiple energy carrier supply. At the PUPIN campus, SCADA View6000 system [13] is operating as the main BMS providing supervision and

control of underlying energy related systems. Currently, it supervises both electricity production (by PV plant) and consumption, while monitoring of the thermal energy consumption (hot water produced by burning the fuel oil-mazut) is envisioned for the near future by integrating already installed digitalized flow meters/calorimeters. Nevertheless, contrary to the existing solutions, the proposed interface was envisioned in flexible and scalable manner so it could be easily extended to encompass subsequently defined metering points.

In order to properly validate and test the developed interface, before the actual deployment, a software component was developed responsible for the low-level (measurement) data simulation and scheduling. In other words, this data simulation and scheduling component, i.e. a data emulator provides the emulation of signals coming from the field sensors, SCADAs, etc. In this way, early prototyping and testing of the EMS and energy optimization modules for multi-carrier interconnected entities was made possible [10]. Manual definition of different data types/signals was also supported. At the same time, this component provides a flexible way for use case scenarios definition. This included the definition of low-level signals, patterns as a set of low-level signals and high-level use case scenarios definition as a set of signal patterns. Data/signals emulated by data emulator component are made accessible to all EMS components and provided in a web-service based manner. Moreover, it supports flexible time resolution/granularity of the output data.

The remainder of this paper is organized as follows. Section 2 specifies the main functionalities which should be supported by the interface module by analysing the interface rationale and managed field-level data. Then, the technological background of the interface was defined through the investigation of interface deployment diagram in Section 3. Section 4 specifies some of the main functionalities, technological background and integration design of the data emulator in terms of field-data simulation and scheduling. Integration and validation of the interface module were investigated subsequently in Section 5. Finally, some concluding remarks were presented in Section 6.

2. SPECIFICATION OF INTERFACE MODULE FUNCTIONALITIES

In order to provide the energy data for further processing, the proposed interface should enable a two-way communication between the energy data acquisition system (such as SCADA system) and EMS of the target infrastructure. This interface module was intended to provide all the information related to the energy consumption of target infrastructure to the rest of the EMS as well as to the energy optimization module [11]. Apart from the information on consumption of energy carriers used to supply the target infrastructure this considered also the acquisition of energy production data (if production units are deployed).

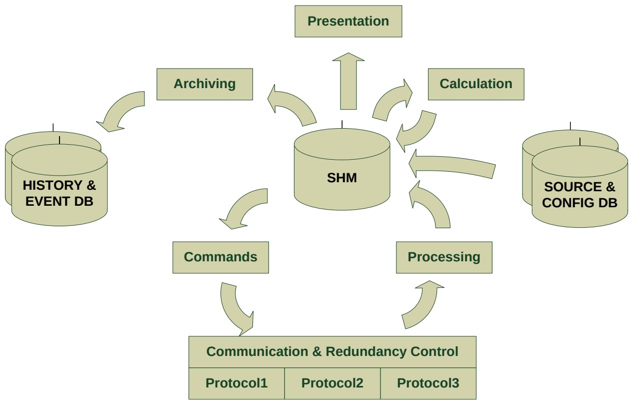


Figure 1: SCADA system architecture (View6000)

For demonstration purposes, the PUPIN campus was taken as a multiple energy carrier supply test-bed platform where SCADA View6000 system [13], shown in Figure 1, is operating as the main BMS. Currently, only electrical energy production and consumption, and some environmental parameters are monitored by the SCADA system at PUPIN premises, while monitoring of the thermal energy consumption was envisioned to be provided by instalment and integration of additional metering equipment (such as temperature sensors, hot water flow meters and digitalized fuel oil-mazut flow meters/calorimeters). Nevertheless, the proposed interface was designed in flexible and scalable manner so it could be easily extended to encompass subsequently defined metering points.

SCADA View6000 system is capable of integrating various subsystems of automatic control (such as power supply, escalators, fire protection, emergency lighting, etc.) and it provides unique facility management informational environment [14]. It also provides visualization and possibility to archive acquired data and signal values. More precisely, it can control and monitor distributed, heterogeneous subsystems communicating with diversity of field-level devices using different proprietary communication protocols and different types of communication links. Some of the relevant features of the SCADA View6000 are the following:

- acquisition and processing of sensor data (deriving new data from the acquired ones),
- accepting command requests and generating programmed sequences of commands,
- generating events (whenever something irregular was detected, alarm went off, etc.),
- logging sensor readings and derived events (within EVENT/HISTORY database),
- various communication tasks such as selection of the best redundant communication link, and
- presenting relevant information to the operator and other stakeholders.

Configuration data such as, information about field-level devices, their properties, provenance of data, i.e. semantics, are stored in Source/Configuration database, which is read at the SCADA system start-up. It holds

semantics of all incoming signals, i.e. it defines a way for processing the incoming raw signals and formulas for determining signal's attributes. After the system start-up, all the configuration parameters as well as the last acquired signal values (from the field) are kept in SCADA shared memory. This shared memory is accessed by different SCADA components as it is presented in Figure 1. Real time data could be the raw sensor readings, derived/aggregated values or set-point data manually defined by the operator. An archiving component stores all the data from the shared memory upon the request (or automatically at regular time intervals) to the Event/History database.

SCADA system at PUPIN campus is processing the raw data which are triggered by the arrival of the data from the communication lines (from remote terminal units - RTU). Both digital and analogue values are monitored by SCADA system. Data read by sensors are converted to the digital value (raw data) and sent to the SCADA along with the information about its source (address). Using the configuration (source) database, SCADA can semantically process (convert or calculate) acquired raw signals according to the information about the source of the data. Finally, the signal attributes are determined after the raw value is processed.

Having in mind the previously listed features of the underlying SCADA system at PUPIN campus, data types which should be managed by the proposed interface were identified as the following:

- Measured data. Data values of measured parameters read from sensors which are already filtered and/or aggregated by the SCADA system. These data are only read. Values of the sensor readings should be accompanied with a unique signal ID indicating that specific reading point.
- Set points. Data values of set points. This data can be only written. Set point values should be accompanied with the corresponding signal ID indicating a specific set point in the system.

3. INTERFACE MODULE DEPLOYMENT

As previously stated, the objective of the proposed interface is to provide the communication with SCADA system. Deployment diagram of interface module is shown in Figure 2. All measured data related to the energy production and consumption are acquired by the interface module and further delivered to the rest of the EMS and energy optimization module in a flexible and secured way. Acquired data are accessible in a web-service based manner supporting different communication protocols (such as FTP, SOAP), and delivered in data polling fashion. In case of PUPIN test-bed platform, communication between SCADA View6000 system and filed-level devices is performed using View6000 legacy protocols and configuration parameters.

Independently on the data delivery, energy consumption data monitored through SCADA View6000 system are

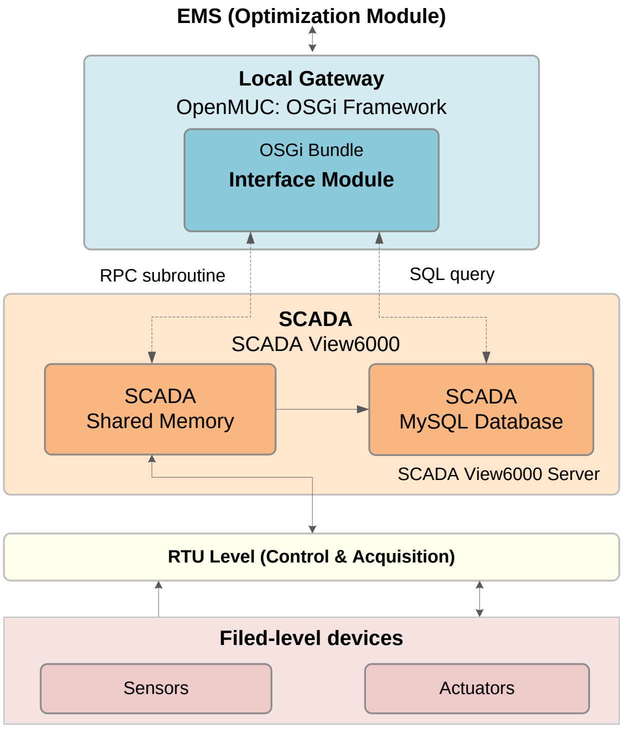


Figure 2: SCADA interface deployment diagram

fetched directly from the SCADA system. In case of data polling, reading of the acquired data is performed by executing the designated "read" method over the available control parameters of the interface module.

At the first place, this is performed by extracting the acquired data directly from the SCADA View6000 History/Event database (through designated remote procedure call - RPC) where all the field-level signals are stored in runtime. By querying the History/Event database (representing the replica of the SCADA memory) any field-level signal value could be acquired and forwarded further to the EMS and energy optimization module at regular time intervals. This requires local communication with SCADA View6000 system deployed at the main PUPIN server. Moreover, the flexible time intervals for fetching the data are supported as well.

When it comes to definition of set-point values and sending the control actions through the SCADA system to the field-level devices, interface was envisioned to support execution of such control actions. Control actions could be performed by executing the designated "write" method over the available control parameters of the related interface module. With respect to the mentioned, control signal could be triggered only within the SCADA View6000 shared memory itself. This further required the development of the interface module capable to trigger the corresponding control signal/set-point value within the SCADA shared memory (again via designated RPC thread). This case required also establishment of the local communication of the interface module with the SCADA system, deployed at the main server at the PUPIN campus.

All the data acquired from the field-level devices/sensors are forwarded from the interface through the Energy Gateway, implemented with the OpenMUC framework [12], to the rest of the EMS. Communication with the Energy Gateway is performed through the OpenMUC data manager using its predefined plug-in interface. Interface module only manages the last values of readings, and therefore there is no information persistence. On the other hand, historical and configuration data are managed by designated sub-component of the Energy Gateway. The deployment of the interface provided the possibility for adjustment of the sampling frequency of monitored parameters in order to meet the requirements of the energy optimization analysis [10],[11]. Furthermore, having in mind that in case of critical infrastructures, open access to the Web is rarely available (due to the security requirements of facility operation), additional constraints should be taken into account such as provision of restricted VLAN for communication between interface module and EMS deployed at the site.

4. DATA SIMULATION AND SCHEDULING

The objective of the data simulation and scheduling component, i.e. data emulator, was to generate and provide the low-level data, such as data related to the energy production and consumption (e.g. electricity, thermal energy, fuel oil, etc.), for the purpose of the verification and testing of the EMS and energy optimization module before the actual deployment within the target infrastructure. In other words, it deals with the simulation and scheduling of the low-level data fed into the system. This includes, first of all, simulation of the data-point values coming from the field-level devices, such as sensors, but also high-level data/parameters such as already filtered and aggregated data-point values coming from SCADA system. By delivering the artificially generated low-level data, early prototyping and calibration of the overall EMS (including its components and energy optimization algorithms) was possible to perform under different use case scenarios. Therefore, data emulator was defined flexible enough to be capable of generating different data types, data formats and protocols. Moreover, it supports flexible time resolution, i.e. granularity of the generated data fed into the EMS.

Data emulator provides, at the first place, a flexible and intuitive way (from the perspective of the end-user, i.e. operator) for definition of various use case scenarios over the chosen time span. This includes the possibility of defining different types of data-points and signals which will be simulated and fed into the EMS (through the Energy Gateway). Low-level data-points and signal values are defined manually. More precisely, the operator have to enter and specify the values of the desired data-points, their specific parameters, time point of generation, etc. This procedure is mainly driven and facilitated by the graphical user interface (GUI) of the data emulator (through corresponding option menus, drop-down lists, default values, etc.) offering some predefined information/parameters for specific data-point types. Such predefined information is currently embedded within the data emulator itself, but it could be also additionally extracted from the EMS central repository/facility model holding various facility/infrastructure related data [15],[16].

Based on the manually defined low-level signals, data emulator provides the possibility to assemble specific signal patterns defined as a set of low-level signals. Such signal patterns could be further utilized (by organizing them over the chosen time span) to assemble the high-level use case scenario which is defined as a sequence of signal patterns. In this way, different, complex use case scenarios could be easily defined by the operator just by definition of the low-level data-point values through the corresponding GUI. Definition of the low-level signal values, signal patterns and high-level use case scenarios is performed off-line. After the definition of the desired use case scenario (stored locally within scenario archive), data emulator provides the possibility to "play" the scenario, i.e. to generate all the low-level signal values according to defined scenario and feed them into the EMS (through the Energy Gateway) for further evaluation.

Data types managed by data emulator are simulated data-point values of measured parameters and set points values. Low-level data-point/signal values generated by data emulator are delivered to the rest of the EMS components and provided in a web-service based manner (supporting different communication protocols such as FTP and SOAP). In terms of generated data delivery, data pushing approach was taken into account. The generated data were fed into the system through the corresponding Energy Gateway (implemented with OpenMUC framework) interface. Communication with EMS Energy Gateway is performed also through the OpenMUC data manager using its predefined plug-in interface.

Logic of the data emulator was wrapped into the bundle-JAR file (based on OSGi specifications [17]) which was deployed under the OpenMUC framework. Data emulator bundle was firing data-point values from previously generated scenario file (for instance, manually specified by the facility operator). It also implements the corresponding Scenario Editor GUI which was developed (in Swing) as a standalone application aimed for user-friendly definition of scenarios through the configuration files that contained all the information needed for scenario simulation. The most important advantage of using the Scenario Editor would be the reusability of data-point value patterns (as well as signal and device patterns), not only within one scenario, but also in other scenarios defined later on, while at the same time providing intuitive and user-friendly approach. In this way, data emulator delivers a convenient way for scenario definition and testing of EMS and energy optimization module.

To have a clear view on the data emulator and its features, integration design and data flow between Scenario Editor GUI component and data emulator bundle component

were illustrated in Figure 3. The interaction of these two components (GUI and logic) was envisioned to implement the following action flow:

1) Scenario Editor GUI component is accessible to the operator sitting in front of the EMS cockpit (main user interface) while providing a way for testing EMS in different scenarios (action (1) in diagram: starting the Scenario Editor GUI by the operator)
2) By using the Scenario Editor GUI, operator is capable of defining the data-point values, signals and devices as part of different scenarios (action (2) in diagram: saving the specified scenario into the file by the operator)
3) Desired scenario file is played by activating the data emulator under the OpenMUC framework (action (3) in diagram: starting/activating the data emulator bundle through the OpenMUC framework)
4) Data emulator bundle should "fetch" and read the selected scenario file stored locally or remotely (action (4) in diagram: initialization of bundle with data-point values stored within the scenario file)
5) Data emulator emulates the data-point values according to the scenario file (interaction (5) in diagram: firing the data-point values through the OpenMUC framework)

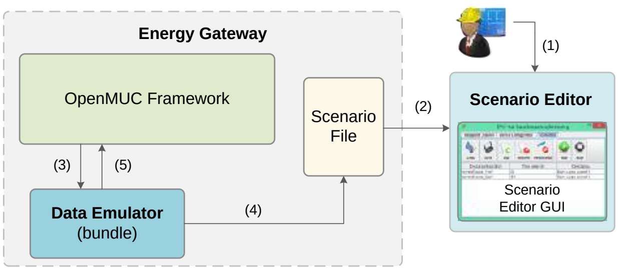


Figure 3: Integration design and data flow of data emulator.

5. SCADA INTERFACE INTEGRATION AND VALIDATION

As previously elaborated, the proposed interface was developed to provide two-way communication with the main BMS system of the target infrastructure. As such, the interface module, together with data simulation and scheduling component, was deployed and validated at PUPIN test-bed platform with multiple energy carrier supply for acquisition of different metering points including electrical and thermal energy production and consumption, and meteorological data, monitored by the SCADA View6000 system.

Interface module was implemented as bundle-JAR file (based on OSGi specifications [17]) which was deployed under the Energy Gateway (OpenMUC framework). For better insight into the data flow, Figure 4 represents simple schematics of interface towards the energy monitoring infrastructure depicting the communication and integration architecture deployed to acquire the designated metering points. As it can be noticed, at the

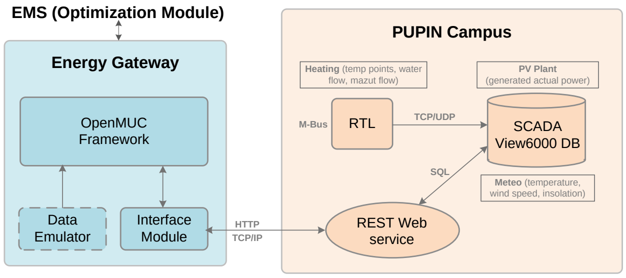


Figure 4. Interface module communication and integration.

gateway side, the interface forwards all the acquired metering values towards the EMS for further processing.

At the same time, it communicates with the REST web service (via internet connection using HTTP/TCP/IP) residing at EMS server set at PUPIN premises which performs the acquisition of metering points from SCADA View6000 system supervising the PUPIN premises. Apart from the interface module, data emulator was also deployed under the Energy Gateway to provide data simulation and scheduling required for system validation.

SCADA View6000 system is responsible for supervision of metering points related to the PV plant and energy production (total actual power generated by the plant), heating system, i.e. thermal energy production and consumption (mazut level/consumption, hot water temperature/flow, radiator and air temperature points) and meteorological parameters acquired by PUPIN local meteorological station (outside temperature, wind speed, solar irradiation). As it can be seen from Figure 4, designated RTL unit (ATLAS-MAX/RTL which is an advanced PLC system running on Real-Time Linux) at the field side of PUPIN campus is responsible for acquisition of metering data using M-Bus communication protocol and communication with SCADA View6000 system over TCP/UDP protocol. Additionally, corresponding OpenMUC Gateway parameters have been defined in order to support acquisition of indicated metering points.

For retrieving the metering values stored by the SCADA View6000 system in real time, REST web service was set at designated EMS server. Through the corresponding SQL queries carrying the information about the specific channel/variable IDs, desired metering values were retrieved and forwarded to the interface bundle under the OpenMUC Energy Gateway framework. Finally, the proposed interface was implemented in a scalable and adaptable manner so it could be easily replicated and deployed practically within any complex infrastructure of interest.

6. CONCLUSIONS

The main objective of this paper was development of interface which supported the operation of EMS components by delivering the real-time energy production and consumption data from legacy SCADA systems. This

was achieved through the means of the proposed interface module, leveraged upon OpenMUC framework, which offers on-site acquired data primarily towards energy optimization module of EMS which should further take control actions. Therefore, this interface was also responsible for accepting the control decisions, derived from the optimization components, and forwarding them back to the SCADA system for execution via asset control signals. The data are extracted directly from SCADA system real-time database and then offered to other EMS components either directly (if the system is deployed locally) or through the means of web-service technology (REST services). For demonstration of the proposed interface, PUPIN campus was taken as a test-bed platform with SCADA View6000 serving as the main BMS.

Considering that EMS requires involvement of large number of different components and systems, its testing would not be possible on a live system, operating at the demonstration site. Therefore, a software component was developed that provides for a low-level data/signal emulation which would replace signals from the field sensors, monitoring equipment, SCADAs, etc. The developed data emulator offers definition of flexible use case scenarios – including definition of low level signals, signal patterns as well as high-level use case scenarios. Finally, the proposed interface together with the data simulation and scheduling component were developed in such a way to enable bridging between different protocols and technologies. The proposed interface was envisioned as generic and scalable enough to be easily adjusted and applied at any target infrastructure in order to support energy data provision for high-level optimisation and energy management purposes.

ACKNOWLEDGEMENTS

The research presented in this paper is partly financed by the European Union (FP7 EPIC-HUB project, Pr. No: 600067), and partly by the Ministry of Science and Technological Development of Republic of Serbia (SOFIA project, Pr. No: TR-32010).

REFERENCES

[1] Kunold I., Kuller M., Bauer J., Karaoglan N., "A system concept of an energy information system in flats using wireless technologies and smart metering devices," 2011 IEEE 6th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), vol. 2, pp. 812-816, Prague, 15-17.09.2011.

[2] Delimustafic D., Islambegovic J., Aksamovic A., Masic S., "Model of a hybrid renewable energy system: Control, supervision and energy distribution," 2011 IEEE International Symposium on Industrial Electronics (ISIE), pp. 1081=1086, Gdansk, 27-30.06.2011.

[3] Tuladhar A., "Power management of an off-grid PV inverter system with generators and battery banks," 2011 IEEE Power and Energy Society General Meeting, p. 1-5, San Diego, CA, 24-29.07.2011.

[4] Savoie C., Chu V., "Gateway system integrates multiserver/multiclient structure," IEEE Computer Applications in Power, vol. 8, no. 2, pp. 10-14, 1995.

[5] Valsalam S.R., Sathyan A., Shankar S.S., "Distributed SCADA system for optimization of power generation," 2008. INDICON 2008. Annual IEEE India Conference, vol. 1, pp. 212-217, Kanpur, 11-13.12.2008.

[6] Teo C.Y., Tan C.H., Chutatape O., "A prototype for an integrated energy automation system," Conference Record of the 1990 IEEE Industry Applications Society Annual Meeting, vol. 2, pp. 1811-1818, Seattle, USA, 7-12.10.1990.

[7] Ung G.W., Lee P.H., Koh L.M., Choo F.H., "A flexible data acquisition system for energy information," 2010 Conference Proceedings IPEC, pp. 853-857, Singapore, 27-29.10.2010.

[8] Brobak B., Jones L., "SCADA systems for large wind farms: Grid integration and market requirements," 2001 Power Engineering Society Summer Meeting, vol. 1, pp. 20-21, Vancouver, Canada, 15-19.07.2001.

[9] Vankayala V., Vaahedi E., Cave D., Huang M., "Opening Up for Interoperability," IEEE Power and Energy Magazine, vol. 6, no. 2, pp. 61-69, 2008.

[10] Beccuti G., Demiray T., Batic M., Tomasevic N., Vranes S., "Energy Hub Modelling and Optimisation: An Analytical Case-Study", PowerTech 2015 Conference, p. 1-6, Eindhoven 29.06.-02.07.2015.

[11] Batic M., Tomasevic N., Vranes S., "Software Module for Integrated Energy Dispatch Optimization", ICIST 2015, 5th international Conference on Information Society and Technology, pp. 83-88, Kopaonik, 08-11.03.2015.

[12] OPENMUC, Fraunhofer ISE, Germany, www.openmuc.org

[13] SCADA View6000 – Project VIEW Center, Institute Mihajlo Pupin, Serbia, www.view4.rs

[14] Tomasevic N., Konecni G., "Generic message format for integration of SCADA-enabled emergency management systems", 17th Conference and Exhibition YU INFO, pp. 71-76, Kopaonik, 06.03.-09.03.2011.

[15] Tomasevic N., Batic M., Blanes L., Keane M., Vranes S., "Ontology-based Facility Data Model for Energy Management," Advanced Engineering Informatics, vol. 29, no. 4, pp. 971-984, 2015.

[16] Tomasevic N., Batic M., Vranes S., "Ontology-enabled airport energy management," ICIST 2013, 3rd International conference on information society technology, pp. 112-117, Kopaonik, 03-06.03.2013.

[17] OSGi - Open Services Gateway initiative, www.osgi.org

ICT Platform for Holistic Energy Management of Neighbourhoods

Marko Batić, Nikola Tomašević, Sanja Vraneš
School of Electrical Engineering, University of Belgrade, Institute Mihajlo Pupin, Belgrade, Serbia
marko.batic@pupin.rs, nikola.tomasevic@pupin.rs, sanja.vranes@pupin.rs

***Abstract* — Fostering of ICT technologies in energy domain has garnered a significant degree of attention in the last few years, especially in the context of improving energy efficiency of large infrastructures. One such approach is presented in this paper and focuses on empowering the neighborhoods with advanced energy management functionalities through the utilization of ICT infrastructure devoted to smart energy metering, supervision, control and global management of energy systems. In particular, this paper reveals the conceptual system architecture, core energy modelling framework and optimization principles used for the development of the neighborhood energy management solution under the EPIC-Hub project.**

I. Introduction

Managing energy infrastructures becomes more and more challenging task considering high operation costs and lack of availability of energy supply in the peak time periods. The problem may be typically solved by following one of the two major pathways. The first one entails the replacement of legacy energy assets, typically associated with poor efficiency performance and lack of controllability, whereas the second one considers improvement of energy management strategies and utilization of existing equipment in a more productive fashion. The former requires high investments and usually does not represent a viable solution for the majority of infrastructures. The latter, on the other hand, may be achieved by employment of contemporary ICT solutions, or simply the retrofit of existing ones, in order to get better insight in the operation of energy assets and provide for their better management. The focus of this paper is to propose such ICT enabled Energy Management System (EMS) which may be expanded from single entities towards whole neighbourhoods and districts. Furthermore, the scope of the proposed solution is placed around complex multi-carrier energy infrastructures with multiple energy conversion options, available energy storages and controllable demand.

We argue that by employing the smart sensing equipment for monitoring of energy flows, advance data analytics for forecasting, and smart energy management strategies for energy dispatch, the facility operators would be able to significantly decrease their energy related costs. Seamlessly integrated through an ICT platform, the trained personnel would be equipped with a set of monitoring and management tools which would bridge the gap between poorly controlled legacy energy systems and dynamic energy pricing context.

An exhaustive survey of the relevant and recent ongoing research initiatives in this domain was performed and the most prominent solutions and corresponding projects were recognized in the following.

A. *State of the Art*

ENERsip project provides a set of ICT tools for near real-time optimization of energy generation and consumption in buildings and neighborhoods [1]. Detailed architecture of the proposed system for in-house monitoring and control was introduced in [2]. Following is the COOPERATE project with its main objective of integrating diverse local monitoring and control systems to achieve energy services at a neighborhood scale by leveraging upon the flexibility and power of distributed computing [3]. An ICT based integration of tools for energy optimization and end user involvement was conducted under the EEPOS project with the ambition of improving the management of local renewable energy generation and energy consumption on the district level [4]. The overall concept and system architecture has been elaborated in [5]. Another perspective to solving the problem of development and operation of energy positive neighborhoods based on intelligent software agents was brought by IDEAS project. Contrary to proprietary solutions developed by previous projects, the ODYSSEUS project developed an Open Dynamic System (ODYS) for tackling the problem of dynamic energy supply for given demand and available storages in urban areas [7]. The ODYSSEUS's pilot experimentation performed at neighborhood level in the XI Municipality of Rome was elaborated in detail in [8]. By putting the emphasize on the business model perspective of the neighborhood energy management scenarios, both ORIGIN and SMARTKEY projects aim at delivering better business decisions based on ICT solutions driven by real-time fine-grained data. In parallel, the NOBEL project aims at integrating various ICT technologies for delivering more efficient distributed monitoring and control system for distribution system operators (DSOs) and prosumers. An overview of all involved actors and envisioned NOBEL infrastructure has been presented in [12] while the evaluation of the agent based monitoring and brokerage system was reported in [13]. Lastly, there are several initiatives for enabling smart neighborhoods by leveraging on the deployment of ICT solutions in residential sector (smart homes) and principles of distributed max-consensus negotiation [14] and decentralized coordination [15].

B. *Selected approach*

Although similar to the aforementioned projects at the conceptual level, the solution proposed in this paper may be clearly distinguished by flexible and modular ICT architecture and energy modelling framework, having as a

result higher level of flexibility and applicability for various energy management scenarios. The presented solution was developed under the EPIC-Hub project [16] which focuses on integrating the flexible approach of Energy Hub [17] with the development of a service oriented middleware solution which connects low level devices and energy assets with application level energy management tools. Using a flexible paradigm for middleware development allows for seamless integration of software components from physically remote locations which enables application of various neighbourhood services ranging from monitoring and analysis towards energy management.

The remainder of the paper starts with Section II which introduces the system architecture and reveals its key components. The Section III briefly elaborates on the modelling framework used for representing the energy infrastructure. The neighbourhood energy management concept is elaborated in Section IV by revealing a typical optimization scenario. Finally, the paper is concluded and the impact of the proposed platform is discussed in Section V.

II. System Architecture

The purpose of this section is to provide a global overview of all the entities belonging to the architectural viewpoint descriptions as part of subsystems of the overall neighborhood energy environment. Furthermore, the following descriptions will not provide any formal representation of software or hardware systems but they will define the “enterprise logic” components playing a role in most common energy management, energy monitoring and energy trading scenarios.

The starting point for the organization of relevant components has been identification of responsibilities within the employed Energy Hub framework, the contexts of operations defined in EPIC-Hub scenarios and the constraints set by existing infrastructures, ICT systems, actors and organizations. Specifically, the foreseen scenarios consider both single Hub and neighborhood/district deployment. To account for this the EPIC-Hub platform was implemented according to a flexible and comprehensive three tier architecture (Figure 1) segmented into the common Tangible, Business Logic and Presentation layers. Flexibility of the proposed architecture lies in its ability to support an arbitrary number of single Hubs which, then, constitute a neighborhood/district. This is reflected in the Business Logic layer, where the actual business scenarios and corresponding algorithms are used based on the use case. Having said this, it should be noted that the presented architecture refers to a full-blown, neighborhood, use case while the relevant components are distinguished with “Local” and “Neighborhood” segments.

A. Tangible layer

This layer represents the physical part of the proposed EPIC-Hub solution entailing a set of available energy assets including energy generation, conversion and storage units as well as ICT infrastructure devoted to energy metering, energy systems supervision, control and global management (e.g. monitoring platforms, EMS or BMS).

The list of potential energy assets and their corresponding features are out of the scope for this paper, as they will be considered purely as input data providers for the EPIC-Hub platform. However, their interfaces towards the data acquisition equipment and monitoring platforms will be detailed in the following. The reason for this is that the main objective of the EPIC-Hub platform is not to promote more efficient energy assets but in fact to highlight the need for ICT retrofit of energy systems which may lead to greater operational efficiency of existing energy assets. The retrofit should actually affect the existing (legacy) management and monitoring systems by empowering them with intelligent predictive control algorithms.

The data acquisition and control is based on typical SCADA system which collects the data from a wide range of low level devices, i.e. field level controllers such as Remote Terminal Units (RTUs) and/or Programmable Logic Controllers (PLCs). The former are typically placed inside the DC and AC cabinets, for monitoring electricity flows, whereas the latter may be used for monitoring of signals coming from various sensors. The deployment of EPIC-Hub solution would typically require information about heat flows, relevant meteorological parameters (solar irradiation, wind speed and velocity, temperature of solar panels), fuel consumption etc.

Given the diversity of employed sensing equipment, the SCADA system must be able to speak the corresponding communication protocols they use, such as BAC Net, ModBus, OPC, KNX, DLMS etc. Furthermore, SCADA system should be able not just to communicate with all aforementioned devices but also to acquire and store data and even more to act on the switches, thus allowing for the remote control of actuators. In this case the actuators are typically tied with existing energy assets. Hence, the control of PV plant, for example, is done through the corresponding inverters, entering heat flows are controlled via remotely controlled servo motors that open/close corresponding valves, the HVAC system is controlled through change of set points for indoor temperature etc.

B. Business logic layer

The main feature of this layer lies in the application of advanced energy scheduling and optimization module, responsible for deriving control actions and offering decision support for planning activities. Also, it has important capabilities for data manipulation (data normalization, conversion, aggregation) and data analytics (used for consumption forecasting).

The former is leveraged upon a custom designed optimization module which allows for both energy infrastructure operation optimization, focusing on minimization of energy costs, as well as for planning optimization aiming at discovering optimal topology and unit sizes for available energy assets, given pricing/regulation context and end user’s energy consumption habits.

The optimization module is a core of EPIC-Hub solution and it consists of three main parts. Starting with a Pre-processing stage, where the key component is the Forecaster module which is based on data analytics. It accounts for the weather forecast, historical energy demand data, models of different renewable energy sources and applicable pricing schemes to produce all input data required by the energy dispatch optimization. The following stage is Scheduling/ Optimization which

contains an overall mathematical model for a given energy infrastructure used for simultaneous Energy Hub and Demand Side Management optimization, as described in detail in [18] and [19]. The optimization process itself is constrained with a number of internal and external Energy Regulations. Finally, in the Management and Control stage, an assessment of optimization outputs is performed and necessary control actions are derived.

When it comes to particular technology used for the development of the optimization module, it was first prototyped in Matlab® environment by defining a linear program aiming at minimization of hub's operation costs. The optimization problem itself is solved by porting the model to IBM ILOG CPLEX®. Having in mind that this module offers the core functionality of the overall solution, a seamless integration with the rest of software components was required. Since the rest of the system was designed with respect to the object oriented paradigm, the development of business logic of Scheduling and Optimization module was based on Java Enterprise Applications (EAR). This was elaborated in detail in [20].

When it comes to the data exchange between Automation and Control level (SCADA) and the rest of EPIC-Hub applications, a flexible Communication Layer

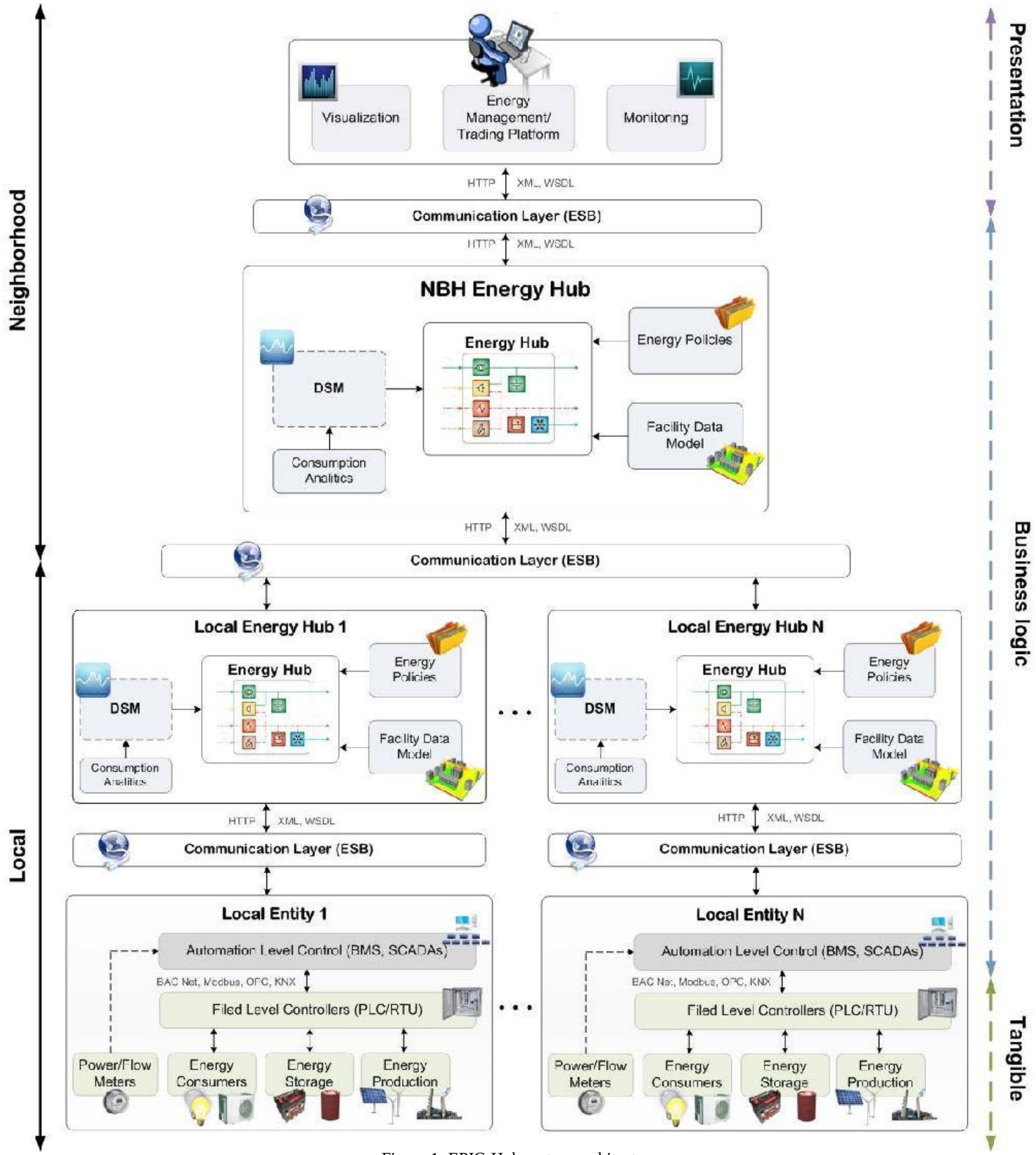


Figure 1: EPIC-Hub system architecture

based on the Enterprise Service Bus (ESB) was used. Again, given the number and variety of stakeholders of this communication layer, a development of proprietary software interfaces between heterogeneous measurement equipment and such data middleware was needed. These interfaces were also referred to as energy gateways which may act as independent software artefacts offering publishing of energy related data towards the middleware while supporting a wide range of communication protocols and frameworks.

The middleware itself is operating as a Web Service, described in an appropriate WSDL format, using proprietary XML-based Canonical Data Model (CDM) for messages and HTTP protocol for communication. Unlike traditional data collection and information exchange mechanisms, the selected service-oriented approach for middleware development allows the EPIC-Hub solution to adapt to multiple deployment scenarios.

C. Presentation layer

Apart from the neighborhood energy optimization related functionalities described previously, the overall management of the EPIC-Hub platform is deployed through a set of user interfaces (UI) and presentation services offered at the presentation layer.

The main purpose of this layer is to allow, at the neighborhood level, the operation of so-called Energy management/trading functionalities. Also, various monitoring and visualization functionalities are considered. The main objective of such layer is to provide the supervision and control functionalities to the appointed neighborhood operator/manager. Although not depicted in Figure 1, the similar layer is considered at the level of each single entity. Together, they facilitate the trading/bidding process which occurs between local entities through a neighborhood intermediate. Particular UI components are derived from scenario-based functions and common energy information system features.

III. Energy Infrastructure Modelling

The framework for modelling of energy infrastructure is based on the concept of energy hub, which represents a flexible and scalable form suitable for the formulation and solution of generic optimization problems related to energy management. Given its holistic approach, it accounts for available energy supply carriers, energy conversion options and existing storage units. The chosen approach was adapted from [22] and can be schematized as depicted in Figure 2.

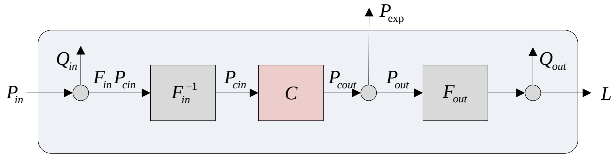


Figure 2: Block schematic of energy hub

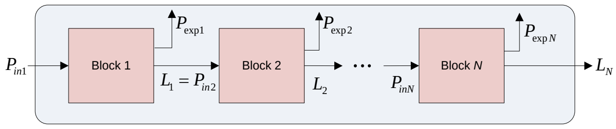


Figure 3: Multi-block schematic of energy hub

The basic modelling block features energy input, sequentially followed by conversion and output stages. Once the energy flows enter the hub (Pin) they can be either stored immediately at input stage (Qin), if storage facilities are available, or dispatched through the dispatch element (Fcin) to the converter stage (C) as Pcin. Once the energy conversion is performed the output (Pcout) can then be exported (Pexp) and the net remaining output (Pout) is forwarded (via the dispatch element Fout) to the output stage where it is either stored (Qout) or immediately employed to satisfy the load demand (L).

The above description refers to the modelling of energy hubs which can be schematized with a single block. However, if the case is about more elaborate and complex topologies, which cannot be represented with a single conversion and/or dispatch stage, the flexibility of the energy hub approach comes into play by placing several blocks in a sequence, as shown in Figure 3. The input to a successive block is taken to be equal to the output of the preceding block, so that for two consecutive blocks the first sees the second as a load.

IV. Energy Management for Neighbourhood

Once the Energy Hub approach is defined on a single entity level, the issue remains to extend this concept towards the neighbourhood. There are several challenges that oppose this objective. The first one lies in the physical limitation of energy interconnections between different entities. Even if they are found in the vicinity of each other, they normally do not share a common energy infrastructure except from the existing electricity of hot water distribution network. However the latter are owned by corresponding authorities and may not be freely used to dispatch energy from one entity to another. The second major obstacle in establishing the neighbourhood energy management is in fact related to the consensus of all beneficiaries around an acceptable business model. Since it is a common situation that different entities act as separate legal subjects, each of them has the objective of achieving the best running performance, including the energy related costs. However, neighbourhood optimization approaches typically aim at operation optimality of neighbourhood as a whole, while the operation of each entity is subjected to the neighbourhood guidelines which, expectedly, are not always the best option for the corresponding entity. As a consequence, there is very little motivation for separate entities to participate in the neighbourhood scheme unless the overall benefits are shared among the stakeholders in a fair manner. This however is not an easy task due to the very nature of the predictive energy management problem. First, entities must accept sub-optimal operation at certain times (for the sake of neighbourhood) in order to gain some benefits in the future. However, since the overall optimization scheme is highly dependent on the forecast of various parameters, such as weather conditions (implying generation from renewables) or consumption profile (depending on weather, number of people, operational requirements etc.), the expected future benefits carry a significant amount of uncertainty. Hence, entities are required to accept the participation in the neighbourhood scheme without being certain of the actual benefits they will receive at the end of contracted period, which may vary from one to an arbitrary number of days or even months. Therefore, EPIC-Hub solution proposes a

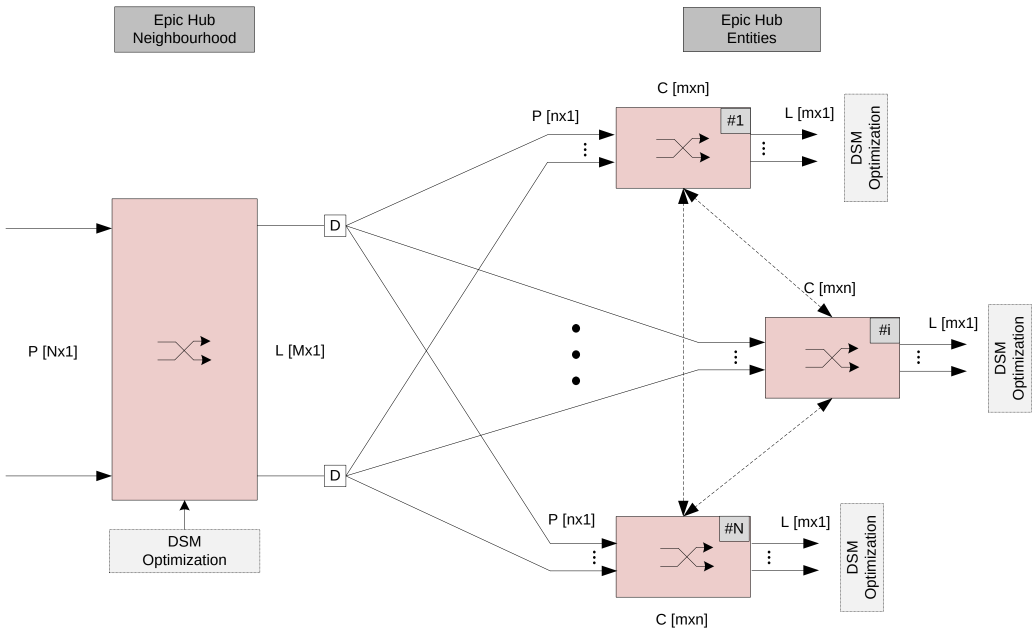


Figure 4: Neighborhood deployment scenario

novel neighborhood scenario, leveraged upon the developed architecture and related technological developments, which aims at tackling these issues. The overall concept is depicted in Figure 4 briefly outlined in the following.

Starting from a set of separate physical entities, each one is represented with an energy hub (numbered with #i), which may be formulated in the form of single or multi block hub, a so-called neighbourhood is formed on voluntarily basis. A neighbourhood may be represented by either a logical and/or physical entity at which a holistic energy management optimization is performed. Picking one option over the other should be governed by the availability of some common energy assets and/or infrastructure, e.g. a CHP plant or a common storage facility. Prior to elaborating the proposed neighbourhood concept and the corresponding business scenario, one should be aware that separate hubs remain empowered with optimization functionalities, which include additional DSM optimization on top of typical Energy Hub optimization.

The entire optimization process is divided into two major steps. The first step refers to the "upwards" direction, going from the separate hubs towards the neighbourhood authority. It starts with the request for energy supply or available storage space from each hub (*P[nx1]*), based on demand (*L[mx1]*), which is derived as an output from the energy hub optimization associated with "local" DSM strategies. These supply vectors are summed accordingly and transferred to the demand side of the neighbourhood (*L[Mx1]*) where another round of energy hub and DSM optimization is applied while taking into account also the energy assets available at neighbourhood level. A necessary neighbourhood supply (*P[Nx1]*) is found as the output, revealing the net energy import from the external grid. However, as a lateral consequence of the neighbourhood optimization the demand of neighbourhood is also altered (due to DSM), yielding a slightly changed profile (*L' [Mx1]*). This has a direct impact on the supply of local entities and, consequently, fulfilment of their energy demand.

At this point, the second step starts, now going from neighbourhood level "downwards" to local entities. The main objective of this "reverse" optimization is to modify the supply of each local entity (*P'[nx1]*) while offering them a quantifiable benefit (financial, environmental etc.) which is calculated for a time span for which optimization is performed. Each local entity is offered with the option to follow the recommendation of the neighbourhood authority or to disregard it, thus relying solely on its own optimization and assets. In the extreme situation, when all local entities decide not to follow the recommendations, there is still one degree of freedom for the neighbourhood optimization, i.e. to optimally operate common energy assets located at the neighbourhood level, completely independent from the local entities. Another issue in this process is how to distribute available energy coming from the neighbourhood (*L' [Mx1]*) and even more how to split the benefits among local entities which participated in the neighbourhood scheme. Different heuristics may be applied for this purpose. For instance, the available energy (*L'*) may be proportionally distributed among the local entities according to the share of their prior energy request in the total energy demand. On the other hand, the split of potential revenues should be tied with the effort of each local entity, i.e. proportional with the share of deviation in their energy supply. After complying with recommendations coming from the neighbourhood level, each local entity follows their proprietary optimization objectives (e.g. operational costs, environmental impact etc.) based on which an actual dispatch of energy is performed.

V. CONCLUSION

With respect to the increasing need for energy management solutions in the context of growing energy related costs, one such solution was proposed by the EPIC-Hub project, presented in this paper. The project aims at delivering a flexible and scalable solution for

holistic energy management at both single entity and neighbourhood/district level. The solution is based on an ICT platform which, based on a service oriented middleware, takes the advantage of integrating various data sources with state of the art energy management paradigms. The latter is based on an existing concept named Energy Hub which offers multi-carrier optimization as well as flexible modelling framework of energy infrastructures. This paper, however documents, apart from the general system architecture, another extension of this concept and proposes an advancement of the optimization strategy in order to tackle energy management challenges on the neighbourhood level. The proposed neighbourhood concept takes the advantage of diverse energy assets located at each entity and allows for online optimization of common energy infrastructure. Depending on available degrees of freedom for the optimization procedure, i.e. the number of energy carriers, conversion capabilities and variable energy pricing, this framework can reduce the operation costs significantly. However, these savings do not come without a compromise from the users' side, which are required to comply with certain changes in the demand profile and/or sharing of their energy production and storage assets.

Finally, having all the aforementioned in mind, it may be concluded that the challenge of neighbourhood optimization represent a demanding task due to several aspects both from technical and organizational/economical perspective. Nevertheless, the contemporary ICT solutions paved the way towards this challenging objective and provided framework for seamless integration of all relevant components. In particular, they play the role of bridging the gap between separated physical entities, and offering a platform for data exchange enabling various energy management services ranging from energy monitoring to analytics and control. Still, some high level organizational issues in the context of neighbourhoods, such as having a consensus on distribution of potential benefits or who should bear the additional cost for the sake of neighbourhood, remained to be further elaborated.

Acknowledgment

The research presented in this paper is partly financed by the European Union (FP7 EPIC-HUB project, Pr. No: 600067), and partly by the Ministry of Science and Technological Development of Republic of Serbia (SOFIA project, Pr. No: TR-32010).

References

[1] Energy Saving Information Platform (ENERsip), EU Fp7 Project, https://sites.google.com/a/enersip-project.eu/enersip-project/.

[2] Carreiro, A.M.; Lopez, G.L.; Moura, P.S.; Moreno, J.I.; de Almeida, A.T.; Malaquias, J.L., "In-house monitoring and control network for the Smart Grid of the future," in *Innovative Smart Grid Technologies (ISGT Europe), 2011 2nd IEEE PES International Conference and Exhibition on* , vol., no., pp.1-7, 5-7 Dec. 2011.

[3] Control and Optimisation for Energy Positive Neighborhoods (COOPERATE), EU Fp7 Project, http://www.cooperate-fp7.eu/index.php/home.html.

[4] Energy management and decision support systems for Energy POSitive neighborhoods (EEPOS), EU Fp7 Project, http://eepos-project.eu/.

[5] Klebow, B.; Purvins, A.; Piira, K.; Lappalainen, V.; Judex, F., "EEPOS automation and energy management system for neighborhoods with high penetration of distributed renewable energy sources: A concept," in *Intelligent Energy Systems (IWIES), 2013 IEEE International Workshop on*, vol., no., pp.89-94, 14-14 Nov. 2013.

[6] Intelligent Neighborhood Energy Allocation & Supervision (IDEAS), EU Fp7 Project, http://www.ideasproject.eu/IDEAS_wordpress/index.html.

[7] Open Dynamic System for Holistic energy Management of the dynamics of energy supply, demand and storage in urban areas (ODYSSEUS), EU Fp7 Project, http://www.odysseus-project.eu/.

[8] Santoro, R.; Braccini, A.; Vecchi, C.; Fantini, C., "Odysseus — Open Dynamic System for holystic energy management — Rome XI district case study: The "Municipality of Rome" XI odysseus pilot cases case study: Technical features, requirements and constraints for improvement and optimization of the energy system Rome municipality neighborhoods," in *Engineering, Technology and Innovation (ICE), 2014 International ICE Conference on* , vol., no., pp.1-7, 23-25 June 2014.

[9] Orchestration of Renewable Integrated Generation in Neighborhoods (ORIGIN), www.origin-energy.eu.

[10] SMARTgrid KeY nEighbourhood indicator cockpit (SMARTKEY), http://www.smartkey.eu.

[11] Neighbourhood Oriented Brokerage Electricity and monitoring system (NOBEL), http://www.ict-nobel.eu/.

[12] S. Karnouskos, M. Serrano, P. J. Marrón, Antonio Marqués, "Prosumer interactions for efficient energy management in smart grid neighborhoods", Proceedings of the CIB W78-W102 2011: International Conference – Sophia Antipolis, France, 26-28 October.

[13] E. Bekiaris, L. Prentza, "Evaluation of an Agent Based Monitoring and Brokerage System for Neighborhood Electricity Usage Optimization", Journal of Energy and Power Engineering 7 (2013) 1915-1921.

[14] Cabras, M.; Pilloni, V.; Atzori, L., "A novel Smart Home Energy Management system: Cooperative neighborhood and adaptive renewable energy usage," in *Communications (ICC), 2015 IEEE International Conference on* , vol., no., pp.716-721, 8-12 June 2015.

[15] Yuanxiong Guo; Miao Pan; Yuguang Fang; Khargonekar, P.P., "Decentralized Coordination of Energy Utilization for Residential Households in the Smart Grid," in *Smart Grid, IEEE Transactions on* , vol.4, no.3, pp.1341-1350, Sept. 2013.

[16] Energy Positive Neighbourhoods Infrastructure Middleware based on Energy-Hub Concept (EPIC-Hub), http://www.epichub.eu/.

[17] P. Favre-Perrod, M. Geidl, B. Klöckl and G. Koeppel, "*A Vision of Future Energy Networks*", presented at the IEEE PES Inaugural Conference and Exposition in Africa, Durban, South Africa, 2005.

[18] Marko Batic, Nikola Tomasevic, Sanja Vranes, "*Integrated Energy Dispatch Approach Based on Energy Hub and DSM*" ICIST 2014, 4th International Conference on Information Society and Technology, ISBN: 978-86-85525-14-8, pp. 67-72, Kopaonik, 09-13.03.2014.

[19] Nikola Tomasevic, Marko Batic, Sanja Vranes, "*Genetic Algorithm Based Energy Demand-Side Management*" ICIST 2014, 4th International Conference on Information Society and Technology, ISBN: 978-86-85525-14-8, pp. 61-66, Kopaonik, 09-13.03.2014.

[20] Marko Batic, Nikola Tomasevic, Sanja Vranes, "*Software Module for Integrated Energy Dispatch Optimization*" ICIST 2015, 5th International Conference on Information Society and Technology, ISBN: 978-86-85525-14-8, pp. 83-88, Kopaonik, 08-11.03.2015.

[21] EPIC-Hub Project Deliverable, "*D3.1 E Reference Architecture andPrinciples*", Lead contributor Thales Italia, 2013.

[22] Almassalkhi M., Hiskens I.A. "Optimization framework for the analysis of large-scale networks of energy hubs". Proceedings of the Power Systems Computation Conference, Stockholm, August 2011.

[23] EPIC-Hub Project Deliverable, "*D2.2 Energy Hub Models of the System and Tools for Analysis and Optimization*", Lead contributor Eidgenoessische Technische Hochschule Zuerich (ETH), 2014.

Correlation of variables with electricity consumption data

Aleksandra Dedinec*, Aleksandar Dedinec**
* Faculty of Computer Science and Engineering,
University St. Cyril and Methodius, Skopje, Macedonia,
** Research Center for Energy and Sustainable Development,
Macedonian Academy of Sciences and Arts, Skopje, Macedonia
e-mail : aleksandra.kanevche@finki, ukim.mk, dedinec@manu.edu.mk,

Abstract— **In this paper the correlation between different variables with the hourly consumption of electricity is analyzed. Since the correlation analyses is a basis for predicting, the results of this study may be used as an input to any model for forecasting in the field of machine learning, such as neural networks and support vector machines, as well as to the various statistical models for prediction. In order to calculate the correlation between the variables, in this paper the two main methods for correlation are used: Pearson's correlation, which measures the linear correlation between two variables and Spearman's rank correlation, which analyzes the increasing and decreasing trends of the variables that are not necessarily linear. As a case study the electricity consumption in Macedonia is used. Actually, the hourly data for electricity consumption, as well as the hourly temperature data for the period from 2008 to 2014 are considered in this paper. The results show that the electricity consumption in the current hour is mostly correlated to the consumption in the same hour the previous day, the same hour-same day combination of the previous week and the consumption in the previous day. Additionally, the results show a great correlation between the temperature data and the electricity consumption.**

I. INTRODUCTION

One of the main creators of the society in the 21st century is for sure the energy. The way of consuming the energy, its price and the manner of providing and accessing the energy resources will greatly affect the quality of people's lives, which in turn answers the question whether the economy of a country will be highly developed or not. On the other hand, smart use of energy and the introduction of new renewable energy sources of energy will be a feature of strong economies. The attractiveness of this topic in the world is huge and it is evidenced by the huge number of published papers. One part of the area, which is very important, is the prediction. The forecasting in the field of energy is very important, because it is an inert field, and the choices that we are making today may have a huge impact on the future generations.

Energy forecasting is present in all of the energy segments and there are three main kinds of forecasting, related to the time period: short, medium and long term. Short-term (less than week) forecasting is becoming more attractive with the introduction of new technologies in the energy field such as the renewable energy sources and the smart grids. The models for electricity forecasting are becoming more attractive for many reasons such as in terms of liberalized electricity market, especially from a financial point of view.

The most important step in forecasting, in any area, is the selection of the input variables on which the output or the predicted variable will depend on. As the choice of input variables is better, the quality of the prediction will be greater. Additionally, as the time scale of the prediction is shorter, the precise choice of input variables has an increasing importance. One of the primary ways of determining the input parameters is by using the correlation between the output variable with various other input variables. Therefore, in this paper, the two main correlation methods are used: Pearson and Spearman coefficients.

The correlation as a method can be used for multiple purposes such as forecasting, validation, reliability and verification of a certain theory. In forecasting, the correlation can be used to see if certain variables were dependent in the past then there is a high probability that they are also going to be dependent in the future, so prediction may be made based on them. In validation, the correlation can be used to check whether the used model gives good results, by comparing the outputs with the actual data.

There are a lot of research papers that use the correlation as a method. For example, in [1] application of various linear and non- linear machine-learning algorithms for solar energy prediction is made. In the same paper it is shown that with proper selection of the input variables using correlation the accuracy of the model can be improved. Different statistical indicators, where one of them is correlation, are used in [2] in order to appraise the performance of adaptive neuro-fuzzy methodology used for selection of the most important variable for diffuse solar radiation. The correlation coefficient between the actual and simulated wind-speed data obtained with artificial neural networks are used in order to calculated the precision of predicted wind-speed [3]. On the other hand prediction of the wind farm using the weather data, analyzes of the important parameters and their correlation is done [4]. Appling regression model, the consumption in the future is predicted for a supermarket in UK [5]. The dependency of the input variable is calculated

using the correlation method. Selection of the key variables in the statistical models is very important because these models are as good as the input data are good [6]. In [6] the consumption of the building is predicted using regularization and the correlation is used in order to eliminate all data that are perfectly correlated. Extreme machine learning (ELM) method correlation method for parameters was applied in [7] in order to estimate building energy consumption. The results of the ELM were validated also by using the correlation method and some other statistical methods.

The goal of this paper is to address the issue of selection of input variables that will be used to forecast the hourly consumption of electricity, by using the method of correlation. The results of this selection can then be used as an input to any forecasting model, whether it is a statistical method or a model in the machine learning field. The selection of input variables is made for real hourly data on electricity consumption in Macedonia for the period from 2008-2014. Additionally, the average hourly temperature data in Macedonia for the same period are also used.

The paper is organized as follows. In the second section the methods that are used for correlation assessment are described. Short review of the data used is given in section III. The results and a discussion are presented in the following section and the last section concludes the paper.

II. METHOD

As it is stated in the introduction, correlation is a method by which the relationship between two variables can be determined. The correlation can be categorized into three groups:

- Positive correlation (the correlation coefficient is greater than 0) – the values of the variables increase or decrease together
- Negative correlation (the correlation coefficient is less than 0) – as the value of one variable decreases, the value of the other increases
- Zero correlation – the values of the variables are not correlated

In this paper the two primary methods for correlation are used: the Pearson’s and the Spearman’s correlation coefficients [8].

A. Pearson correlation

One of the most commonly used methods is the Pearson’s method used in statistics to determine the degree of linear correlation between two variables. This Pearson’s coefficient quantifies the relationship between two continuous variables that are linearly related. The Pearson correlation coefficient is given by the eq.(1).

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right]}\sqrt{\left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]}} . \quad (1)$$

Where:

$r =$ Pearson r correlation coefficient

$n =$ number of value in each data set

$\sum x_i y_i =$ sum of the products of paired scores

$\sum x_i =$ sum of x scores

$\sum y_i =$ sum of y scores

$\sum x_i^2 =$ sum of squared x scores

$\sum y_i^2 =$ sum of squared y scores

B. Sperman rank correlation

Spearman's rank correlation coefficient is a nonparametric measure of statistical dependence between two variables. When compared to the Pearson’s method, the Spearman’s coefficient assesses how well the relationship between two variables can be described using a monotonic function. Additionally, the Spearman’s coefficient is appropriate for both continuous and discrete variables, including ordinal variables.

The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the ranked variables. For a sample of size n, the n raw scores X_i, Y_i are converted to ranks x_i, y_i and is computed from:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} . \quad (2)$$

where:

$\rho =$ Sperman rank correlation

$d_i =$ the difference between the ranks of corresponding values x_i and y_i

$n =$ number of value in each data set

So, as it is for the Pearson’s method, if Y tends to increase when X increases, the Spearman correlation coefficient is positive. If Y tends to decrease when X increases, the Spearman correlation coefficient is negative. A Spearman correlation of zero indicates that there is no tendency for Y to either increase or decrease when X increases.

III. INPUT DATA

As input data, in this paper, the hourly electricity consumption is used, as well as the hourly temperature data in the Republic of Macedonia, for the period from 2008-2014. Therefore, in this section short overview of these data is provided.

The total electricity load in Republic of Macedonia on monthly basis for the years 2008-2014 is presented in Figure 1. It can be noticed that the highest consumption is during the heating season, which means that high share of the electricity consumption is used for heating.

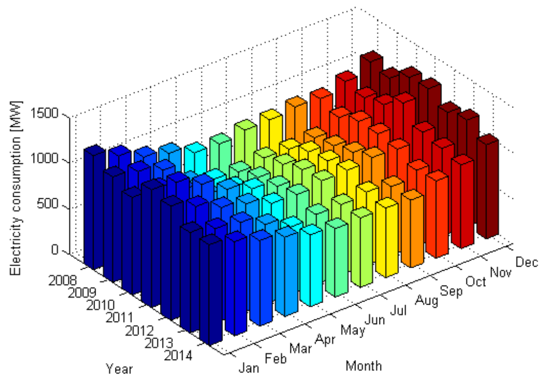


Figure 1. Electricity consumption in Republic of Macedonia on monthly basis for years 2008-2014

The average hourly consumption in RM during the working days in 2014 is presented in Figure 2. As it is shown, there are daily patterns that in a certain way present the consumption behavior of the population in Macedonia. Namely, there are mainly two peaks during the day – the first one starting from 5 pm and the second one starting from 10 pm.

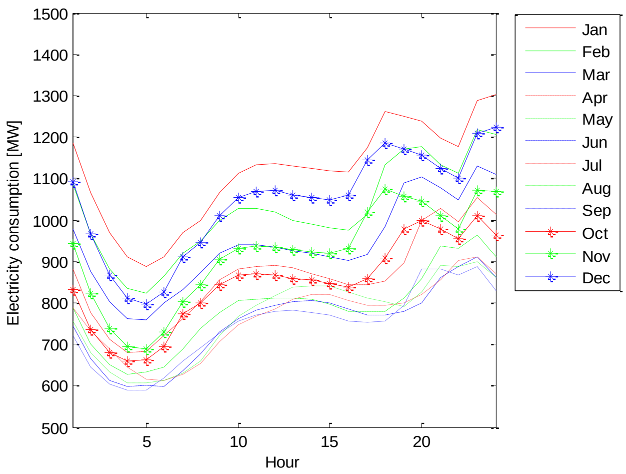


Figure 2. Average hourly consumption in working days for 2014 for RM

In Figure 3 the average monthly temperatures in Macedonia for the period from 2008 to 2014 are shown.

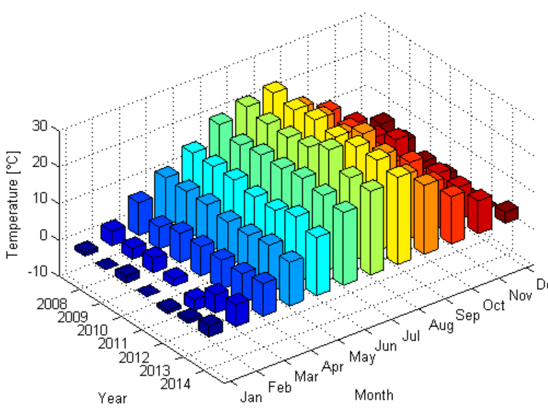


Figure 3. Average monthly temperatures in Macedonia for the period from 2008 to 2014 [9]

IV. Results

The Pearson's and the Spearman's correlation coefficient methods were applied to the electricity consumption data of Macedonia. By using literature review [10]-[13] and by own analyzes of the data, the output variable was compared against eight other variables that may affect the prediction of the electricity load:

1. Temperature
2. Cheap tariff flag
3. Hour of day
4. Day of week
5. Holiday flag
6. Previous day's average load
7. Load of the same hour of the previous day
8. Load for the same hour – day combination of the previous week

In Figure 4 the correlation between the variables load for the same hour – day combination of the previous week, load of the same hour of the previous day and the current load, by using the Pearson's correlation method are presented. Histograms of the variables appear along the matrix diagonal and scatter plots of the variable pairs appear off diagonal. The slopes of the least-squares references lines in the scatter plots are equal to the displayed correlation coefficients. As it is shown, there is a high correlation between these variables, and the coefficient is above 0.9. It can be noticed that, not only there is a correlation between the input variables and the output variable, but also there is a correlation among the input variables. This may mean that when predicting the output, if we need to reduce the size of the input, one of these two input variables may be excluded. Similar results are obtained by using the Spearman's correlation coefficient, as it is presented in Figure 5. The only difference is in the correlation coefficient among the input variables, but in both cases its value is high. The Spearman's method did not give any significant additional results because these three variables are mostly linearly correlated.

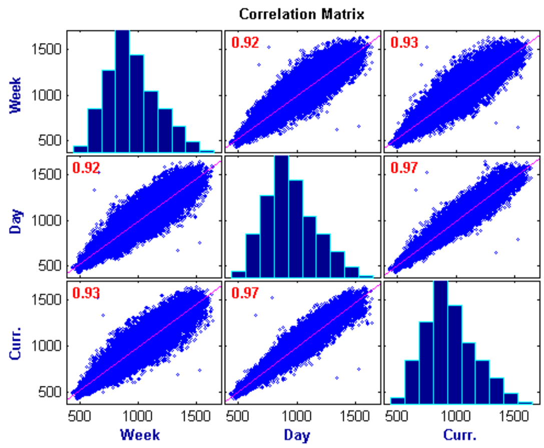


Figure 4. Correlation between the variables: Load for the same hour – day combination of the previous week, Load of the same hour of the previous day and the current load, by using the Pearson's correlation method

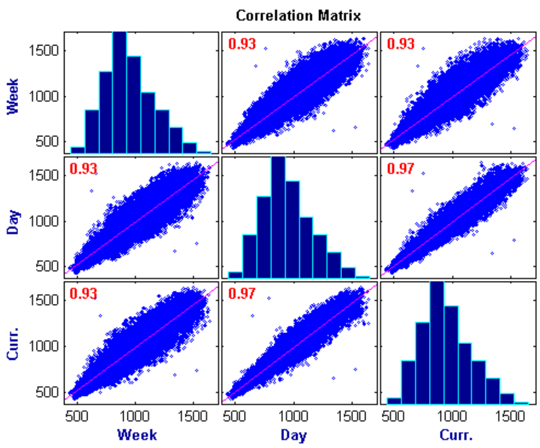


Figure 5. Correlation between the variables: Load for the same hour – day combination of the previous week, Load of the same hour of the previous day and the current load, by using the Spearman's correlation method

The correlation between the temperature data and the electricity load data in Macedonia is presented in Figure 6. It is shown that these two variables are greatly correlated, but they have negative correlation. This means that as one of the variables is increasing, the other is decreasing and vice versa. This may be explained by the fact that, as the temperature is decreasing, more electricity is consumed for heating which is in winter time, but in summer when the temperatures are high, the electricity consumption is low. The same result is obtained by using the Spearman's coefficient.

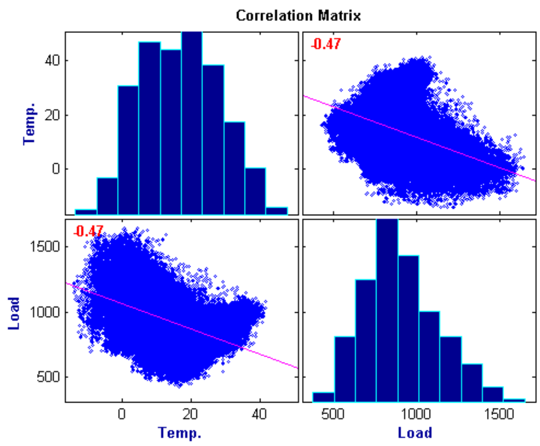


Figure 6. Correlation between the temperature and the electricity load, by using the Pearson's coefficient

Another important input variable is the average load of the previous day, as it is notable in Figure 7. This input variable is compared to the average load of the current or the predicted day. The correlation coefficient is 0.98 which implies that these variables are strongly correlated, as it is also obvious on the histograms.

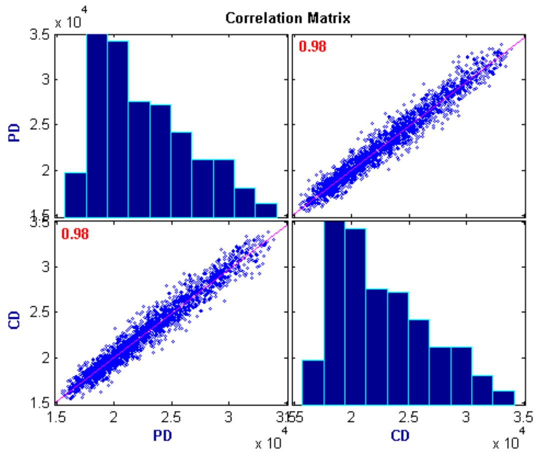


Figure 7. Correlation between previous day's average load and the current day electricity load, by using the Pearson's coefficient

Figure 8 shows the correlation among the maximum daily temperature, maximum load for the same day of the previous week, maximum load of the previous day and the maximum load of the current day, by using Pearson's coefficient. The dark red color and the dark blue color show high correlation. As all of the three input variables are highly correlated to the maximum load of the predicted day, it is obvious that good results will be obtained when using them for forecasting the daily peak electricity load.

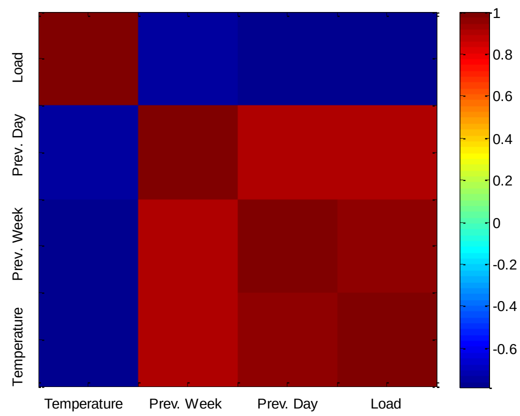


Figure 8. Correlation between the maximum daily temperature, maximum load for the same day of the previous week, maximum load of the previous day and the maximum load of the current day, by using Pearson's coefficient

The last input variables whose correlation is examined against the electricity load are: the cheap tariff flag, the hour of day, the day of week and the holiday flag. Their correlation by using the Pearson's coefficient is presented in Figure 9. All of these input variables are not continuous and Pearson's method is used only for continuous data. However, as presented in Figure 10, the Spearman's coefficient did not significantly change the results.

Although, the correlation between these variables is much lower than between the output and the previously mentioned variables, there still exists correlation. The correlation coefficients highlighted in red indicate which pairs of variables have significant correlations. So, it can

be concluded that all of the analyzed variables are correlated to the electricity consumption in Macedonia, but with different extend.

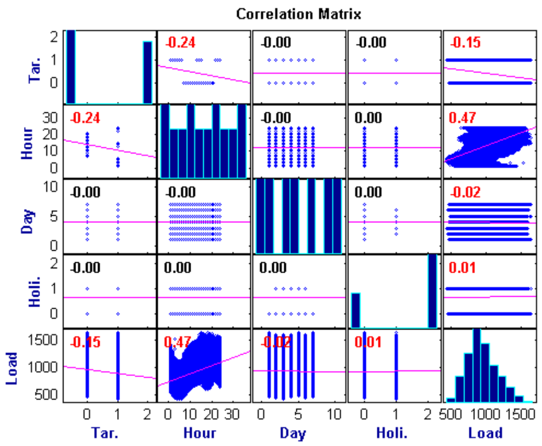


Figure 9. Correlation between the cheap tariff flag, the hour of day, day of week, holiday flag and the current hour electricity load, by using Pearson's coefficient

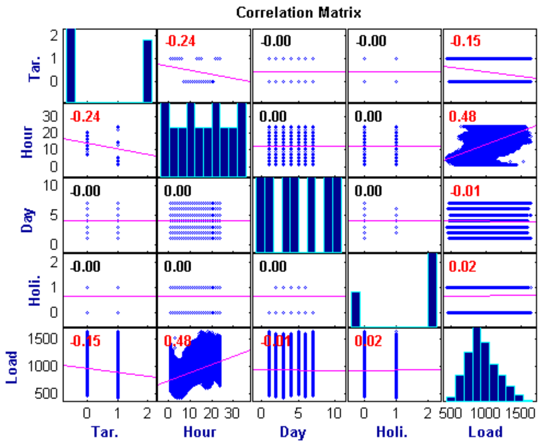


Figure 10. Correlation between the cheap tariff flag, the hour of day, day of week, holiday flag and the current hour electricity load, by using Spearman's coefficient

V. Conclusion

This paper analyzes the correlation between the consumption of electricity in the Republic of Macedonia and eight other variables. The correlation or the degree to which a variable changes its value in terms of another variable, either positively or negatively, is essentially a measure of the usefulness of one variable in predicting the other. Therefore, the correlation can be used as an indicative measure of the relationship between the variables. From this paper, it can be concluded that the correlation between the consumption in the current (or next) hour is mostly correlated (above 0.9) with the historical data for the same hour the previous day, the same hour-same day combination of the previous week and the consumption in the previous day. Significant correlation of 0.47 exists between the temperature and the consumption of electricity in Macedonia. The remaining variables analyzed in this paper, for which the used methods of correlation showed a loose correlation, have discrete values. The fact that their relationship with the predicted variable is not linear or monotonic function does not mean that they will not give additional precision in the forecasting results. However, this mainly depends on the model that is being used for forecasting purposes. In fact, if the model used significantly increases its complexity by adding more variables then the variables with correlation coefficient of less than 0.1 are the first one that would be excluded from the calculations. Otherwise, if the complexity of the model does not increase significantly with the addition of new variables, then it would be desirable to use all of the mentioned variables in this paper, as they all have some correlation with the hourly consumption of electricity.

Additionally, the results of this paper showed that the maximum daily temperature, the maximum load of the same day-previous week and the maximum consumption in the previous day are the variables that are mostly correlated to the daily peak electricity consumption.

Acknowledgment

This work was partially financed by the Faculty of Computer Science and Engineering at the "Ss. Cyril and Methodius" University.

References

[1] S. K. Aggarwal and L. M. Saini, "Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society) 2013-14 Solar Energy Prediction Contest," *Energy*, vol. 78, pp. 247–256, 2014.

[2] K. Mohammadi, S. Shamshirband, D. Petković, and H. Khorasanizadeh, "Determining the most important variables for diffuse solar radiation prediction using adaptive neuro-fuzzy methodology; Case study: City of Kerman, Iran," *Renew. Sustain. Energy Rev.*, vol. 53, pp. 1570–1579, 2016.

[3] J. Koo, G. D. Han, H. J. Choi, and J. H. Shim, "Wind-speed prediction and analysis based on geological and distance variables using an artificial neural network: A case study in South Korea," *Energy*, vol. 93, pp. 1296–1302, 2015.

[4] E. Vladislavleva, T. Friedrich, F. Neumann, and M. Wagner, "Predicting the energy output of wind farms based on weather data: Important variables and their correlation," *Renew. Energy*, vol. 50, pp. 236–243, 2013.

[5] M. R. Braun, H. Altan, and S. B. M. Beck, "Using regression analysis to predict the future energy consumption of a supermarket in the UK," *Appl. Energy*, vol. 130, pp. 305–313, 2014.

[6] D. Hsu, "Identifying key variables and interactions in statistical models of building energy consumption using regularization," *Energy*, vol. 83, pp. 144–155, 2015.

[7] S. Naji, A. Keivani, S. Shamshirband, U. J. Alengaram, M. Z. Jumaat, Z. Mansor, and M. Lee, "Estimating building energy consumption using extreme learning machine method," *Energy*, vol. 97, pp. 506–516, 2016.

[8] J. L. Myers and A. Well, "Research design and statistical analysis," 2010.

[9] Weather Undergraund - Weather History & Data Archive n.d. http://www.wunderground.com/history/

[10] Tang N, Zhang D-J. Application of a Load Forecasting Model Based on Improved Grey Neural Network in the Smart Grid. Energy Procedia 2011;12:180–4. doi:10.1016/j.egypro.2011.10.025.

[11] Jovanović S, Savić S, Bojić M, Djordjević Z, Nikolić D. The impact of the mean daily air temperature change on electricity consumption. Energy 2015;88:604–9. doi:10.1016/j.energy.2015.06.001.

[12] Badri a., Ameli Z, Birjandi AM. Application of Artificial Neural Networks and Fuzzy logic Methods for Short Term Load Forecasting. Energy Procedia 2012;14:1883–8.

doi:10.1016/j.egypro.2011.12.1183.

[13] Mahmoud TS, Habibi D, Hassan MY, Bass O. Modelling self-optimised short term load forecasting for medium voltage loads using tunning fuzzy systems and Artificial Neural Networks. Energy Convers Manag 2015;106:1396–408. doi:10.1016/j.enconman.2015.10.066.

Risk Analysis of Smart Grid Enterprise Integration

Aleksandar Janjic [*], Lazar Velimirovic [**], Miomir Stankovic [***]
[*] Faculty of Electronic Engineering, Nis, Serbia
[**] Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia
[***] Faculty of Occupational Safety, Nis, Serbia
Aleksandar.Janjic@elfak.ni.ac.rs; lazar.velimirovic@mi.sanu.ac.rs; miomir.stankovic@znrfak.ni.ac.rs

***Abstract*— The application domain of Smart Grid is a matter of a great interest of scientific and industrial communities. A comprehensive model and systematic assessment is necessary to select the most suitable Enterprise Architecture (EA) development or integration project from many alternatives. In this paper, a Fuzzy Influence Diagrams with linguistic probabilities is proposed for the Smart Grid EA conceptual model and risk analysis. The proposal relies on a methodology devoted to decision-making perspectives related to EA and risk assessment on Smart Grid initiatives based on fuzzy reasoning and influence diagrams (FID).**

I. Introduction

Electric utilities are turning their attention to Smart Grid deployment including a wide range of infrastructure and application-oriented initiatives in areas as diverse as distribution automation, energy management, and electric vehicle integration. The challenge of integrating all of these Smart Grid systems will place enormous demands on the supporting IT infrastructure of the utility, and many utilities are turning to the discipline of EA for a solution [1]. System interoperability, information management and data integration are among the key requirements for achieving the benefits of Smart Grid. For instance, the use of Advanced Metering Infrastructure can reduce the time for outage detection and service restoration. However, this will require integration of Outage Management System (OMS) with a number of other applications, including Meter Data Management System (MDMS), GIS, work management system, SCADA/DMS and distribution automation functions.

The concept of Smart Grid affects both the topology of the grid and the IT-infrastructure, leading to various heterogeneous systems, data models, protocols, and interfaces on an enterprise level. The discipline of EA proposes the use of models to support decision-making on enterprise-wide information system issues. The challenges arise when a company's information systems need integration. Therefore, a comprehensive model and systematic assessment is necessary to select the most suitable EA development or integration project from many alternatives. In this paper, the new analysis tool for the selection of optimal EA for the Smart Grid implementation has been presented. EA analysis of complex Smart Grid solutions has been carried out using Fuzzy Influence Diagrams. They differ from the conventional ones in their ability to cope with the uncertainty associated with the use of language and in their ability to represent multiple levels of abstraction. The contribution of this paper is the successful implementation of the probabilistic inference mechanism of Bayesian networks and Fuzzy influence diagrams to the Smart Grid conceptual model. This kind of modeling allows the analysis of a wide range of important properties of EA models, which is illustrated on an example of risk analysis of different Smart Grid solutions.

II. EA Model Analysis Using Fuzzy Influence Diagrams

The development of Smart Grid raises many economic questions. The impact of Smart Grid on the development of the economy, market regulation of electricity sales, reduction of the risk for investors and increase of competitiveness is processed in [2]. Also, the issue of increasing the use of renewable energy sources and their optimization within the system of Smart Grid in order to increase the economy competitiveness is considered.

To fulfill these requirements, EA should be model-based, in the sense that diagrammatic descriptions of the systems and their environment constitute the core of the approach. A number of EA initiatives have been proposed, including The Open Group Architecture Framework, Zachman Framework, Intelligrid, and more [3], [4], [5]. Employing the probabilistic inference mechanism of Bayesian networks, extended influence diagrams are also proposed for the analysis of a wide range of important properties of EA models [6]. However, although the concept of influence diagram has been successfully applied to different areas, different Smart Grid architectures have not been modeled in such a way.

A. EA for the Smart Grid Requirements

Smart grid EA encompasses all four levels according to Zachman terminology:

- Conceptual - models the actual business as the stakeholder conceptually thinks the business is, or may want the business to be; defines the roles/services that are required to satisfy the future needs.
- Logical - models of the "services" of the business uses, logical representations of the business that define the logical implementation of the business.
- Physical - the specifications for the applications and personnel necessary to accomplish the task.
- Implementation - software product, personnel, and discrete procedures selected to perform the actual work.

Achieving interoperability in such a massively scaled, distributed system requires architectural guidance, which is provided by the Smart Grid architectural model based on influence diagrams. In this conceptual model of an enterprise, the information represented in the models is normally associated with a degree of uncertainty. The EA model may be viewed as the data upon which the algorithm operates. The result of executing the algorithm on the data is the value of the utility node, i.e. how "good" the model is with respect to the assessed concept (information security, availability, etc.) [6]. Therefore, EA analysis, as the application of property assessment criteria on enterprise architectural models, is the necessary step in the EA model evaluation.

Furthermore, in the field of EA, it is necessary to make decisions about different alternatives on information systems. Making of rational decisions concerning information systems in an organization should be supported by conducting risk analyses on EA models. Dependencies and influences between different systems are not obvious, and for the support of decision making in such complex environments, the concept of a model-based EA analysis is used.

Generally, a conceptual architecture defines abstract roles and services necessary to support Smart Grid requirements without delving into application details or interface specifications. It identifies key constructs, their relationships, and architectural mechanisms. The required inputs necessary to define a conceptual architecture are the organization's goals and requirements. In influence diagrams, decision nodes represent a choice between alternatives. In EA analysis, these alternatives are concretized by different EA scenarios. The risk calculation of different scenarios is performed using fuzzy influence diagram.

III. Smart Grid Scenario Risk Assessment

The overall risk is calculated with exhaustive enumeration of all possible states of nature, and their expected value of risk. Let suppose a system in which risk value node has X_n parent nodes, with different number of discrete states. Fuzzy probability of the chance node X_i being in the state j is expressed as $FP(X_i = x_{ij})$. Fuzzy value of possible consequences in the state x_{ij} is represented by $FD(X_i = x_{ij})$. The expected value of risk is then calculated:

$$R \cong \sum_j \sum_i FP(X = x_i) \otimes FD\left(X = x_{ij}\right). \qquad (1)$$

EA analysis is the application of property assessment criteria on EA models. One investigated property might be the information security of an information system and a criterion for assessment of this property might be the estimated risk level of such an intrusion.

In the following example, Fuzzy Influence Diagram will be used for the risk assessment of two different Smart Grid scenarios. Decision about the introduction of two alternative vendors of Smart Grid application has to be brought, having in mind different constraints, stakeholders, their goals and interdependencies between them. The conceptual model is presented on Figure 1. Decision node (1) presents two possible scenarios of Smart Grid technology introduction (S1 and S2). The chance node (2) depicts the possible impact on information security of these scenarios, with three possible states: Information security low (ISL), medium (ISM) and high (ISH). Appropriate fuzzy conditional probabilities are presented in Table I. Two different stakeholders are presented in this model, including network owner and customers, with their particular goals: profit and satisfaction, presented in chance nodes (3) and (4) respectively.

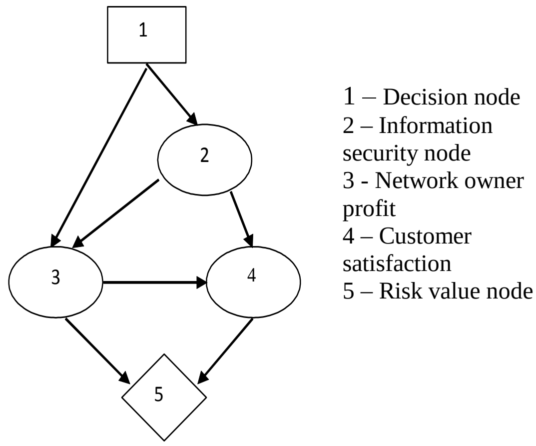


Figure 1. Fuzzy influence diagram for the conceptual model of Smart Grid risk assessment

TABLE I.
FUZZY CONDITIONAL PROBABILITIES FOR THE INFORMATION SECURITY NODE (2)

	Scenario 1 (S1)	Scenario 2 (S2)
Information security low (ISL)	H	EL
Information security medium (ISM)	VL	L
Information security high (ISH)	EL	MH

Network owner profit depends on both: selected scenario and the level of information security and encompasses three possible states (low (PL), medium (PM) and high profit (PH). Fuzzy conditional probabilities for this node are presented in Table II.

TABLE II.
FUZZY CONDITIONAL PROBABILITIES FOR THE NETWORK OWNER PROFIT NODE (3)

States of node 3	States of nodes 1 and 2					
	S1 ISL	S1 ISM	S1 ISH	S2 ISL	S2 ISM	S2 ISH
Low profit (PL)	L	EL	EL	M	L	EL
Medium profit (PM)	MH	VL	EL	ML	M	MH
High profit (PH)	EL	H	VH	EL	VL	L

On the other hand, customer satisfaction is inversely related to the state of network owner profit (the bigger the profit, the customer satisfaction is lower). Three possible states of customer satisfaction are Customer satisfaction is also depending on the state of information security, and

conditional probabilities for this node are presented in Table III.

TABLE III.
FUZZY CONDITIONAL PROBABILITIES FOR THE CUSTOMER SATISFACTION NODE (4)

States of node 4	States of node 2 and 3		
	PL ISL	PL ISM	PL ISH
CL	VL	EL	EL
CM	L	VL	EL
CH	M	H	VH
States of node 4	States of node 2 and 3		
	PM ISL	PM ISM	PM ISH
CL	L	VL	EL
CM	MH	M	ML
CH	EL	L	M
States of node 4	States of node 2 and 3		
	PH ISL	PH ISM	PH ISH
CL	M	L	EL
CM	ML	M	MH
CH	EL	VL	L

Using the expressions (1), values of fuzzy risk for both alternatives are calculated and presented on Figure 2 [7].

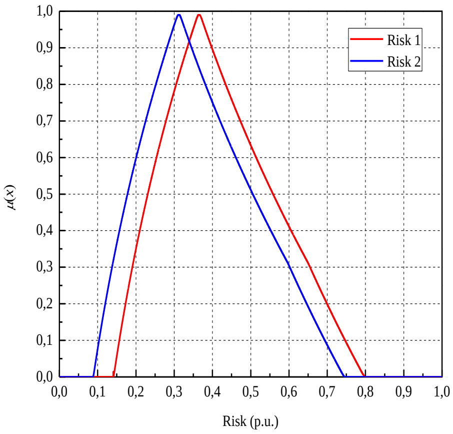


Figure 2. Fuzzy risks for two Smart Grid alternatives

It is obvious that the alternative with low information security risk (S1), and moderate customer satisfaction dominates the alternative with higher profit for the utility.

IV. CONCLUSION

Investments made in Smart Grid technologies are faced with risk that the diverse Smart Grid technologies will become prematurely obsolete or, worse, be implemented without adequate security measures. Furthermore, a Smart Grid cyber security strategy must include requirements to mitigate risks and privacy issues pertaining to Smart Grid customers and uses of their data.

The implementation of Smart Grid requires an overarching architecture to accommodate regulatory, societal and technological changes in an uncertain environment. Utilities are managing much more data and information in real time, using completely “of the shelf” applications requiring internal and external interoperability. A critical step in this integration is the development of the EA semantic model enabling data and information services. In this paper, fuzzy influence diagrams are proposed for the conceptual modeling of Smart Grids and the decision making support in the assessment of different Smart Grid alternatives and integration framework. The probabilistic inference mechanism of Bayesian networks and influence diagrams with linguistic probabilities is adequate to the EA model viewed as the data upon which the algorithm operates. Therefore, EA analysis as the application of property assessment criteria on enterprise architectural models is the necessary step in the EA model evaluation.

The advantage of this kind of modeling is the consistent representation of the Smart Grid conceptual level, consisting of several domains, each of which containing many applications and roles that are connected by associations, through interfaces. A vast range of properties on enterprise architectural model can be assessed using this model. Furthermore, this kind of modeling allows the analysis of a wide range of important properties of EA models, which is illustrated on an example of risk analysis of different Smart Grid solutions

ACKNOWLEDGMENT

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia under Grant III 42006 and Grant III 44006.

REFERENCES

[1] A. Ipakchi, “Implementing the Smart Grid: Enterprise Information Integration,” *Grid-Interop Forum 2007*, K. Parekh, Utility Enterprise Information Management Strategy, 2007, pp. 122-127.

[2] C. D. Clastres, “Smart grids: Another step towards competition, energy security and climate change objectives,” *Energy Policy*, 39(9), 2011, pp. 5399–5408.

[3] TOGAF, Introduction The Open Group Architecture Framework, 2009.

[4] J. A. Zachman, “A Framework for Information Systems Architecture,” *IBM Systems Journal*, 26(3), 1987, pp. 276-292.

[5] Electric Power Research Institute, IntelliGrid Program, 2014.

[6] P. Johnson, E. Johansson, T. Sommestad, J. Ullberg, “A Tool for Enterprise Architecture Analysis,” *11th IEEE International Enterprise Distributed Object Computing Conference*, Annapolis, MD, 15-19 Oct. 2007, pp. 142.

[7] A. Janjic, M. Stankovic, L. Velimirovic, Multi-criteria Influence Diagrams – A Tool for the Sequential Group Risk Assessment, in Granular Computing and Decision-Making, Studies in Big Data, 10, 2015, pp 165-193.

An implementation of Citizen Observatory tools used in the CITI-SENSE project for air quality studies in Belgrade

Milena Jovašević-Stojanović*, Alena Bartonova**, Dušan Topalović*, Miloš Davidović*, Philipp Schneider**
and the CITI-SENSE consortium***
* Vinča Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia
** NILU Norwegian Institute for Air Research, Kjeller, Norway
*** http://www.citi-sense.eu/Project/Consortium.aspx
mjovst@vinca.rs, alena.bartonova@nilu.no, dusan.topalovic@vinca.rs, davidovic@vinca.rs, ps@nilu.no,
citisense.beograd@vinca.rs

Abstract—According to Serbian Environmental Agency (SEPA), more than 30% of citizens of Serbia were exposed to air that is considered not healthy in 2014. An important element in the work to reduce atmospheric pollution is involvement of the citizens, through awareness raising and other means such as adherence to pollution reduction measures.

The aim of our work is to raise awareness of pollution gradients and levels by developing a user-friendly citizen's observatory for air quality. This would allow the general public to generate and access information about air quality that is directly relevant to them – near their homes or work places. Citizen observatory is a quickly developing concept: it is a combination of citizen science approaches with an ICT infrastructure, to objectively monitor or subjectively assess the state of the environment (air quality), and to provide this information in a manner that can be useful to both individual citizens, their groups or the authorities.

Monitoring techniques for air quality are numerous, and those that are suitable for citizen's own measurements are just starting to be accessible. Our citizen observatory (co.citi-sense.eu) provides access to a selection of products that were developed in collaboration with users, and that allow the users to participate in air quality assessments. A practical implementation of the citizen observatory requires establishing operative connections between monitoring and data collection systems, a central information platform, and user products such as apps or web pages. Practical challenges include quality control and assurance of the input information, use of existing communication infrastructures, and not least identification of the interests and needs of the citizens: oftentimes, the awareness level of the problem addressed, and an understanding of the issues, are the limiting elements for the use of such systems. This paper shows some of the products intended for the public of Belgrade. The work is a result of the CITI-SENSE consortium efforts and can be accessed at co.citi-sense.eu*.

* CITI-SENSE, www.citi-sense.eu, is a collaborative project co-funded by the European Union's Seventh Framework Programme for Research, Technological Development and Innovation under grant agreement no 308524.

I. Introduction

Air pollution stems from both anthropogenic and natural emissions that undergo further changes in the atmosphere. It is a mixture of mixtures that is not constant in level and composition, and varies through space and time.

In order to warn against the harmful consequences of exposure to main pollutants (such as respirable particulates matter RPM , nitrogen dioxide NO_2, sulfur dioxide SO_2 and ozone O_3, and to protect human health, World Health Organization WHO established air quality guidelines [1]. A decade after, air pollution is the single largest environmental health risk in Europe [2]. Premature death attributable to air pollution happen most due heart disease and stroke, followed by lung diseases and cancer [3]. In addition, air pollution is associated with increase in incidence of numerous additional diseases. The International Agency for Cancer Risk IARC designated outdoor air pollution as a Group 1 carcinogenic substance, i.e., proven human carcinogen [4].

Numerous publications estimated that level of regulated air pollutants in most European cities are far above the air quality guidelines values [2]. As such, citizens are at risk to be exposed to potentially harmful levels of air pollutants.

Owing to increased focus and legislative commitments, more and more cities provide timely air quality information to the public through printed and electronic media including web pages and mobile apps. Question is how useful is such information, as the most important issues for citizens, i.e., air quality data at individual level, is still a rarity. The information on the content of ambient air and related hazards is currently mostly generic, and seldom personally relevant. It would be necessary to offer information to a person about air quality level in microenvironment and on the route s/he frequents, and what does that mean for her/him. It is of ultimate importance for citizens to recognize the problem and to change behavior related to their contribution and their exposure to air pollution.

In cities, dominant local sources of air pollution are road traffic along with domestic combustion [2]. Studies show that traffic related air pollution may cause major

adverse health effects in the population living at or near air polluted roadways [5]. Residential heating with wood and coal is an important source of outdoor air pollution, but also cause indoor air pollution through infiltration. Looking at Europe, central Europe [6], including parts of Balkan peninsula, is the region with the highest levels of outdoor fine particulate matter, PM with an aerodynamic diameter of less than 2.5 um ($PM_{2.5}$) that can be traced to both traffic and residential heating with solid fuels.

The highest population density in Serbia is in the area covered by the Master Plan of Belgrade, where 2/3 of inhabitants of the Belgrade Metropolitan live [7,8]. Average fleet age of passenger cars in 2008 was under 12 years in central zone of city and between 12 and 14 years in other zones that belong to Master Plan of Belgrade area [9]. In the region of Belgrade, the total number of registered motor vehicles in 2011 was 556,070 (85% passenger cars, 9% tracks).

The Public Utility Company “Beogradske elektrane” installed heating capacity of 2656.5 MW produced by 15 district heating facilities and 240 MW by other local heating facilities [7]. There is lack of data about type of fuel that is used by citizens of Belgrade for residential heating for houses that are not connected to district and local heating system. But, it is estimated that in 2004, about 60% of the population of Serbia used wood and lignite coal as their major source of energy for heating, domestic hot water and cooking [10].

The CITI-SENSE project aims to develop a mechanism through which the public can easily be involved, a set of Citizen Observatories [11]. Using a combination of citizen science and environmental monitoring approaches, we have developed technological tools for public involvement, and we are testing these tools to investigate their potential for a large scale public use. This article describes the tools we intend to use in Belgrade.

II. THE CITI-SENSE CONCEPT

CITI-SENSE is developing “citizens’ observatories” to empower citizens to contribute to and participate in environmental governance, to enable them to support and influence community and societal priorities and associated decision making. It has been working on developing and testing sensors for distributed monitoring of environmental exposure and health associated with outdoor air quality and the physical environment, as well as the quality of indoor environment in schools [12].

The aims of CITI-SENSE are to raise environmental awareness in citizens, raise user participation in societal environmental decisions and provide feedback on the impact that citizens had in decisions. Using the CITI-SENSE approach, these aims can only be achieved after user-friendly technological tools are in place.

The concept of CITI-SENSE rests on three pillars: (i) technological platforms for distributed monitoring; (ii) information and communication technologies; (iii) societal involvement. Overview of the project is in Figure 1. Three multi-center case studies focus on a range of services related to environmental issues of societal concern:

- combined environmental exposure and health associated with ambient (outdoor and indoor) air quality,
- noise and development of public spaces,
- and indoor air at schools.

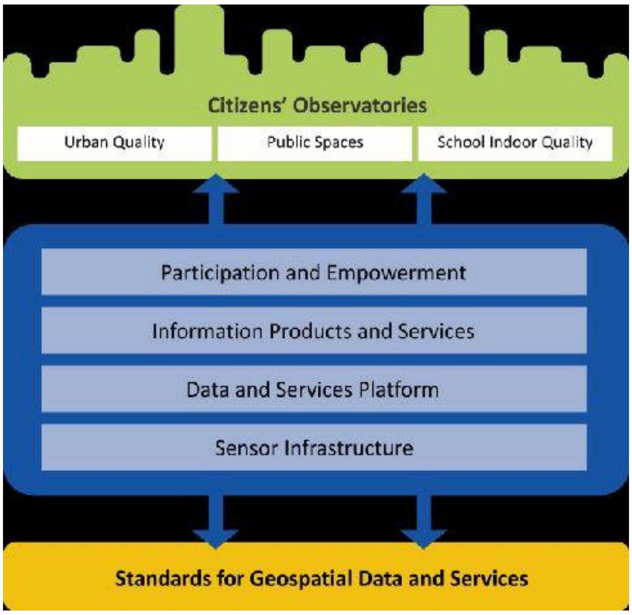


Figure 1. Schematic overview of the project elements and partner involvement in CITI-SENSE. Source: http://co.citi-sense.eu.

The CITI-SENSE project is developing infrastructure to present in near-real-time the levels of selected air pollutants in eight European cities, including the area of the Belgrade Master Plan. This will be done using data fusion algorithms utilizing statistical information and own collected data from low-cost sensors distributed over an observation area. The project is also enabling citizens to participate in monitoring outdoor and indoor air by using personal sensors, and turning citizens into sensors through filling questionnaires and by providing their own perception about the air quality in their surroundings. The highly spatially resolved data on air quality and perception is geo-located and allows for simultaneous visualization of information on a map. All data collected will be available to the citizens in a user-friendly and visually informative layout, using both web services and mobile phone apps.

Adolescents spend significant part of their time indoors, mainly at home, but also in schools. Ratio between the inhalation rate and body weight is higher for adolescents than for adults [13]. Also, children’s tissues and organs developing rapidly so the susceptibility of children to air pollution is greater than in adults [14]. CITI-SENSE project made devices and tools, and offer children from elementary schools and students from secondary school to learn more about the importance and levels of air pollution in school microenvironment by following, collecting and analyzing data from installed low-cost cost sensors. The main aim is to empower them to be able to understand the issue, and to propose measures for improvement of the indoor air quality in schools, as well as to perform research studies.

Figure 2 shows the schematics of a general CITI-SENSE platform data flow. Within CITI-SENSE, there are nine pilot cities (across all use cases), including Belgrade, that will employ one or more end user ‘products’ developed within the project. These products are building on the various support services, such as sensor platforms, GIS, WMS, or mathematical modeling

that actually enable the products to function. The main products fall into two basic types:

a) a web application and

b) a smart phone/mobile device application.

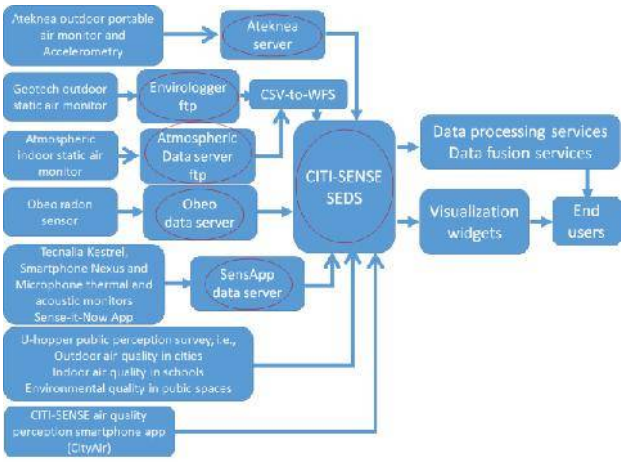


Figure 2. CITI-SENSE platform data flow. Source: http:co.citi-sense.eu and [15].

III. CITI-SENSE OUTDOOR AIR QUALITY STUDY

There are two ways used for personal exposure assessment:

- Direct assessment – a person carries a portable sensor device that detects concentrations and activity level while on move through the urban environment.
- Indirect assessment - using a network of static sensors distributed over the city. The sensor data is combined with statistical model using data fusion techniques, to provide air quality maps for the city for each hour with sufficient measurements. These maps can then be used to estimate individual exposure along a given pathway through the city.

A. Direct assessment

In CITI-SENSE, a Little Environmental Observatory - LEO (Figure 3) is used for direct personal exposure assessment (developed by Ateknea). The LEO is a portable sensor pack (80x96x44 mm) equipped with Alphasense electrochemical sensors for measuring NO, NO_2 and O_3. It also provides information about the current temperature and relative humidity. ExpoApp (Figure 4) is a smartphone application for Android devices that communicates with the LEO sensor via bluetooth to read the data and upload it to Ateknea's platform. ExpoApp also collects information about physical activity by using the accelerometer already in the smartphone. The Ateknea platform then preprocess the data before they are sent onwards to the CITI-SENSE platform.

The near-real-time and historical measured values of all mobile sensors in each of the cities are available on a web portal as shown in Figure 5.

Figure 3. LEO personal device

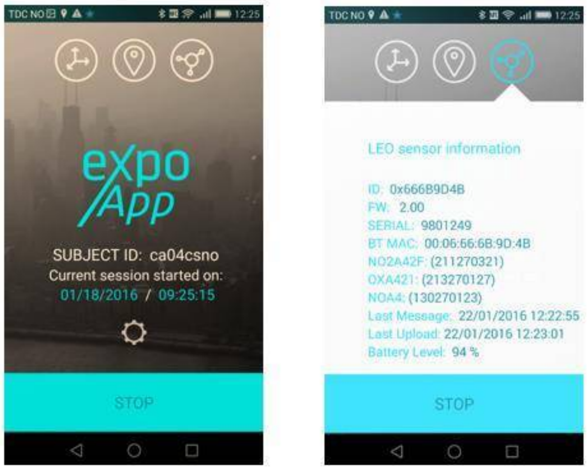


Figure 4. Mobile application for LEO personal device-ExpoApp

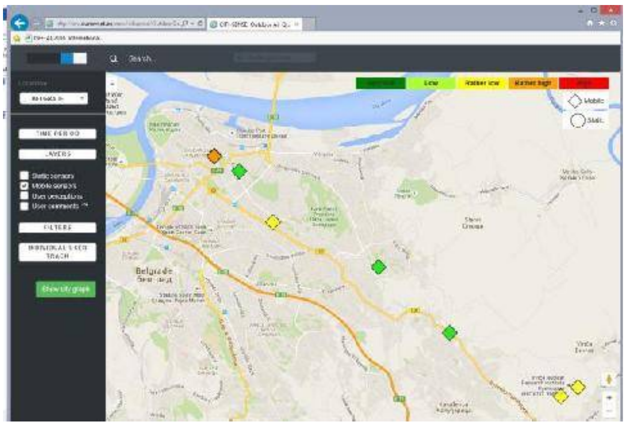


Figure 5. Looking up LEO data

B. Indirect assessment

Current air quality monitoring networks aim at compliance monitoring and consist of a prescribed number of stations for selected locations. They employ rigorous standardized QA/QC protocols. These reference and equivalent ambient PM and gaseous monitoring units do not capture spatial gradients in the area for which they are representative, and cannot by themselves provide individualized personal information [16]. The low-cost sensors deployed with support of the CITI-SENSE project have a significant potential for improving high-resolution mapping of air quality in the urban environment. The

procedure of creating near-real-time maps [17,18] consists of:

- Creation of a basemap that provides information about general spatial patterns:
- Establishing network of low-cost sensor that provide information about current status of atmosphere, air pollutants and meteorological parameters level, at sampling locations
- Fused map that is **v**alue-added product providing a best guess of current state of atmosphere for the entire domain.

A static basemap is created for each city and each air pollutant of interest to show the long-term spatial patterns.

For development of a basemap it is the best to use urban-scale dispersion model, as it is used for Oslo. For most cities however, the detailed input information is not available, and we apply Land Use Regression (LUR) modelling as an alternative technique. LUR is a statistical modelling technique used to spatially extrapolate concentration of air pollutant over limited observed area based on values of predictor variables. Underlying principle is that the concentration of air pollutants is strongly correlated to the predictor variables, and assumption that we know the values of predictor variables anywhere in the area of interest. Predictor variables may include traffic intensity, road length in buffer around monitoring site, or other variables that are locally available in GIS [19].

European multicity model, developed by [20], extended land use approach to model $PM_{2.5}$ and NO_2 pollutants for several European cities. This model was used to create basemaps over area of Belgrade Master Plan. LUR models often used to predict long-term average concentrations of air pollutants, e.g. LUR modeling of annual average NO_2 as it is shown at Figure 6 [21]. Figure 7 presents an example of fused map for NO_2 over Belgrade Master Plan area, calculated data from Local monitoring network for NO_2 consisting of 14 sampling sites.

Figure 8. shows photos of static nodes developed for use in CITI-SENSE main and pilot campaigns, AQMesh form Geotech (battery power supply) and EB700 from Dunavnet (mains power supply). Both nodes have integrated Alphasense electrochemical sensors for gases and optical sensor for RPM. In total, 34 nodes are distributed over Belgrade Master Plan area, 25 AQMesh and 9 EB700 nodes.

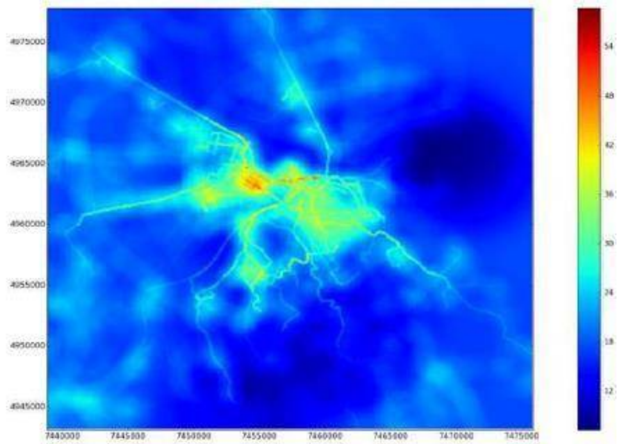


Figure 6. Basemap for annual average of NO_2 over Belgrade Master Plan area

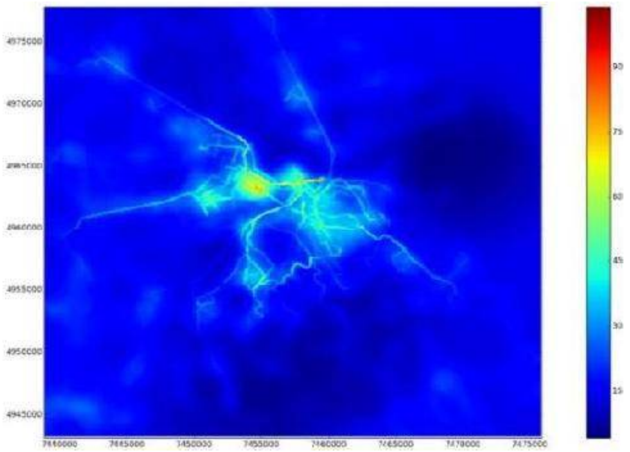


Figure 8. An example of NO_2 fused map for Belgrade

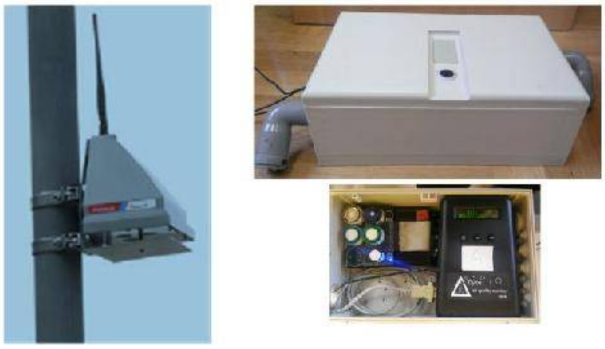


Figure 8. AQMesh (Geotech) and EB700 (Dunavnet) device with low-cost sensors for gases and RPM

Calibration of the statisc nodes in Belgrade was performed by co-locating with a reference instruments at an Automatic Monitoring Station (ATM) that is part of the state network run by the Serbian Environmental Protection Agency (SEPA).

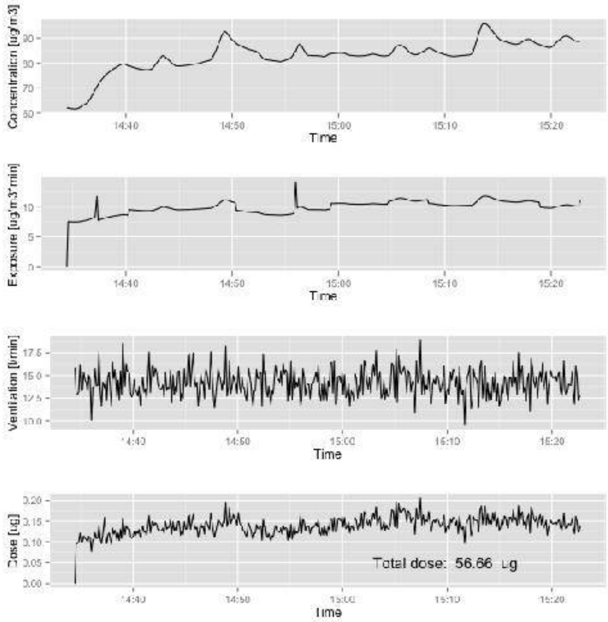


Figure 9. Personal exposure and dose assessment according to observed and/or modelled concentration and approximated ventilation rate [17]

IV. CONCLUSION

The above tools are intended for use in Belgrade during a trial period in spring 2016. The technological challenges were numerous, but at this time, a working prototype exists. Currently, the main challenge is connected with the use of the system: how to make the system known to the public, and how to engage both the interested public, and the authorities who are one of the intended recipients of the collected information. Only by a collaboration between the public and the authorities (on all levels) may we achieve progress towards better quality of air.

ACKNOWLEDGMENT

The low-cost sensors and IC technologies described here evolved through the work undertaken by the CITI-SENSE Consortium. CITI-SENSE is a collaborative project partly funded by the EU FP7-ENV-2012 under grant agreement no 308524 in period 2012-2016.

REFERENCES

[1] WHO, ”Air Quality Guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide”. WHO Regional Office for Europe, Copenhagen. 2006.

[2] EEA, Air quality in Europe — 2015 report. Copenhagen, *EEA Report No 5/2015*,2015.

[3] WHO, “Burden of disease from Ambient Air Pollution for 2012 - Summary of results” World Health Organization 2014., http://www.who.int/phe/health_topics/outdoorair/databases/AAP_ BoD_results_March2014. pdf , accessed February 2016.

[4] IARC press release, https://www.iarc.fr/en/media-centre/iarcnews /pdf/pr221_E.pdf , accessed February 2016.

[5] HEI Panel on the Health Effects of Traffic-Related Air Pollution, “Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects”. *HEI Special Report 17*. Health Effects Institute, Boston, MA., 2010.

[6] WHO, “Residential heating with wood and coal: health impacts and policy options in Europe and North America”, World Health Organization 2015.

[7] Statistical Office of Belgrade Citizens, 2014. Statistical Yearbook of Belgrade 2013. https://zis.beograd.gov.rs/images/ZIS/Files/ Godisnjak/G_2013S_03.pdf , assessed February 2016.

[8] Statistical Office of the Republic of Serbia, 2014. 2011 CENSUS ATLAS. www.popis2011.stat.rs, assessed February 2016.

[9] M. Pavlović,, “Future Advancements of End of Life Vehicles System in Serbia”. Recycling Association of Motor Vehicles of Serbia, 2009. http://www.iswa.org/uploads/tx_ iswaknowledgebase/s406.pdf , accessed February 2016 .

[10] A. Kovacević, “Stuck in the Past: Energy, Environment and Poverty in Serbia and Montenegro” United Nations Development Programme, Belgrade, 2004.

[11] H.-Y. Liu, M. Kobernus, D. Broday, A. Bartonova, “A conceptual approach to a citizens' observatory - supporting community-based, environmental governance” *Environmental Health* 2014. vol. 13 pp. 1-13, doi: 10.1186/1476-069X-13-107

[12] M. Engelken-Jorge, J Moreno, H Keune, W Verheyden, A Bartonova, CITI-SENSE Consortium. Developing citizens' observatories for environmental monitoring and citizen empowerment: challenges and future scenarios. *In: Proceedings of the Conference for E-Democracy and Open Governement (CeDEM14)*: 21–23 May 2014. Danube University Krems, Austria. Edited by Parycek P, Edelmann N. pp.49–60. http://www.donau-uni.ac.at/imperia/md/content/department/gpa/ zeg/bilder/cedem/cedem14/cedem14_proceedings.pdf , accessed February 2016.

[13] G. Viegi, M. Simoni, A. Scognamiglio, S. Baldacci, F.Pistelli, I. Carrozzi, I.Annesi-Maesano, Indoor air pollution and airway disease, Int. pollution and airway disease, *Int. J. Tuberc. Lung Dis*., 2004. Vol. 8. pp. 1401–1415.

[14] E.M. Faustman, S.M. Slibernagel, R.A. Fenske, T.M. Burbacher and R.A. Ponce, Mechanisms underlying children's susceptibility to environmental toxicants, *Environ Health Perspect*. 2000. vol. 108. pp. 13–21.

[15] M. Kobernus, AJ. Berre, M. Gonzalez, H.-Y. Liu, M. Fredriksen, R. Rombouts, A. Bartanova, “A Practical Approach to an Integrated Citizens' Observatory: The CITI-SENSE Framework” *In: Proceeding of the Workshop "Environmental Information Systems and Services - Infrastructures and Platforms 2013 - with Citizens* Observatories, Linked Open Data and SEIS/SDI Best Practices", co-located with the International Symposium on Environmental Software Systems 2013 (ISESS 2013). ENVIP 2013, Neusiedl am See, Austria, October 10, 2013. Edited by Arne J. Berre and Sven Schade. http://ceur-ws.org/Vol-1322/paper_1.pdf, , accessed February 2016.

[16] M. Jovašević-Stojanović, A. Bartonova, D. Topalović, I. Lazović B. Pokrić, Z. Ristovski Z. “On the use of small and cheaper sensors and devices for indicative citizen-based monitoring of respirable particulate matter”. *Environmental Pollution* 2015. vol. 206 pp. 696-704.

[17] P. Schneider, N. Castel. W. Lahoz, „Making sense of crowdsources observations: Data fusion technique for real-time mapping of urban air quality“, ESA eo open science 2.0, 12-14 October 2015, Frascati, Italy, http://www.eoscience20.org/

[18] W.A. Lahoz, P. Schneider Data assimilation: making sense of Earth Observation. *Frontiers in Environmental Science*, 2014 vol 2 pp 1–28.

[19] W. Liu, X. Li, Z. Chen, G. Zeng, T. León, J. Liang, et al.,”Land use regression models coupled with meteorology to model spatial and temporal variability of NO2 and PM10 in Changsha, China” *Atmospheric Environment*, 2015 vol 116 pp. 272-280.

[20] Wang M. Wang, R. Beelen, T. Bellander, M. Birk, G. Cesaroni, M. Cirach, et al., “Performance of multi-city land use regression models for nitrogen dioxide and fine particles” *Environmental health perspectives* 2015 vol 122 pp. 843-849.

[21] M. Davidović, D. Topalović , M. Jovašević-Stojanović, “Use of European Multicity Model for Creation of NO2 and PM2.5 base maps for Belgrade”, COST Action TD1105-EuNetAir, WGs Meeting, Belgrade, 13 - 14 October 2015, http://www.eunetair.it/, assessed February 2016.

Software development for incremental integration of GIS and power network analysis system

Milorad Ljeskovac[*], Imre Lendak[**], Igor Tartalja[*]
[*] University of Belgrade, School of Electrical Engineering, Belgrade, Serbia
[**] University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
mljeskovac@gmail.com, lendak@uns.ac.rs, tartalja@rcub.bg.ac.rs

***Abstract*—The paper discusses an automation procedure for integration of Geographic Information System (GIS) and Power Network Analysis System (PNAS). The purpose of PNAS is to provide analysis and optimization of electrical power distribution system. Its functionality is based on a detailed knowledge of electrical power system model. Automation procedure problems are identified and concepts of an implemented software solution are presented. The core of the solution is the software for implementation of an internal model based on IEC 61970-301 CIM standard. Second part of the solution, which this paper is focused on, is a software layer that detects difference between two relevant GIS states just before successive synchronizations with PNAS. Using "divide and conquer" strategy, the detected difference is split into groups of connected elements. Those groups are used for update of PNAS to make it consistent with GIS. The paper also explains an algorithm for detection of the difference between two states and a procedure for formation of groups, as well as the conditions and limitations of automation procedure for incremental integration.**

I. Introduction

Manual integration of Geographic Information System (GIS) and Power Network Analysis System (PNAS) cannot keep the pace with dynamics of changes inside Power System (PS). Frequent changes inside PS are accumulating inside GIS and there is a need for their periodic transfer into PNAS. The transfer has to be efficient and reliable. Even though GIS and PNAS are often found inside PS, and there is a need for their integration, commercial GIS systems do not offer support for automated integration with PNAS system.

Fully automatic integration of these two systems is hard to achieve. A detailed list of all the problems, which have to be resolved for their integration, can be found in [1]. While fully automatic solution is yet to come, there is a need for partial automatization of integration through simplification of its procedure and reduction of volume and need for manual data entry. This solution has to fit seamlessly into already existing business processes where GIS is the central source of data. This paper presents a solution that fulfils those requirements. Although the solution does not guarantee a full automatic integration, its usage is far more efficient and reliable than manual integration. It significantly reduces the volume of data that has to be manually synchronized.

The remaining part of the paper is organized in the following sections. *Problem* explains the role of GIS and PNAS systems inside PS, the need for their integration and the problems that occur on the way. *Incremental integration* describes the procedure for detection of difference between two relevant GIS states just before successive synchronizations of two systems, and presents the way of splitting of the detected difference into smaller groups, which are used for PNAS update. *Discussion of results* evaluates the solution and gives a short overview of its usage by two clients. The *Conclusion* summarizes the most important experiences and provides directions for further research and development.

II. Problem Statement

In this section, the roles of GIS and PNAS inside PS will be described. General integration problems of these two systems will be presented and a concrete problem, which this paper addresses, will be explained.

The role of GIS is management and visualization of spatial data. Data are objects in space, which represent inventory of one PS and include substations, conducting equipment (inside or outside of substations) and conducting lines. It is widely used in versatile fields. In PS it represents the central point for many technical and administrative business processes.

PNAS is specific for PS. It helps to manage and maintain the system. Its role is to increase the overall efficiency of the system by anticipating and preventing the problems that lead to disconnection of customers. To be able to fulfil its basic functionality it needs a detailed knowledge of PS electrical model. It also needs all information about equipment status changes and measurement results inside processing units in a relatively short time interval after the change has occurred.

What is common for both systems is that they are modeling PS, but in different ways and for different purposes. It would be ideal for both systems to work on one, shared data model. In the absence of the shared model, the problem is usually resolved through system integration, which is the subject of this paper.

The paper describes a developed procedure for incremental integration. The integration implies a transfer of changes from spatial GIS model into electrical PNAS model. The procedure also checks the validity of data that represent transferred changes. Frequent synchronizations indirectly contribute to the overall validity of GIS data. GIS elements are manually supplemented with the data that represent their electrical characteristics. Supplemented data are not part of GIS data model, and GIS has limited abilities for checking their completeness and validity. Connections among GIS elements are defined by spatial connectivity rules that exist among

those objects. These rules are not sufficient to guarantee a full and precise electrical connectivity of elements. Other business processes do not depend on supplemented electrical properties, and the problems with these data stay unnoticed all the way down to the moment of integration with PNAS. Validity and proper connectivity of exported electrical data is a precondition for successful synchronization. It has to be provided by proper planning and discipline during manual GIS data entry. During synchronization, discovered errors in the exported data have to be fixed first in GIS, and then the whole export has to be done again.

Integration uses internal data model based on IEC 61970-301 [2] CIM standard. CIM is described by its UML model and defines dictionary and basic ontology needed for data exchange in the context of electrical power transmission and distribution inside a PS. It is used for derivation of other specifications like XML and RDF schemas, which are needed for software development in integrations of different applications.

CIM fundamentals are given in the above-mentioned document [2], which focuses on needs in energy transmission where related applications include energy management system, SCADA usage, planning and optimizations. IEC 61970-501 [3] and IEC 61970-552 [4] define CIM/RDF and CIM/XML format for power transmission. IEC 61968-13 [5] defines CIM/RDF format for power distribution. IEC 61968 [6] series of standards extend IEC 61970-301 to meet the needs of energy distribution where related applications among others include distribution management system, outage management system, consumption planning and metering, inventory and corporate resources control, and GIS usage.

III. INCREMENTAL INTEGRATION

GIS data export and assumptions, which the presented solution is based upon, will be explained first. After that, the initial state for procedure of incremental integration and detection of difference between two relevant GIS states will be defined and explained. Problems that emerge during update of PNAS will be analyzed and a solution how to solve them, based on splitting detected difference made from added and changed elements into groups of connected elements, will be presented. Only parts of incremental integration, which are independent of PNAS used, will be explained.

Integration requires GIS data to be represented as a set of CIM/XML files. It is assumed that GIS supports CIM/XML as one of its export formats. Each exported file represents one unit called feeder. Feeder represents topologically connected elements fed from the same exit on high voltage substation. Feeder elements can be substations, conducting equipment (inside or outside a substation), and conducting lines.

Procedure of incremental integration relies on next few assumptions about GIS export:

1. One file represents one complete feeder. Name of the feeder is the same as the file name.
2. Each element can be inside one file only. The same element in more files indicates incomplete or invalid connectivity.
3. Feeder data model adjusts to PNAS data model according to substation description. Deviation from given substation prototype is treated as export error.

In order to simplify PNAS update, a prototypical structure of exported substation is defined (Fig. 1). Substation is modelled as container for simple CIM Bay elements. Each bay can contain more pieces of simple conducting (e.g. breakers or fuses) or measurement equipment. There are four bay categories. First bay category connects conducting line on input to primary busbar, and forth bay category connects consumer to secondary busbar. Second bay category connects transformer's primary side to primary busbar and third bay category connects transformer's secondary side to secondary busbar. A number in the top left corner (of red squares) in the picture indicates bay's category.

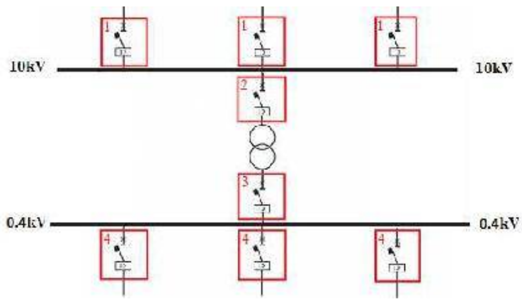


Figure 1. Structure of prototypical substation with indicated bay categories

Before the synchronization between GIS and PNAS can start, we need to align those two systems first. It can be done manually or by some bulk export of data from GIS to PNAS. It is important that after the initial alignment of systems all relevant feeders' files be exported from GIS. This is definition of initial state, which is needed for start of incremental integration. After each successful synchronization, which represents an increment in integration, states of those systems will be aligned again.

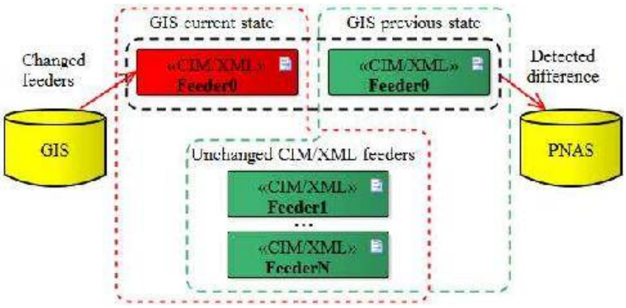


Figure 2. Role of feeders in incremental integration

Fig. 2 illustrates the role that feeders play in detection of differences between two relevant GIS states at the moment of synchronization with PNAS. Changes inside GIS affect only certain number of feeders. For incremental integration, it is enough to export only those feeders that have been changed. Current state of the changed GIS is described by exported all feeders (changed and unchanged from previous synchronization). All unchanged feeders from previous synchronization describe previous GIS state, which corresponds to the

current state of PNAS. After successful integration, all GIS feeders, with changed feeders included, are kept for the next synchronization where they will play the role of previous GIS state.

XML parser uses a defined XML schema to validate all feeders that belong to one relevant GIS state and, after that creates set of CIM/XML objects, which correspond to simple elements of conducting and measurement equipment, bays and substations. The set can be represented as an array of subsets, each containing objects of the same type because comparison has sense only with objects of the same type. Object can be accessed through its unique identifier, provided by GIS, which is the central source of all data. After all feeders are parsed, created objects are linked together according to the relations that are kept inside CIM/XML attributes. Containment relations (substation-bay, bay-conducting and measurement equipment) and CIM topology, which then can be used for finding of conducting and measurement equipment neighbors, are created this way. The same procedure is repeated for both compared states. For each object, the procedure sets an attribute for identification of the state.

Comparison of two relevant GIS states is performed by finding symmetric difference [7] of the two sets:

$$\Delta(m1, m2) = (m1 \backslash m2) \cup (m2 \backslash m1)$$

Total difference, is a union of symmetric differences of all subsets, each of which contain objects of the same type. It is calculated with generic function, which, as its arguments, accepts two subsets of objects of the same type from two relevant GIS states and produces their symmetric difference. After applying generic function on all existing subsets, resulting detected difference can be represented with three sets: added, changed and removed elements. By traversing a subset from the current state and comparing it with corresponding subset of the same type from previous state, generic function detects the sets of:

1. added elements - where unique identifier exists in the current but not in previous state
2. changed elements - where unique identifier exists in both states but elements are not the same.

By traversing a subset from previous state and comparing it with corresponding subset of the same type from the current state, generic function detects the set of:

3. deleted elements - where unique identifier exists in previous but not in the current state.

Detected difference represents a list of object references from the previous (for deleted elements) or the current state (for added and changed elements). Detection of difference in case of addition, change or deletion of an element inside substation automatically adds the substation reference in the set of changed elements. Only current state objects, which represent changed elements, have internal reference to a corresponding object from the previous state. During analysis of a detected difference, referred object can provide the information about the state it belongs to. If it belongs to the previous state, it is a deleted element. If it belongs to the current state, it can be determined whether it is a new or changed element based on existence of corresponding object internal reference. Elements in the detected difference, which are references to objects inside substation, are removed from the difference because the difference already contains the substation reference. The synchronization of changed substation is performed by its deletion and recreation. This will also synchronize changed and added elements inside the substation.

By updating PNAS with the detected difference, its state will become consistent with current GIS state. Update could be done by sending deleted, added and changed elements one by one. Taking into account that accessing PNAS initiates a CORBA request, this solution would be inefficient. On the other side, such a solution would provide the simplest diagnostics when something goes wrong because it is a strait forward action to find problematic element. Another possibility is to send the whole detected difference inside one request, which would be efficient, but if something goes wrong the diagnostics would be much more complex. Deleting non-existing element inside PNAS is not a problem and can be ignored. Updating PNAS with the detected difference that includes added and changed elements can fail in at least the following two cases:

1. emerging of a new equipment structure inside one of CIM/XML bays, which does not exist in PNAS
2. manual change inside PNAS which is not synchronized with GIS

These types of problems are too rare and their automation cannot justify increase in complexity of solution and potential introduction of dependency on the used PNAS. An error during the update needs to be found and fixed manually. As the number of elements in update request grows larger, finding the error gets harder. In order to make this process more efficient and user friendly, the solution applies “divide and conquer” strategy, and splits the detected difference of added and changed elements into isolated groups of connected elements. This reduces the scope of error search to just one group instead of the whole detected difference. Update of PNAS, after removal of all deleted elements, is done individually with each group of added and changed elements instead of the whole detected difference. Errors are fixed by reorganization of data inside GIS or PNAS. After data reorganization, the whole synchronization procedure is repeated.

Forming of groups with connected added and changed elements (only “groups“ in the rest of the text) represents a procedure of exhaustion of two resulting sets produced at the end of difference detection phase: the set of added and the set of changed elements (“detected difference“ in the rest of the text). Each iteration extracts one group, isolated from the rest of detected difference. Isolation assumes that the group elements do not have neighbors or that all their neighbors are unchanged. This procedure changes only two resulting sets, which at the end become completely empty. Connectivity of elements is defined by topological relations inside CIM/XML and, after their creation, does not change. Each element can provide set of all its neighbors. Each group starts by random selection of a starting element, which recursively pulls other elements from the detected difference. A group can be

trivial with just one element but it can also include all the elements from detected difference. Elements of extracted group are removed from detected difference and the procedure is recursively repeated until detected difference is not fully exhausted.

Fig. 3 illustrates topology of one detected difference with two isolated groups. Line segments represent conducting lines, squares substations and circles conducting equipment outside substation, which terminates conducting lines. A selected starting element is specified, and numbers represent order of each of its neighbors.

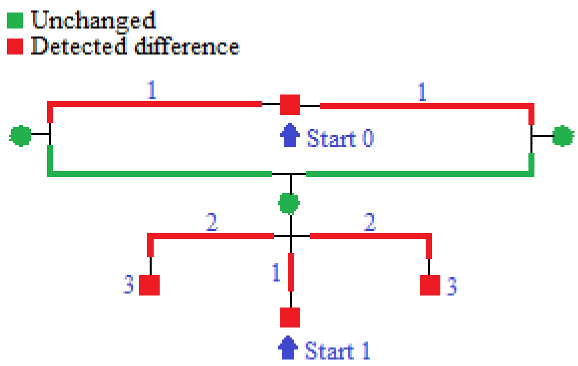


Figure 3. Topology of connected groups inside detected difference

Algorithm for creation of groups with connected elements is given by the following description:

1. Select an arbitrary element from detected difference. It is a new starting element. After selection, remove it from the detected difference.
2. Based on topological connectivity, the search starts for all elements that are starting element's neighbors and they belong to the detected difference. In other words, red neighbors are searched for. They are neighbors of the 1st order. After the search, remove them from the detected difference.
3. Neighbors of the 1st order are treated as starting elements and step 2 is recursively repeated for them. This gives the neighbors of the 2nd order. Eventually, recursion will stop when neighbors of order `k` become an empty set (they either do not exist nor they are added nor they are changed). All neighbors in range `i=1..k-1` represent one group of connected elements extracted from the detected difference.
4. The whole procedure starting from step 1 is repeated unless the detected difference is an empty set.

Resulting groups do not depend on the elements that are selected as starting elements. Order of groups does not affect update of PNAS, but on the client's request, the groups are arranged in such a way that first come all the groups with substations, then all the groups with conducting equipment and at the end the groups with only conducting lines. To provide this order of groups, before selection of the first starting element and after removal of all substation internal elements, the remaining elements are to be arranged according to the following order:

1. added substations
2. changed substations
3. added conducting equipment outside substation
4. changed conducting equipment outside substation
5. added conducting lines
6. changed conducting lines

The starting element is now the first element in the remaining detected difference.

The solution has been implemented for Windows platform, as one DLL component, in C++ programming language. The component is integrated in desktop application developed in PyQt programming language. The overall solution for incremental integration has 42k physical lines of code and 380 classes, out of which 194 classes belong to CIM standard realization.

IV. Discussion of Results

The solution does not guarantee a full automatic integration because the update of PNAS can fail and the manual intervention is needed to resolve this problem. The most frequent reason for that are changes inside PNAS, which are not synchronized with GIS. On the other hand, the solution is quite general and does not depend on the used PNAS. The fieldwork has shown that the most important thing for a successful resolving of manual interventions is education of a customer and his focus on correct entry of GIS data, because that is how the level of automatization is really controlled. Insisting on correct entry of GIS data does not additionally complicate the existing procedures. This shows that the procedure of incremental integration simply fits into already existing business processes based on a GIS system, which is the central point of the whole software system.

Presented solution is deployed and continuously used by two clients. After setting GIS up to the level that is required by integration, the usage of the application becomes an integral part of GIS maintenance. Thanks to the extensive validation of exported feeders, the application contributed to more systematic and more uniform GIS data entry. Splitting of detected difference into groups helps to resolve remaining problems faster and simplifies troubleshooting.

Tab. 1 presents weekly success rate of the system synchronizations for two client installations within one month. The integration is performed once a day or more frequently if problems appear. Installation 1 is being used for a longer period than Installation 2 and has less frequent and more stable integrations. Installation 2 has around 3 times more feeders than Installation 1, with changes of larger range that include a significant number of feeders, which ultimately give more integration problems. On the sites of both clients, unsuccessful integrations are resolved by entering the corrections in GIS and then repeating the integration.

- S – number of successful integrations
- U – number of unsuccessful integrations
- F – number of changed feeders

Table 1. Success rate of five weekly integrations for two clients

Week		I			II			III			IV			V		
Installation	Feeders	S	U	F	S	U	F	S	U	F	S	U	F	S	U	F
1	500	1	1	27	3	0	55	1	0	40	0	0	0	2	1	56
2	1400	3	1	76	8	3	113	7	2	63	5	1	76	6	2	422

V. Conclusion

Efficient integration of GIS and PNAS is a practical problem inside the PS domain. This paper explains the need for a solution of this problem and presents one possible way of solving the problem. Suggested solution does not achieve a full automatic integration, but has proved to be practically useful because it fits into existing business processes and increases the level of their automatization and consequently improves their efficiency and reliability. The solution implements functional requirements by using widely accepted mechanisms for integration development: incremental integration supported by standard defined exchange of data. CIM standard proved itself as a very good specification for internal data model realization, which showed to be the central point for success of the whole application.

A problem of fully automatic integration remains as a challenge for further research and development. A promising solution would be to do an export of only CIM/XML changes of GIS states and avoid the need for full state comparison. In order to make something like this possible, it is necessary to get adequate support from GIS, which needs to know how to generate difference of its two relevant states. Except CIM/XML export format, the solution should also offer other export formats supported by different GIS systems.

References

[1] Trussell L.V., "GIS Based Distribution Simulation and Analysis," *16th International Conference and Exhibition on Electricity Distribution,* 2001.

[2] "IEC 61970-301, Energy management system application program interface (EMS-API) - Part 301: Common information model (CIM) base," [Online]. Available: https://webstore.iec.ch/publication/6210.

[3] "IEC 61970-501, Energy management system application program interface (EMS-API) - Part 501: Common Information Model Resource Description Framework (CIM RDF) schema," [Online]. Available: https://webstore.iec.ch/publication/6215.

[4] "IEC 61970-552, Energy management system application program interface (EMS-API) - Part 552: CIMXML Model exchange format," [Online]. Available: https://webstore.iec.ch/publication/6216.

[5] "IEC 61968-13, Application integration at electric utilities - System interfaces for distribution management - Part 13: CIM RDF Model exchange format for distribution," [Online]. Available: https://webstore.iec.ch/publication/6200.

[6] "IEC 61968: Common Information Model (CIM) / Distribution Management," [Online]. Available: http://www.iec.ch/smartgrid/standards/.

[7] Förtsch S., Westfechtel B., "Differencing and Merging of Software Diagrams - State of the Art and Challenges," *Proceedings of the Second International Conference on Software and Data Technologies,* pp. 90-99, 2007.

Statistical Metadata Management in European e-Government Systems

Valentina Janev, Vuk Mijović, and Sanja Vraneš
“Mihajlo Pupin” Institute, University of Belgrade, Volgina 15, 11060 Belgrade, Serbia
Valentina.Janev, Vuk.Mijovic, Sanja.Vranes@pupin.rs

***Abstract*—The goal of this paper is, from one side, informative i.e. to introduce the existing activities in the European Commission related to metadata management in e-government systems, and from the other, to describe our experience with metadata management for processing statistical data in RDF format gained through development of Linked Data tools is recent EU projects LOD2, GeoKnow and SHARE-PSI. The statistical domain has been selected in this analysis due to its relevance according to the EC Guidelines (2014/C 240/01) for implementing the revised European Directive on the Public Sector Information (2013/37/EU). The lessons learned shortly described in this paper have been included in the SHARE-PSI collection of Best practices for sharing public sector information.**

Keywords: *Linked Data, metadata, RDF Data Cube Vocabulary, PSI Directive, Best Practices*

I. Introduction

The Directive on the re-use of Public Sector Information (known as the 'PSI Directive'), which revised the Directive 2003/98/EC and entered into force on 17th of July 2013, provides a common legal framework for a European market for government-held data (public sector information) [1,2].

The PSI Directive is a legislative document and does not specify technical aspects of its implementation. Article 5, point 1 of the PSI Directive says “*Public sector bodies shall make their documents available in any pre-existing format or language, and, where possible and appropriate, in open and machine-readable format together with their metadata. Both the format and the metadata should, in so far as possible, comply with formal open standards*.”

Analyzing the metadata management requirements and existing solutions in EU Institutions and Member States, the authors [3] have found that ‘activities around metadata governance and management appear to be in an early phase’.

A. Related Work

The most common definition of metadata is “data about data.” Metadata management can be defined as “as a set of high-level processes for structuring the different phases of the lifecycle of structural metadata including design and development of syntax and semantics, updating the structural metadata, harmonisation with other metadata sets and documentation” [1] Commercial software providers e.g. IBM, make difference between business, technical, and operational metadata. Hence, metadata management includes tools, processes, and environment that enable organization to answer different questions related to resources they own. Samsung Electronics [4], for instance, is looking at three types of issues related to metadata management: (1) metadata definition and management, (2) metadata design tools, and (3) metadata standards.

B. Paper structure

The goal of this paper is, from one side, informative i.e. to introduce the existing activities in the European Commission related to metadata management based on our knowledge obtained through participation in recent EU projects LOD2, GeoKnow and SHARE-PSI. From the other, it describes our experience with metadata management for processing statistical data in RDF format gained through development of Linked Data tools.

Section 2 introduces the European holistic approach to interoperability in eGovernment Services. Next, Section 3 shows the latest trends to semantic interoperability in public administration across Europe based on the Linked Data approach. Using an example from Serbia, Sections 4 further analyses the challenges of metadata management of statistical data and points to tools developed in the Mihajlo Pupin Institute. The Pupin team contributed also to sharing the existing experiences with European partners as is described in Section 5. Section 6 concludes the paper.

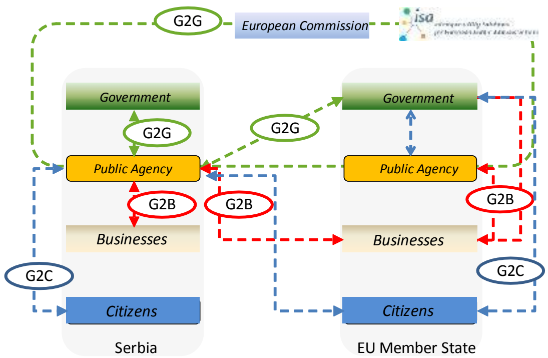


Figure 1. E-Government services across Europe

II. Problem Statement

A. Holistic Approach to PSI Re-use and Interoperability in European eGovernment Services

Since 1995, the European Commission has conducted several interoperability solutions programmes, where the last one "Interoperability Solutions for European Public Administrations" (ISA) will be active during the next five

years (2016-2020) under the name ISA[2]. The holistic approach (G2G, G2C, G2B, see Figure 1) foresees four levels of interoperability, namely legal, organizational, semantic and technical interoperability. In our research, we take special interest in methods and tools for semantic data interoperability that support the implementation of the PSI Directive in the best possible way and "*make documents available through open and machine-readable formats together with their metadata, at the best level of precision and granularity, in a format that ensures interoperability, re-use and accessibility*". Up to now, the ISA programme has provided a wide range of supporting tools, see Repositories of reusable software, standards and specifications.

B. Other Recommendations: What Truly Open Data Should Look Like?

In order to share datasets between users and platforms, the datasets need to be accessible (regulated by license), discoverable (described with metadata) and retrievable (modelled and stored in a recognizable format). According to the World Bank Group definition "*Data is open if it is technically open (available in a machine-readable standard format, which means it can be retrieved and meaningfully processed by a computer application) and legally open (explicitly licensed in a way that permits commercial and non-commercial use and re-use without restrictions)*" ("Open Data Essentials"). According to the PSI Directive, open data can be charged at marginal costs i.e. it does not have to be free of charge. Acceptable file formats for publishing data are CSV, XML, JSON, plain text, HTML and others. Recommended by the W3C consortium, international Web standards community, is the RDF format that provides a convenient way to directly interconnect existing open data based on the use of URIs as identifiers.

C. What do we need for efficient data sharing and re-use?

The data can be exposed for download and/or exploration in different ways. Although there are "Best Practices for Publishing Linked Data" (2014), the metadata of published datasets can be of low quality leading to the questions such as:

- Is the open data ready for exploration? Is the metadata complete? What about the granularity? Do we have enough information about the domain/region the data is describing?
- Is it possible to fuse heterogeneous data and formats used by different publishers and what are the necessary steps? Are there standard approaches / services for querying government portals?
- What is the quality of the data / metadata, i.e., do we have a complete description of the public datasets? Does the publisher track changes on data and schema level? Is the publisher reliable and trustful?

In order to that make the use of open data more efficient and less time-consuming, standardized approaches and tools are needed e.g. the Linked Data tools that work on top of commonly accepted models for describing the underlying semantics.

III. Linked Data Approach for Open Data

A. Linked Data Principles

The Linked Data principles have been defined back in 2006 [5], while nowdays the term Linked Data[2] is used to refer to a set of best practices for publishing and connecting structured data on the Web. These principles are underpinned by the graph-based data model for publishing structured data on the Web – the Resource Description Framework (RDF) [6], and consist of the following: (1) using URIs as names for things, (2) making the URIs resolvable (HTTP URIs) so that others can look up those names, (3) when someone looks up a URI, providing useful information using the standards (RDF, SPARQL), and (4) including links to other URIs, so that they can discover other things on the Web.

These best practices have been adopted by an increasing number of data providers over the past five years, leading to the creation of a global data space that contains thousands of datasets and billions of assertions - the Linked Open Data cloud[3]. The government data respresents a big portion of this cloud.

Some governments around the world[4] [7,8] have adopted the approach and publish their data as Linked Data using the standards and recommendations issued by the Government Linked Data[5] (GLD, Working Group,one of the main providers of Semantic Web standards.

B. Metadata on the Web

Metadata, or structured data about data, improves discovery of, and access to such information[6]. The effective use of metadata among applications, however, requires common conventions about semantics, syntax, and structure. The Resource Description Framework (RDF) is indeed the infrastructure that enables the encoding, exchange, and reuse of structured metadata.

In the last twenty years standardization organizations such as the World Wide Web Consortium (W3C) are working on defining conventions e.g. for describing government data[7] (see RDF Data Cube Vocabulary[8], Data Catalog Vocabulary[9] and the Organization Vocabulary[10]).

C. Re-using ISA Vocabularies for Providing Metadata

The ISA programme supports the development of tools, services and frameworks in the area of e-Government through more than 40 actions[11]. In the area of metadata management, the programme recommends using Core Vocabularies (Core Person, Registered organisation, Core

---

[2] http://linkeddata.org/
[3] http://lod-cloud.net
[4] http://linkedopendata.jp/
[5] http://www.w3.org/2011/gld/)
[6]

[7] https://www.w3.org/blog/news/archives/3591
[8] https://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/
[9] https://www.w3.org/TR/2014/REC-vocab-dcat-20140116/
[10] https://www.w3.org/TR/2014/REC-vocab-org-20140116/
[11] http://ec.europa.eu/isa/ready-to-use-solutions/index_en.htm

Location, Core Public service)[12] as '*simplified, re-usable and extensible data models that capture the fundamental characteristics of an entity in a context-neutral fashion'*. These vocabularies should support the description of the base registries that are maintained by EU public administrations (i.e. *a base registry is a trusted, authentic source of information under the control of an appointed public administration.*) Moreover, they should support harmonization of base registries across Europe, as well as additional registries, e.g. see *The notion of Linked Land Administration* [9][13].

IV. EXAMPLE: METADATA MANAGEMENT IN STATISTICAL DATA PROCESSING

A. *SDMX and RDF Data Cube standards*

In January 2014, W3C recommended the *RDF Data Cube* vocabulary[14], as a standard vocabulary for modeling statistical data. The vocabulary focuses purely on the publication of multi-dimensional data on the Web. The model builds upon the core of the *SDMX 2.0 Information Model*[15] realized in 2001 by the Statistical Data and Metadata Exchange (SDMX[16]) Initiative with the aim to achieve greater efficiencies in statistical practice.

SDMX Information model differentiates between

- "structural metadata" - those metadata acting as identifiers and descriptors of the data, such as names of variables or dimensions of statistical cubes.
- "Reference metadata" - metadata that describe the contents and the quality of the statistical data (concepts used, metadata, describing methods used for the generation of the data, and metadata, describing the different quality dimensions of the resulting statistics, e.g. timeliness, accuracy).

B. *Structural Metadata (DSDs)*

Each data set has a set of *structural metadata*. These descriptions are referred to in SDMX as Data Structure Definitions (DSD), which include information about how concepts are associated with the measures, dimensions, and attributes of a data "cube," along with information about the representation of data and related identifying and descriptive (structural) metadata. DSD also specifies which code lists (conceptual schemas, see Figure bellow) provide possible values for the dimensions, as well as the possible values for the attributes, either as code lists or as free text fields. A data structure definition can be used to describe time series data, cross-sectional and multidimensional table data.

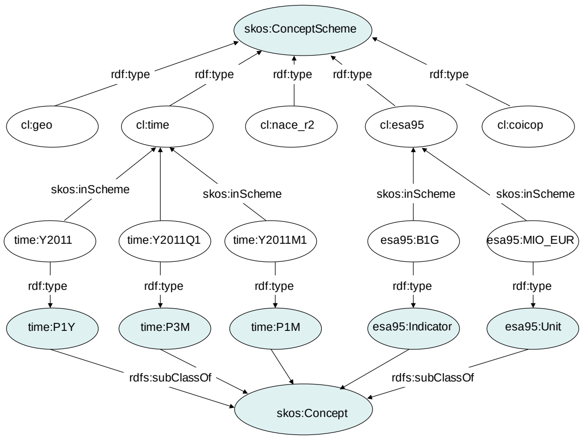


Figure 3. Example of concepts and instances in code lists

Once defined on a national level, and then registered in the EU JOINUP platform, code lists can be used for publishing statistical data coming from different public sector agencies. For more information on how to define a DSD for a statistical dataset, we refer to results from the LOD2 project see D9.5.1.

C. *Quality Assessment of Structural Metadata of RDF Data Cubes*

According to W3C recommendations, a statistical dataset in RDF should be modeled with the *RDF Data Cube* vocabulary and should adhere to the integrity constraints defined in the standard. To that aim, we developed a specialized tool (RDF Data Cube Validation Tool, 2014) to be used prior to publishing the statistical data in RDF format. The RDF *Data Cube Validation component* checks if the statistical dataset (RDF graph) is valid according to the integrity constraints defined in the RDF Data Cube specification (http://www.w3.org/TR/vocab-data-cube). Each constraint in the W3C document is expressed as narrative prose, and where possible, a SPARQL ASK query or query template that returns true if the graph contains one or more Data Cube instances which violate the corresponding constraint. Our tool runs slightly modified versions of these queries which allow it to not only show if the constraint is violated or not, but also to list the offending resources, provide information about the underlying issue and if possible offer a quick solution in order to repair the structural metadata.

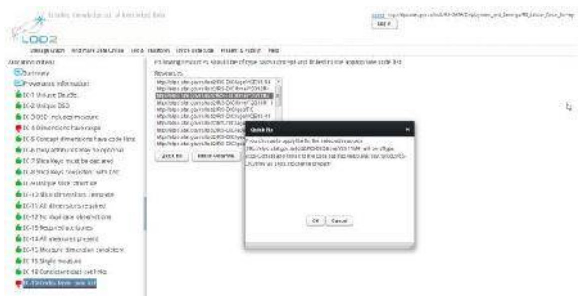


Figure 4. RDF Data Validation Tool - GUI

D. *Improving Quality by Storing and Re-using Metadata*

In order to support reuse of code lists and DSD

---

12 https://joinup.ec.europa.eu/site/core_vocabularies/Core_Vocabularies_user_handbook/ISA%20Hanbook%20for%20using%20Core%20Vocabularies.pdf

13 http://www.yildiz.edu.tr/~volkan/INSPIRE_2014.pdf

14 http://www.w3.org/TR/vocab-data-cube/.

15 *SDMX Content-oriented Guidelines: Cross-domain code lists.*(2009). Retrieved from http://sdmx.org/wp-content/uploads/2009/01/02_sdmx_cog_annex_2_cl_2009.pdf

16 http://www.sdmx.org/

descriptions defined by one publisher (e.g. Statistical Office of the Republic of Serbia), we have designed a new service that stores the DSDs that have been used in published open data and offers a possibility to

- build and maintain a repository of unique DSDs,
- create a new DSD based on the underlying statistical dataset,
- refer to, and reuse a suitable DSD from the repository.

Thus the tool has potential to uniform the creation and re-use of structural metadata (DSDs) across public agencies, reduce duplicates and mismatches and improve harmonization on national level (we assume here that public agencies in one country will use the same code lists).

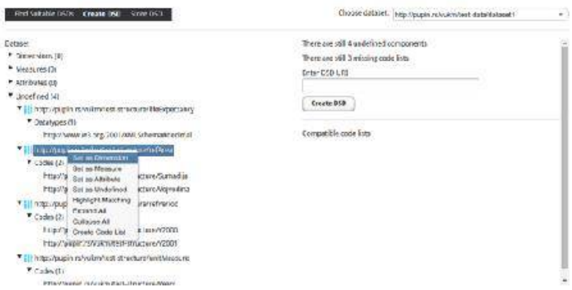


Figure 5. DSD Repository - GUI

V. BEST PRACTICES FOR OPEN DATA

A. *About the SHARE-PSI project*

Financed by the EU Competitiveness and Innovation Framework Programme 2007-2013, in the last two years, the SHARE-PSI network is involved in analysis of the implementation of the PSI Directive across Europe. The network is composed of experts coming from different types of organizations (government, academia, SME/Enterprise, citizen groups, standardization bodies) from many EU countries. Through a series of public workshops, the experts were involved in discussing EU policies and writing best practices for implementation of the PSI Directive. In the project framework, a collection of Best Practices[17] was elaborated that should serve the EU member states to support the activities around PSI Directive implementation.

Curious about the adoption of the Linked Data concepts in European e-government systems, we carried out an anlaysis which produced findings that are presented bellow.

B. *Findings about EU OGD Initiatives (related to metadata management)*

1) Semantic Asset Repositories

According to our research, there are differences in the effort of establishing semantic metadata repositories on country level (see e.g. Germany[18], Denmark[19], and Estonia[20]), as well as in the amount of resources that are published and shared with other member states via the JoinUp platform (see federated repositories[21]).

Currently in the EU, still ongoing is the adoption of the ISA Core Vocabularies and Core Public Service Vocabulary on national level (see e.g. implementation in Flanders[22] or Italy [10]) and the exchange of data inside a country (see Estonian metadata reference architecture)[23].

2) Federation of Data Catalogues

The DCAT-AP is a specification based on the Data Catalogue vocabulary (DCAT) for describing public sector datasets in Europe. Its basic use is to enable cross-data portal search for data sets and make public sector data better searchable across borders and sectors. This can be achieved by the exchange of descriptions of datasets among data portals. Nowadays, an increasing number of EU Member States and EEA countries are providing exports to the DCAT-AP or have adopted it as the national interoperability solution. The European Data Portal[24] implements the DCAT-AP and thus provides a single point of access to datasets described in national catalogs (376,383 datasets retrieved on January 5th 2016).

The hottest issue regarding federation of public data in in EU is the quality of metadata associated with the national catalogues [11].

C. *SHARE-PSI Best Practices*

Having strong technical background in semantic technologies, besides the activities in the SHARE-PSI network, our team was involved in other European projects that delivered open source tools for Linked Open Data publishing, processing and exploitation. In LOD2 project we tested the LOD2 tools with Serbian government data and the knowledge gain was communicated at the SHARE-PSI workshops [12].

Additionally, we contributed to publishing more than 100 datasets from the Statistical Office of the Republic of Serbia to the Serbian CKAN. Based on that experience we contributed to formulation of the following Best Practices:

- Enable quality assessment of open data[25] (PUPIN contributed with the experience with the RDF Data Cube Validation);
- Enable feedback channels for improving the quality of existing government data[26];
- Publishing Statistical Data In Linked Data Format[27] (PUPIN contributed with the experience with Statistical Workbench [12]).

While the first two Best Practices are well recognized and already in practice across EU countries, the third one still has a status *draft*, meaning that consensus across EU is needed.

---

17 https://www.w3.org/2013/share-psi/bp/
[18] XRepository, https://www.xrepository.deutschland-online.de/
[19] Digitalisér.dk, http://digitaliser.dk/
[20] RIHA, https://riha.eesti.ee/riha/main

[21] https://joinup.ec.europa.eu/catalogue/repository
[22] https://www.openray.org/catalogue/asset_release/oslo-open-standards-local-administrations-flanders-version-10
[23] http://www.w3.org/2013/share-psi/workshop/berlin/EEmetadataPilot
[24] http://www.europeandataportal.eu/
[25] https://www.w3.org/2013/share-psi/bp/eqa/
[26] https://www.w3.org/2013/share-psi/bp/ef/
[27] https://www.w3.org/2013/share-psi/wiki/Best_Practices/Publishing_Statistical_Data_In_Linked_Data_Format

VI. Conclusion

According to the Serbian e-Government Strategy, Serbia foresees to implement the PSI Directive in the next period 2016-2020. The Directive envisions publishing of the public/private datasets in machine readable format, thus, making sharing, using and linking of information easy and efficient.

This paper introduced the latest open data and interoperability initiatives in the EU, including ISA recommendations, and described how Linked Data technologies can be used to publish open data on the Web in a machine readable format that makes it easily accessible, and discoverable. In this process, metadata plays an important role as it provides a way to describe the actual contents of the dataset which can then be published on well-known portals and catalogues, thus allowing data consumers to easily discover datasets that satisfy their specific criteria. Described principles were demonstrated on statistical data, however the approach (enhancing the data with metadata, quality assessment, reuse of metadata on national level) is generic and, using domain-specific vocabularies, applicable to other areas as well. In the future, a significant effort will be put into further adaptation of the EC recommendations for building interoperable tools and services, while taking into consideration different aspects, such as scalability, flexibility and ease-of-use/friendliness.

Acknowledgment

The research presented in this paper is partly financed by the European Union (CIP SHARE-PSI 2.0 project, Pr. No: 621012; FP7 GeoKnow, Pr. No: 318159), and partly by the Ministry of Science and Technological Development of the Republic of Serbia (SOFIA project, Pr. No: TR-32010).

References

[1] European legislation on reuse of public sector information. (2013, June 23). Official Journal of the European Union L 175/1. Retrieved from European Commission, http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32013L0037&from=EN

[2] Guidelines on recommended standard licences, datasets and charging for the reuse of documents (2014/C 240/01). Official Journal of the European Union C240/1-10 24.7.2014.

[3] Makx Dekkers, Stijn Goedertier, Audrius Leipus, Nikolaos Loutas, Metadata management requirements and existing solutions in EU Institutions and Member States, SC17DI06692, http://ec.europa.eu/isa/documents/metadata-management-requirements-and-existing-solutions-in-eu-institutions-and-member-states_en.pdf

[4] Won Kim, On Metadata Management Technology:Status and Issues, JOURNAL OF OBJECT TECHNOLOGY, Vol. 4, No.2, March-April

[5] Berners-Lee, Tim. (2006). Design Issues: Linked Data. Retrieved from http://www.w3.org/DesignIssues/LinkedData.html

[6] F. Manola, E. Miller, B. McBride,"RDF Primer", 2004, February 10, Retrieved from http://www.w3.org/TR/rdf-primer/ 2005

[7] Hendler, J., Holm, J., Musialek, C. ; Thomas, G. (2012) US Government Linked Open Data: Semantic.data.gov, Intelligent Systems, IEEE (Volume:27 , Issue: 3), IEEE Computer Society.

[8] Wood, David (Ed.), Linking Government Data, Springer-Verlag New York, 2011

[9] Volkan Çağdaş, Erik Stubkjær, Supplementing INSPIRE through e-Government Core Vocabularies, http://inspire.ec.europa.eu/events/conferences/inspire_2014/pdfs/20.06_4_09.00_Volkan_%C3%87a%C4%9Fda%C5%9F.pdf

[10] Ciasullo, G., Lodi, G., Rotundo, A. (2015) Core Public Service Vocabulary: The Italian Application Profile. SHARE-PSI Workshop. Retrieved from http://www.w3.org/2013/share-psi/wiki/images/7/73/AgID_BerlinWorkshop.pdf

[11] Carera, W. (2015) The Role of the European Data Portal, http://www.w3.org/2013/share-psi/wiki/images/4/46/Share_PSI_2_0_EDP_paper_v1_1.pdf

[12] V.Janev, Publishing and Consuming Linked Open Data with the LOD Statistical Workbench, SHARE-PSI Workshop, Samos, Greece, 2014 https://www.w3.org/2013/share-psi/workshop/samos/agenda

Transformation and Analysis of Spatio-Temporal Statistical Linked Open Data with ESTA-LD

Vuk Mijović[*], Valentina Janev[**], Dejan Paunović[**]
[*] School of Electrical Engineering, University of Belgrade, Institute Mihailo Pupin, Belgrade, Serbia
[**] University of Belgrade, Institute Mihailo Pupin, Belgrade, Serbia
{Vuk.Mijovic, Valentina.Janev, Dejan.Paunovic}@pupin.rs

***Abstract*—Recent open data initiatives have contributed to opening of non-sensitive governmental data, thereby accelerating the growth of the LOD cloud and the open data space in general. A lot of this data is statistical in nature and refers to different geographical regions (or countries), as well as various performance indicators and their evolution through time. Publishing such information as Linked Data provides many opportunities in terms of data aggregation/integration and creation of information mashups. However, due to Linked Data being relatively new field, currently there is a lack of tools that enable exploration and analysis of linked geospatial statistical datasets. This paper presents ESTA-LD (Exploratory Spatio-Temporal Analysis), a tool that enables exploration and visualization of statistical data in linked data format, with an emphasis on spatial and temporal dimensions, thus allowing to visualize how different regions compare against each other, as well as how they evolved through time. Additionally, the paper discusses best practices for modeling spatial and temporal dimensions so that they conform to the established standards for representing space and time in Linked Data format.**

I. Introduction

As stated by the OECD (The Organization for Economic Co-operation and Development), *"Open Government Data (OGD) is fast becoming a political objective and commitment for many countries"*. In the recent years, various OGD initiatives, such as the Open Government Partnership[1], have pushed governments to open up their data by insisting on opening non-sensitive information, such as core public data on transport, education, infrastructure, health, environment, etc. Moreover, the vision for ICT-driven public sector innovation [1] refers to the use of technologies for the creation and implementation of new and improved processes, products, services and methods of delivery in the public sector. Consequently, the amount of the public sector information, which is mostly statistical in nature and often refers to different geographical regions and points in time, has increased significantly in the recent years, and this trend is very likely to continue.

In parallel, the wider adoption of standards for representing and querying semantic information, such as RDF(s) and SPARQL, along with increased functionalities and improved robustness of modern RDF stores, have established Linked Data and Semantic Web technologies in the areas of data and knowledge management [2]. However, these technologies are still quite novel, and a lot of the tooling and standards are either missing, still in development, or not yet widely accepted. For example, the RDF Data Cube vocabulary [3] which enables modeling of statistical data as Linked Data is a W3C recommendation since January 2014, and the GeoSPARQL [4] standard that supports representing and querying geospatial data on the Semantic Web was published in June 2012, while the Spatial Data on the Web Working Group is still working on clarifying and formalizing the relevant standards landscape with respect to integrating spatial information with other data on the Web, discovery of different facts related to places, and identifying and assessing existing methods and tools in order to create a set of best practices. As a consequence, the tools that are based on these standards are scarce, and representation of spatio-temporal concepts may vary across different datasets.

This paper describes ESTA-LD (Exploratory Spatio-Temporal Analysis of Linked Data), a tool that enables exploration and analysis of spatio-temporal statistical linked open data. The RDF Data Cube vocabulary, which is the basis of this tool, is discussed in Section 2, along with the best practices for representing spatial and temporal information as Linked Data and transformation services that help to transform different kinds of spatial and temporal dimensions into a form that is compliant with ESTA-LD. The tool itself, and its functionalities are given in Section 3, while the conclusions and outlook on future work are given in Section 4.

The work described in this paper builds upon and extends previous efforts elaborated in [5, 6].

II. Creating Spatio-Temporal Statistical Datasets and Transforming them with ESTA-LD

This section will discuss modeling of spatio-temporal statistical datasets as Linked Data. First, standard well-known vocabularies will be introduced, followed by recommendations based on these standards. Finally, the section will describe the approach taken in ESTA-LD and introduce its services that help to transform different spatial and temporal dimensions into an expected form.

A. Modeling statistical data as Linked Data

The best way to represent statistical data as Linked Data is to use the RDF Data Cube vocabulary, a well-known vocabulary recommended by the W3C for modelling statistical data. This vocabulary is based on the *SDMX 2.0 Information Model* [7] which is the result of the Statistical

[1] http://www.opengovpartnership.org/

Data and Metadata Exchange (SDMX[2]) Initiative, an international initiative that aims at standardizing and modernizing ("industrializing") the mechanisms and processes for the exchange of statistical data and metadata among international organizations and their member countries. Having in mind that SDMX is an industry standard, and backed up by influential organizations such as Eurostat, World Bank, UN, etc., it is of crucial importance that RDF Data Cube vocabulary is compatible with SDMX. Additionally, the linked data approach brings the following benefits:

- Individual observations, as well as groups of observations become web addressable, thus enabling third parties to link to this data,
- Data can be easily combined and integrated with other datasets, making it integral part of the broader web of linked data,
- Non-proprietary machine readable means of publication with out-of the-box web API for programmatic access,
- Reuse of standardized tools and components.

Each RDF Data Cube consists of two parts: structure definition, and (sets of) observations. The Data Structure Definition (DSD) provides the cube's structure by capturing specification of dimensions, attributes, and measures. Dimensions are used to define what the observation applies to (e.g. country or region, year, etc.) and they serve to identify an observation. Therefore, a statistical data set can be seen as a multi-dimensional space, or hyper-cube, indexed by those dimensions, hence the name cube (although the name cube should not be taken literally as the statistical dataset can contain any number of dimensions, not just exactly three). Attributes and measures on the other hand provide metadata. Measures are used to denote what is being measured (e.g. population or economic activity), while attributes are used to provide additional information about the measured values, such as information on how the observations were measured as well as information that helps to interpret the measured values (e.g. units). The explicit structure captured by the DSD can then be reused across multiple datasets and serve as a basis for, validation, discovery, and visualization.

However, in order to spur reuse and discoverability, the data structure definition should be based on common, well-known concepts. To tackle this issue, the SDMX standard includes a set of content oriented guidelines (COG) which define a set of common statistical concepts and associated code lists that are meant to be reused across datasets. Thanks to the efforts of the community group, these guidelines are also available in linked data format and they can be used as a basis for modeling spatial and temporal dimensions. Although they are not part of the vocabulary and do not form a Data Cube specification, these resources are widely used in existing Data Cube publications and their reuse in newly published datasets is highly recommended.

B. *Representing space and time*

One of the earliest efforts for representing spatial information as linked data is the Basic Geo Vocabulary by W3C. This vocabulary does not address many of the issues covered in the professional GIS world, however it provides a namespace for representing latitude, longitude and other information about spatially-located things, using the WGS84 CRS as the standard reference datum.

Since then, GeoSPARQL has emerged as a promising standard [2]. The goal of this vocabulary is to ensure consistent representation of geospatial semantic data across the Web, thus allowing both vendors and users to achieve uniform access to geospatial RDF data. To this end, GeoSPARQL defines an extension to the SPARQL query language for processing geospatial data, as well as a vocabulary for representing geospatial data in RDF. The vocabulary is concise and among other things, it enables to represent features and geometries, which is of crucial importance for spatial visualizations, such as ESTA-LD. Following is an example of specifying a geometry for a particular entity using GeoSPARQL:

```
PREFIX geo: <http://www.opengis.net/ont/geosparql# >
eg:area1 geo:hasDefaultGeometry eg:geom1 .
eg:geom1 geo:asWKT  "MULTIPOLYGON(…)" ^^geo:wktLiteral
```

Therefore, GeoSPARQL enables to explicitly link a spatial entity to a corresponding serialization that can be encoded as WKT (Well Known Text) or GML (Geography Markup Language). Alternatively, one can refer to geographical regions by referencing well-known data sources, such as GeoNames. This approach is simpler and less verbose, however in this case the dataset doesn't contain the underlying geometry and is therefore not self-sufficient and requires any tool that operates on top of it to acquire the geometries from other sources.

In order to represent time, the two most common approaches are the use of the OWL Time ontology and using the XSD date and time data types. The OWL time ontology presents an ontology of temporal concepts, and at the moment, it is still a W3C draft. It provides a vocabulary for expressing facts about topological relations among instants and intervals, together with information about durations, and about datetime information. On the other hand, the XSD date and time data types can be used to represent more basic temporal concepts, such as points in time, days, months, and years, however they cannot be used to denote custom intervals (e.g. a period from 15th of January till 19th of March). Although they are clearly less expressive than the OWL time ontology, XSD date and time data types are widely used and supported in existing tools and libraries, thus requiring little to no effort for type transformation when third party libraries are used.

C. *ESTA-LD approach*

First and foremost, ESTA-LD is based on the RDF Data Cube vocabulary, and any dataset based on it can be visualized on a chart. However, ESTA-LD offers additional features (visualizations) for datasets containing a spatial and/or temporal dimension. In order for these features to be enabled, the dataset needs to abide to certain principles that were described earlier. Namely, spatial entities need to be linked to their corresponding geometries using GeoSPARQL vocabulary, while the serialization needs to be encoded as a WKT string. On the other hand, with regards to the temporal dimension, ESTA-LD can handle values encoded as XSD date and time data types. Furthermore, even though it is not

[2] http://www.sdmx.org/

required, the use linked data version of the content oriented guidelines for specifying and explicitly indicating spatial and temporal dimensions is highly encouraged. An example observation, along with the definition of the temporal and the spatial dimension is given in Figure 1.

```
@prefix qb: <http://purl.org/linked-data/cube#>
@prefix geo: <http://www.opengis.net/ont/geosparql#>

eg:refArea a rdf:Property, qb:DimensionProperty;
  rdfs:label "reference area"@en;
  rdfs:subPropertyOf sdmx-dimension:refArea ;
  rdfs:range eg:Area, geo:Feature ;
  qb:concept sdmx-concept:refArea .

eg:refPeriod a rdf:Property, qb:DimensionProperty;
  rdfs:label "reference period"@en;
  rdfs:subPropertyOf sdmx-dimension:refPeriod ;
  rdfs:range xsd:gYearMonth ;
  qb:concept sdmx-concept:refPeriod .
sdmx-concept:refPeriod a sdmx:TimeRole.

eg:obs1 a qb:Observation ;Blah,
  qb:dataset eg:ds1 ;
  eg:refArea eg:area1 ;
  eg:refPeriod "2014-12"^^xsd:gYearMonth ;
  sdmx-measure:obsValue "123623" .

eg:area1 geo:hasDefaultGeometry eg:geom1 .
eg:geom1 geo:asWKT "MULTIPOLYGON(…)"^^geo:wktLiteral .
```

Figure 1 Spatio-Temporal Data Cube Example

This example shows the spatial dimension `eg:refArea` that is derived from the dimension `sdmx-dimension:refArea` and associated with the concept `sdmx-concept:refArea`, both of which are available in the linked data version of the content oriented guidelines. Similarly, the temporal dimension is derived from `sdmx-dimension:refPeriod` and associated with the concept `sdmx-concept:refPeriod`. Finally, the observation uses the defined dimensions to refer to the particular time period and the geographical region, which is in turn linked to the geometry and its WKT serialization by using the GeoSPARQL vocabulary.

D. *ESTA-LD Data Cube Transformation Services*

The modelling principles for temporal and spatial dimensions are still in early stages and therefore not so well known and widespread, meaning that many Data Cubes may vary slightly when it comes to modelling these two dimensions. In other words, there is a reasonable chance that there are spatio-temporal datasets that do not clearly express the presence of spatial and temporal dimensions or that values of spatial and temporal dimensions are represented in a custom ("non-standard") way, thus requiring slight modifications in order to enable all of ESTA-LD's functionalities. To address this issue, ESTA-LD is accompanied with an "Inspect and Prepare" component that provides services for transforming spatial and temporal dimensions. This component provides a visual representation of the structure of the chosen data cube. The structure is displayed as a tree that shows dimensions, attributes, and measures, as well as their ranges, code lists (if a code list is used to represent values of that particular dimension/measure/attribute), and values that appear in the dataset. Furthermore, the tree view can be used to select any of the available dimensions and initiate transformation services.

1) *Transforming Temporal Dimensions*

Many temporal dimensions may miss a link to the concept representing a time role. Furthermore, in some cases, organizations may decide to use their own code lists to represent time. However, even if this is the case, the URIs representing time points usually contain all the information needed to derive the actual time. Take for example the code list used by the Serbian statistical office, where code URIs take the following form: http://elpo.stat.gov.rs/RS-DIC/time/Y2009M12. This URI clearly denotes the December of 2009, and it can be parsed in order to transform it to an XSD literal such as "2009-12"^^xsd:gYearMonth. To achieve this with ESTA-LD's *Inspect and Prepare* view (see Figure 2), one only needs to provide the pattern by which to parse the URIs and the target type, after which the component transforms all values, changes the range of the dimension, removes a link to the code list since code list is not used any more, and links the dimension to the concept representing the time role. The target type is selected from the drop-down list, while the pattern is provided in the text

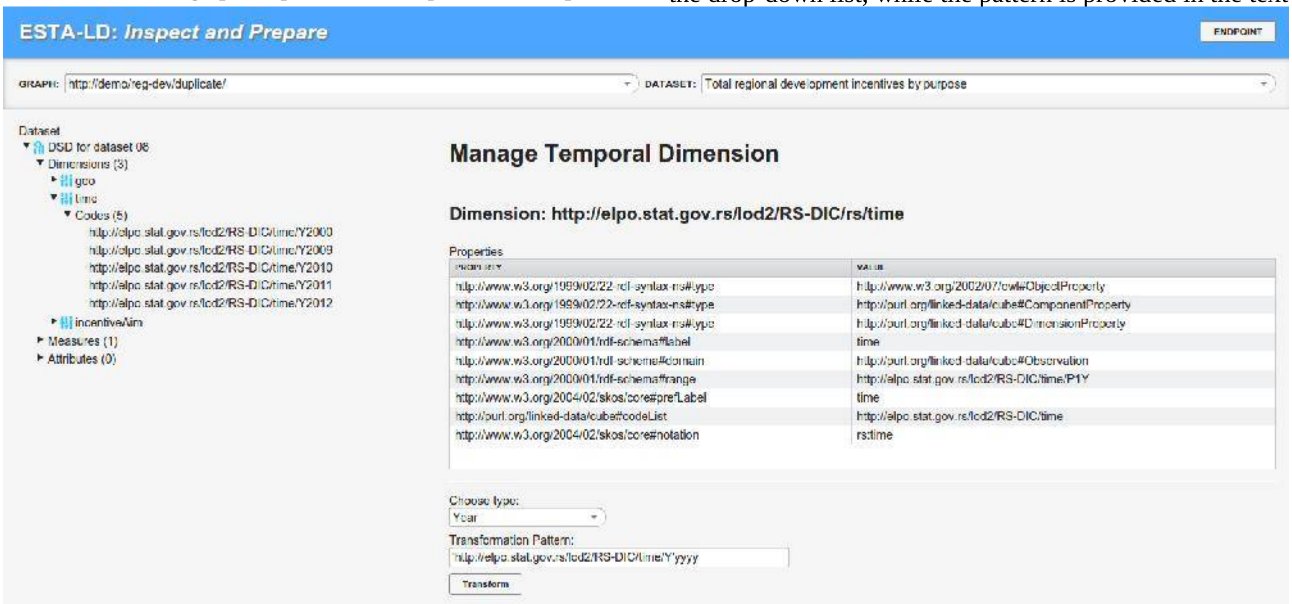


Figure 2 ESTA-LD Inspect and Prepare Component - Transformation of the Temporal Dimension

field.

2) Transforming Spatial Dimensions

In many cases, a Data Cube may contain a spatial dimension but miss the polygons which are required for visualization on the choropleth map. If this is the case, ESTA-LD's *Inspect and Prepare component* can be used to enrich the Cube with polygons acquired from LinkedGeoData. All that is needed on the user's part is to supply a pattern that the tool will use to extract one of the identifiers that can be used for the lookup, as well as to specify identifier's type, which can be one of the following: name, two-letter code (ISO 3166-1 alpha2) and three-letter code (ISO 3166-1 alpha 3). Similarly to the temporal dimension transformation service, the pattern is supplied in a text field, while the identifier type can be selected in a drop-down list. At the moment, this functionality can only be used to acquire polygons of countries.

III. ESTA-LD

ESTA-LD is a tool for visualizing statistical data in linked data format, i.e. data cubes modeled with the RDF Data Cube vocabulary. However, unlike other tools that treat any data cube in the same manner, such as CubeViz [8], ESTA-LD distinguishes spatial and temporal dimensions from the rest, and provides specialized visualizations based on their specific properties. Namely, if a Cube contains observations related to different geographic regions, i.e. it contains a spatial dimension, then the data can be visualized on a map where regions are colored in different grades/shades of blue based on observation values, thus giving intuitive insight into the disparities across regions. On the other hand, if a Cube contains measurements at different points in time, all measurements are organized on the time axis where a user can choose any time interval he or she wants to analyze and/or slide through time, thereby gaining insights into the evolution of the indicator under analysis over time.

A. Architecture and Implementation

ESTA-LD is a web application that can be deployed on any servlet container. Furthermore, it can operate on top of any SPARQL endpoint and accepts query string parameters for specifying the default endpoint and graph, thus ensuring that it can be used as a standalone tool, but at the same time easily integrated into other environments such as the GeoKnow Generator. It is based on the following frameworks and libraries:

- Vaadin: a Java framework for building web applications,
- Sesame: an open-source framework for querying and analysing RDF data,
- Leaflet: an open source JavaScript library for mobile-friendly interactive maps,
- Highcharts: a charting library written in pure HTML5/JavaScript, offering intuitive, interactive charts to a web site or web application,
- wellknown: a Javascript library for parsing and stringifying Well-Known Text into GeoJSON,
- jQuery: a Javascript library that makes things like HTML document traversal and manipulation, event handling, animation, and Ajax much simpler with an easy-to-use API that works across a multitude of browsers

Most of the user interface components, including the drop-down lists that are used to associate dimensions to particular values, are implemented in Vaadin, while the choropleth map and the chart are implemented using LeafletJS and Highcharts respectively, due to lack of adequate UI components in Vaadin. Sesame is used to query the selected SPARQL endpoint for available data cubes, as well as for the selected cube's structure, including contained dimensions, attributes, and measures. This information is used to populate Vaadin drop-down lists which allow the user to specify desired visualization, i.e. which dimension(s) will be visualized, and the values of other dimensions. After the user specifies the desired visualization, selection is passed to the Javascript layer, and jQuery is used to query the endpoint for observations that satisfy the selected criteria. After the endpoint returns the desired observations, the data is transformed and fed to the Leaflet map and Highcharts chart in the expected format. During the transformation process, wellknown is used to parse the WKT strings returned by the endpoint into GeoJSON that is required by Leaflet.

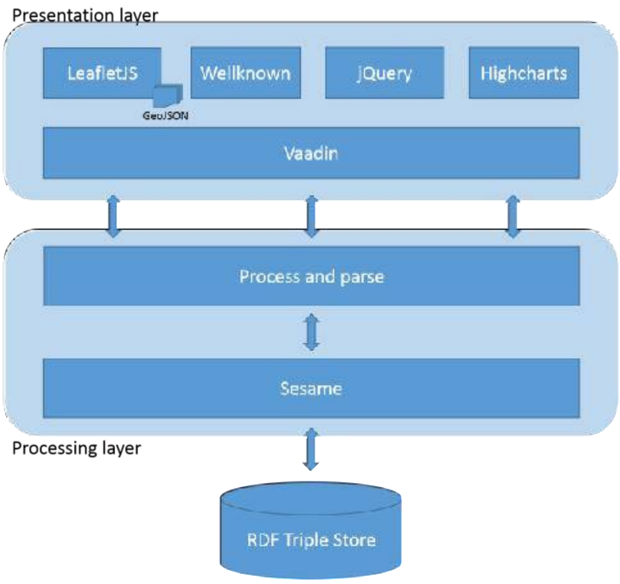


Figure 3 ESTA-LD Architecture

B. Chart Functionalities

The chart allows to analyze how observations vary across the selected dimension. If two dimensions are selected, values of the first dimension are laid on the X axis as categories, while values of the second dimension can be chosen in the legend as series (see Figure 4), thus allowing comparison between the selected dimensions. When two dimensions are visualized, it is also possible to swap series and categories, as well as to stack the values of the selected series. Furthermore, the chart allows to switch the axes, and to change its size by dragging the separator on the left side and showing/hiding the parameters section. Finally, in case the cube contains multiple measures, any two measures can be visualized in parallel in order to enable comparison, as shown in Figure 5. This example shows that measure comparison can be used to find correlations between different measures.

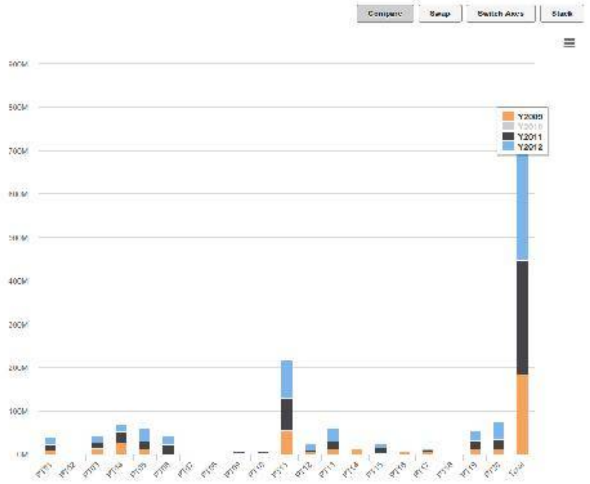


Figure 4 Chart Visualization - Two Dimensions

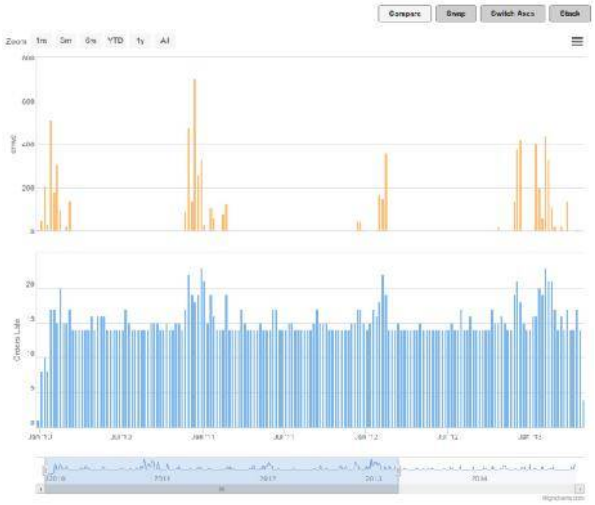


Figure 5 Time Chart - Comparing Two Measures

C. *Spatio-Temporal Visualization/Analysis*

The choropleth map always visualizes the spatial dimension, i.e. it shows the same information as the chart would if the only selected dimension were the spatial dimension. However, while the chart visualization would place a separate bar for each region, the choropleth map paints regions on the map in different shades of blue based on observation values. There are nine shades available, and each one represents a certain value range, where ranges are calculated based on the maximum and minimum observation value. This way, it is much easier to note disparities across different geographical regions (see Figure 6). The map also allows the user to select a particular region on the map, thus changing the region to be visualized in the chart. Finally, the choropleth map supports multiple hierarchy levels. Namely, if the hierarchical structure of geographical regions is given using the SKOS vocabulary, the tool allows to select which hierarchy level is to be visualized.

If the dataset contains a temporal dimension, and it is selected for visualization, the tool uses a specialized time chart that includes a timeline at the bottom. The timeline can be used to specify a particular time window to be visualized which is very useful when the underlying dataset contains huge number of time points, such as for example, daily data for a period of four years. For convenience, the time chart provides a shortcut for setting the duration of the time window to commonly used values such as 1 month, 3 months, 6 months, and 1 year. It is also possible to drag the time window, thus gaining an insight into the evolution of the selected indicator through time (see Figure 6).

Finally, the choropleth map and the time chart are fully synchronized. This means that whenever a region is selected on the map, the time chart is updated to show how the chosen measure evolved through time in that particular region. Similarly, whenever a time window is changed in the time chart, the map is updated to show only the selected period in time. Moreover, this happens immediately as the time window gets changed. As a

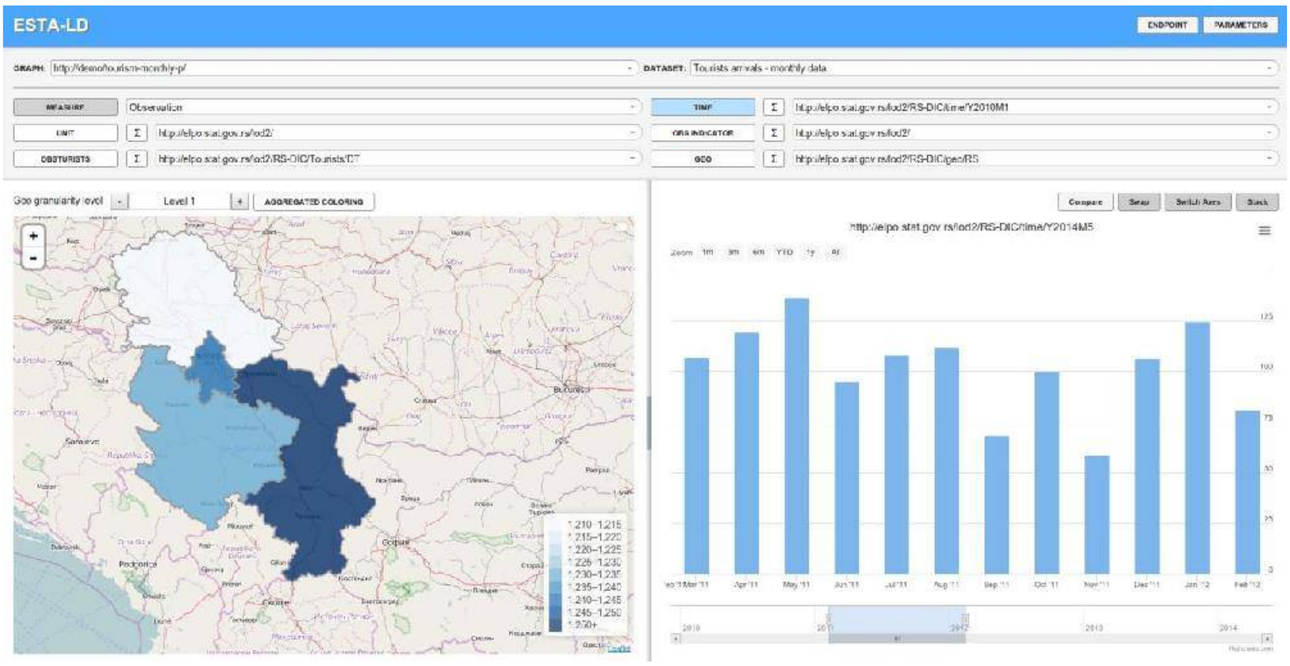


Figure 6 ESTA-LD Spatio-Temporal Visualization

consequence, by dragging the time window it is possible to get an insight into how different regions evolved over time. To further support this functionality, the map supports "aggregated coloring" which ensures that value ranges for different shades of blue do not change unless the duration of the window changes. Namely, without the aggregated coloring, whenever the time window is moved, the underlying set of observations to be visualized on the map changes, and with it the maximum and minimum observation value change as well. Consequently, value ranges for different shades get recalculated whenever the time window is dragged, making it impossible to determine how a particular region fairs against the previously selected period (since now every shade represents a different value range). With aggregated coloring employed, the tool calculates the minimum and maximum values based on every possible time window of the same duration and calculates the value ranges accordingly. This way, dragging of the time window doesn't impact the coloring scheme, making it unnecessary to recalculate the value ranges unless the duration of the time window changes. Therefore, it not only provides insight into disparities across regions through time, but also into the evolution of the chosen measure in each region that is shown on the map.

IV. Conclusions and Future Work

ESTA-LD is a tool that enables exploration and analysis of statistical linked open data. While it can visualize any statistical dataset on a chart, the tool puts an emphasis on spatial and temporal dimensions in order to enable spatio-temporal analysis. Namely, if the dataset contains a spatial and a temporal dimension, it is visualized on the choropleth map and the time chart respectively. Furthermore, these two views are synchronized, meaning that every selection in one of the views updates the other, thus providing insights into disparities of the chosen indicator across different geographical regions as well as their evolution through time. Furthermore, this paper showed how statistical data can be modeled as linked open data and discussed different approaches to representing spatio-temporal information within statistical data cubes, including the approach adopted by ESTA-LD. Having in mind that linked data is a relatively new technology where standards for representing spatial and temporal concepts are still shaping up, the paper described how ESTA-LD's *Inspect and Prepare* component can be used to transform different types of spatial and temporal dimensions to a form that is compliant with ESTA-LD.

In the future, ESTA-LD will be extended with additional types of graphs, and possibly with a data structure definition (DSD) repository that would reduce replication of information since DSDs can be reused and shared across different datasets. Finally, we intend to examine how to best leverage enrichment of statistical datasets with external sources of information such as DBpedia in order to provide advanced search and filtering capabilities over the data cubes.

Acknowledgment

The research presented in this paper is partly financed by the European Union (FP7 GeoKnow, Pr. No: 318159; CIP SHARE-PSI 2.0, Pr. No: 621012), and partly by the Ministry of Science and Technological Development of the Republic of Serbia (SOFIA project, Pr. No: TR-32010).

References

[1] EC Digital Agenda, "Orientation paper: research and innovation at EU level under Horizon 2020 in support of ICT-driven public sector.",http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=2588, May 2013.

[2] J. Lehman, et al., "The GeoKnow Handbook", http://svn.aksw.org/projects/GeoKnow/Public/GeoKnow-Handbook.pdf, Accessed in December 2014.

[3] R. Cyganiak, D. Reynolds, J. Tennison, "The RDF Data Cube vocabulary", http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/, January 2014.

[4] M. Perry, J. Herring, "OGC GeoSPARQL – A Geographic Query Language for RDF Data", http://www.opengis.net/doc/IS/geosparql/1.0, July 2012.

[5] D. Paunović, V. Janev, V. Mijović, "Exploratory Spatio-Temporal Analysis tool for Linked Data", *In Proceedings of 1st International Conference on Electrical, Electronic and Computing Engineering*, RTI1.2.1-6., June 2014, Vrnjačka Banja, Serbia.

[6] V. Mijović, V. Janev, D. Paunović: "ESTA-LD: Enabling Spatio-Temporal Analysis of Linked Statistical Data", *In Proceedings of the 5th International Conference on Information Society Technology and Management, Information Society of the Republic of Serbia*, pp 133-137, March 2015, Kopaonik, Serbia.

[7] SDMX Standards, "Information Model: UML Conceptual Design", https://sdmx.org/wp-content/uploads/SDMX_2-1-1_SECTION_2_InformationModel_201108.pdf , July 2011.

[8] M. Martin, K. Abicht, C. Stadler, A. Ngonga, T. Soru, S. Auer, "CubeViz: Exploration and Visualization of Statistical Linked Data", *In Proceedings of the 24th International Conference on World Wide Web*, pp 219-222, May 2015, Florence, Italy.

Management of Accreditation Documents in Serbian Higher Education using ontology based on ISO 82045 standards

Nikola Nikolić[*], Goran Savić[*], Robert Molnar[*], Stevan Gostojić[*], Branko Milosavljević[*]
[*] University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
{nikola.nikolic, savicg, rmolnar, gostojic, mbranko}@uns.ac.rs

***Abstract*— The paper deals with managing accreditation documents in Serbian higher education. We propose domain ontology for the semantic description of accreditation documents. The ontology has been designed as an extension of a generic document ontology which enables customization of document centric systems (DCS). The generic document ontology has been based on ISO 82045 families of standards and provides a general classification of documents according to their structure. By inheriting this high-level ontology, the ontology proposed in this paper introduces concepts for representing accreditation documents in Serbian higher education. The proposed ontology allows defining specific features and services for advanced search and machine-processing of accreditation data. An evaluation of the proposed ontology has been carried out on the case study of accreditation documents for the Software engineering and information technologies study program at the Faculty of Technical Sciences, University of Novi Sad.**

I. Introduction

Accreditation in higher education is a quality assurance process where educational institutions are validated by an official supervisor whether they meet specific standards. The standards define methods and procedures for the work quality assurance in various fields such as study programs; teaching process; teachers and staff, scientific, artistic and professional research; textbooks, literature, library and information resources; quality control and so on. Among all mentioned accreditation components, this paper is primary focused on standards which regulate a study program, its curriculum, and teaching staff.

The accreditation process involves managing different documents which represent accreditation data. To avoid manual management of these documents, a document management system (DMS) can be used to provide storage and retrieval of such documents. DMSs commonly provide storage, tracking, versioning, metadata, security, as well as indexing, retrieval capabilities, integration, validation and searching [1, 2]. Most of current DMS implementations enable generic document management with scarce support of domain-specific semantics.

Semantically-driven DMSs rely on a semantic structure which describes documents data allowing the existence of complex services that "understands" the nature of documents [3, 4]. Still, these systems are commonly designed for a general-purpose document management allowing domain-neutral semantics and features only. Among all necessary data and services, such DMSs contain only those that are common to all domains.

This paper proposes a novel approach for a domain-specific semantically-driven management of accreditation documents in Serbian higher education. The proposed solution should provide a complex management of accreditation documents, which has not been provided by the general-purpose DMSs. We represent accreditation documents formally using a model that describes their content, semantics and organization. Such an approach enables the establishment of domain-specific services for document management. Our solution relies on the semantically-driven DMS presented in [5]. The system provides semantic document management based on the techniques of the semantic web [6].Although it relies on the generic domain-neutral ontology, it may be customized for different domains by creating domain-specific ontologies. In this paper, we have created an ontology which represents accreditation documentation in Serbian higher education. The ontology may be used as a basis for the implementation of services for advanced search and another machine reasoning of the accreditation documents. As a case study, we have formally represented data on curriculum and teaching staff from the Software engineering and information technologies study program which at the Faculty of Technical Sciences, University of Novi Sad [7].

The rest of the paper is structured as follows. The next chapter presents other researches in this field. Chapter three gives a description of the generic ontology for document representation. Then, domain ontology for representing accreditation documents is presented. Chapter five presents a case study on a representative study program from the University of Novi Sad. Finally, the last section gives paper's summary and outlining plans for the further research.

II. Related work

In this chapter, we present other researches on semantically-driven document management. Clowes et al. in [8] claim that semantic document model is a hybrid model composed of two parts: 1) a document model which is used to present the architecture and the structure of a document and 2) a semantic model which is used to add the semantic data on the document, i.e. to represent meaning and relationships of the structure elements. As a case study, they use the Tactical Data Links [9] domain, which is used as military message standard. The paper proposes a document model which includes junction points used to attach the semantic model. The semantic

model must be specifically developed for each domain. Regarding document types, the presented model is mainly focused on textual and structured documents.

Health Level 7 [10] is a non-profit standard-developing organization providing a comprehensive framework and related standards for the electronic health-care data and document management. Clinical Document Architecture (CDA) is one of their popular markup standards for representation of the clinical documents by specifying the structure and semantics of such documents. A clinical document has several characteristics described in [11] and all of the CDA documents derive their semantic content from the HL7 Reference Information Model. The standards cover clinical/medical concepts required to fill huge medical documentation. Some semantic parts are deliberately omitted due to the complexity and enriching the model with missing semantics is expected in future versions. Although the model is incomplete and domain-specific, it gives a valuable modeling example for documents from other domains.

The system presented in [5] introduces semantics into DMS and WfMS (Workflow Management System). Authors explain that semantic layer should consist of two sublayers: a domain-free layer which models abstract documents and a layer which will provide concepts from a concrete domain. The domain-free layer has been described by the generic document management ontology presented in [13]. This ontology has been developed to enable different customizations in document-centric systems. The ontology represents a generic document architecture and structure, which can be extended to describe some specific domain. The ontology models document management concepts as they have been defined by ISO 82045 family of standards. The ISO 82045 family of standards [12] formally specifies concepts related to document structure, metadata, life-cycle and versioning.

In this paper, we use this idea of two abstraction layers to represent accreditation documents. The first (higher) abstraction layer which represents generic document management concepts is formally modeled using the generic document management ontology presented in [13]. The second (lower) abstraction layer represents the domain of accreditation documentation, which is the main subject of this paper.

The following section presents the generic document management ontology while the ontology of accreditation documents is presented in section 4.

III. Generic document management ontology

This chapter presents the generic document management ontology (GDMO) which has been used in our solution as a basis for describing accreditation documents. As mentioned, these ontology models document management concepts as they have been defined by ISO 82045 family of standards. In addition, it relies on two other ontologies recommended by the W3C - PROV-O [14] and Time [15] ontologies. GDMO covers the most generic cross-domain document concepts specified by the ISO 82045 family of the standards, such as a document, a part of the document, document metadata, document version, as well as document identification, classification and document format. The graph nodes represent ontology classes while object properties are displayed as graph links. The key ontology concepts and their semantic relationships are presented in Figure 1.

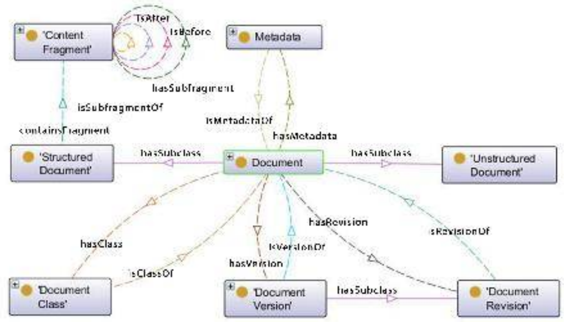


Figure 1. GDMO concepts

The Document is the main concept in the model and it is considered as an FRBR work entity [16], which is defined as a distinct intellectual or artistic creation. A document categorization can be performed based on the structure of its content. The structured document is represented by *Structured Document* class that contains structured content which is represented by individuals of the *ContentFragment* class. Each fragment may have its own subfragments. This hierarchy between fragments is modeled using the object properties *hasSubfragment* and *isSubfragmentOf*. In addition, fragments at the same hierarchy level can be ordered using the *isBeforeFragment* and *isAfterFragment* object properties. The unstructured document is represented by the *Unstructured Document*. The content of documents of this type is defined within *unstructuredContent* data property of the *UnstructuredDocument* class.

The document can be classified by an arbitrary classification scheme. The generic classification is presented with the *DocumentClass* class. For a specific domain, this class must be specialized using a domain-specific classification of documents. Documents may contain metadata that is presented with the *Metadata* class. The given model is neutral with respect to the representation of the document metadata. The paper [19] presents a metadata model based on the ebRIM (ebXML Registry Information Model) specification [20] that can be used to additionally describe the semantics of the documents formally represented by this generic ontology.

During a document life cycle, the content of the document is being changed. Any modification of the content presents a new version of the document. The model provides document version tracking through the *Document Version* class and its subclass *Document Revision*. Only the official document revision is presented underclass *Document Revision*. To define data required for versioning, the PROV-O ontology has been used [14]. A relation of a document with its version has been defined by the *isVersion* and *hasVersion* object properties.

Similarly, a document and its revision are associated with the *hasRevision* and *isRevision* object properties.

Depending on document structure, a document is an instance of exactly one document type. Besides structured and unstructured document which are described in figure 1, additional document types are supported, i.e. single document, compound document, aggregated document and document set. These additional types are shown in Figure 2.

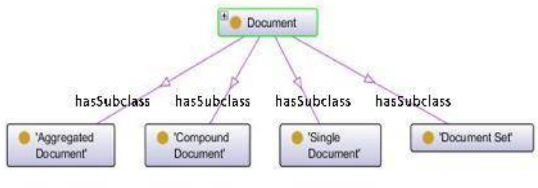


Figure 2. Document subclasses

The basic unit of document management is a document of the type of *Single Document* class. An aggregated document has been represented by the *Aggregated Document* class. An aggregated document is a document which contains metadata and other documents. The document which contains metadata and other documents without metadata is a compound document represented by the *Compound Document* class. A collection of documents has been represented using the *Document Set* class.

A formal definition of these document types can be found in [13]. The definition has given as a plain text, as well as OWL expressions using Manchester syntax [17].

The generic document ontology has been used in [5] and [13] to represent legislative documents. For that purpose, a domain-specific ontology for representing legislative domain has been developed. Still, the ontology models document's structure only. In this paper, we present a domain-specific ontology for accreditation documents which describes both their content and structure.

IV. Ontology for Accreditation Documents

In the previous chapter, we have described formal ways of representing and managing documents at a generic domain-neutral level. In this chapter, we introduce a semantic structure of accreditation documents. The semantic structure has been implemented as an extension of the generic document ontology presented in the previous chapter.

Before describing the semantic structure, we are going to describe the content of an accreditation document briefly. The competent authority of the Republic of Serbia has defined guidance for creating accreditation documentation [21]. They have stipulated twelve standards that must be followed in the accreditation documentation. The standards cover various fields such as study program; teaching process; teachers and staff; scientific, artistic and professional research; textbooks, literature, library and information resources; quality control and so on. Each of these standards must be separately met by appropriate data. This paper's focus has been set on standards which regulate study programs and their curriculum (standard no. 5) and teaching staff (standard no. 9).

The Curriculum Standard is composed of a list of courses and their specifications. Each course contains details about a semester, course type, title, ECTS points, the number of weekly lectures, course objectives, etc. Also, course content, course and evaluation methods, literature and teachers are described. The Teaching staff Standard contains a list of teachers who are involved in teaching process at the particular study program. Teachers are represented by general personal data, lists of qualifications and references, and a list of courses which they teach.

In this paper, the data proposed by the mentioned accreditation standards have been formally represented using ontology. In the following text, we present the ontology classes, object properties and data properties related to these data.

The graph shown in Figure 3 presents the proposed

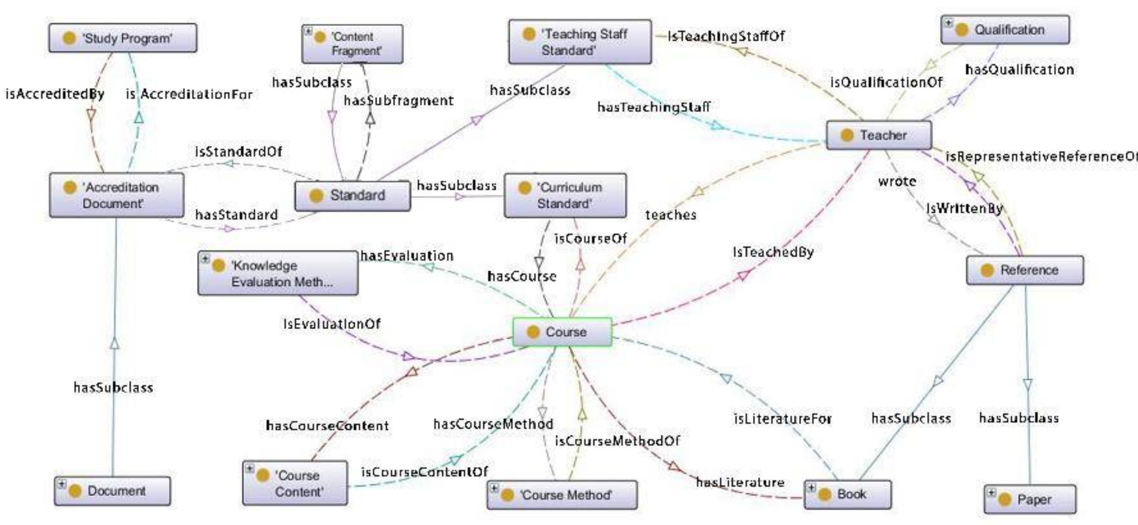


Figure 3. Semantic structure for accreditation documents

ontology of accreditation documents. The classes and properties proposed by the generic document management ontology serve as supertypes for the concepts introduced in the ontology of accreditation documents. Due to limited space, most of these classes and properties are not displayed.

Class *Document* represents an abstraction of all types of documents and it is defined within the generic document management ontology. For representing accreditation documents we have introduced the *AccreditaionDocument* class as a subtype of the *Document* class. The figure shows that an accreditation document is related to a corresponding study program using the object properties *isAccreditationFor* and *isAccreditedBy*. Given that accreditation documents are structured and composed of standards, we can notice the object properties *hasStandard* and *isStandardOf* which defines the relation between *Accreditation Document* and *Standard* as a special type of fragments of an accreditation document. The *ContentFragment* class represents the document content meaning that our ontology describes both the content and structure of the accreditation documents, in contrast to the generic document ontology which describes the structure only.

In order to represent the content of an accreditation document, the *Content Fragment* class has been inherited by the *Standard* class that represents all the thirteen accreditation standards at the generic level. Keeping in mind the focus of this paper, the *Curriculum Standard* and *Teaching Staff Standard* classes have been derived from the *Standard* class as new subtypes. All other accreditation standards have simple textual content and can be represented as an individual of the *Standard* class which may be related to multiple subfragments with *hasSubfragment* object property.

Data about the specific course in the curriculum are represented by the *Course* class. The object properties *hasCourse* and it inverse property *isCourseOf* defines which courses are contained within the curriculum. The course content is defined as a textual value represented as the data property of the *CourseContent* class. Data about instructional methods used within the course are represented by the *Course methods* class. Student's knowledge in the course is evaluated using the methods represented with the *Knowledge Evaluation Methods* class.

According to the requirements of the study program, the *Teaching Staff Standard* class defines teaching staff that has required professional and academic qualifications to participate in the course. A single *Teacher* may be related to multiple courses and vice versa. The *Reference* class describes all teachers' publications where some of them can be used as a recommended literature for the course. The ontology distinguishes two types of publications – scientific papers and books.

The next section presents the instances of the concepts presented in this section on a case study of the representative accreditation documents.

V. CASE STUDY

This section presents an evaluation of the ontology for accreditation documents presented in the previous section. The ontology has been evaluated on the accreditation documents of the study program of Software engineering and information technologies at the Faculty of Technical Sciences, University of Novi Sad. We have represented data from these documents using the classes and properties defined by the proposed ontology for accreditation documents. The ontology, together with the corresponding individuals of the Software engineering and information technologies study program, is publicly available at http://informatika.ftn.uns.ac.rs/files/faculty/NikolaNikolic/icist2016/accreditation-document-ontology.owl. The ontology data are illustrated in figure 4.

Keeping in mind the complexity of the represented accreditation document, we have chosen a limited set of

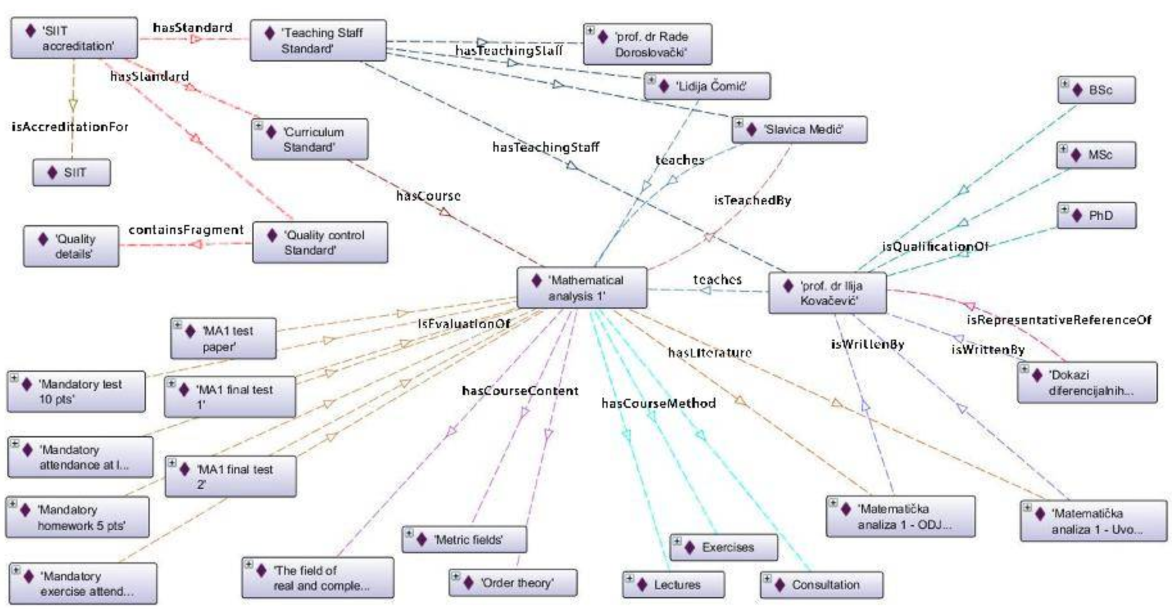


Figure 4. Semantic structure for accreditation documents with individuals

individuals to present in the figure. Accreditation document is represented as the *SIIT accreditation* individual of the *AccreditationDocument* class. This accreditation document has been related the corresponding study program (represented by the SIIT individual) using the *isAccreditationFor* object property.

Three instances of the *Standard* class can be noticed in the figure. *Teaching Staff Standard* and *Curriculum Standard* are individuals of the specific types of standards (subclasses of the *Standard* class) that support the semantic representation of the document content. *Quality Control Standard* has a simple textual content and it is represented as an individual of the generic *Standard* class. The content of this standard has been represented using the generic *Content Fragment* class. Regardless of the type of standards, accreditation documents can be searched and classified by any type of standard. Still, the special type of standards provides detailed semantic structure and document management services.

Among all the courses the study program contains, the figure presents the course of Mathematical analysis 1. According to that, the course details such as teaching methods, teachers, course content, course methods and literature are presented with the appropriate individuals.

The course of "Mathematical analysis 1" has seven mandatory and non-mandatory knowledge evaluation methods. These methods are not specific for this course only and can be related to other courses too. Three teachers are involved in this course. A major teacher "prof. dr Ilija Kovačević" has three qualifications and references, where two of them are books used as an official literature for this course.

Based on the presented semantic representation of the evaluated study program, a semantically-driven DMS can support semantic search services of the accreditation document of this study program. Furthermore, additional knowledge on the accreditation document can be obtained using a semantic reasoner. Using SPARQL queries, DMS can execute complex queries to discern relationships between documents and their parts. For example, one can get accreditation documents involving teachers with competencies and references in a particular scientific field.

VI. Conclusion

In this paper, we have presented a method for the formal representation of the semantics of accreditation documents in Serbian higher education. Accreditation documents have been semantically described using a domain ontology based on ISO 82045 standard. Ontology for describing accreditation documents relies on a generic document ontology for document management which enables domain customization of document management systems. Such an approach provides using domain-specific document management concepts and services for managing accreditation documents.

The generic ontology allows document classification according to its structure, as well as representing document's identifiers, metadata, and life cycle. The generic ontology has been extended with a semantic layer for describing the domain of accreditation documents. The domain ontology specifies an additional set of classes and object properties for representing data describing the structure and content of an accreditation document. Among all the data represented by an accreditation document, the papers' focus has been set on standards for describing curriculum and teaching staff. As a case study, we have developed domain ontology which represents curriculum and teaching staff standard from the accreditation document of the study program of Software engineering and information technologies at the Faculty of Technical Sciences, University of Novi Sad.

Future work will be focused on the integration of the proposed ontology with the semantically-driven DMS. Our plan is to develop domain-specific services for the management of accreditation documentation. The purpose of these services will be to help users while creating accreditation documents. The system should automatically validate an accreditation document by checking its consistency with the official legislative norms. This should facilitate the accreditation process for an educational institution.

Acknowledgment

Results presented in this paper are part of the research conducted within the Grant No. III-47003, Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

[1] H. Zantout, F. Marir, "*Document Management Systems from current capabilities towards intelligent information retrieval: an overview*", International Journal of Information Management, vol.19, Issue 6, pp. 471-484, 1999.

[2] A. Azad, "*Implementing Electronic Document and Record Management Systems*", chapter 14, Auerbach Publications, ISBN-10: 084938059, 2007

[3] Castillo-Barrera, FE Durán-Limón, HA Médina-Ramírez, C Rodriguez-Rocha, B (2013). A method for building ontology-based electronic document management systems for quality standards—the case study of the ISO/TS 16949: 2002 automotive standard. Applied intelligence 38(1):99—113. doi 10.1007/s10489-012-0360-1.

[4] Simon, E Ciorăscu, I Stoffel, K. (2007). An Ontological Document Management System. In Abramowicz, W and Mayr, HC (eds) Technologies for Business Information Systems, Springer Netherlands, pp. 313—325.

[5] Gostojic, S Sladic, G Milosavljevic, B Zaric, M Konjovic, Z (2014). Semantic-Driven Document and Workflow Management, International Conference on Applied Internet and Information Technologies (AIIT).

[6] Berners-Lee, T, Hendler, JLassila, O (2008). The Semantic Web", Scientific American.

[7] Faculty of Technical Sciences, "Dokumentacija za akreditaciju studijskog programa: Softversko inženjerstvo i informacione tehnolog*ije (In Serbian)*", http://www.ftn.uns.ac.rs/n1808344681/

[8] D. Clowes, R. Dawson, S. Probets, "*Extending document models to incorporate semantic information for complex standards*", Computer Standards & Interfaces, vol. 36, Issue 1, pp. 97-109, November 2013.

[9] U.S. Department of Defense, "*Department of defense interface standard – tactical data link (tdl) 16 – message standard*", Tech. Rep. MIL-STD-6016C, US Department of Defense, 2004.

[10] Health Level 7 International, [online] Available at: https://www.hl7.org/

[11] R.H. Dolin, L. Alschuler, C. Beebe et al., "*The HL7 Clinical Document Architecture*", Journal of the American Medical informatics Association, vol. 8, No. 6, pp. 552-569, November 2001.

[12] International Organization for Standardization (ISO), "*ISO IEC 82045-1: Document Management - Part 1: Principles and Methods*", ISO, Geneva, Switzerland, 2001.

[13] R. Molnar, S. Gostojić, G. Sladić, G. Savić, Z. Konjović, "*Enabling Customization of Document-CentricSystems Using Document Management Ontology*", 5. International Conference on Information Science and Technology (ICIST), Kopaonik, 8-11 Mart, 2015, pp. 267-271, ISBN 978-86-85525-16-2.

[14] W3C (2013). PROV-O: The PROV Ontology, [online] Available at: http://www.w3.org/TR/prov-o/

[15] Time ontology, [online] Available at: http://www.w3.org/TR/owl-time/

[16] IFLA FRBR, [online] Available at: http://archive.ifla.org/VII/s13/frbr/frbr1.htm

[17] I.CFogaraši, G.Sladić, S.Gostojić, M.Segedinac, B.Milosavljević, "*A Meta-metadata Ontology Based on ebRIMSpecification*", 5. International Conference on Information Science and Technology (ICIST), Kopaonik, 8-11 Mart, 2015, pp. 213-218, ISBN 978-86-85525-16-2.

[18] OASIS ebXML RegRep Version 4.0. Registry Information Model (ebRIM). (2012). [online] Available at: http://docs.oasis-open.org/regrep/regrep-core/v4.0/os/regrepcore-rim-v4.0-os.pdf [Accessed 20 Nov. 2014].

[19] OWL 2 Web Ontology Language Manchester Syntax (Second Edition), [online] Available at: http://www.w3.org/TR/owl2-manchester-syntax/

[20] Ministry of Education and Sport, "*Akreditacija u visokom obrazovanju (In Serbian)*", Ministry of Education and Sport of the Republic of Serbia, [online] Available at: http://www.ftn.uns.ac.rs/1379016505/knjiga-i-uputstva-za-akreditaciju

Facebook profiles clustering

Branko Arsić[*], Milan Bašić[**], Petar Spalević[***], Miloš Ilić[***], Mladen Veinović[****]
[*] Faculty of Science, Kragujevac, Serbia
[**] Faculty of Sciences and Mathematics, Niš, Serbia
[***] Faculty of Technical Sciences, Kosovska Mitrovica, Serbia
[****] Department of Informatics and Computing, Singidunum University, Belgrade, Serbia

brankoarsic@kg.ac.rs, basic_milan@yahoo.com, petar.spalevic@pr.ac.rs, milos.ilic@pr.ac.rs, mveinovic@singidunum.ac.rs

***Abstract*— Internet social networks may be an abundant source of opportunities giving space to the "parallel world" which can and, in many ways, does surpass the realty. People share data about almost every aspect of their lives, starting with giving opinions and comments on global problems and events, friends tagging at locations up to the point of multimedia personalized content. Therefore, decentralized mini-campaigns about educational, cultural, political and sports novelties could be conducted. In this paper we have applied clustering algorithm to social network profiles with the aim of obtaining separate groups of people with different opinions about political views and parties. For network case, where some centroids are interconnected, we have implemented edge constraints into classical k-means algorithm. This approach enables fast and effective information analysis about the present state of affairs, but also discovers new tendencies in observed political sphere. All profile data, friendships, fanpage likes and statuses with interactions are collected by already developed software for neurolinguistics social network analysis - "Symbols".**

I. Introduction

In recent years, social media are said to have had an impact on the public discourse and social communication. Social networks, such as Facebook, Twitter and LinkedIn have been becoming very popular during the last few years. People experience various life events, happy or unfortunate life circumstances and all these negative and/or positive impressions are almost immediately shared online, winning inner peace and friends' support or opinion to the others. A great variety of stances is to be found online, independently from the subject of discussion. This permanently enlarges pool of comments on brands, events, educational or health system and could be used as a baseline for research in quality and service improvement [1]. Nonetheless, social network potentials are widely recognized. Many companies, schools, public institutions, political parties, popular individuals and groups have already created online profiles for gathering and analyzing the data [2]. These data are, afterwards, useful in numerous areas such as marketing, public relations, and any type of a thorough research of public opinion [3].

It is certain that, apart from web crawlers that are crucial for forum research, social networks can yield material for sophisticated analyze in the field of marketing and branding [4]. An advantageous approach to grouping people based on their interests comes from the knowledge of their personal data, such as one`s location, birthday, job and education.

In particular, social media are increasingly used in political context [5][6]. Potential voters share their impressions daily in the form of statuses about upcoming events and present state of affairs, their problems, political stances, agreements or disagreements with political activities, plans, and such like daily subjects. In order to meet the citizens' needs, politicians and spin-doctors extract and analyze the information of interest from the available statuses. Twitter is favorite amongst politicians and other known personalities, and thus seems better for collecting and comparing public opinions. Facebook is the most used social network in Serbia, hence we focused our online political study on Facebook. Moreover, Facebook offers the way of entering into direct dialog with citizens and encouraging political discussions, while Twitter streams short flurry of information while the fresh ones rush in continuously. Two more important differences between Facebook and Twitter are: real life friends vs. connecting with strangers and undirected vs. directed edges between profiles. The undirected edges for nodes equality were also the milestone for Facebook selection, too. The unique possibilities of public opinion research through internet, such as real-time data access, knowledge about people's changing preferences and access to their status messages provide prospect for innovation in this field, contrasting to classical offline ways.

In this paper, we present a procedure for finding and analyzing valuable information related to the specific political parties. Our approach is based on Facebook profiles clustering according to their common friends and interests. Clustering techniques can help us to understand relations between profiles and create a global picture of their traits, and eventually conclude how politicians can have impact on them. For this purpose, we adopted well-known clustering algorithm "k-means" for dividing social network profiling separate groups, thus providing a room for profiling potential voters. In precise, algorithm k-means is adjusted for graph clustering process in order to form several connected components respecting the similarity between nodes. Collecting and filtering is done by already developed software for neurolinguistics social network analysis - "Symbols"[a], which is described in more details, in Section 3. Other approaches are also present and they are focused on analyzing the structure of the social networks and profiles centrality (e.g. see [7, 8, 9, 10]).

[a] http://symbolsresearch.com

The remainder of the paper is structured as follows. Section 2 gives an overview of the literature. Section 3 presents the details of our software "Symbols". Recent surveys of Facebook popularity in Serbia are highlighted in Section 4. Section 5 describes our research methodology. Section 6 extends the standard k-means from vectors to the nodes of graph. The results are presented in Section 7, while Section 8 concludes the study.

II. Related work

Much of real data could be presented as a network (graph). Objects can be presented as nodes, and relations among them as graph's edges. Based on Facebook users' relationships and fanpage likes we have created a network out of Facebook profiles. The problem of data clustering with constraints is now surpassed with graph-based clustering. In this way each element which is be clustered is represented as a node in a graph and the distance between two elements is modeled by a certain weight on the edge thus linking the nodes [11]. The stronger the relation between objects, the higher the weight is (smaller is the distance), and vice-versa. Graph based clustering is a well-studied topic in the literature, and various approaches have been proposed so far.

In paper [12], the graph edit distance and the weighted mean of a pair of graphs were used for cluster graph-based data under an extension of self-organizing maps (SOMs). In order to determine cluster representatives, the authors in [13] conducted the clustering of attributed graphs by means of Function Described Graphs (FDGs). In later approaches the notion of set median graph [14] was presented. It has been used to represent the center of each cluster. However, better presentation of each cluster data is obtained by the generalized median graph concept [14]. Given a set of graphs, the generalized median graph is defined as a graph that has the minimum sum of distances to all graphs in the set. However, median graph approaches are suffering from exponential computational complexity or are restricted to special types of graphs [15]. It would seem that spectral clustering algorithm [16] appears as a much better solution. This method uses the eigenvectors of the adjacency and other graph matrices to find clusters in data sets represented by graphs. k-means clustering algorithm for graphs was introduced [17], bearing in mind the simplicity and speed of algorithms. In this paper we suggested an extension of classical k-means algorithm for Euclidean spaces [18][19], but implemented in the case of graph (see Section 5).

III. "Symbols" data collection

In this section we give a brief overview of Symbols software and its possibilities. As "glue" between our software and Facebook API we developed a Facebook application SSNA (Software for Social Network Analyses). When users start this app, they are asked for the private data access permission. Upon their agreement, the app calls Facebook API on behalf of users after which valid security token for the next two months is obtained. The data encompasses the following network records:

1) The friendship network: ego network includes the SSNA app users (egos) as nodes and friendship relations between them;
2) The communication network:
   (a) Like relations: by clicking a "like" button, Facebook users can value another person's content (posts, photos, videos);
   (b) Comment relations: Facebook users can leave comments on another person's content;
   (c) Post relations: Facebook users can post on the "wall" of another person to leave non-private messages.
3) Affinity network: Attachments to various fanpages and groups implicating support and agreement within their niche.

This software offers graphical presentation of statistical data for selected political parties based on social network statuses and likes, and many more.

IV. Facebook in Serbia

According to the last researches of Ministry of Trade, Tourism and Telecommunications in Republic of Serbia, 93.4% of Internet users aged 16 to 24 have a profile on the social networks (Facebook, Twitter). Our research paper is based on Facebook audience, because most of the world's population are friendly oriented according to this global Internet social network. Facebook Advertisement service presents potential reach of 3,600,000 people from Serbia for the promotion. If we are to believe the self-reported information from people in their Facebook profiles, about 45% of them are women and 55% are men. Information are only available for people aged 18 and older. The largest age group is currently form 18 to 24 with total of 1 440 000 users, followed by the users in the age form 25 to 34. Faculty (College) level educated people participate in about 66%, whilst high school students participate in about 32%. At the same time, percentage for single and married relationship status is 38% to 42%.

V. Methodology

Our research focuses on the political parties' prevalence in the whole of territory of the Republic of Serbia. According to our figures, the total number of grabbed fanpages is 663925 and it corresponds to a total of 78758 profiles. Among these fanpages, 4095 are placed by their creators in the sphere of politics, while 771 pages have more than three likes. Profiles and fanpages are used for graph construction. Profiles represent graph nodes, while fanpages determine a measure for similarity between profiles, i.e. weight of the edges.

Last social research shows that people on the Internet social networks, such as Facebook, mark interactions with small number of friends compared to the total number of friends (about 8%), while the remaining ones are "passive". Members of the mentioned minority have similar interests, common friends, and acquaintances from diverse events. This kind of Internet behavior leads us toward taking into consideration common pages as well as common friends in order to create graph with strong edges. We have taken into consideration the limited number of pages for every political party according to total number of page likes, because a very large number of fanpages can yield misleading results. Bearing this in mind, we selected ten most numerous fanpages of each political party by searching keywords in the title related to their name, abbreviation and leaders. Lets denote this set of fanpages with S. We limited our examination to the four most popular political parties at this moment.

VI. ADOPTED k-MEANS ALGORITHM

The concept of a sample mean is defined as the mean of the observed samples. The sample mean is well-defined for vector spaces only, and we are often forced to present objects by definite discrete structures such as strings, graphs, sets, etc., where sample mean is not always possible to define. The k-means algorithm is a popular clustering method because of its simplicity and speed [20][21]. Algorithm 1 describes k-means for vectors in order to point out changes with our adaptation for graphs.

Algorithm 1: *k-means algorithm for Euclidian space.*

```
1. Choose initial centroids Y = {y_1,...,y_k} ⊂ X, where X
   is a set of all vectors and |Y| = k.
2. repeat
3.     assign each x ∈ X to its closest centroid
       y = y(x) = argmin_{y∈Y} ‖ x − y ‖^2 of a cluster C(y)
4.     recompute each centroid y ∈ Y as the mean of
       all vectors from C(y)
5. until some termination criterion is satisfied;
```

As previously mentioned, we did not consider only friends connections for graph construction, but also the same interests and common friends in order to make stronger connections among people. We say that two friends are connected if they have more than three fanpages (four and five have been also tested) and more than four common friends; otherwise we disconnect the edge in graph. Through the same interests and acquaintances, created edges represent strong relations between active friends (Fig. 1). In accordance with these rules, we obtained a graph with 428 nodes and 4448 edges (more than three fanpages and four friends in common, Fig. 2). In a spirit of k-means algorithm, for similarity between connected nodes we used the following function:

$$sim = \frac{1}{\alpha \times \sigma(u,v) + \beta \times \phi(u,v)}$$

where $\sigma(u,v)$ presents structural similarity between nodes [22] and $\phi(u,v)$ the number of chosen fanpages in common for profiles u and v and then divided by total number of pages (40 in our case). The smaller the value of similarity function, the closer the nodes are. Parameters α and β can be used to favour one of the parameters. Here, we considered that $\alpha = \beta = 1$. If we obtained a disconnected graph, we would choose two arbitrary nodes from any separated components and make an edge between them with the smallest similarity value, and so on until the connected graph is obtained. For cluster centers determination we used betweenness centrality as an indicator of a node's centrality in a network [23]. We chose this measure because betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes thus matching the nature of a problem. A node with high betweenness centrality has a large influence on the transfer of items through the network. The Algorithm 2 presents an adaptation of Algorithm 1 for graph paradigm.

Algorithm 2: *k-means algorithm for graphs.*

```
1. Choose initial centroids Y = {y_1,...,y_k} ⊂ nodes(G),
   where nodes(G) is a set of graph nodes and |Y| = k.
2. repeat
3.     assign each x ∈ nodes(G) to its closest
       centroid y(x) = argmin_{y∈Y} Σ_{e∈shortest_path} sim(e)
       of a cluster C(y)
4.     replace each centroid y ∈ Y with the node
       which corresponds to the maximal value of
       betweenness centrality of all nodes from C(y)
5. until number of iterations is equal to t;
```

The first step in data clustering is determining a number of clusters k. Generally speaking, number of clusters k is determined in advance according to data sample. The problem we have been solving suggests the fixed cluster number with the value 4. First step is to randomly choose four nodes. In every loop step, an association of all nodes to the nearest centroid is performed. The nearest centroid is determined as a minimal sum of weights along the shortest path between a node and centroids. The next step includes betweenness centrality calculation for every current cluster and the replacing centroids according to the largest coordinate. Calculating the betweenness centrality of all the vertices in a graph is very complex. It is precisely $\Theta(|nodes(G)|^3)$ time-consuming, because it involves calculation of the shortest paths between all pairs of vertices in a graph. We have noticed in numerous experiments that after a few iterations centroids remain the same. This feature has a good influence on algorithm

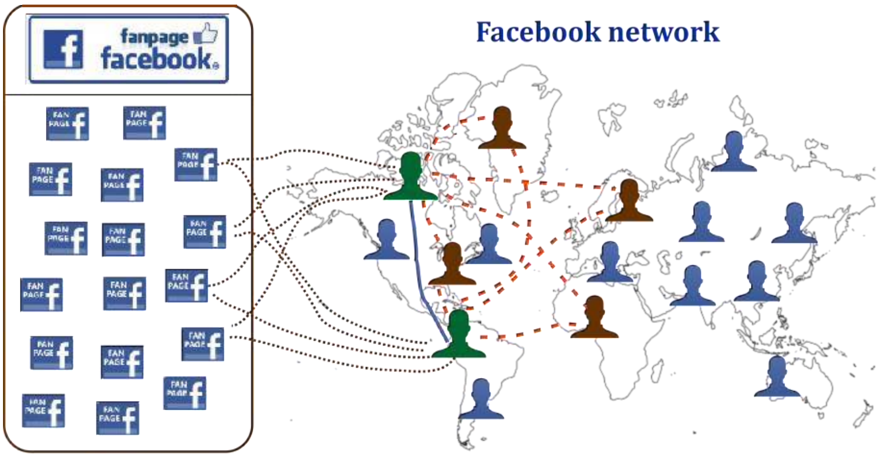


Figure 1. Friends (green) with four fanpages and four friends in common.

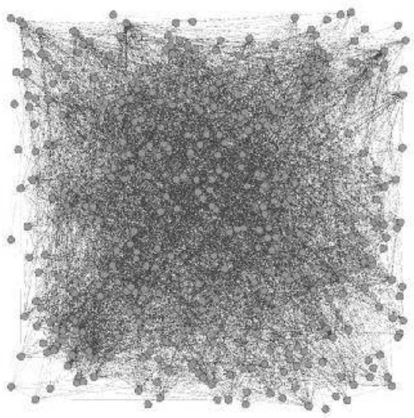


Figure 2. Facebook profiles network, 428 nodes and 4448 edges.

complexity, because we do not need to execute a large number of iterations. Experimental results suggest us to set the number of iterations t from two to four. The calculation of shortest paths between graph nodes in the third step of Algorithm 2 are used for betweenness centrality calculations in the fourth step. This is also relaxation for algorithm calculation complexity. In the following section we give an overview of the experimental results.

VII. RESULTS AND DISCUSSION

This section is dedicated to experimental results obtained by applying Algorithm 2 on the data collected. Our experiments on profiles are divided into three groups according to the number of fanpages in common: with more than three, four and five fanpages in common. Firstly, we fixed number of clusters to $k = 4$ (number of the most popular political parties in Serbia). Secondly, after the algorithm for clustering is performed in graph constructed of Facebook profiles, for each cluster we have listed all fanpages from S liked by its profiles. Simultaneously, with respect to the cluster, we calculated number of likes for each fanpage listed. A list sample is presented in Table 1.

Based on this list, we determine which political party each cluster represents. Sometimes, it happens that cluster consists of inadequate fanpages, the ones which do not belong to an expected party. If so, the problem of noise is solved with the percentage of contribution calculation for the most dominant fanpages belonging to a political party. If this figure is higher than 80% we relate a cluster with the corresponding party. On the contrary, we mark cluster as ”mixed” if the ratio is less than 80% (see Table 1). In almost all cases we had one ”mixed” and three ”clean” clusters. Tables 2, 3 and 4 show the results of experiments for five algorithm starts per group, the percentage of contribution and the number of nodes in the cluster.

TABLE I.
FANPAGES WHICH BELONG TO PROFILES FROM ONE CLUSTER

Fanpage name	Number of likes	Political party
Fanpage 1	2	Party 1
Fanpage 2	2	Party 1
Fanpage 3	2	Party 1
Fanpage 4	2	Party 1
Fanpage 5	2	Party 1
Fanpage 6	2	Party 2

TABLE II.
NUMBER OF THE FANPAGES IN COMMON IS GREATER THAN 3. THE NUMBERS OF NODES AND EDGES ARE 428 AND 4448, RESPECTIVELY

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1.	MIXED - 278	98,84% - 89	82,30% - 19	92,68% - 42
2.	MIXED - 375	88,95% - 17	100,0% - 6	95.13% - 30
3.	MIXED - 320	92,54% - 47	86,27% - 17	85,25% - 44
4.	MIXED - 335	97,46% - 12	92,41% - 43	95,56% - 38
5.	MIXED - 325	92,37% - 47	97,43% - 12	84,47% - 44

TABLE III.
NUMBER OF THE FANPAGES IN COMMON IS GREATER THAN 4. THE NUMBERS OF NODES AND EDGES ARE 213 AND 1141, RESPECTIVELY

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1.	MIXED - 142	98,68% - 32	92,42% - 30	82,30% - 9
2.	MIXED - 187	100% - 17	100% - 2	95.93% - 20
3.	MIXED - 167	88,04% - 16	95,93% - 20	92,30% - 10
4.	MIXED – 150	95,93% - 20	96,55% - 40	100% - 3
5.	MIXED - 180	87,32% - 12	95,40% - 18	100% - 3

TABLE IV.
NUMBER OF THE FANPAGES IN COMMON IS GREATER THAN 5. THE NUMBERS OF NODES AND EDGES ARE 93 AND 298, RESPECTIVELY

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1.	MIXED – 68	66,67% - 2	100% - 10	95,18% - 13
2.	MIXED – 75	92,30% - 2	90,67% - 13	58,33% - 3
3.	MIXED - 71	88,89% - 6	100% - 3	95,18% - 13
4.	MIXED – 65	95,18% - 13	100% - 2	99,12% - 13
5.	MIXED - 74	91,30% - 12	100% - 2	96,15% - 5

The largest clusters, consisting of profiles affiliated with different political parties at the same time were indecisive ones. This anomaly can be explained as a consequence of numerous coalitions, both local and global. In this cluster, we noticed that the fanpages of two specific political parties cover the largest part of all fanpages listed population. The two of them dominate alternately, but at all times one political party fanpages contribute between 45% and 60% of the fanpages set, depending on the contents of other corresponding clusters. Even though these results are consistent with the results of online polls conducted on -

"Tvoj stav"[b], and may contain valuable information useful for additional comments, we shall avoid drawing generalized conclusions and will not deal with such clusters. Finally, with these clusters we are able to make a voter's profile for a political party in a simple way.

VIII. Conclusion

People share contents about almost every aspect of their life, from opinions on global problems, comments on events, to criticism of political parties and their leaders. These daily online activities encourage the opinion exchange, thus creating political clusters aimed at inspiring certain political actions and coaxing new voters. The goal of this research was to study network ties between profiles according to their common interests. In this paper, we presented a novel graph-based clustering approach which relies on classical k-means algorithm. The algorithm was tested on real Facebook data, and we showed that similar conclusions could be obtained in a faster way when compared to the research conducted by marketing agencies engaged for the same purpose and tasks. We determined three clear clusters for chosen political parties, so that we could distinguish them. The fourth cluster (mixed) consists of about 50% of all the profiles, and this problem remains unsolved. In the future, our efforts would be oriented to its splitting, because undecided group of voters seems to hide important information. The algorithm k-means++ should be a good start [24]. With small modification the same algorithm could be tested on Twitter data. An application upgrade for Twitter profiles will also be our tendency for the future research.

Acknowledgment

This paper was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (scientific projects OI174033, III44006, ON174013 and TR35026).

References

[1] C. C. Aggarwal, "An introduction to social network data analytics," in *Social Network Data Analytics*, C. C. Aggarwal, Eds. Springer US, 2011, pp. 1–15.

[2] S. A. Catanese, P. De Meo, E. Ferrara, G. Fiumara and A. Provetti, "Crawling facebook for social network analysis purposes," In *Proceedings of the international conference on web intelligence, mining and semantics*, ACM, pp. 52, 2011.

[3] M. Burke, R. Kraut and C. Marlow, "Social capital on Facebook: Differentiating uses and users," In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 571–580, 2011.

[4] B. Arsić, P. Spalević, L. Bojić and A. Crnišanin, "Social Networks in Logistics System Decision-Making," In *Proceedings ot the 2nd Logistics International Conference*, pp. 166–171, 2015.

[5] D. Zeng, H. Chen, R. Lusch and S. H. Li, "Social media analytics and intelligence," *Intelligent Systems*, IEEE, vol. 25, no. 6, pp. 13–16, 2010.

[6] S. Wattal, D. Schuff, M. Mandviwalla and C. B. Williams, "Web 2.0 and politics: the 2008 US presidential election and an e-politics research agenda," *Mis Quarterly*, vol. 34, pp. 669–688, 2010.

[7] S. Catanese, P. De Meo, E. Ferrara, G. Fiumara and A. Provetti, "Extraction and analysis of facebook friendship relations," In *Computational Social Networks*, A. Abraham, Eds. Springer London, 2012, pp. 291–324.

[8] M. G. Everett and S. P. Borgatti, "The centrality of groups and classes," *The Journal of mathematical sociology*, vol. 23, pp. 181–201, 1999.

[9] J. Scott, *Social network analysis*. Sage, London, 2012.

[10] J. Sun and J. Tang, "A survey of models and algorithms for social influence analysis," in *Social Network Data Analytics*, C. C. Aggarwal, Eds. Springer US, 2011, pp. 177–214.

[11] A. K. Jain, M. N. Murty and P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.

[12] S. Günter and H. Bunke, "Self-organizing map for clustering in the graph domain," *Pattern Recognition Letters*, vol. 23, pp. 405–417, 2002.

[13] F. Serratosa, R. Alquézar and A. Sanfeliu, "Synthesis of function-described graphs and clustering of attributed graphs," *International journal of pattern recognition and artificial intelligence*, vol. 16, pp. 621–655, 2002.

[14] X. Jiang, A. Münger and H. Bunke, "An median graphs: properties, algorithms, and applications," *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, vol. 23, pp. 1144–1151, 2001.

[15] H. Bunke, A. Münger and X. Jiang, "Combinatorial search versus genetic algorithms: A case study based on the generalized median graph problem," *Pattern recognition letters*, vol. 20, pp. 1271–1277, 1999.

[16] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, pp. 395–416, 2007.

[17] A. Schenker, *Graph-theoretic techniques for web content mining*, World Scientific, 2005.

[18] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, vol. 24, pp. 881–892, 2002.

[19] P. Berkhin, "A survey of clustering data mining techniques," In *Grouping multidimensional data*, J. Kogan, N. Charles and T. Marc, Eds. Springer Berlin Heidelberg, 2006, pp. 25–71.

[20] D. Arthur and S. Vassilvitskii, "How Slow is the k-means Method?" In *Proceedings of the twenty-second annual symposium on Computational geometry*, ACM, pp. 144–153, 2006.

[21] S. Har-Peled and B. Sadri, "How fast is the k-means method?," *Algorithmica*, vol. 41, pp. 185–202, 2005.

[22] X. Xu, N. Yuruk, Z. Feng and T. A. Schweiger, "Scan: a structural clustering algorithm for networks," In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 824–833, 2007.

[23] L. C. Freeman, "A set of measures of centrality based upon betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[24] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics*, pp. 1027–1035, 2007.

[b] http://www.tvojstav.com/page/analysis#analize_mdr

Proof of Concept for Comparison and Classification of Online Social Network Friends Based on Tie Strength Calculation Model

Juraj Ilić, Luka Humski, Damir Pintar, Mihaela Vranić, Zoran Skočir
University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
{juraj.ilic, luka.humski, damir.pintar, mihaela.vranic, zoran.skocir}@fer.hr

***Abstract*— Facebook, the popular online social network, is used on daily basis by over 1 billion people each day. Its users use it for message exchange, sharing photos, publishing statuses etc. Recent research shows that a possibility exists of determining the connection level (or strength of their friendship, tie strength) between users based on analyzing their interaction on Facebook. The aim of this paper is to explore, as a proof of concept, the possibility of using a model for calculating strength of friendship to compare and classify ego-user's Facebook friends. A survey, which involved more than 2500 people and collected a significant amount of data, was conducted through a developed web application. Analysis of collected data revealed that the model can determine with a high level of accuracy the stronger connections of ego-user and classify ego-user's friends into several groups according to the estimated strength of their friendship. Conducted research is the base for creating an enriched social graph – graph which shows all kinds of relations between people and their intensity. Results of this research have plenty of potential uses, one of which is specifically improvement of the education process, especially in the segment of e-learning.**

I. Introduction

In today's era of rapid development of technology and information, the Internet has emerged as the largest global computer network. Facilitating everyday life for billions of people, Internet is used for easy discovery of information and performing various tasks, quickly becoming one of the main platforms for communication between people. Internet's ability to connect people in short time encouraged the emergence of a large number of online social networks (OSN), one of the most popular ones being Facebook. Facebook users share information about themselves and interact with other users in various ways. The assumption, which is confirmed by many recent researches, is that the data about the interaction on the online social network can be used to evaluate the level of connection between people in real life. On Facebook the complexity of real life relationships is reduced to only one type of relationship – "friendship". The nature of this relationship is merely binary in nature – ego-user is either connected with someone or not. There is no information about strength of friendship. The question which may arise is how to use the data published by the ego-user or their friends on online social networks to distinguish between strong and weak friendships and somehow evaluate the strength of connection between ego-user and his friends.

The aim of this paper is to verify the next hypothesis by using a model for calculating tie strength introduced in [1]:

1. Higher tie strength calculated by the model for two observed individuals is closely related with the strength of real-life relationship between said individuals.

This hypothesis will be explored in detail by investigating the next subhypothesis:

1. If the ego-user is asked to pick a better friend from selected pair of social network friends, he will select the one for whom the model calculate the higher tie strength.
2. Bigger gap in relationship intensity between two social network friends will result in a larger probability of model providing the correct output.
3. Model results will allow classification of ego-user's friends in 3 groups: *best friends*, *friends* and *acquaintances*. The ordering of these groups will be reflected by the decrease of calculated tie strength (*friends* will have lower tie strength than *best friends*, but higher than *acquaintances*).

To verify the hypothesis and its subhypotheses, we held a survey through a developed web application the purpose of which was to collect data about interaction of ego-user and his friends on Facebook, as well as to collect ego-user's answers on pertinent questions such as: select a better friend out of a selected pair, or distribute friends in one of the following groups – *best friends*, *friends* and *acquaintances*. Ego-users' answers are considered as ground truth. Friendship strength will then be calculated based on data about interaction between ego-user and his Facebook friends, and subsequent research will check for the overlap percentage between ground truth and model outputs, i.e. the level of model accuracy will be determined.

The paper is organized as follows: in section II related work is described; section III introduces a model for calculating tie strength and describes carried out research; in section IV the results of research are presented; section V provides a discussion of these results and in section VI a conclusion is given and ideas for future research are elaborated upon.

II. Related work

Social network (or social graph) is a social structure composed of entities connected by specific relations (friendship, common interests, belonging to the same

group, etc.). By the network theory, social network is composed of nodes and ties. Nodes are entities and ties represent relations between them. An analysis of social networks is not based on entities, but on their interaction.

Nowadays, online social networks, such as Facebook and Twitter, are widely used by people for message exchange, sharing photos or publishing statuses. By using a more concise definition of what social network represents, online social networks can be considered as applications for social networks management.

Information about connections between online social network users is usually relatively poor, most commonly represented in a binary fashion– two users are either friends or not. Social graph that can be created from this information does not differentiate between strong and weak ties, i.e. there is no information about tie (friendship) strength [2].

In recent years, a lot of published papers are dealing with determining the strength of connection between two users on the basis of data about their interaction on online social networks. It is assumed that the interaction between strongly connected users will be more frequent then between those who are not mutually close.

Research being carried out in this area can be categorized by answering the next 3 questions: *What?*, *On what basis?* and *How?*. *What?* refers to what should be achieved by analysis, i.e. what are the analysis' objectives. These objectives can be different: identifying close friends [4][5], calculating trust between users [3][6], searching for the perpetrator and the victims [7], predicting health indicators [11] or recommending content [12][13]. Question *On what basis?* refers to data that are going to be analyzed. Example of parameters that can be analyzed on online social network Facebook are: private messages [3][10], personal interests (music, movies) [3], political views [3] or the frequency of interaction in general [3][4][5][6]. Question *How?* refers to mathematical algorithms and models used to correlate collected data and the objective of analysis. Commonly used are simple linear models, models based on machine learning, optimization algorithms such as genetic algorithms, etc.

A common characteristic of these researches is collecting two types of information: data from online social networks about users, their actions and interaction, and users' (subjective) assessment of observed relationship which is considered as ground truth. Based on those two types of data scientists are trying to construct a model which will be able to calculate the intensity of a connection between two people based on data about their interaction.

Recent research differs in a way how the researchers find ego-user's opinion about his friends, i.e. how they extract ground truth from ego-user. Generally it is done through surveys where users are asked about the type of relation which is the subject of analysis. There are some questions like: *Would you feel uncomfortable if you have to borrow 100$ from him?* [3] or *Would you believe to the information that this user shares with you?* [14]. Users are also asked to select their close friends [4][5] or a few of their best friends [6][8], new friends are recommended to them and a check is performed if the recommendation was accepted [9] or if the algorithm will successfully detect existing friends in a wider group of people [10][12].

The need for knowing the intensity of relations between users appears in different areas. Telecoms are trying, by analyzing social network where ties mean influence between users, to detect possible churners (users that are likely to change network) [15][16][17]. Usually information for building that kind of social network is fetched from call detail records (CDR). Enterprises would like to see a social network of their employees where tie strength means level of cooperation and communication between them [18]. That kind of social network is built by a process of analyzing communication of employees through different corporation communication channels. Also, it is interesting to build a social network where tie strength describes similarity of consumer interests (similar interest in music, movies, theater, art, etc.) or level of trust between users [6][13]. All of those are different, but correlated relations.

In the context of educational data mining, building and analyzing of social networks is important for understanding and analyzing connection between students or course participants [19]. Interaction of participants in collaborative tasks can be analyzed and, as a result of this analysis social network can be constructed. Instructor can subsequently use this social network to find which participants are most important for the propagation of information, i.e. who is central node in social network [20]. If those participants acquire certain course knowledge, it is also likely that this knowledge will be more easily transferred and acquired by other participants.

III. METHODOLOGY

A. Model for calculating friendship strength on an online social network

Model introduced in [1] is used to calculate tie strength, i.e. strength of friendship. Friendship is calculated based on the analysis of interaction between users and includes, with certain (differing) level of significance, all communication parameters (such as "like"s, private messages, mutual photos, etc.).

Friendship is shown as a one-way connection from *user A* to *user B*, where *user A* is the *ego-user*, and *user B* is his network friend. Friendship weight between *user A* and *user B* is not necessarily equal in both ways.

Friendship weight is calculated as sum of multiplication of communication parameters, with the corresponding weight of communication parameters:

$$\begin{aligned} &friendship\_weight = w(likes) \times number\_of\_likes \\ &+ w(comments) \times number\_of\_comments \\ &+ w(messages) \times number\_of\_messages \\ &+ w(tags) \times number\_of\_tags + ... \end{aligned} \tag{1}$$

The weight of communication parameter p for the *ego-user* A – $w(p,A)$ depends on two factors:

- The general significance of each communication parameter $w_g(p)$
- The specific significance of each communication parameter for each user $w_s(p,A)$,

and is calculated with formula (2).

$$w(p,A) = w_g(p) \cdot w_s(p,A) \tag{2}$$

The general significance of each communication parameter is equal for each user and is being defined experimentally as it is described in [1].

The specific significance of each communication parameter is different for each user, because each user uses different communication parameters in a different ratio. For example, some users communicate mostly via private messages, while others prefer *liking* everything that appears on their *News Feed.*

The specific significance of each communication parameter p for *ego-user A* – $w_s(p,A)$ is inversely proportioned to this parameter's usage frequency (if the user is a frequent *liker*, each *like* is individually worth less), is calculated by formula (3), in which $n_p(A)$ defines the quantity of the communication parameter p between *ego-user* and all of his friends. The overall communication of *ego-user A* - $n(A)$ is the sum of all communication parameters of *ego-user A* (the total number of messages, likes etc.)

$$w_s(p,A) = (1 - \frac{n_p(A)}{n(A)}) \tag{3}$$

The specific significance of communication parameter shows how much the specific communication parameter is important in user's communication and how large is its part in overall user's communication (etc. is the user a "*liker*" or a "message sender").

The final formula of friendship strength form *user A* to *user B* is:

$$friendship\_weight_{(A \to B)} = \sum_p w(p,A) \cdot n_p(A \to B) \tag{4}$$

in which $n_p(A \to B)$ defines the amount of communication parameters, in other words, how many communication parameter units were exchanged between *user A* and *user B*. (e.g. number of messages between *user A* and *user B*).

B. *Division of friends into subgroups*

To test research hypothesis we divided ego-user's Facebook friends into 9 subgroups (**Figure 1**). Division is based on calculation of model described in previous subsection. First step is to make an ordered (by tie strength) list of ego-user's friends. The 1st subgroup consists of friends with the highest strength of friendship and the 9th subgroup holds friends with lowest strength of friendship. First subgroup is filled with 1% the best friends of ego-user, second group is filled with following 1%, third with following 1%, fourth with following 2%, fifth with following 5%, sixth with following 10%, seventh with following 20%, eighth with following 30% and ninth with following 30%. It is mandatory to have each subgroup filled with at least one friend and each friend cannot be assigned to two different subgroups. Thus, total ordered list of friends is divided into 9 disjunctive subgroups with ordered the same as in the initial list. Subgroups are different-sized because of the assumption that strength of friendship is mostly distinguish between ego-user and his close friends. With lower friendship strengths, relations are mutually more similar, i.e. ego-user in survey is not able to decide which of his acquaintances is closer to him.

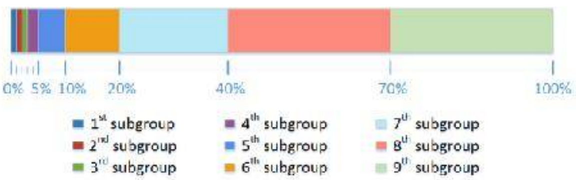


Figure 1. Subgroups of friends

C. *Description of the survey*

Survey is held through a web application. Application, on one side, fetches data about interaction between ego-user (examinee in survey) and his friends and, on the other side, through survey records ego-user's (subjective) opinion about strength of friendship between him and his Facebook friends (ground truth). Survey is divided into 2 sections of questions. In first section ego-user should compare his two friends (**Figure 2**) and select better in pair. In total he should make 24 comparisons. Each friend is randomly chosen from one of 9 subgroups. Although ego-user compares friends, they are actually representatives of their subgroups. **Table 1** shows which subgroups are compared.

Figure 2. Comparing friends – application screenshot

1. – 9.	1. – 3.	2. – 5.	5. – 7.
2. – 8.	1. – 4.	3. – 5.	5. – 8.
3. – 7.	2. – 3.	4. – 5.	5. – 9.
4. – 6. (twice)	2. – 4.	7. – 9.	4. – 7.
5. – 6. (twice)	3. – 4.	8. – 9.	
1. – 2.	1. – 5.	7. – 8.	

Table 1. Compared subgroups of friends

In second section ego-users are asked to classify their friends into 3 groups: *the best friends*, *friends* and *acquaintances* (**Figure 3**). In total they should classify 34 friends – it is chosen randomly 4 friends from first 7 subgroups and 3 from 8th and 3 from 9th subgroup. As in first section, in this section offered friends are representatives of their subgroups.

Tie strength between ego-user and his friends is being calculated by using model introduced in subsection III-A and based on data about interaction between users fetched by application. All examinees approved fetching data about them and their interaction with their friends.

Figure 3. Classifying friends into groups – application screenshot

Answers of ego-users in survey are considered as ground truth and it is analyzed if ego-users answers are matched with results of model. In first section it is considered that by model better friend is that one in pair who has higher tie strength (by calculation of model) to ego-user. In second section it is expected that friends with higher tie strength to ego-users will be classified in higher subgroup, i.e. *the best friends* will be from 1st, 2nd and 3rd subgroup, but friends from 8th and 9th subgroup will be *acquaintances*.

IV. Results

A. Demographic structure

In survey were included more than 3 300 examinees. 2 626 examinees have successfully finished survey. We fetched data about their interaction with more than unique 650 000 of their Facebook friends and analyzed 1 400 000 friendships. Examinees are divided into groups by age, i.e. occupation (**Figure 5**): elementary school (43 participants), secondary school (593 participants), faculties (1 466 participants), employed (444 participants) and unemployed (80) participants. Since we plan to use these results in the future for improvement of educational system, most of examinees were faculty students. Individually most examinees were from electrotechnical faculties in Zagreb (Croatia), Osijek (Croatia) and Belgrade (Serbia). In survey 57.7% of examinees were men and 42.3% women. Questions were written in Croatian language so survey involved only people that understand that language – mostly citizens of the former Yugoslav republics.

B. Comparing friends

In this subsection user's answers in first section of survey are presented where examinees were asked to choose the better friend from two friends in pair. Each friend was representative for one subgroup so comparison of friends can be interpreted as a comparison of subgroups. **Figure 4** and **Table 2** show results, i.e. the percent in which user chose first or second friend in pair as better.

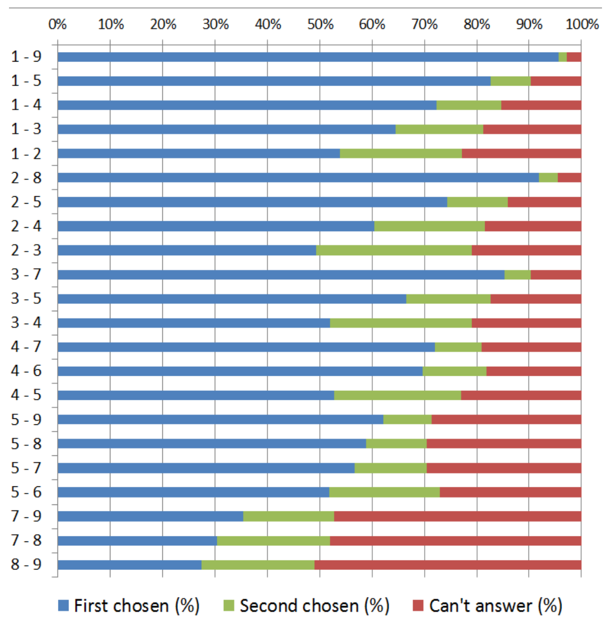


Figure 4. Comparison of pairs of subgroups

Pairs of subgroups	First chosen (%)	Second chosen (%)	Can't answer (%)	Total pairs
1 – 9	95.65%	1.53%	2.82%	2 622
1 – 5	82.64%	7.69%	9.67%	2 615
1 – 4	72.28%	12.41%	15.31%	2 619
1 – 3	64.47%	16.73%	18.80%	2 612
1 – 2	53.88%	23.27%	22.85%	2 617
2 – 8	91.91%	3.51%	4.58%	2 620
2 – 5	74.32%	11.67%	14.01%	2 605
2 – 4	60.50%	21.07%	18.43%	2 610
2 – 3	49.41%	29.72%	20.87%	2 611
3 – 7	85.33%	5.04%	9.63%	2 618
3 – 5	66.55%	16.09%	17.36%	2 604
3 – 4	52.04%	26.96%	21.00%	2 600
4 – 7	71.97%	8.97%	19.06%	2 608
4 – 6	69.64%	12.19%	18.17%	5 234
4 – 5	52.79%	24.18%	23.03%	2 601
5 – 9	62.14%	9.25%	28.60%	2 594
5 – 8	58.81%	11.66%	29.52%	2 598
5 – 7	56.76%	13.72%	29.52%	2 595
5 – 6	51.87%	21.06%	27.06%	5 203
7 – 9	35.47%	17.39%	47.14%	2 588
7 – 8	30.38%	21.63%	48.00%	2 594
8 – 9	27.52%	21.50%	50.98%	2 591

Table 2. Comparison of pairs of subgroups

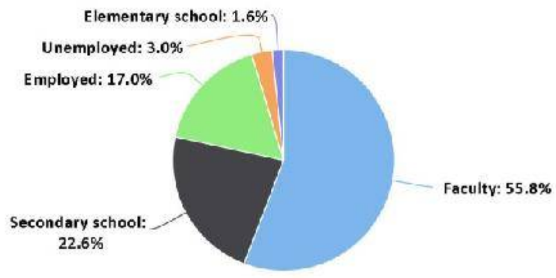


Figure 5. The distribution of participants by occupation

C. Classifying friends into groups

In this subsection ego-user's classifications of their friends into 3 groups are presented: *the best friends*, *friends* and *acquaintances*. Each friend was a representative for one subgroup so classifying of friends can be interpreted as classification of subgroups. **Figure 6** and **Table 3** show results per subgroups.

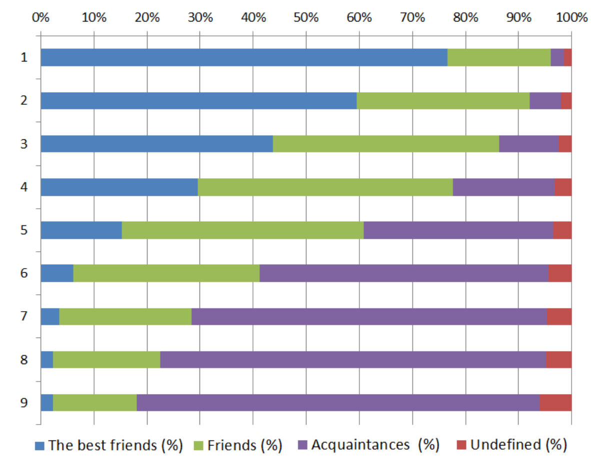


Figure 6. Classifying friends into groups

Subgroup	The best friends (%)	Friends (%)	Acquaintances (%)	Undefined (%)	Total classified
1	76.58%	19.47%	2.53%	1.42%	8 664
2	59.52%	32.56%	5.87%	2.05%	7 989
3	43.77%	42.64%	11.08%	2.51%	7 953
4	29.54%	48.02%	19.29%	3.14%	10 016
5	15.34%	45.46%	35.66%	3.53%	11 376
6	6.13%	35.04%	54.47%	4.35%	11 885
7	3.40%	24.95%	67.03%	4.63%	12 062
8	2.30%	20.14%	72.77%	4.79%	9 532
9	2.28%	15.77%	75.91%	6.04%	9 399

Table 3. Classifying friends into groups

V. Discussion

A. Comparing friends

Results of first section of questions in survey, where examinees were asked to select better friend in pair, are shown in **Figure 4** and **Table 2**. The fact that the higher percent is tied with a chosen friend from higher subgroup suggests that model work properly and confirms the hypothesis that in most cases model is able to detect which friend in pair is better to the ego-user. Furthermore, big differences in percentage are visible between subgroups whenever the subgroups are relatively far apart. It confirms the hypothesis that online social network contains both strong and weak friendships. By comparing subgroups 1 and 9, a friend who represents the 1st subgroup is in 95.65% cases chosen as better, but if we compare subgroups 1 and 2, 1st subgroup is chosen in only 53.88% of cases – though still more than double in comparison with cases where friend from 2nd group is selected as better than friend from 1st subgroup. It shows that the first subgroup truly contains the closest friends, but also that the bigger the real-life difference in tie strength is, the larger is the probability that the model will give correct output. Friends from 1st subgroup are chosen as better in 70% of cases (exclude pairs 1-3 and 1-2). That shows the ability of the model to distinguish strong from weak friendships, but indicates possible problems in the correct ordering of close friends. It is most evident in comparing 2nd and 3rd subgroups where 2nd subgroup is chosen in only 49.41% cases and 3rd in 29.72% cases.

The biggest percentage of answers *can't answer* is for subgroup pairs 7-9, 7-8 and 8-9, which is around 50%. This is understandable since these subgroups contain ego-user's friends with whom he communicates relatively rarely so ego-user is in a difficult position to state which of these Facebook friends is his better friend in real life – both are seen as merely acquaintances. Also, examinees were in 22.85% of the cases unable distinguish between subgroups 1 and 2 which indicates that it is also hard to decide which friend is better if both are ego-user's close friends.

Taking all this into account, it can be stated that these results confirm subhypothesis (1) and (2).

B. Classifying friends into groups

In the second section of questions in survey examinees were asked to classify friends into 3 groups: *the best friends*, *friends* and *acquaintances*. An assumption is that it is possible classify ego-user's friends into groups based on strength of friendship. As friends in main friends list are ordered by strength of friendship, the question is where the border between groups is.

Results are shown in **Figure 6** and **Table 3** and they are in accordance with hypothesis that with rising number of the subgroup, percentage of *the best friends* is decreasing, but percentage of *acquaintances* is increasing. 76.58% friends from 1st subgroup are classified as *the best friends* and only 19.47% as *friends*. It suggests that it is really possible to find ego-user's best friends by using described model. 59.52% friends from 2nd subgroup are classified as *the best friends* and 32.56% as *friends*. It shows that first two subgroups are mostly filled with ego-user's closest friends. In 3rd subgroup is 43.77% of *the best friends* and 42.64% of *friends*. As it is a very small difference we can conclude that 3rd subgroup is the bordering subgroup between groups of *the best friends* and *friends*. Subgroups 3, 4 and 5 are mostly filled with *friends*. Subgroup 6 is first subgroup where *acquaintances* are majority so we can conclude that border between *friends* and *acquaintances* is between 5th and 6th subgroup. It means that about 90% of ego-users Facebook friends are in fact his *acquaintances*.

These results confirm subhypothesis (3).

VI. Conclusion and Future Work

This paper describes research aimed to examine, as a proof of concept, the possibility of using a model for calculating tie strength between ego-user and his Facebook friends, based on analyzing their interaction on Facebook, to compare pairs of friends and to classify friends into predefined groups: *best friends*, *friends* and *acquaintances*. Results show that in most cases this is possible – although perfect detection and classification cannot realistically be expected.

For research purposes a survey was held which included a total of 2 626 examinees. This survey allowed collection of a large amount of data about Facebook users. This dataset is planned also to be used as a referent data set in future for similar types of researches. By using a data set described in this paper as a referent data set we plan to explore new approaches for calculating tie strength based on algorithms for supervised learning. The ultimate goal is to build an enriched social graph which will contain information about different relations between people (friendship, influence, sharing interests, etc.) and information about intensity of each relation. We will explore the possibility of applying enriched social graph in education with special emphasize on using its results in e-learning solutions.

Acknowledgment

The authors acknowledge the support of research project "Leveraging data mining methods and open technologies for enhancement of the e-learning infrastructure" (UIP-2014-09-2051) funded by the Croatian Science Foundation.

REFERENCES

[1] M. Majic, J. Skorin, L. Humski, Z. Skocir, "Using the interaction on social networks to predict real life friendship," in *22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2014*, pp.378-382, 17-19 Sept. 2014.

[2] R. Xiang, J. Neville, M. Rogati, „Modeling Relationship Strength in Online Social Networks“, *19th International Conference on World Wide Web*, Raleigh, NC, USA, pp. 981-990, 2010.

[3] E. Gilbert, K. Karahalios, „Predicting Tie Strength With social Media“, University of Illinois, *Conference on Human Factors in Computing Systems, CHI,* 2009.

[4] M. Madeira, A. Joshi, „Analyzing close friend interactions in social media“, Social Computing (SocialCom), 2013 International Conference, pp 932-935, 8-14 September 2013.

[5] I. Kahanada, J. Neville, „Using transactional Information to Predict Link Strength in Online social Networks“, *Third International Conference on Webloga and Social Media, ICWSM*y, 2009.

[6] V. Podobnik, D. Štriga, A. Jandras, I. Lovrek, "How to Calculate Trust between Social Network Users?", 20th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, 11 – 13 September 2012.

[7] M. Mouttapa, T. Valente, P. Gallaher, „Social Network Predictors of Bullying and Victimization“, Adolescence, Vol. 39, No. 154, Libra Publishers, San Diego, 2004.

[8] M. Majić, J. Skorin, L. Humski, Z. Skočir, „Using the interaction on social networks to predict real life friendship“, *22th International Conference on Software, Telecommunications, and Computer Networks (SoftCOM)*, Split, 17 - 19 September 2014.

[9] Jin Xie, Xing Li, "Make best use of social networks via more valuable friend recommendations", *2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, 2012, pp.1112-1115, 21 – 23 April 2012.

[10] Kuan-Hsi chen, T. Liang, „Friendship Prediction on Social Network Users“, *International Conference on Social Computing (SocialCom), 2013,* pp. 397-384, 8 – 14 September 2013.

[11] S. M. Khan, J. M. Shaikh, „Predicting students blood pressure by Artificial Neuron Network“, *Scienct and Information Conference (sAI), 2014. pp. 430-437, 27 - 29 August 2014.*

[12] J. Naruchitparames, M. Hadi, S. Louis, „Friend Recommendations in Social Networks using Genetic Algorithms and Network topology“, *Evolutionary Computation (CEC), 2011, pp. 2207-2214, 5. - 8. June 2011.*

[13] B. Nie, H. Zhang, Y. Liu, „Social Interaction Based Video Recommendation: Recomending YouTube videos to Facebook User“, *Computer Communication Workshops (INFOCOM WKSHPS), 2014., pp. 97-102, April 27* – May 2, 2014.

[14] E. Khadangi, A. Bagheri, „Comparing MLP, SVM, and KNN for predicting trust between users in Facebook“, *International Conference on Computer and knowledge engineering*, studeni 2013., pp. 477-481

[15] K. Dasgupta, R. Singh, B. Viswanathan, A. Joshi, „Social Ties and their Relevance to Churn in Mobile Telecom Networks“, *EDBT'08*, Nantes (Francuska), 25. – 30. March 2008.

[16] C. Phadke, H. Uzunalioglu, V. B. Mendiratta, D. Kushnir, D. Doran, „Prediction of Subscriber Churn Using Social Network Analysis“, Bell Labs Technical Journal, Vol. 17, No. 4, pp. 63-76, 2013.

[17] G. Benedek, A. Lubloy, G. Vastag, „The importance of Social Embeddedness: Churn Models at Mobile Providers“, *Decision Sciences – a journal of the decision sciences institute*, Vol. 45, No. 1, 2014.

[18] L. Humski i ostali, „Building implicit corporate social networks: The case of a multinational company“, *12th International Conference on Telecommunications (ConTEL)*, pp. 31-38, 26 – 28 June 2013.

[19] C. Romero, S. Ventura, „Data mining in education“, *WIREs Data Mining Knowledge Discovery*, Vol. 3, pp. 12–27, 2013.

[20] R. Rabbany, M. Takeffoli, O. R. Zaïane, „Analyzing Participation of Students in Online Courses Using Social Network Analysis Techniques“, *Proceedings of the 4th International Conference on Educational Data Mining*, Eindhoven, The Netherlands, July 6-8, 2011.

Enabling Open Data Dissemination in Process Driven Information Systems

Miroslav Zaric*
* University of Novi Sad/Faculty of Technical Sciences, Novi Sad, Serbia
miroslavzaric@uns.ac.rs

Abstract — **Open Data movement is gaining momentum in recent years. As open data promise free access to valuable data, as well as its usage, it has become an important concept for ensuring transparency, involvement and monitoring of certain processes. As such open data is of special importance in science and in government. At the same time yet another trend is also visible - adoption of process driven information systems. In this paper an approach to systematic enabling of open data dissemination from process driven information system is explored.**

I. INTRODUCTION

As defined in [1] open data *is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike*. Open data is especially important in science, where certain information should be made available in order to promote scientific research, exchange of ideas, and maximize positive outcomes of scientific research on modern society. One other field where open data is gaining ground is open government data. Government institutions are creating vast amount of data at any given moment. Making (potentially large) portions of that data openly available can bring multiple benefits, such as transparency of government procedures, improved accountability and enhanced involvement of citizens in government. Furthermore, enabling free access to open government data can bring new business opportunities, through creation of new applications that could provide new value to the customers (citizens). In many cases, especially local, governments don't have neither human capacity nor funding to explore, evaluate innovative ways to utilize data that has been accumulated through their every-day operation.

Process driven information systems, primarily workflow management systems (WfMS - introduced in late 1980's), and later business process management systems (BPMs) have been gaining wider acceptance steadily. Though there were different languages for specifying workflow/process models in recent years systems are converging on the use of BPMN [2] as standard notation. Since these systems are increasingly found in larger enterprise suites, such as document management suites, ERPs, CRMs, there is also an ISO/IEC standard [3] governing operational view of BPM systems.

The driving force behind the idea of process driven systems is process modeling, and promise of greater efficiency achieved through workflow automation (routing, coordination between participants and task distribution). Process driven architectures are well adopted in many business suites, they also prove as a welcome tool to enhance productivity in complex multidisciplinary project environments [4], and are well suited for many administrative procedures usually implemented in government institutions [5,6].

Giving the fact that open data is becoming more interesting, and in some cases mandatory by law, or at least preferred by "good practice", and the fact that process driven systems are readily available and embedded in different enterprise systems, there should be some consistent way of specifying correlation between these two concepts.

In this paper an approach to systematically handle open data dissemination from process driven information system is presented. Although the paper discusses case of handling public procurement data, and initial deployment is performed on Activiti [7] process engine, main concepts are independent from specific usage scenario, and are mainly focused on introducing open data dissemination at process model level.

II. ACCESSING DATA IN BUSINESS PROCESS MANAGEMENT SYSTEMS

Virtually all business process management systems handles two groups of data:

- Data defined by BPM implementation environment and derived from deployed process model. This data represent the core data model needed to represent process models, process instances, activities, users, roles and other concepts related to operational process management. Although BPM implementations may display subtle differences, this data largely conforms to conceptual model given in [8]. There is almost 1:1 mapping between concepts given in [8] and entities available in [2]. Typically BPM implementations are using relational databases and object-relational mapping systems to maintain these data and track process state.
- User data - created, manipulated and maintained during process enactment. This data represent actual data processed when process instance is running. Data model is usually complex and suited for specific processing goal. This data may have it's own data storage, or can use the same persistence layer as process engine itself. This user data may be represented at process model level as *Data object*. BPMN specification has defined data object for process model, but support for it in implemented BPM systems is yet not standardized.

While user data (data object) is always directly exposed and available during process execution (in accordance with user roles and access permissions), it is obvious that this data is not the only source of valuable information. For process designers and management personal data recorded by process engine itself may represent valuable source of new information. During process execution, process instance state changes are recorded (during process advancement). By accessing this information process designers and managers can infer new knowledge about process behavior, possibly new process models, detect common process pitfalls, or places where process remodeling could further improve system performance. Techniques used to extract knowledge from BPM systems are commonly known as *process mining* [8,9]. Some data gathered by examining process execution history logs may well be suited to be exposed as open data, and may be used as valuable data set for data further analysis. Therefore not only user data objects are candidates for exposing as open data, but also process details.

However exposing process data in this way, by accessing data storage and/or execution history data tends to be completely decoupled from any given process model, and hence completely out of its control. In process driven environment such thing is rarely desirable. It would probably be beneficial if at least user data object exposure as open data is controlled by process model.

BPMN provides enough concepts to allow embedding open data dissemination as an integral part of relevant process. In fact, there are several ways to implement this, with different impact on process model. Furthermore, embedding data dissemination as integral part of process models, where such dissemination is desirable, may lower the burden of later data gathering, formatting and preparation for publishing as open data.

As presented in [10], one of the problems with overall strategy to make more government data accessible as open data, comes from the fact that most of the data is produced by local levels of government, and with no additional funding for IT services for the task of preparing open data, this tends to create problems.

To explore possibilities of embedding open data dissemination as part of business process model, the rest of the paper will concentrate on case of exposing public procurement data. Main reason for this is that public procurement is well-defined procedure, yet subject to changes in legal requirements and as such is a good candidate to be implemented in BPM systems. Furthermore, most of the readers understand the basics of such procedures, although details can be quite complex. However, full model of public procurement process is quite detailed, large, and will not be presented here, as process model specifics are not of prevailing relevance.

Simplified, procurement procedure requires public entities to publicly declare their procurement need in a public tender. Interested parties pickup tender documentation and decide if they are willing to participate. To participate they need to send a bid to the public entity before the stated deadline. Public entity than performs selection to award the contract to the party with best bid (the scoring system should be also clearly stated in the tender documentation). After the selection procedure, bidders have certain rights to challenge the selection result in a defined period of time. Depending on the results of this challenge, public entity either sign the contract with winning bidder, or the procedure should be annulled. Actual process has much more details, but for basic understanding this simplified description is sufficient.

Currently, there is an international data standard in development for public procurement purposes, named Open Contracting Data Standard [11]. The standard specifies data needed in any phase of public contracting process: planning, tender, award, contract, implementation. Hence, contracting process is more exhaustive than procurement process as it also includes contract implementation monitoring and conclusion. In the procurement process data represented by this standard would be data object during process execution.

III. Enabling Open Data in Business Process Models

As stated earlier both working data object, relevant in process executions, and process execution details may be offered as open data. However, since process business management system aims at enforcing timely and coordinated execution of business activities, time aspect of data exposure should be taken into consideration.

To illustrate this, we can analyze public procurement data. Although it is of public interest to have insight in public procurement procedures, not all the data, and not at all times should be readily available as open data. In fact if some data (such as details of a bid) were exposed prior to tender deadline, whole procurement procedure would have to be aborted since this constitutes violation of law. Hence, it is not only important which data will be exposed as open data, but also when it will be exposed.

In scenarios where no attention has been given to the process model, data gathering and exposure (export to open data storage) will be performed outside of the process control. In that case, the logic of selecting data for "harvesting" from process data storage, is entirely in an outside entity (*export module*). In this case, one solution is to gather, and deliver as open data, only the data gathered from completed process instances.

However, this solution also have its shortcomings - although this approach can guarantee that no data will be exposed before process instance is completed, it will also prevent some data, that should already be readily available, from being visible. In the case of procurement procedure, tender details should be available as soon as it is approved. It is obvious that timing of publicly opening data is important, and furthermore directly influenced by process model.

In other words some data should be exposed as open data as soon as process execution reaches certain point in process ("milestone").

Since process model can change (sometimes even frequently), rules that apply to data exposure may also change, but in this scenario no such change is directly conveyed to the data export module. Additional effort is therefore needed to reconcile behaviors of process management system and data export module. Obviously such approach is prone to errors, or at least requires constant monitoring and adjustment.

There are two domains in which action may be taken to put data exposure under process control.

- First, process model could be supplied with information about which data should be exposed. As stated earlier, BPMN specifies the data object to convey information about data needed for process execution. BPMN allows for describing data object state in each step of the process. However, implementation of data object concept is varying between BPMN systems. If BPM system has support for data object, a logical solution would be to tie the information whether data object (or its constituents) should be made open data. In this case, a special flag property would be assigned to the data object marking it for exposure as open data. However this solution is inadequate from timing perspective.
- Second dimension is timing of exposing data as open data. Process models are composed of activities, gateways and control flow objects (nodes of process graph). Each of these objects has unique id – usually assigned by process designer in order to have some meaningful value. If not assigned, these unique values are created during process model deployment. During process execution, process engine tracks the nodes that have been reached. Some engines are using tokens for this purpose, while others are using *open node* list approach to achieve this. Using the node id, it is possible to detect if certain point in process is reached. If data object is marked for exposure as open data, besides the simple flag it may also has assigned id of the node that must be reached before this data is made openly accessible. Furthermore, since all process engines provide timer functions it would also be possible to specify moment in time when data should be published.

Therefore, in order to achieve controlled publication of open data, any standard data object, representing process working data, would be extended by special *openData* property, marking it for open data publication, and optional *openDataCondition* property stating the node id and/or moment in time when the data should be accessible. Using these markers it is now possible to enforce data publication as open data using the process engine itself.

Since process execution data (such as logged user, performed activity and other) are available at any moment to the process engine, we can treat it in the same manner as user data, practically creating another data object (as process variable), populated with process execution data.

IV. IMPLEMENTATIONAL ISSUES

For test implementation Activiti BPM is used. It is a Java based open source BPM solution. It is BPMN conformant, and provides both implementation engine (process engine), as well as graphic editor for process modeling.

Although Activiti supports data object at model level, this support is limited. Data object may be defined at model level for the process model, but not at node level. Furthermore data objects may be only of simple types. To circumvent this situation it is possible to create xml schema describing complex data objects. But, since no direct support for data object is available at node level of the model (for example at task nodes), object representing procurement data, extended by *openData* marker and *openDataCondition* is created as standard POJO and used as process variable. In this manner this object is readily available to process engine and to relevant task instances. Therefore, this approach solves the problem of identifying which data needs to be exposed as open data.

The exact moment during process execution in which data will be published may be determined and expressed in model in several ways:

- Explicitly using automatic (service) tasks
- Using the event listeners on certain elements in model
- Explicitly using intermediate or boundary events

Each of these approaches have some benefits and shortcomings.

Using explicit service tasks in the model – in this approach process model is amended by specific service tasks, fig.1. In this case data export process is becoming a main activity in certain phases of the process. It is easily understandable from the model.

***Figure 1**. Data publication as service task*

As stated in [11], task nodes in process model should represent main activities needed to advance process toward its end state. In the case of publication of procurement tender documentation the use of service task is justifiable since public availability of this documentation is a prerequisite for process continuation. If service task is used, complete specification of component needed to acomplish the task must be specified. In this case it ammounts to specifying JavaDelegate class responsible for perfroming this task. JavaDelegate class is then able to access process instance, and its variables. As added flexibility, it also possible to use expression to specify delegate class. In test case this JavaDelegate class was responsible for accessing procurement tender data object and transfering it to external data storage. Most common option for achieving any communication with external systems is through web services. In this case since the place of the service task is defined by the model, data export to public domain (open data) will happen as soon as process execution reaches this task. Downside of this approach is that if used unproportionally, for non-principal activities, it tends to clutter the model, making it hardly readable. Additionaly, it is tightly coupled with implementation class.

Using event listeners on certain elements of the model – this approach does not add any visible elements

to the process model. Although this may be the positive side, since the model is not expanded by additional elements, it is at the same time also it drawback. This implementation is not comprehensible just from viewing the model, and for unfamiliar process designer/developer it may take more time to identify all points at which certain actions are performed. Additional annotation on the model may help remedy this deficiency (fig 2).

In this case information about what needs to be done is attached as specification of listener attached to an element in the model or the process model as a whole.

In this case we have different events to listen for:

- for process, automatic tasks, gateways: start, end
- for user tasks: start, assignment, complete, all
- for transitions: take

As with the service task implementations, listener class needs to be specified.

Positive side of this approach is that it is very flexible in regarding the moment in which action will be performed. In our case

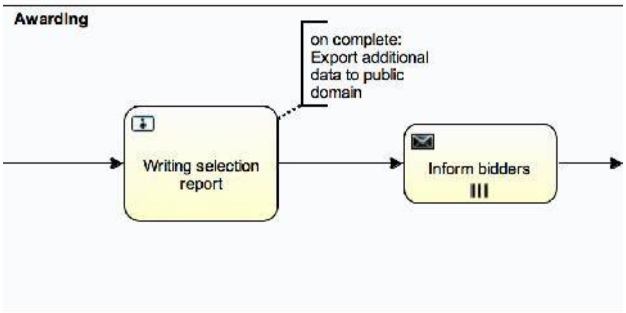


***Figure 2**. Annotating additional actions performed on task completion*

Using explicit events

BPMN defines events as basic concept. Events are concept used to enable process to communicate with its environment. Process may be listening for an event (in this case process model will contain catching event), meaning the event will be triggered by outside source, or process may be the one responsible for creating the event (throwing event). Events may be start, end and intermediate. Furthermore there are different types, in correspondence with the nature of event.

In this approach process model is amended by intermediate throwing events - fig 4, and possibly boundary events. When process execution reaches the intermediate event it will generate event to the execution environment. Other processes or parts of the same process may be registered to listen for certain type of events. This approach provides additional flexibility since it may be possible to allow multiple listeners to react to the same event.

BPMN allows different throwing events, but for this purpose signal and message throwing event may be of interest. Difference between message and signal event is subtle but important, while message events must be specifically aimed at certain receiver, signaling event is more general purpose, signal is emitted to anyone capable of receiving it, making it more versatile.

Activiti modeler does not support using message throwing event – its implementation is covered by signaling throwing event.

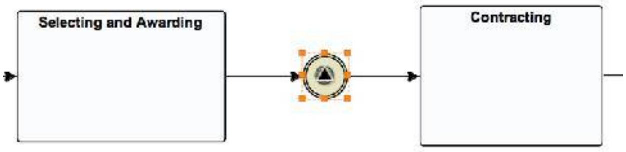


***Figure 3**. Intermediate signaling event as trigger source for data export*

Usage of event based triggering for publication of data is appropriate whenever the task is not considered as principal task required for process model.

Another approach would allow for using combination of intermediate events and boundary events. Boundary events allow for task or sub-process to listen and react to some events that may happen during its execution. It allows for alternative (or additional path, if event is non canceling) of process. This approach proves to be useful when multiple outcomes may result during the sub-process execution.

Similar to previous solutions a listener class must be registered to receive certain type of signal.

All the solutions use delegation principle to accomplish their goals. Basic principle is relaying on process engine to signal the data export module that it is appropriate moment to export certain data.

In test model all principles has been used, but in later refinement, model was streamlined to use service tasks for critical activities, and signaling events for non-critical, while event listeners attached to tasks and transitions has been removed. Primary reason for such decision was to make model explicit in regards to the points of data export. Data export was implemented by exporting to external database. Nevertheless, any other data export is easily achievable, once the data is extracted form process engine at appropriate moment.

By implementing these steps, data export from the running process has been brought under its control. Changes introduced to the model had no affect on the way process is executed form the users point of view. This way introduction of data export was transparent to the users. And since BPM system is used to control the process execution, it was later easy to adapt model and rearrange elements related to data export to better fit the process model.

One obvious improvement may be implemented - creation of generalized process definition specifically aimed at data export. This process definition could then be used as *called activity* in different processes as a standardized pattern for data export.

V. CONCLUSIONS

Business process management systems are common solution in enterprise systems. Their primary goal is to enable deployment, enactment and monitoring of processes and enhance process productivity. They accomplish this goal by implementing explicit process

models that are used to steer the process execution during enactment.

Open data, as a recent trend, offers perspective of free access and free usage of vast amount of data. Main sources of open data could be government institutions. The driving force for opening data stores as open data is to enforce transparency and to promote possible development of new services based on open data access. In recent time, local governments have taken lead in offering their data as open source. This gave rise to development of different apps aimed at helping citizens in various areas of life.

Open data may be sourced from different kind of information systems, but it is almost inevitable that some data will be extracted from BPM systems deployed in various enterprise systems. Gathering data and publishing it to open domain is not always straightforward task, it often requires additional effort from IT departments. Often special care needs to be given not only to the problem of data that should be published, but also to ensure that data is published in appropriate moment.

If BPM systems are used as data source for open data, it is only natural to embed data export capabilities in the process models. This way process may control which data and when it should be made publically available.

This approach, although it requires additional effort to adjust process models, simplifies later tasks of data exposure by reducing need to gather data on different storage, and in process execution history tables. Additionally it ensures that required data is available when certain conditions in the process execution are reached.

In this paper an approach to embedding data extraction features as integral part of process model has been discussed. Several different options, possible in BPMN, and available in current BPM systems have been employed and discussed.

Further enforcement of this principle (of process controlling data export) may be achieved by creating a standard *called activity*, that would be used whenever data export is needed from running process.

References

[1] Open data handbook, accessible at http://opendatahandbook.org/guide/en/what-is-open-data/.

[2] Business Process Model and Notation, Specification, Object Modelling Group, http://www.bpmn.org/, visited January 2016.

[3] ISO/IEC 15944-8:2012, Information technology — Business Operational View, accessible at: https://www.iso.org/obp/ui/#iso:std:iso-iec:15944:-8:ed-1:v1:en

[4] Nikola Vitković, Mohammed Rashid, Miodrag Manić, Dragan Mišić, Miroslav Trajanović, Jelena Milovanović, Stojanka Arsić: "Model-Based System for the creation and application of modified cloverleaf plate fixator", in Proceedings of 5th International Conference on Information Society and Technology (ICIST 2015), p.22 – 26, Kopaonik, Serbia

[5] Miroslav Zaric, Milan Segedinac, Goran Sladic, Zora Konjovic, "A Flexible System for Request Processing in Government Institutions", ACTA POLYTECHNICA HUNGARICA, vol. 11, br. 6, p. 207-227, 2014.

[6] Miroslav Zaric, Zoran Miškov, Goran Sladic: "A Flexible, Process-Aware Contract Management System" in Proceedings of 5th International Conference on Information Society and Technology (ICIST 2015), p.22 – 26, Kopaonik, Serbia

[7] Activiti BPM Platform, available at: http://activiti.org/

[8] Aalst, W. van der, Weijters, A., & Maruster, L.: "Workflow Mining: Discovering Process Models from Event Logs", IEEE Transactions on Knowledge and Data Engineering, 16 (9), 1128–1142 (2004)

[9] Agrawal, R., Gunopulos, D., Leymann, F., "Mining Process Models from Workflow Logs", in: Advances in Database Technology - Pro- ceedings of the 6th International Conference on Extending Database Technology, ed. Fèlix Saltor, Isidro Ramos and Gustovo Alonso, Valencia, 23-27 (1998).

[10] Melissa Lee, Esteve Almirall, And Jonathan Wareham: "Open Data and Civic Apps: First-Generation Failures, Second-Generation Improvements", Communications of the ACM, vol. 59, no. 1, january 2016

[11] Mathias Weske: "Business Process Management: Concepts, Languages, Architectures", 2nd ed. 2012, XV, 403 p. 300 illus. Springer-Verlag Berlin Heidelberg 2012, Springer Link

[12] Open Contracting Partnership, "The Open Contracting Data Standard", http://standard.open-contracting.org/, accessed: January, 2015.

[13] Michael zur Muehlen: "Process-driven Management Information Systems - Combining Data Warehouses and Workflow Technology", Proceedings of the Fourth International Conference on Electronic Commerce Research (ICECR-4), Dallas, TX, November 8-11, 2001., pp. 550-566

The Development of Speech Technologies in Serbia Within the European Research

Nataša Vujnović Sedlar*, Slobodan Morača*, Vlado Delić*
* Faculty of Technical Sciences, Novi Sad, Serbia
natasav@uns.ac.rs, moraca@uns.ac.rs, vdelic@uns.ac.rs

Abstract—**This paper presents potentials, needs and problems of creating an adequate research area for the development of speech technologies in Serbia within the European Research Area. These technologies are a very important part of the strategy Europe 2020, because of Europe's awareness that speech technologies are one of the missing parts that will help Europe to build a unified digital market, overcome language barriers and increase the fluctuation of employees. The authors of the paper made a comparative analysis of the European actions in the European Research Area and projects in Serbia dedicated to the development of speech and language technologies. The analyses shows how European programs influence the development of speech technologies in Serbia.**

Key words - **European Research Area, European programs, Europe strategy, speech and language technologies**

I. Introduction

One of the characteristics of the European Union, as a multicultural and multinational community, is the diversity of languages that are spoken within its border. From the beginning, European leaders have been aware of that and presented the European languages as an inalienable part of its cultural heritage. Besides these, there is also a political reason why Europe keeps language diversity as an integral part of its policy since its establishment [1]. However, the linguistic diversity requires a significant investment and large budget funds have been spent on translation services only to ensure the availability of formal documents in 24 [2] official languages. On the other hand, it is of the great importance to ensure access and usability of information in the mother tongue languages to all European citizens.

Language diversity in Europe is supported by various European policies. In the line with the Multilingual Policy [3], Europe supports various activities in education, research programs, language learning and development of language and speech technology. The development and implementation of speech and language technologies, which should break languages barriers between European citizens and push raising European economy, have been helped through the Information Society Policy and the Research and Innovation Policy. Regarding to language and speech technologies, the Policy of Information Society is in correlation with the Multilingual Policy [4], or it can be said that these two policies have the same aim to provide the availability of content in different languages at all European languages and wider across all communication channels and sources of information.

European Commission presume that language and speech technology will bring new quality of social and business life for citizens of Europe and because of that includes development of these technologies in some of the European programmes to ensure market introduction technology readiness level while these technologies should have great influence at business collaboration as well as staff fluctuation, better knowledge sharing etc. [5].

That's why Europe has already invested significant resources in the development of these technologies through the Framework Programme and the entire European Research Area and still does it. The subject of this paper is to make correlations between the projects for development of speech and languages technologies in Serbia and project supported in the European research area from this topic and to understand the impact that the European research area has on development of these technologies in Serbia. Also, this paper indicates the weak points of the Serbian project and suggests improvements for better involvement of the Serbian projects in European research area.

II. Language and Speech Technologies

The human language appears in both oral and written form. Speech as an oral form of the language is the oldest and most natural way of verbal communication. On the other hand, text like written form of language allows storage of human knowledge and keeps it from forgetting. Language technology deals with language aspects - sentences, grammar and meaning of sentences in the field of information communication technology are important for the development of speech technologies and word processing. It is a branch of ICT technology that is based on knowledge of linguistics and other interdisciplinary fields of science like mathematics, computer sciences, telecommunications, signal processing, and others.

A. Language, speech and IT technologies

Voice machines have intrigued humans for a long time. According to the current knowledge, the first attempts of creating such machines dating from the 13th century, when the German philosopher Albertus Magnus and the English scientist Roger Bacon [6, 7] made metal heads that produce voice separately from each other. Todays in the era of digitalization, the interest of scientists for the realization of the machines that will understand, recognize speech and translate it from one language to another is much bigger and in relations with the needs of a modern person.

Nowadays, world characterize an abundance of information that is important for different areas of people's lives, not just business life, but also private and social aspects of life. Beside that it should be stressed that information-exchanging channels have been significantly expanded in comparison with the time of first produced voice machines. Information has a global character; there are no limits or borders. Language, as a most natural holder of information, has found itself in a new context, a new environment, which requires study of all aspects of language from the new perspective, the information technology perspective.

Language technology, or as it is often called human language technology (HLT), is the information-communication technology that deals with language, the very complex medium for communication, in a new digital environment. Language technology provides great support to the development of speech technology and text processing technology.

It should also be noted that speech technologies and text processing are intertwining and overlapping with other information communication technologies. For example, multimedia presentation of information collects pictures, music, speech, gestures, facial expression and other forms of information presentations. All that defines the meaning of spoken text and because of that these technologies cannot be studied separately.

In the domain of language technology, researchers are engaged in different research fields such as automatic translation, Automatic Text Summarization, automatic text analysis, optical character recognition, spoken dialogue, speech recognition and speech synthesis. Beside that, researchers have been faced with various problems such as the segmentation of written text, speech segmentation, solving ambiguous meaning of words, syntactic ambiguity solving, overcoming the imperfections of the input data. They also have to take into account the context and the speaker's intentions. One of the biggest problems here is the dependence of these technologies on the language. There is a large degree of inability of applying methods developed for one language to other languages.

B. Development and implementation of language and speech technologies in Europe

The development and implementation of these technologies in Europe differs from language to language. META-NET [8], Network of Excellence dedicated to fostering the technological development of a multilingual European information society, has implemented a series of reports entitled Europe's Languages in the Digital Age [9]. The document treated the state of language and speech technologies for 30 languages, whereby the following areas have been observed:

- Automatic translation - also taking into account: the quality of existing technology, the number of covered language pairs, coverage of linguistic phenomena and domains, the quality and size of the parallel corpus, the amount and variety of applications.
- Speech processing - in which is observed the quality of existing speech technologies, domain coverage, the number and size of the existing corpus, the volume and variety of available applications.
- Text analysis - with an emphasis on the quality and coverage of existing technologies in the field of morphology, syntax and semantics, completeness linguistic phenomena and domains, amount and variety of applications, the quality and size of the corpus, the quality and coverage of lexical resources.
- Resources - the quality and size of text, voice and parallel corpus, the quality/coverage lexical resources and grammar.

Although the language and speech technology can solve the complex issue of multilingualism in Europe, its development is uneven and in some languages such as Lithuanian and Irish it is at the beginning. In addition, it should be noted that the automatic translators are [10] tools that will contribute to the unity of the European market, because it attempts to help bridging the language barrier that currently exist, cannot been fully used yet. Europe's awareness of the necessary development of these technologies allocates significant resources for their development through research funds [11]. A great part of the funds for the development of speech and language technology comes from national funds. These national projects generate very good results that should eventually be implemented and upgraded in European programmes.

The development of language and speech technology not only takes place at research institutions, but also at innovative companies, most of which are small and medium enterprises. There are about 500 European companies that actively participate in the development and/or implementation of language and speech technology. Typically, most of them are focused on national markets and are not included in the European value chain.

III. EU SUPPORT TO THE DEVELOPMENT OF LANGUAGE AND SPEECH TECHNOLOGIES

A. EU support through the Framework Programmes

European Union has thrived to support development of languages technologies and to take leading place in that development through various funding programmes, mostly through the Framework programmes. Within the Framework Programmes it can be identified several major research sub-areas of language and speech technologies that are financed:

- Automated Translation;
- Multilingual Content Authoring & Management;
- Speech Technology & Interactive Services;
- Content Analytics;
- Language Resources;
- Collaborative Platforms.

In order to make overview of the language and speech technologies research directions, supported by European Commission through Framework programmes, authors analysed funded projects available in CORDIS base [13]. Usually analysed projects did not address only one topic, but more similar ones. Relative percentage participation of each topic in relation to total number of funded projects in the last three Framework programmes is presented in the table below.

%	Automated Translation	Multilingual Content Authoring & Management	Speech Technology & Interactive Services	Content Analytics	Language Resources	Collaborative Platforms
FP5 - IST	11,34	41,24	62,89	30,93	32,99	10,31
FP6 - IST	7,41	24,07	79,63	53,70	33,33	5,56
FP7 - ICT	40,00	21,54	29,23	23,08	26,15	15,38

Table 1. Relative percentage participation of each topic

The Automated translation topic, where the projects are mostly devoted to overcoming the machine translation problems and developing different cross-lingual applications, has got up to 40% funding rate in 7th Framework programme, while in 5th and 6th Framework programme funding rate of the approved projects for this topic was 11,34% and 7,41% respectively.

The funding rate for the topic called Speech technology and interactive services has been significantly decreased in last Framework programme. Until the 7th Framework programme it was the most supported topic, but the Automatic translation has taken the lead. In 5th Framework programme this topic was supported in 62,89% projects, in 6th Framework programme in 79,63% projects, while in 7th Framework programme only 29,23% projects with this topic was funded.

Topic Multilingual Content Authoring & Management as a topic of interest for European Union after 5th Framework programme has dropped double of the over all approved projects. Its participation percentage in 5th Framework programme was 41,24% of all funded projects, and then EU support to this topic dropped at 24,07% in the 6th Framework programme. Similar percentage, 21,54%, remained in 7th Framework programme. It is important to stress that in 4th and 5th Framework programme this topic has got special support through specific programmes Multilingual Information Society (MLIS) [14].

Content Analytics as a topic of interest for the development of language and speech technologies, through these three framework programmes, reached its peak in 6th Framework programme, when percentage of funded project related to this topic was 53,70%. Peak in the 6th Framework programme was reached owing to the calls of Semantic-based Knowledge Systems and Proactive Initiatives.

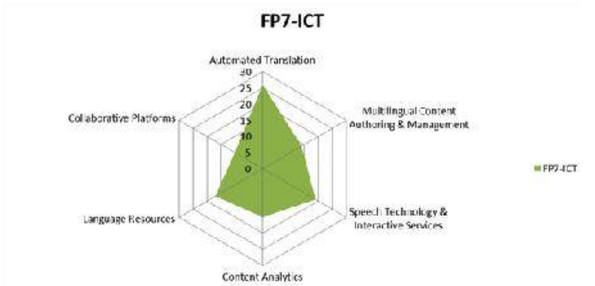


Figure 1. Representation of topics in Framework programme 7

The Language Resources topic has been uniformly supported during each of the three observed framework programs. Nevertheless funding percentages were reached 32,99% in the 5th, 33,33% in the 6th, and up to 26,15% in the 7th Framework programme. Main aim of European Commission by supporting this kind of project was to overcome evident lack of language resources for European languages.

Projects Collaborative Platforms have goal to improve further research to ensure leading position in this domain, or to prepare research community for the next research phase related to the new development directions and the problems linked to multilingualism of Europe and usage of new resources. Percentage partake of this topic in the 5th Framework programme was 10,31%, in the 6th 5,56%, and in the 7th 15,38%.

The analysis represents a big change in direction of support research topics related to language and speech technologies. The primate status of the topic Speech technology and Interactive Services has now been overthrone by the topic Automated Translation. This trend was detected also in the calls established in the Work programme for 2014 and 2015 in new Framework programme Horizon 2020[15]. Enormous change can be seen in the Work programme for 2016 and 2017, where calls explicitly dedicated to the language technologies cannot be registered like in previous calls.

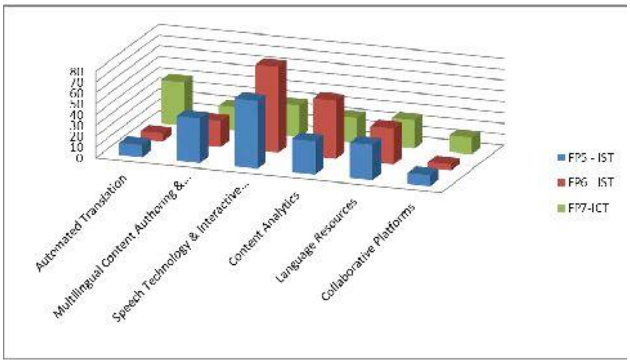


Figure 2. Representation of topics in Framework programmes

IV. The Language and Speech Technologies in Serbia

The development of language and speech technologies in Serbia does not have a long history. Some of the first attempts that deal with computer linguistic were made in 1970's when the first software tool for discovering and correction of spelling mistakes was made. Besides this tool a more recent software RAS has been introduced. The software has processed Serbian text on the computer, which offers correction of the text: dividing words into syllables, conversion of the codes pages and sorting out punctuation.

It should be noticed that the development of language technologies in Serbia is mostly scattered in research organisations. The development is taking place at the faculties of Philosophy as well as at Mathematical and Electrotechnical faculties of the University of Novi Sad and Belgrade. There is a certain level of cooperation between language technology research groups established at these faculties, but with evident problem of correlation among the achieved research foregrounds. The stated cooperations are still insufficient to make a serious progress in this area.

In the area of speech technologies the greatest accomplishment has made by the Faculty of Technical Sciences of the University of Novi Sad. They have managed to put in practice speech synthesis and speech

recognition for Serbian, Macedonian and Croatian languages. Besides that, the Faculty has made accentuation morphological dictionaries for Serbian and Croatian language, for over 4 and 3 million words respectively. The Faculty cooperates closely with AlfaNum company which is making a transfer of these technologies for commercial usage.

V. Language and Speech Technologies for Serbian Language in ERA

The directions of development of language and speech technologies in Serbia match the four topics of interest to the European Union, and to the Speech Technology and Interactive Services, Content Analytics, Language Resources and Collaborative Platforms. The acceptable level of development compared to the technology of English language, as a reference point against which to measure the development of language and speech technologies to other languages, is made only in the area of synthesis. The level of development of tools and resources for Serbian language is quite low, and the volume and quality of text, voice and parallel corpus, and the quality of lexical resources and grammar should seriously be increased. Serbia when it is still working on developing core technologies, and scientific developments outside the domain of relatively rare.

In all of this it should be taken into consideration that the language and speech technology developed in Serbia thanks to national research programmes. Support the development of core technologies through the European Programme occurred in the period when Serbia could legitimately participate in these programs. Scientists from Serbia who work in this specialized field currently take part in collaborative platforms within the Framework Programme, EUREKA program, COST and SCOPES program. Since 2007 there has been a group from the Faculty of Technical Sciences that has submitted seven applications for the participation in the framework programs mainly from the topic of Speech technology and interactive resources, where the two received a score of 13 and 13.5 points, Serbia failed to take significant participation in the ERA.

VI. Discussion

Since 2007, beginning of the Seventh Framework Programme, Europe has changed the priority of topics funded through the Framework Programme. This change of course was retained in the new Framework Programme Horizon 2020. At the moment Serbian researcher have to make an extra effort and try to reorganize the project development of language and speech technologies in order to leverage the research policy in Europe in this field. Also, Serbian researchers have recognized importance of their inclusion and chance for the funding of further research through financial support of the framework programme. In addition they find the exchange of experience and knowledge with researchers from abroad necessary for their further research work.

To realize high quality alignment directions of the development of language and speech technologies Serbia with Europe, it is necessary:

- Increase the amount and quality of text, voice and parallel corpus, and the quality of lexical resources and grammar
- Bring together interdisciplinary teams to improve the situation
- Develop co-ordination of research activities in this field in order to reduce the gap in the development of speech and language technology for Serbian language compared to English, German and French.

Serbia should also seek their chance at commercializing existing technologies whether through direct marketing of advanced technology on the market or through other program of the European Commission funded the research and development of technologies to be quickly commercialized.

Acknowledgment

The presented research was performed as a part of the project “Development of Dialogue Systems for Serbian and Other South Slavic Languages” (TR32035), funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

[1] Gianni Lazzari (2006), "*Human Language Technologies for Europe*", available at: http://cordis.europa.eu/documents/documentlibrary/90834371EN6.pdf

[2] http://ec.europa.eu/languages/policy/language-policy/official_languages_en.htm

[3] Council of the European Union (2008), "*Council Resolution on a European strategy for multilingualism*", available at: http://www.ceatl.eu/wp-content/uploads/2010/09/EU_Council_multilingualism_en.pdf

[4] European Commission (2010), "*A Digital Agenda for Europe*", available at: http://ec.europa.eu/information_society/digital-agenda/publications/.

[5] The Language Rich Europe Consortium (2011), "*Towards a Language Rich Europe*". Multilingual Essays on Language Policies and Practices, British Council, available at: http://www.language-rich.eu/fileadmin/content/pdf/LRE_FINAL_WEB.pdf.

[6] David Lindsay (1997), "*Talking Head*" Invention. & Technology, pp. 57-63

[7] http://www.haskins.yale.edu/featured/heads/simulacra.html

[8] http://www.meta-net.eu/

[9] http://www.meta-net.eu/whitepapers/overview

[10] LT-Innovate (2013), "*Status and Potential of the European Language Technology Markets*", available at: http://www.lt-innovate.eu/

[11] http://cordis.europa.eu/fp7/ict/language-technologies/

[12] COM(2005) 596 - The 2005 Commission communication A new framework strategy for multilingualism

[13] http://cordis.europa.eu/fp7/ict/language-technologies/portfolio_en.html

[14] Kimmo Rossi (2013), "*Language technology in the European programmes and policies*", 6th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznań, Poland, available at: http://www.ltc.amu.edu.pl/pdf/abstract-kimmo-rossi.pdf

[15] http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/main/h2020-wp1415-leit-ict_en.pdf

DSL for web application development

Danijela Boberić Krstićev[*], Danijela Tešendić[*], Milan Jović[*], Željko Bajić[*]
[*]Faculty of Sciences, University of Novi Sad, Serbia
dboberic@uns.ac.rs, tesendic@uns.ac.rs, milan.jovic@dmi.uns.ac.rs, zeljko.bajic@dmi.uns.ac.rs

***Abstract* – This paper addresses some issues of generative programing. The main aim of this paper is to present a DSL designed for developing web applications. It will be possible to easily generate whole web application by defining its functionalities in proposed DSL. Also, we will present possibilities to implement such DSL in Scala language. In addition, advantages of Scala language for this purpose will be discussed.**

I. INTRODUCTION

Domain specific language (DSL) is a programming language that's designed to solve a specific problem from some domain and it has limited expressive syntax. DSL offers a vocabulary that's specific to the domain it addresses. From the standpoint of an end user, domain specific language is a solution to his realistic problems. DSL provides a clear interface for all the requirements of the end user and should be intuitive for the user who is domain expert rather than programmer. Because nonprogrammers need to be able to use them, DSLs must be more intuitive to users than general-purpose programming languages need to be. Also, syntax of general-purpose programming languages is much richer than the one of DSL. DSLs are widespread no matter whether we brand them as DSLs or not. Some of the most commonly used DSLs are SQL, LaTeX, Ant, CSS or similar to them.

DSLs are mostly used to bridge the gap between business and technology and between stakeholders and programmers. DSLs speaks the language of the domain and provide a humane interface to target users. Domain of a DSL can also relate to specific field of software development. Software developers don't have to take a part in development of every aspect of software. For example, web developers don't have to be expert in parallelism and architecture to properly optimize software for modern heterogeneous hardware. In this case, concepts of parallel programing can be abstracted by DSL. This DSL can be further used by high-level developers.

Usage of DSLs can increase programmer productivity by providing extremely high-level abstractions tailored to a particular domain of software development. If we see development of an information system as a domain, we can develop DSLs, which will in an easy and simple way provide possibility to specify and generate some parts of system. During the development of various information systems it has been noticed that a large amount of time is spent on initial setup and development of basic functionalities. Furthermore, almost all information systems have a lot of similarities in their architecture. For example, if we talk about data layer, systems usually have database and persistence layer with entities, their relations and code lists. Usually, for all entities, it is necessary to provide input and binding data, modifying existing data and searching by various criteria. In the case of large systems, which contain over hundreds of entities, this work takes an enormous amount of time. Similar examples can be found in other layers of systems. Because of these uniformities in structure, we can talk about automation and code generation of some parts of information systems. Those features can be achieved by developing appropriate DSLs.

These DSLs can be useful to speed up initial setup of projects. By using DSLs, we can specify basic system functionalities and automatically generate programming code. This programming code represents skeleton code which can be further enhanced and refined. Also, these DSLs can be used for rapid development of system prototypes used during requirements analysis and design phases. Usage of prototypes simplifies gathering and discovery of requirements from stakeholders and improves quality of final software solution.

This paper addresses some issues of generative programing. The main aim of this paper is to present a DSL designed for developing web applications. It will be possible to easily generate whole web application by defining its functionalities in proposed DSL. Also, we will present possibilities to implement such DSL in Scala language. In addition, advantages of Scala language for this purpose will be discussed.

The presentation of this paper proceeds in 6 sections. At the beginning of the paper, we give brief overview of recent research in the field of DSLs used for specifying and implementing an information system as well as DSLs written in Scala language. Discussion on different approaches for implementing DSLs in Scala in given in Section 3. We continue by describing syntax of proposed DSL. Conclusion remarks as well as some plans for further research are presented in the last sections.

II. RELATED WORK

When an application and its components have well-defined structure and behaviour, developing process can be automated. Programs that perform that automation are called application generators. Application generators can be seen as compilers for DSLs used to define application's functionalities. Application generators supports code generation of applications with similar architecture, but with different semantics. Compared to traditional software development, generators offer a better potential for optimization and can provide better error-checking.

One approach to develop a web application is to start from a formal specification of an application and to generate application code. In this case, specification is used as an input for application generator. This approach is known as Model Driven Development. There are a lot of tools supporting this paradigm and usually they have graphical environment for application's specification. One example of those tools is AndroMDA [1] which transforms UML classes, use cases and state chart diagrams into deployable components for chosen platform. It can

generate code for multiple programming languages such as Java, .NET, PHP and many others. Similar tool is WebRatio [2] which uses Interaction Flow Modelling Language for modelling user interface of an application and generates fully functional application.

Also, it is possible to define database model of an application as a starting point for further code generation of the application. The authors of the paper [3] describes framework that fetches database metadata via JDBC and converts it into an XML document which is further transformed into HTML, XHTML or any other sort of web page using an XSLT. The same approach is used by ASP.NET Dynamic Data [4], but in comparison with previous solution it is more customisable. There are various templates of web pages that can be selected as the starting point of the new web application.

As we previously mentioned, development of a web application is mostly straightforward process. There have been many efforts that have focused on automation of that process by defining specific DSLs. Good example is WebDSL [5]. WebDSL is a DSL for developing dynamic web applications. Applications written in this DSL are translated to Java web applications. The WebDSL generator is implemented using Stratego/XT, SDF, and Spoofax. In addition, DSLs are used for creating mobile applications. MobiClaud DSL [6] is DSL designed for generating hybrid applications spanning across mobile devices as well as computing clouds. MobiCloud DSL closely resembles the MVC design.

Choice of general-purpose language for writing DSLs is substantially determined by language's built-in support. Ruby is an example of languages suitable for writing DSLs. Ruby on Rails [7] is the most popular DSL written in Ruby and designed for web development. Also, authors of the paper [8] combined several other DSLs written in Ruby to create web application. Scala is another language with good built-in support for writing DSLs. There is a lot of examples of DSLs written in Scala. Regarding that Scala allows programmers to create natural looking DSLs, we choose it for implementing DSL presented in this paper.

III. BUILT-IN SCALA SUPPORT FOR DSL DEVELOPMENT

DSLs can be classified regarding the way they are implemented. Generally speaking, DSLs can be divided into internal and external DSLs. Internal DSLs are also known as embedded DSLs because they’re implemented within a host language. Host language has to be flexible and to offer the possibility of extension. An internal DSL uses same syntax as a host language. An internal DSL program is, in essence, a program written in the host language and uses the entire infrastructure of the host but it is limited to the syntactic structure of the host language.

A DSL designed as an independent language without using the infrastructure of an existing host language is called an external DSL. It has its own syntax, semantics, and language infrastructure implemented separately by the developer. External DSL is developed as an independent language with its own syntax. Unlike internal, external DSLs provide a greater syntactic freedom and the ability to use any syntax. With an external DSL, it is necessary to learn about parsers, grammars, and compilers, while implementing an internal DSL is just like writing API using facilities of a known language.

Before we delve further into describing syntax of our DSL, it is worth addressing the advantages of Scala language for developing embedded, as well as external DSLs. A number of books and papers have already discuses Scala’s features and support for developing DSLs [9-12] and in this section we will summarise that. Scala is an object-functional language that runs on the JVM and has great interoperability with Java. Scala provides functional, as well as object-oriented, abstraction mechanisms that help in design of concise and expressive DSLs.

When an external DSL is designed, usually some external tools, like ANTLR are used. Based on grammar of external DSL, ANTLR will generate all necessary language implementation infrastructure. Another way to develop external DSLs is to use *parser combinators*. Scala offers an internal DSL which consist of library of *parser combinators* (functions and operators) that serves as building blocks for parsers. That internal DSL is used for designing external DSLs without usage of any external tools.

Although Scala supports development of external DSLs, it has even better support for developing embedded DSLs. First of all, Scala has a nice, concise and flexible syntax.

Scala provides flexibility through infix methods which allow to write *a x*, instead of *a.x*, where is *x* method of *a*. Also, some syntactic constraints like semicolon and parentheses are optional in Scala. Also, case classes in Scala enable usage of a shorthand notation for the constructor. It is not necessary to specify the keyword *new* while instantiating an object of the class. In addition, Scala’s type inference is another factor that contributes to its conciseness. It is, for instance, often not necessary in Scala to specify the type of a variable, since the compiler can deduce the type from the initialization expression of the variable. Also return types of methods can often be omitted since they corresponds to the type of the body, which gets inferred by the compiler.

Next, Scala is object-oriented language in pure form, which means that every value is an object and every operator is method call. For example, expression *1 + 2* is actually invocation of method named + defined in class *Int*. Control flow statements are also expressed in terms of method calls. For example, the expression *if (c) a else b* is defined as the method call *__ifThenElse(c,a,b)*. As everything is method call in Scala, any built-in abstraction may be redefined by a DSL, just like any other method.

In Scala, functions are first-class values, and a function as an argument can be passed to yet another function. Also a function can return another function. Scala supports so called partial functions. Using partial functions, entire parameter list can be replaced with underscore mark. For example, rather than writing *println( ),* it is possible to write *println _*. It is possible to leave out the last argument of a function by declaring it to be implicit. The compiler will look for a matching argument from the enclosing scope of the function. Existing libraries in Scala can be extended without making any changes to them by using implicit type conversions. The implicit conversion function is automatically applied by the compiler.

All those features offer possibility to write DSLs in Scala which closely resemble natural language and they are sufficient for writing DSLs as pure libraries. Also, Scala offers some more options for developing embedded DSLs. Lightweight modular staging (LMS) [13,14] is a means of building new embedded DSLs in Scala and

creating optimizing domain-specific compilers at the library level. LMS generates code by build an intermediate representation of a program and translating nodes to their corresponding implementation in the target language. Currently, LMS generates Scala, C++ and CUDA code. A number of DSL projects [15- 17] are built in Scala by using LMS.

IV. FEATURES OF DSL FOR DEVELOPING WEB APPLICATIONS

In linguistics, of a particular language, syntax examines the rules that determine how words are combined into sentences. In computer science, the syntax is a set of rules that define the combination of symbols in order to get the correct language structure in a given language. If the syntax would not exist, we would have insignificant structures and we would be without any possibility of validating the language.

In this section, we present features of DSL for developing web applications. Proposed DSL is minimalistic and it strives to take after natural language. This DSL bears a resemblance to entity–relationship model (ER model). It is used to describe entities of an application. Entities have attributes of different types and among entities can exist relationships. Those concepts in the context of our DSL are described in following subsections. Based on entities' definition, DSL will generate database model, application's persistence data layer with CRUD operations on those entities, business layer with controller classes and presentation layer with web pages.

A. Dsl Data Types

Definition of data types used in this DSL is given in this subsection. This DSL supports following data types:

- *text[n]* - Represents the text data type of arbitrary length n. If the length of the text is omitted, the default value is 100. This type of data is equivalent to String in the programming language Scala.
- *number[n]* - This data type represents a numeric value. The parameter n specifies the number of decimal units and when it is present type is interpreted as Double. If the parameter n is omitted then it will be interpreted as Integer in Scala.
- *date* -It describes date and time values and it is equivalent to data type Date in Scala.

These data types are used in the construction of other elements of DSL.

B. Entity

An entity is an abstraction of some aspect of the real world that can be uniquely identified and is capable of an independent existence. Keyword *entity* is used to describe a group of data that has common attributes. In order to define an entity it is necessary to declare name of the entity and list of its attributes. Every attribute has name and data type separated by column. An example of an entity is shown in Listing 1, where we have an entity *Person* which has attributes: *Name*, *Surname* and *Date of birth*. Data type of attributes *name* and *surname* is *text* (implicit length is 100 characters), and data type of attribute *Date of birth* is *date*.

```
entity("Person", "Name‶: text, "Surname‶:
text, "Date of birth‶: date)
```

Listing 1. – Entity example

A. Relationships

Relationships capture how entities are related to one another. In our DSL, keyword *relationship* is used to describe the relationship between entities. This construction consists of entities' names and definition of visibility among them. Visibility can have value from the following set: *include*, *field*, *none*. Value *include* is interpreted that instance of the first entity has connection with more instances of the second entity. Value *field* means that instance of the first entity has connection with only one instance of the second entity. Value *none* means that instance of the first entity has no connection with any instance of the second entity. An example of this construction is given in Listing 2. that describes relationship between entities *Artist* and *Album*. Meaning of this relationship is that entity *Artist* references several entities *Album*, and that entity *Album* reference only one entity *Artist*.

```
relationship("Artist", "Album", {include,
field})
```

Listing 2. – Relationship example

B. Relationships With Attributes

Like entities, relationships can also have attributes and this concept is supported with our DSL. This is an extension of *relationship* construction. To support this concept, firstly, it is necessary to define additional entity that will contain the new attributes associated to relationship. The name of that entity is further used in relationship construction. In this construction, name of the new entity is grouped with definitions of visibility among entities. Definition of visibility has the same meaning as in relationship construction without attributes. An example of this kind of relationship is shown in Listing 3. In this example we have a new entity *Song award* which contains information when the award was given and this entity is used in relationship between entities *Song* and *Award*. Visibility between entities *Song* and *Award* is defined with values *include* and *none*. This means that *Song* reference several entities *Award* and *Award* doesn't have reference to entity *Song*.

```
entity("Song award", "Award received‶: date)
relationship("Song", "Award", {"Song award",
include, none})
```

Listing 3. – Relationship with attributes example

C. Codebooks

Codebooks are simple entities that don't have relationship with other entities. Other entities just use values from codebooks. Keyword *codebook* is used to define this concept in our DSL. Definition of codebook contains name of codebook and its attributes, which is similar to entities' definition. Relationship with entities is defined by relationship construction but visibility is defined only for entity's side. An example of this concept is presented in Listing 4. Meaning of this example is that entity *Artist* has connection to one instance of codebook *City*.

```
codebook("City", "Name‶: text)
relationship("Artist", "City", field)
```

Listing 4. –Codebook example

D. Example Of Web Application Specification

In order to better explain usage of proposed DSL, an example of a specification of an simple web application in

that DSL is presented in Listing 5. This application manages music artists and their work and tracks awards which they have received. Artists and their work are defined with entities *Artist*, *Album*, *Song* and corresponding relationships. Regarding the fact that artist can represent musical band, membership to the band is defined with entities *Artist*, *Person*, *Member* and corresponding relationships. Awards received by artists are defined with entities *Artist*, *Award*, *Song award* and corresponding relationships.

```
entity("Artist", "Name": text[60], "Formed":
number, "Active
till":number,"Website":text[100],
"Description": text[2048])
entity("Album", "Title": text[100],
"Production": text[100], "Released": date)
entity("Song", "Song number": number, "Title":
text, "Time": number, "Song text": text[500])
entity("Person", "Name": text, "Surname":
text, "Date of birth": date, "Gender": text)
entity("Award", "Name": text, "Establish
date": date)
entity("Member", "Role": text[100], "Member
since": date)
entity("Song award", "Award received": date)
relationship("Artist", "Album", {include,
field})
relationship("Album", "Song", {include,
field})
relationship("Artist", "Person", {"Member",
include, field})
relationship("Song", "Award", {"Song award",
include, none})
relationship("Artist", "Genre", include)
relationship("Artist", "City", field)
codebook("City", "Name": text)
codebook("Genre", "Name": text)
```

Listing 5. – Web application specification

V. ARCHITECTURE OF GENERATED WEB APPLICATION

The first step in developing DSL is to understand problem domain and according to that to define vocabulary and grammar of new DSL. In the next step it is necessary to choose host language as well as technologies which will be used in implementation of the main concepts of DSL.

In this paper, we presented syntax of our DSL for developing web applications and the next phase in our research is to implement it. In this section we are giving a brief overview how we plan to do that.

Taking into consideration advantages of Scala language for development of DSLs, we chose Scala for implementation of our DSL. We will implement our DSL as embedded DSL. We are planning to embed our DSL in Scala using Lightweight Modular Staging, which will provide a common intermediate representation and basic facilities for optimization and code generation.

Also, generated web applications will be based on Scala web technologies. Generated web applications will have multi-tier architecture consisting of a database layer, persistence layer, business layer, service layer and presentation layer. We will use PostgreSQL database management system for storing data. Persistence layer will be implemented using framework Slick [18] which is a modern database query and access library for Scala. Other layers of application will be implemented using Play web framework [19] which follows MVC architectural pattern.

VI. CONCLUSION

In this paper we presented features of DSL for developing web applications. Proposed DSL is minimalistic and it strives to take after natural language. This DSL bears a resemblance to entity–relationship model. It is used to describe entities of an application. Based on entities' definition, DSL will generate database model, application's persistence data layer with CRUD operations on those entities, business layer with controller classes and presentation layer with web pages.

Presented DSL will be implemented as an embedded DSL in Scala. Paper also gave overview of Scala's built-in support for developing DSLs.

ACKNOWLEDGMENT

This work is partially is supported by the Swiss National Sceince Foundation (SCOPES project IZ74Z0 160453/1, Developing Capacity for Large Scale Productivity Computing).

LITERATURE

[1] AndroMDA Model Driven Architecture Framework, http://www.andromda.org/

[2] WebRatio, http://www.webratio.com/

[3] Elsheh, Mohammed M., and Mick J. Ridley. "Using database metadata and its semantics to generate automatic and dynamic web entry forms in." In *Proceedings of the World Congress on Engineering and Computer Science 2007 WCECS 2007*. 2007.

[4] ASP.NET Dynamic Data, https://msdn.microsoft.com/en-us/library/ee845452.aspx

[5] Visser, Eelco. "WebDSL: A case study in domain-specific language engineering." In *Generative and Transformational Techniques in Software Engineering II*, pp. 291-373. Springer Berlin Heidelberg, 2008.

[6] Ranabahu, Ajith H., Eugene Michael Maximilien, Amit P. Sheth, and Krishnaprasad Thirunarayan. "A domain specific language for enterprise grade cloud-mobile hybrid applications." In *Proceedings of the compilation of the co-located workshops on DSM'11, TMC'11, AGERE! 2011, AOOPES'11, NEAT'11, & VMIL'11*, pp. 77-84. ACM, 2011.

[7] Ruby on Rails, http://rubyonrails.org/

[8] Günther, Sebastian. "Multi-dsl applications with ruby." IEEE software 5 (2010): 25-30.

[9] Odersky, Martin, Lex Spoon, and Bill Venners. Programming in Scala. Artima Inc, 2010.

[10] Ghosh, Debasish. DSLs in action. Manning Publications Co., 2010.

[11] Rompf, Tiark, Nada Amin, Adriaan Moors, Philipp Haller, and Martin Odersky. "Scala-virtualized: Linguistic reuse for deep embeddings." Higher-Order and Symbolic Computation 25, no. 1 (2013): 165-207.

[12] Moors, Adriaan, Tiark Rompf, Philipp Haller, and Martin Odersky. "Scala-virtualized." In Proceedings of the ACM SIGPLAN 2012 workshop on Partial evaluation and program manipulation, pp. 117-120. ACM, 2012.

[13] Lightweight Modular Staging, https://scala-lms.github.io/

[14] Rompf, Tiark, and Martin Odersky. "Lightweight modular staging: a pragmatic approach to runtime code generation and compiled DSLs." Communications of the ACM 55.6 (2012): 121-130.

[15] Brown, Kevin J., Arvind K. Sujeeth, HyoukJoong Lee, Tiark Rompf, Hassan Chafi, Martin Odersky, and Kunle Olukotun. "A heterogeneous parallel framework for domain-specific languages." In Parallel Architectures and Compilation Techniques

(PACT), 2011 International Conference on, pp. 89-100. IEEE, 2011.

[16] Sujeeth, Arvind, HyoukJoong Lee, Kevin Brown, Tiark Rompf, Hassan Chafi, Michael Wu, Anand Atreya, Martin Odersky, and Kunle Olukotun. "OptiML: an implicitly parallel domain-specific language for machine learning." In Proceedings of the 28th International Conference on Machine Learning (ICML-11), pp. 609-616. 2011.

[17] Vogt, Jan Christopher. "Type safe integration of query languages into Scala." PhD diss., Diplomarbeit, RWTH Aachen, Germany, 2011.

[18] Slick, http://slick.typesafe.com/

[19] Play, https://www.playframework.com/

An Approach and DSL in support of Migration from Relational to NoSQL Databases

Branko Terzić, Slavica Kordić, Milan Čeliković, Vladimir Dimitrieski, Ivan Luković,
University of Novi Sad, Faculty of Technical Sciences, 21000 Novi Sad, Serbia
{branko.terzic, slavica, milancel, dimitrieski, ivan}@uns.ac.rs

***Abstract*—In this paper we present a domain specific language, named *MongooseDSL*, used for modeling Mongoose validation schemas and documents. *MongooseDSL* language concepts are specified using Ecore, a widely used language for the specification of meta-models. Besides the meta-model we present the concrete syntax of the language alongside the examples of its usage. This *MongooseDSL* language is a part of the developed *NoSQLMigrator* tool. The tool can be used for migrating data from relational to NoSQL database systems.**

I. Introduction

Relational database management systems are preferred way of storing and managing data in the last few decades. Nowadays, due to the development of technologies, primarily the Internet, there was an increase in the number of different data sources. A lot of data is being generated every second and usually it is unstructured or semi-structured. Requirements for storing and processing such data are beyond the capabilities of traditional relational database management systems. In order to alleviate this problem, NoSQL database systems were introduced comprising a new approach to storing and processing large amounts of data [1]. These systems have built-in mechanisms for processing and analyzing large amounts of data, as well as the ability to save data in various formats, such is JSON. The absence of formally specified database schema in majority of NoSQL systems allows easier handling of variations of input data. This leads to the increase in number of users who are using NoSQL database systems in their applications. Accordingly, the need for reengineering legacy databases and migrating existing data to NoSQL systems is considered as an unavoidable step in such a process.

From the other side, model-driven approaches to software development increase the importance and power of models. Model is no longer just a bare image of a system, taken at the end of design process and used mostly for the communication and documentation purposes. Model-driven software engineering (MDSE) promotes the idea of abstracting implementation details by focusing on: models as first class entities and automated generation of models or code from other models. Each model is expressed by the concepts of a modeling language, which is in turn defined by a meta-model.

A reengineering process, and thus the whole migration processes, can benefit of using meta-models in almost every step. In this paper, we present a part of our research efforts focused on the database reengineering and the data migration process. We have developed a model-driven software tool named *NoSQLMigrator* that aims to fully automatize the migration process. *NoSQLMigrator* provides means to extract data from relational databases and then to validate and insert extracted data into a NoSQL database. Currently, *NoSQLMigrator* supports extracting data from most of the modern relational databases and inserting data into MongoDB database [2]. MongoDB has been chosen as it is one of the most used NoSQL databases. It is a document-oriented database, which stores data as a collection of documents serialized in JavaScript Object Notation (JSON). Migration process in *NoSQLMigrator* is implemented by means of a series of model-to-model (M2M) and model-to-text (M2T) transformations, so as to generate fully functional transaction programs and applications that are executed over a legacy relational database and new NoSQL database. One of the main reasons for the development of such a tool was to make developers’ job easier, and particularly to free them from manual coding and testing. M2M transformations are based on meta-models to which source and target database models conform to. We denote such meta-models as database meta-models. We have developed Relational database schema meta-model in [3]. For the needs of developing the *NoSQLMigrator* tool, we have developed a domain specific language (DSL), named *MongooseDSL*. In this paper, we present both abstract (meta-model) and concrete syntaxes of this language.

MongooseDSL is a modeling language that can be used for modeling of Mongoose validation schemas and documents [4]. Mongoose is an object modeling tool that provides validation and data insert functionalities for MongoDB [5]. Mongoose validation schemas can be used for specifying constraints on data before it is inserted into MongoDB. For inserting the documents into MongoDB we use functions provided by the Mongoose tool. Meta-model of the *MongooseDSL* is also used as a database meta-model in the migration process.

Apart from the Introduction and Conclusion, the paper has four sections. In Section 2 we present the architecture of *NoSQLMigrator*. Abstract syntax of *MongooseDSL* is briefly described in Section 3, while in the fourth section we present the concrete textual syntax of the language. In the Section 5 we give an overview of the related work.

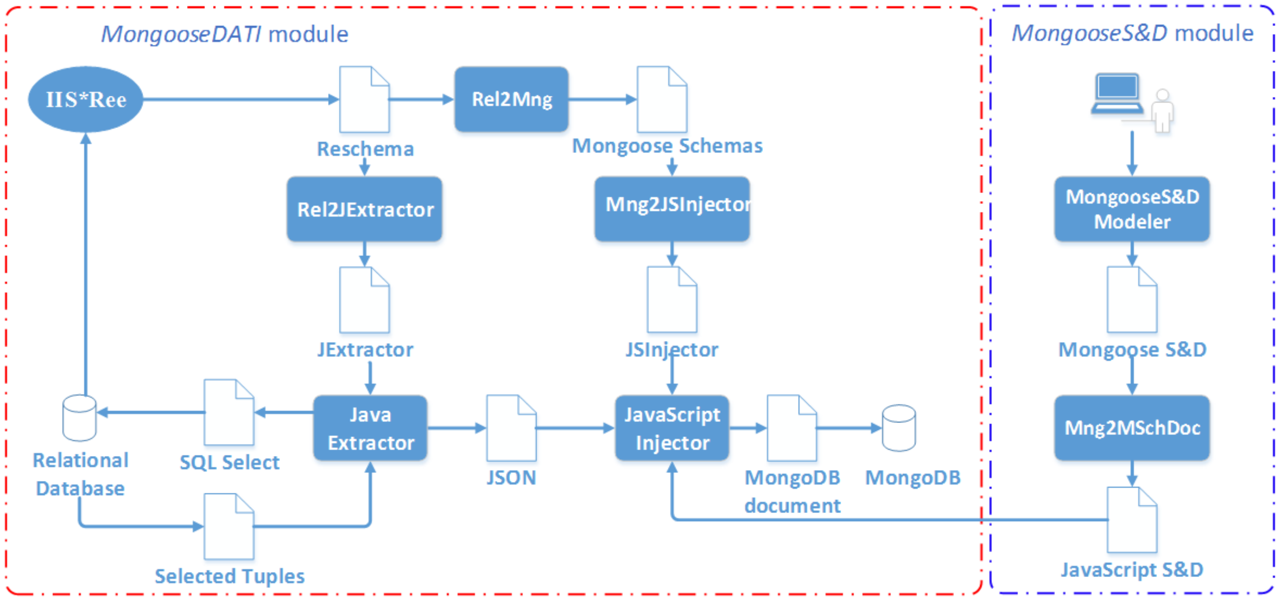


Figure 1. *NoSQLMigrator* architecture

II. The Architecture of NoSQLMigrator

In this section we present the architecture of the *NoSQLMigrator* tool. Its global picture is depicted in Figure 1. *NoSQLMigrator* comprises the following modules: *MongooseDATI* module and *MongooseS&D* module.

The *MongooseDATI* (Mongoose Data Aqusition, Transofrmation, and Injection) module allows user to perform the main part of migration process. *MongooseDATI* module comprises following components: *Rel2Mng, Rel2JExtractor, Mng2JSInjector, Java Extractor, and JavaScript Injector.*

Migration process is divided in four phases. During first phase reengineering of the relational database is done by using the IIS*Ree tool [6]. This tool provides a relational database schema specification according to relational database dictionary data. The specification conforms to meta-model based on standards, typical for the most relational database management systems (SQL:1999, SQL:2003, SQL:2011). In Figure 1. we present this specification as *Reschema* and the entire meta-model can be found in paper [3]. In the second phase of migration process, *Rel2Mng* component performs transformation from *Reschema* to Mongoose validation schemas specification. This specification is presented as *Mongoose Schemas*, and conforms to meta-model of developed *MongooseDSL* language.

The third phase of the migration process involves code generation using *Rel2JExtractor* and *Mng2JSInjector* components. *Rel2JExtractor* provides executable Java code based on *Reschema* specification. *Mng2JSInjector* provides executable JavaScript code based on *Mongoose Schemas* specification. In Figure 1. generated executable Java code is presented as *JExtractor* and generated executable JavaScript code as *JSInjector*. Code in *JExtractor* is used for data extraction from realational database schema. Code in *JSInjector* is used for validation of extracted data. Validation process is performed before data insertion into MongoDB database.

The last phase of data migration is generated code execution. The execution of generated Java code using *Java Extractor* component performs extraction of data from relational database and tranformation of extracted data to JSON documents. After transformation, JSON documents are sent to *JavaScript Injector* component. The execution of generated JavaScript code using *JavaScript Injector* compont results with acceptance of sent data, data validation acording to appropriate Mongoose validation schema and insertion of valid data to MongoDB database.

The *MongooseS&D* (Mongose Schema and Document) module enables user to specify Mongoose validation schemas and documents. *MongooseS&D* comprises *MongooseS&D Modeler* and *Mng2MSchDoc* components. This module provides generation of executable JavaScript code according to user specification. The user specifies Mongoose validation schemas and documents, using *MongooseDSL* concrete syntax within *MongooseS&D Modeler* component. Using *Mng2MSchDoc* component executable JavaScript code is generated based on user specification. Generated code comprises implementation of specified Mongoose validation schemas, documents and functions for insertion of valid data to MongoDB database.

III. *MongooseDSL* Abstract Syntax

In this section, we present the abstract syntax of the *MongooseDSL* language. The abstract syntax is implemented in a form of a meta-model that conforms to the Ecore meta-meta-model [7]. The meta-model is presented in Figure 2. In the rest of this section, we describe each of the *MongooseDSL* concepts with the corresponding meta-model class written in italics inside the parentheses.

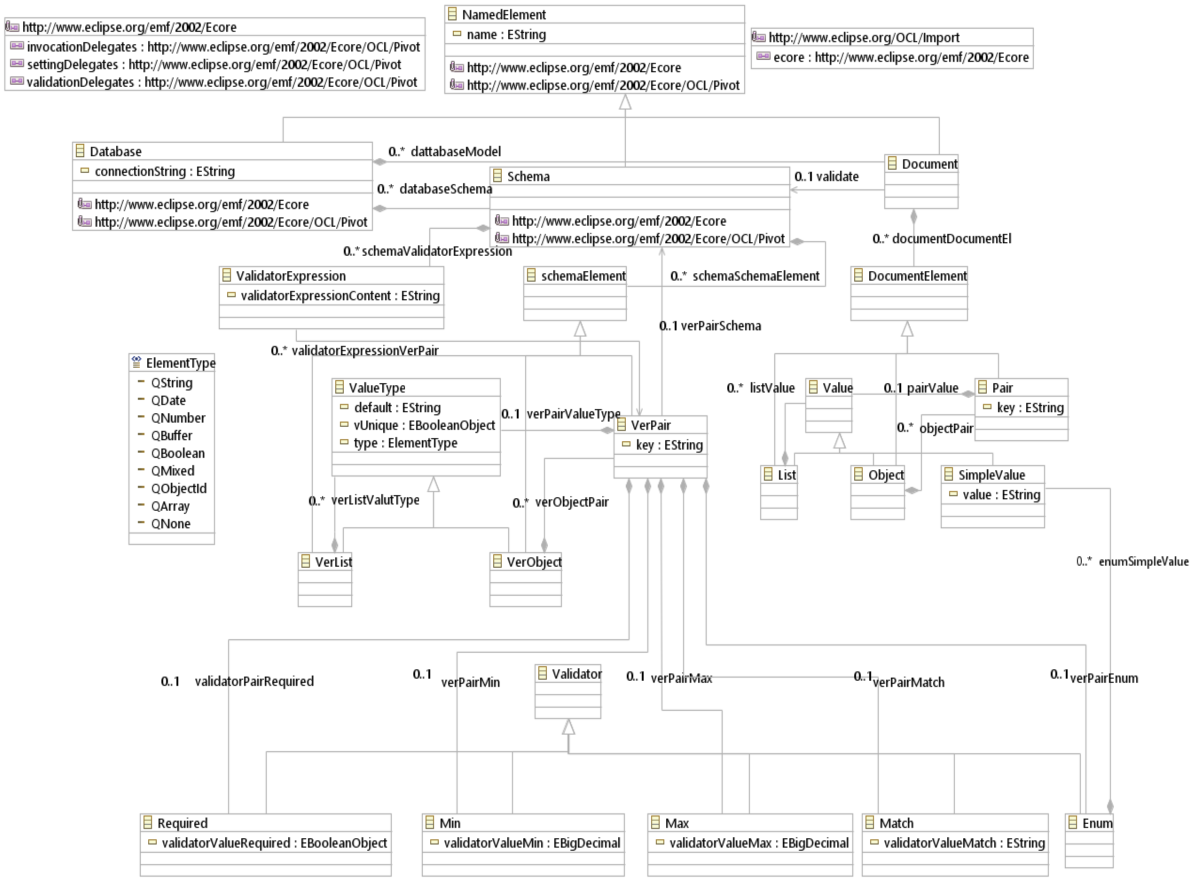


Figure 2. *MongooseDSL* meta-model.

The main language concept is the Mongoose-compliant database (*Database*) that comprises zero or more Mongoose validation schemas and zero or more Mongoose documents that are validated according to a defined schema.

The Mongoose document (*Document*) is a set of key-value pairs (*Pair*) called fields. For each field a key is defined in the *key* attribute of the *Pair* class, while values (*Value*) can be categorized as:

- Simple values (*SimpleValue*). Basic or atomic values that cannot be decomposed to more basic values such as integers, real numbers, etc.
- Complex values (*Object*). These are the values that represent structures comprising of other key-value pairs.
- List of values (*List*) that can contain any of the mentioned values.

Each Mongoose document is validated according to the associated validation schema. This is modeled with the *validate* reference.

The Mongoose validation schema (*Schema*) allows specification of validators for validating Mongoose documents before they are inserted into a database. By specifying such a schema, a user may define the structure, datatypes, and constraints on data. Each schema comprises zero or more schema fields (*VerPair*) that are in fact again key-value pairs. The key is modeled with the *key* attribute of *VerPair*, while the value (*VerType*) represents type constraints on values inserted into the document that is under validation. In *MongooseDSL*, we support creation of simple and complex type constraints. For simple type constraints, a user may choose one of the predefined types (*type* attribute of *VerType*) and a default value to be used if no other value is inserted (*default* attribute of *VerType*). Types are predefined in a form of the enumeration (*ElementType*) and cover the common datatypes found in modern database systems and programming languages. Further, it is possible to set the modeled value as unique (*vUnique* attribute of *VerType*). Complext type constraints may be comprised of other key-value pairs (*VerObject*), or they may represent lists of other simple or complex constraints (*VerList*). A schema field can be also a reference to other Mongoose validation schema allowing a user to decompose complex validation schemas into smaller ones. This is modeled with the *verPairSchema* reference.

Besides defining the type constraint, more detailed constraints on document values may be specified. These constraints are specified in a form of schema validators (*Validator*). *MongooseDSL* supports two types of validators: predefined and user-defined validators. The following predefined validators are supported:

- The validator for specifying if a Mongoose document field should be present in each document validated according the specified schema (*Required*). The attribute *validatorValueRequired* should be set to *true* if the field is required.
- The validator for specifying the minimum value of a document field (*Min*). The minimal

numerical value is specified in the attribute *validatorValueMin.*

- The validator for specifying the maximum value of a document field (*Max*). The maximal numerical value is specified in the attribute *validatorValueMax.*
- The validator for specifying a set of allowed values a Mongoose document field can have (*Enum*). Values are defined as simple values (*SimpleValue*) in the *enumSimpleValue* reference.
- The validator for specifying a regular expression that must be matched by a value of a Mongoose document field being validated (*Match*). The regular expression is defined in the *validatorValueMatch* attribute.

User-defined validators (*ValidatorExpression*) are specified using the validation functions whose body is defined in the *validatorExpressionContent* attribute. These functions implement the whole custom validation process.

IV. MONGOOSEDSL CONCRETE SYNTAX

In this section we present *MongooseDSL* textual concrete syntax. The *MongooseDSL* concrete syntax represents the visual representation of the meta-model concepts. The instances of the *MongooseDSL* concepts and their attribute values are modeled by the production rules specified by concrete syntax.

```
VerPair returns VerPair:
    {VerPair}
    name=EString ":"
        ('[' verPairSchema=[Schema|EString]']')?
        (('(')?verPairValueType=ValueType)?
        ('{ ' subVerPair+=VerPair ( "," subVerPair+=VerPair)* '}' )?
        ( validatorPairRequired=Required(')')?)?
        ( verPairMin=Min(')')?)?
        ( verPairMax=Max(')')?)?
        ( verPairEnum=Enum(')')?)?
        ( verPairMatch=Match(')')?)?
    ;
```

Figure 3. *VerPair* production rule

By means of Eclipse plug-in named Xtext [8, 9], we have generated the concrete syntax of *MongooseDSL*. In Figure 3. we present the production rule for defining Mongoose fields in the validation schema. First the user needs to specify the field name (*name*). After the special character ":", the user can specify the name (*verPairSchema*) of another validation schema used for the field validation. The user may also specify the validation type (*verPairValueType),* or specify the value of Mongoose predefined validator (*validatorPairRequired, validatorPairMin, validatorParirMax, validatorPairEnum* and *validatorPairMatch).*

MongooseDSL does not require knowledge of programming language JavaScript. It is necessary for the user to be familiar with the *MongooseDSL* concepts and the language production rules. The number of *MongooseDSL* concepts is much smaller than number of JavaScript concepts. Executable JavaScript code is generated on the user specification defined by *MongooseDSL*. Generated code comprises complete specification of modeled Mongoose validation schemas and documents. It also contains functions for the operations of validation and insertion in MongoDB database.

In the presented *MongooseDSL* concrete syntax each meta-model concept is presented by its name. The special characters "{", "}", "(" and ")" are used for the representation of edges between the modeling concepts. First the user needs to specify the main concept *Database*, while the other concepts are defined within the main concept. The character "," is used as the delimiter between the concepts within the main concept. The references between the linked concepts are specified by the name of the connected concept within the specified concept. Each of the concepts are represented by the other color. This approach is used because of better overview of the model structure. The main concept *Database* is represented by red color. Schema and Document concepts are presented by green and blue color. Each of the concepts modeled by the Mongoose predefined validator are presented by pink color.

```
Schema<'userSchema'>{
    email:(
        type:QString
        unique:true
        required:true
    ),
    password:(
        type:QString
        unique:true
        required:true
        match:'[/a-zA-z0-9]/'
    ),
    gender:(
        type:QString
        required:true
        enum:{'male', 'female'}
    ),
    name:{
        first:(
            type:QString
            required:true
        ),
        last:(
            type:QString
            required:true
        )
    },
    addres:{
        country:(
            type:QString
            required:true
        ),
        city:(
            type:QString
            required:true
        ),
        zip:(
            type:QNumber
            required:true
            min:1.0
        )
    },
    phone:[
            {
                no:(
                    type:QNumber
                    required:false
                    match:'^(\\+\\d{1,2}\\s)?\\(?\\d{3}\\
                          ?[\\s.-]\\d{3}[\\s.-]\\d{4}$'
                )
            }
    ]
}
```

Figure 4. Mongoose validaiton schema model

In the following part of this section we present a fragment of a model specified using a textual syntax of *MongooseDSL.*

In Figure 4. we present an example of modeled Mongoose validation schema, used for validation of document comprising an internet portal users. This Mongoose validation schema is modeled by the Schema concept. The validation schema name is specified by the attribute *name* within the characters "<" and ">". The schema comprises a filed set, specified within special characters "{" and "}". First we model the filed *email.* The field name is modeled

by specifying the value of *VerPair* concept attribute *key*. The field value is presented by usage *ValueType* concept. Within this concept the user specifies the field type setting the value of *type* attribute, choosing one of the predefined values of *ElementType* concept. The attribute *unique* is used to specify the field uniqueness at the level of collection in MongoDB database. In this field the user can model *required* Mongoose embedded validator, using the *Required* concept. Specifying the value *true* of *validatorValueRequired* attribute, the user defines the *email* field mandatory accorrding to the validation schema. *Password* field is modeled in the the similar way as the *email* field. The value of *Password* field is modeled by the *Match* concept that represents *match* Mongoose predefined validator. The Mongoose document field is validated by the value of *validatorValueMatch* attribute that stores the regular expression. The field is unique at the level of the collection that comprises Mongoose document. The field is also mandatory within this Mongoose document. The field *name* is modeled by the concepts *VerPair* and *VerObject*. The name of the field is specified by the value of the attribute *key*, within the instance of the *VerPair* concept. The value of this field is defined as complex value in the instance of *VerObject*

```
Document<'User_1423412',userSchema>{
    email:'user_1423412@mail.com',
    password:'user_1423412_pass',
    name:{
        first:'Jonh',
        last:'Sattler'
    },
    gender:'male',
    address:{
        country:'Louisiana, USA',
        city:'New Orleans',
        zip:'70129'
    },
    phone:[
        {
            no:'(504) 842-3000'
        },
        {
            no:'(504) 842-3012'
        }
    ]
}
```

Figure 5. Mongoose document model

concept. It comprises two fields *first* and *last*. Both of the fields are modeled as the instances of the *VerPair* and *ValueType* concepts.

The value of attribute *require* specifies that both of theese fields are mandatory. In the field *gender* we modeled field type, *required* Mongoose validator and *enum* Mongoose validator. *Enum* Mongoose validator is modeled by *Enum* concept. The instance of *Enum* concept comprise list of allowed values for an appropriate field of Mongoose document. The value of the field *phone* is modeled as the instance of *VerList* concept. The filed *phone* comprises the list of *VerObject* instances. Each *VerObject* instance contains a field modeled by the *VerPair* and *ValueType* concepts. Each of the fields comprises two Mongoose validators *required and match* are modeled, using *Required* i *Match* concepts.

Mongoose document presented in Figure 5. is modeled by the *Document* concept. The name of the document and and the name of Mongoose validaiton shema used for document validation, are presented within special characters “<” and “>”. The name of the document is specified by the value of attribute *name*. The edges of the document specification are presented by special characters “{” and “}”.

In the document *email* field is modeled as the instance of *Pair* and *Value* concepts. The field name is specified by the value of *Pair* concept attribute *key*. The value of the field is modeled by the attribute *value* in the *SimpleValue* concept instance. The edges of the modeled field are presented by special characters “(” and “)”. The field *password* is modeled in the same way as the field *email*. The field *name* represents complex field comprising two sub-fields. The field name is specified by the value of *Pair* concept attribute *key*. The value of the *name* field comprises two sub-fields *first* and *last*. *First* and *last* sub-fields are specified by the instance of *Object* concept. The name of the field is defined by the attribute *key* of the *Pair* concept. The value of the field is specified by the *value* attribute of the *SimpleValue* concept. The *first* and *last* sub-fields store information about first and last name of registered user. The field *gender* specifies information about gender of registered user. The field name is specified by the value of *Pair* concept attribute *key*. The value of the field is modeled by the attribute *value* in the *SimpleValue* concept instance. The field *phone* contains the list of telephone numbers of the registered user. The value of the field is modeled by the *List* concept. The instance of the *List* concept contains instances of the *Object* concept. Each instance of the *Object* concept represents a telephone number of registered user. Each field in the *Object* instance is modeled by the *Pair* and *SimpleValue* concepts.

V. RELATED WORK

There are many papers describing migration data and services, but to the best of our konowledge there are no approaches to this problem by using *MDSD* (Model Driven Software Development) paradigm. Rocha et al. [10] prsent *NoSQLayer*, a framework capable to support conveniently migrating from relational (i.e., MySQL) to NoSQL DBMS (i.e., MongoDB). Lee et al. [11] describe how to migrate content management system (CMS) data from relational to NoSQL database to provide horizontal scaling and improve access performance. Zhao et al. [12] present a schema conversion model for transforming SQL database to NoSQL, providing high performance of join query with nesting relevant tables, and a graph transforming algorithm for containing all required content of join query in a table by offering correctly nested sequence. Zhao et al. [13] describe approach to migration of data from relational database do HBase NoSQL database and algorithm to find column name corresponding to attribute in relational database.

Many NoSQL database vendors, like MongoDB and Couchbase provide their own mechanisms and tools for data migration from relational to their own databases [14, 15].

VI. Conclusion

In this paper we presented a DSL for Mongoose schema and document specification, named *MongooseDSL*. Through our research we developed the *NoSQLMigrator* tool. It provides a data migration approach based upon the usage of *MongooseDSL*. Our intention was to provide automated mechanism for data migration from most of the relational databases to MongoDB. First of all we needed to create the *MongooseDSL* meta-model specified by Ecore that actually represents the abstract syntax of the language. Then, we created textual notations for *MongooseDSL*. Using textual notation user is able to specify Mongoose validation schemas and documents. *MongooseDSL* does not require knowledge of programming language JavaScript. The number of *MongooseDSL* concepts is much smaller than number of JavaScript concepts.

In our further research, we plan to extend *MongooseDSL* and the *NoSQLMigrator* tool to provide data migration to document-oriented databases of different vendors. We also plan to develop and embed into *MongooseDSL* some graphical notation. Also, another research direction would be to extend *MongooseDSL* with new concepts allowing more detailed specifications of data models. These new concepts should provide new constraint specifications. For example, formal specification of database referential integrity constraint is not implemented in our solution.

VII. References

[1] Pramod J. Sadalage, Martin Fowler, “NoSQL Distilled: a brief guide to the emerging world polyglot persistence”, Crawfordsville, Indiana, pp 35-45, January 2013.

[2] “MongoDB” [Online], Available: https://www.mongodb.com/ [Accessed:01-02-2016]

[3] V. Dimitrieski, M. Čeliković S. Aleksić, S. Ristić, I. Luković, “Extended entity-relationship approach in a multi-paradigm information system modeling tool”, in: Proceedings of the 2014 Federated Conference on Computer Science and Infor Systems, Warsaw, Poland, pp.1611-1620, November 2014.

[4] Mernik, M., Heering, J., Sloane, M. A.: “When and how to develop domain-specific languages.” ACM Computing Surveys. pp. 316-344, December 2005.

[5] “Mongoose“.[Online], Available: http://mongoosejs.com/ [Accessed:01-Jan-2016]

[6] S. Aleksić, “Methods of database schema transformations in support of the information system reengineering process”, Ph.D. thesis, University of Novi Sad (2013).

[7] “Meta-Object Facility” [Online] Available: http://www.omg.org/mof/ [Accessed:01-Jan-2016]

[8] M. Eysholdt, H. Behrens, Xtext: “Implement your language faster than the quick and dirty way”, in: Proceedings of the ACM International Conference Companion on Object Oriented Programming Systems Languages and Applications Companion, OOPSLA '10, ACM, New York, NY, USA, pp. 307-309, 2010.

[9] “Eclipse” [Online], Available: http://projects.eclipse.org/projects/modeling. [Accessed: 02-Jan-2014].

[10] L. Rocha, F. vale, E. Cirilo, D. Barbosa F. Murao, “A Framework for Migration Relational Datasets to NoSQL”, International Conference On Computational Science, Reykjavik, Iceland, pp. 2593–2602, June 2015.

[11] Lee Chao-Hsien, Zheng Yu-Lin, “SQL-to-NoSQL Schema Demoralization and Migration: A Study on Content Management Systems”, Systems, Man, and Cybernetics (SMC), IEEE International Conference, Konwloon Tong, Hong Kong, pp. 2022-2026, October 2015.

[12] G. Zhao, Q. Lin, L. Li, Z. Li, “Schema Conversation Model of SQL Database to NoSQL”, P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2014 Ninth International Conference , Guangdong, pp. 355-362, November 2014.

[13] G. Zhao, L. Li, Z. Li, Q. Lin, “Multiple Nested Schema of HBase for Migration from SQL”, P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2014 Ninth International Conference, Guangdong, pp 338-343, November 2014.

[14] “RDBMS to MongoDB Migration Guide” [Online], Available: https://s3.amazonaws.com/info-mongodb-com/RDBMStoMongoDBMigration.pdf. [Accessed: 02-Jan-2014].

[15] “Moving from Relational to NoSQL: How to Get Started” [Online], Available: http://info.couchbase.com/rs/302-GJY-034/images/Moving_from_Relational_to_NoSQL_Couchbase_2016.pdf. [Accessed: 02-Jan-2014].

Framework for Web application development based on Java technologies and AngularJS

Lazar Nikolić, Gordana Milosavljević, Igor Dejanović
University of Novi Sad, Faculty of Technical Sciences, Serbia
{lazar.nikolic, grist, igord}@uns.ac.rs

***Abstract* – The paper presents a framework for developing single page Web applications based on Java technologies and AngularJS. Unlike traditional code-generation, the frontend of this solution uses a set of built-in generic user interface components capable of adapting to any metadata provided from the backend. The goal is to quickly provide a base for building a fully functional application and to switch focus from frontend development to modeling and backend development.**

I. Introduction

Persistent (data) model of an application usually defines concepts, types, constraints and relationships needed for the application functioning. This model is traditionally used for automatic generation of a persistent layer of the application. But, this model, enhanced with some presentation details can also be used for automatic construction of a presentation layer, in order to minimize the effort needed for its development.

In contrast with solutions presented in [3, 4, 5] that require manual coding or frontend generation based on a persistent model, our approach uses a reusable AngularJS client framework that adapts itself according to the metadata extracted from persistence layer (figure 1). This can greatly reduce the time needed to develop an initial version of the fully functioning application.

Architecture of the framework consists of four layers (figure 1):

- **Persistence layer** – contains class definitions, O/R mappers and provides data manipulation.
- **Service layer** – contains RESTful services for client-server communication.
- **Transformation layer** – contains mechanisms for object serialization and metadata extraction.
- **Presentation layer** – AngularJS client application.

The first three layers form the backend, while Presentation layer forms the frontend. Each part of the architecture will be briefly described in the next chapters.

Although concepts used for building the framework are platform independent, the framework itself is implemented using Java technologies for the backend and AngularJS for the frontend.

II. Related Work

Rapid application development is currently an active field of research. Kroki [11, 12, 13] is a rapid prototyping tool that allows users to actively participate in application development. It uses a Java based application engine to quickly create an executable prototype. However, it does not make use of up-to-date technologies, such as AngularJS.

Our framework uses EJB components at the backend, similar to work presented in [9]. In [9] is presented implementation of intermediate form

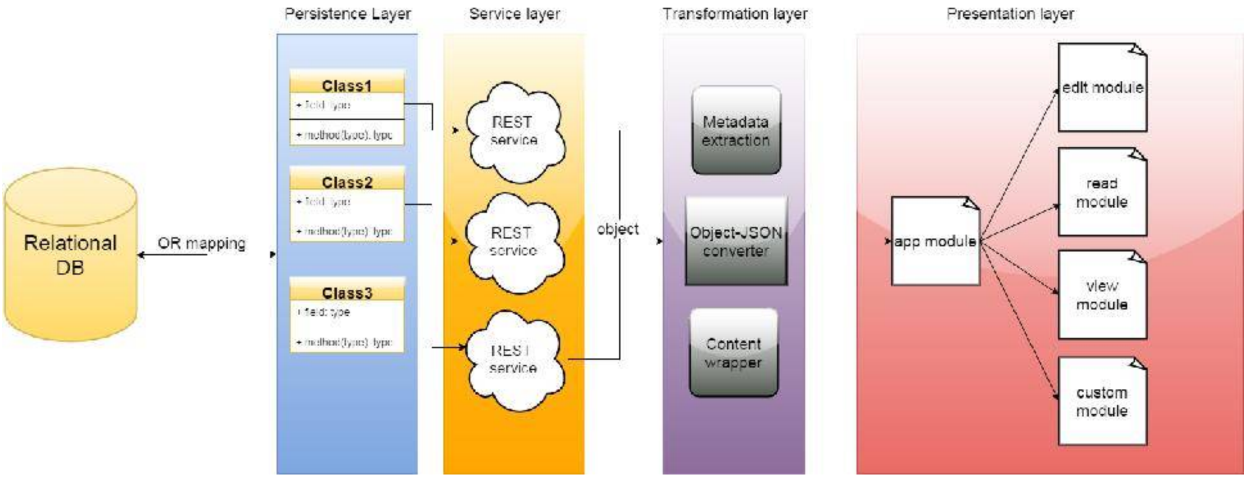


Figure 1. Adaptive framework architecture based on AngularJS

representation based on XML document, similar to the metadata object description used in our framework.

Adaptive user interface frameworks based on pre-built structures were proposed in [10, 14]. Aspects were used to implement adaptation of web application user interface based on runtime context.

Web frameworks are another related field of research. Some of the most prominent Web frameworks currently in use are Ruby on Rails, Django and Play framework. Our framework uses concepts similar to the mentioned frameworks, such as MVC architecture and built-in reusable components. The importance of such frameworks [2] is reflected by the need to quickly emerge on the market and adapt to changing user requirements, as well as their widespread use [15].

III. Backend

In order to present data from the backend, frontend must be provided with metadata, such as column names, column types, constraints, etc., which are contained in database schema. Since the backend has access to database schema, it is where mechanisms for fetching of metadata are contained.

By using Java Persistence API (JPA) [6], database schema can be defined through definition of persistence entities. Entities are Java classes that are mapped to a table in a relational database, with its fields representing database columns. These entities can be coded by hand or automatically generated from aforementioned persistence model.

Metadata necessary for object/relational transformation is provided by using annotations. This means that, instead of directly reading the database schema, it is possible to read class annotations (figure 3).

The backend contains RESTful web services. These services contain components needed to transform Java objects into JSON objects [8]. The reason behind choosing JSON as the serialization format is because the frontend is written in JavaScript, which has a native support for JSON objects. The web services in question are implemented using JAX-WS [7]. Each web service is represented with a java object. Service definitions are formed by using annotations.

Entity classes extend the *Model* class (figure 2). This class defines methods that obtain the metadata by

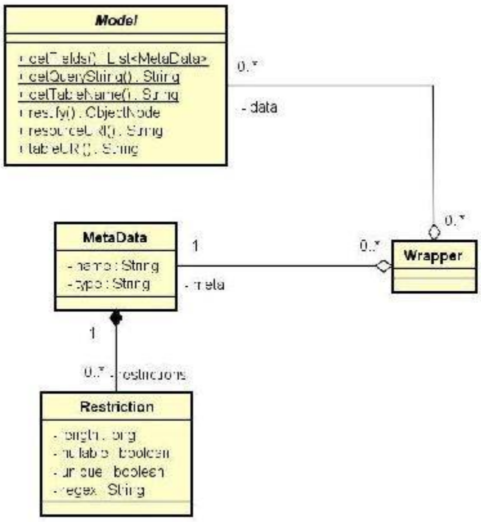


Figure 2. Transformation layer classes

reading JPA annotations of the entity. Once an object is ready to be serialized, its fields are being read during the runtime. Each field contains annotations with database metadata which is extracted and transformed into format suitable for JSON.

Classes that form the transformation layer are shown in figure 2. *Model* class contains path required for endpoint definition of the corresponding RESTful service.

Once formed, the metadata is contained in a *MetaData* object. Type name of primitive fields, such

```java
@Entity
@Table(name = "actor")
public class Actor extends Model {
    @Id
    @GeneratedValue(strategy = GenerationType.IDENTITY)
    @Column(name = "actor_id", unique = true, nullable = false)
    private long id;

    @Column(name = "actor_first_name", unique = true, nullable = false)
    private String firstName;

    @Column(name = "actor_last_name", unique = true, nullable = false)
    private String lastName;

    @JsonIgnore
    @JoinTable(name = "film_actor",
    joinColumns = @JoinColumn(name = "actor_id", referencedColumnName = "actor_id"))
    @ManyToMany(cascade = { ALL }, fetch = FetchType.EAGER)
    private Set<Film> films = new HashSet<Film>();
}
```

```json
{
  "data":[
    {
      "id":1,
      "firstName":"\tArnold",
      "lastName":"Schwarzenegger",
      "films":"/actor/1/films"
    }
  ],
  "meta":[
    {
      "name":"id",
      "type":"id",
      "restriction":{
        "length":255,
        "nullable":false,
        "unique":true,
        "regex":null
      }
    },
    ...
```

Figure 3. Transformation of a model class

as integers and strings, are taken as they are Java. Type of the primary key is always *id*. Associations are treated depending on their multiplicity. One-to-many and many-to-many relations are represented with a collection, while many-to-one is a reference to an object. Collections are called *zoom* fields, while references are called *link* fields, coming from the way of their representation [14]. Following the HATEOAS [19] principles, metadata for such fields should contain URI of the associated object.

Each *MetaData* object contains a collection of *Restriction* object, each containing a set of database restrictions for a single field. This includes all the restrictions supported by JPA annotations, such as *nullable*, *unique*, and *length*, with additional restriction *regex*, needed to define derived types. Once a *MetaData* object is fully formed, it is added to a JSON object with key "meta"; the java object itself is added with key "data" (figure 3). Finally, each *MetaData* object includes URI of the data object based on its type and its primary key. *Wrapper* class is responsible for generating the required JSON object.

IV. FRONTEND

Unlike other solutions [12] that use JSP pages for the frontend, our solution makes use of AngularJS framework. AngularJS enables creation of a single page applications and allows some of the logic (such as validation) to be included on the client side.

The client side application consists of modules (figure 1):

- **app** - routing and definition of modules.
- **view -** detailed view of an object.
- **edit -** editing and creation of objects.
- **read -** listing of all objects.
- **collection/zoom** – enables collection manipulation, such as adding or removing objects from a collection.

A module consists of a controller and multiple views and is responsible for one type of operation. Multiple views for a single controller enable the customization of the user interface, depending on object type. Each view is available on its URL path. This path is used to bind object types to their presentation page. URL paths are defined in *app* module as *routes* formed of controller-view pairs. Controllers and views can be combined to meet a certain requirement (for example, customization of the edit page). Routes for basic CRUD operations are initially generated. New routes can be added by writing a new AngularJS module that extends the *app* module. By adding new routes it is possible to further expand the client application beyond basic functionalities.

The *app* module is responsible for routing of both requests and responses. Backend services are mapped to controllers-view pairs in the module.

AngularJS offers mechanisms for extending HTML with directives. A directive acts like any other HTML tag. Upon execution, AngularJS scans the page and invokes the corresponding controller whenever it encounters a directive. The controller then modifies the HTML page by replacing said directive with new code.

Module controller is responsible for preparing data for presentation. Once an object is received from the backend, it is added to controller scope along with its metadata. This way, it is accessible from the module view. Module controller also contains functionalities for filtering, sorting, and validation. If a controller uses a backend service, its route should be equivalent to the service route. Basic controllers can be extended by using inheritance.

Table 1: Web services and operations for video library example

Operation Route	HTTP method	Count
film	POST GET DELETE	3
actor	POST GET DELETE	3
category	POST GET DELETE	3
id/film/actors	POST GET DELETE	3
id/film/categories	POST GET DELETE	3
id/actor/films	POST GET DELETE	3
id/category/films	POST GET DELETE	3
id/film	POST PUT GET DELETE	4
id/actor	POST PUT GET DELETE	4
id/category	POST PUT GET DELETE	4
Total: 3 services, 33 operations		

Table 2: Routing table for video library example

Paths	Controller	View	Total
/actor, /category, /film	Read	read-edit	3
/actor, /category, /film	Add	add	3
/id/film, /id/actor, /id/category	Edit	edit	3
/id/film/actors, /id/films/categories, /id/category/actor,	Collection	read	3
/id/film/actors/new, /id/films/categories/new, /id/category/actor/new,	Read	zoom	3
/	Main	main	1
Total: 16 routes			

Module view consists of an HTML page, directives and its controllers. AngularJS allows writing of custom directives. One such directive is used for presentation of an object. This directive is the central piece in user interface adaptiveness. The idea is to be agnostic to the received object type. This directive brings two options: the first allows the presentation of all objects fields in one place; the second allows custom layouting of the user interface, by passing a field as a parameter. Directive controller works similarly to a template engine. It generates an HTML component based on passed parameters and puts it in place of the directive. Type-component mapping is shown in figure 5.

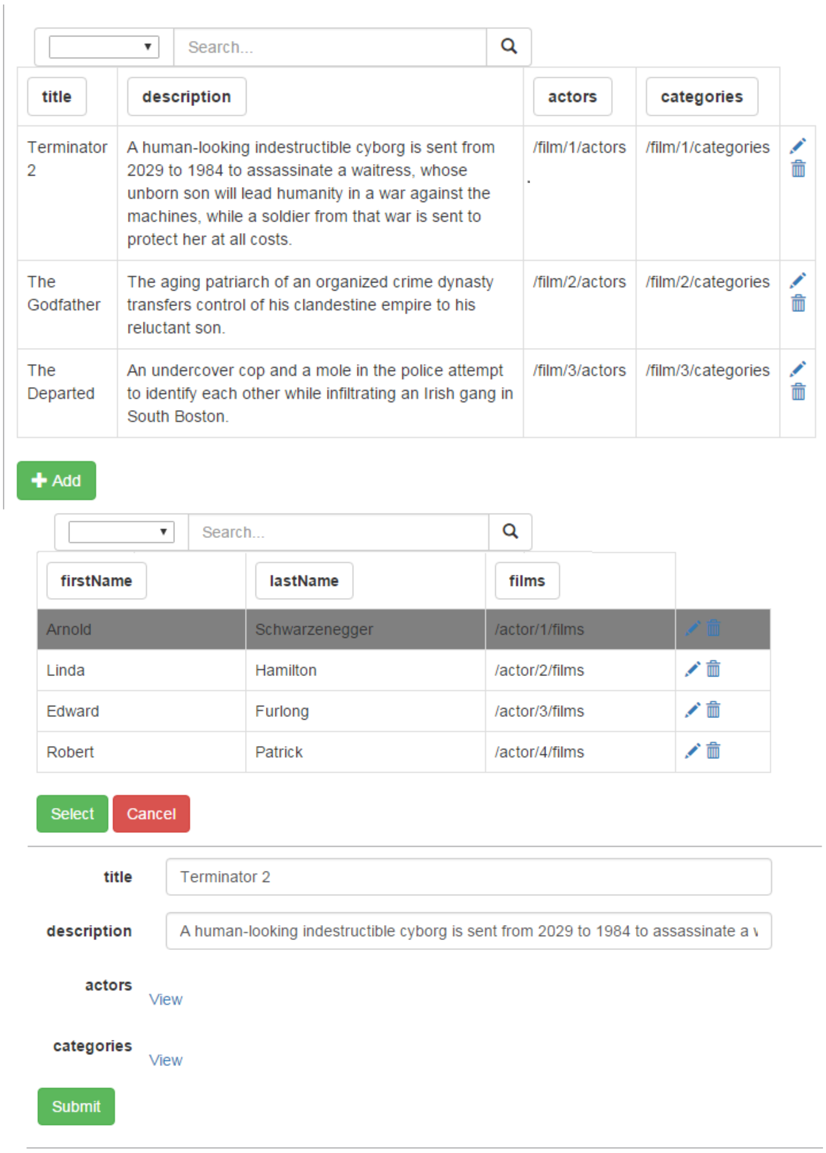


Figure 5. Application screens, top to bottom: table view, zoom picker and edit/create form

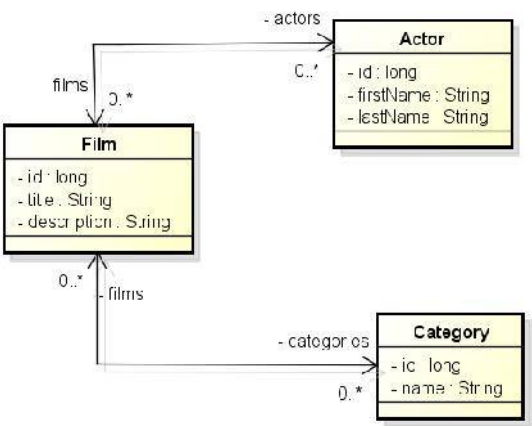


Figure 4. A part of a video library persistent model

V. AN EXAMPLE

The framework presented in this paper was used to build a simple web application based on model in figure 4. This web application provides only CRUD operations to a client. Entities are implemented by generating annotated Java classes, as shown in figure 3. These classes extend Model class, and are called *model classes*. Services are implemented by

generating JAX-WS REST services. Each service contains CRUD operations for model classes. Services communicate with clients via wrappers, which encapsulate request/response data. The client application contains routes. Screen form examples are shown in figure 5.

The web application in question is a Java enterprise application. Developing this web application required generating model classes. Each model class required a corresponding DAO and REST service that provide CRUD operations for the client application. REST services were generated as Java classes, using JAX-WS annotations. Summary is shown in table 1.

Each operation required a route in the client application *app* module. Routes in this module are formed by using ngRoute, which is an AngularJS module. These routes map relative paths to controller/view pairs. Each pair allows using a service operation with the same relative pair, enabling data manipulation or presentation, depending on the assigned controller/view combination.

Basic controllers and views are provided by the framework. These include Add, Edit, Delete, Update, Read, Collection and Zoom controllers; and Edit/Create, Read, Read-Edit, Zoom and Table views. Routes were formed by combining existing controllers and views (table 2). Example routes are shown in listing 1. Expanding the functionalities of an application beyond basic CRUD operations is done by creating new routes. Writing additional routes takes minimal effort, since their form is rather simple. Client application required no additional modifications. Screens from the client application are shown in figure 5.

```
.config(function ($routeProvider) {
  $routeProvider
  .when('/', {
    templateUrl: 'views/main.html',
    controller: 'MainCtrl'
  })
  .when('/:id/film', {
    templateUrl: 'views/edit.html',
    controller: 'EditCtrl'
  })
  .when('/film', {
    templateUrl: 'views/read-
edit.html',
    controller: 'ReadCtrl'
  })
  .when('/:id/film/actors', {
    templateUrl: 'views/read-
edit.html',
    controller: 'CollectionCtrl'
  })
```

Listing 1 A route example

VI. CONCLUSIONS

The paper presented a framework for development of single-page web applications. The framework enables frontend development through the use of adaptive user interface. The presented realization is based on Java technologies for the backend.

The backend generated in this solution is based on Java technologies, although ideas, mechanisms and methods presented by this paper can be easily realized in any other web-based technology.

The goal was to shift focus from frontend development. This way it was possible to develop models and backend services with minimal effort required to build the user interface.

Instead of generating the user interface or manually writing view for each data type, this approach uses built-in components capable of presenting any data at the backend. The components are generic enough to provide basic business operations on provided data.

Once the model is done and RESTful services are implemented, the application is ready for deployment; the user interface offers basic functionalities as soon as the application is deployed. The price for quick startup is limited customization options. The framework does not offer much customization options beyond layouting of the user interface. Additional styles can be loaded from an external css file, but currently there is no mechanism for defining a class for HTML inputs offered by the framework.

Further development can be directed in a way to remove or reduce framework's deficiencies. Potential improvements include: (1) possibility to write new controllers from scratch, (2) more initial controllers and controller-view combinations, (3) HTML class attributes for directives, (4) improvements in layer for object-JSON transformation, (5) ability to define custom constraints and data types etc.

Based on the example in section 5, we concluded that the client application developed using this framework required no modification. This meant that the client application was ready for deployment as soon as the backend was operational.

Model classes, services and basic routes were generated from the persistence model. This greatly reduced the amount of work needed to create the backend and minimized the time needed for project startup. The final goal is to create a full-stack Web development framework, with seamless integration of other frameworks used for its implementation.

VII. REFERENCES

[1] G. Milosavljevic, M. Filipovic, V. Marsenic, D. Pejakovic, I. Dejanovic, Kroki: A mockup-based tool for participatory development of business applications.. SoMeT (p./pp. 235-242), : IEEE. ISBN: 978-1-4799-0419-8, 2013.

[2] Douglas C. Schmidt, Aniruddha Gokhale, and Balachandran Natarajan, Frameworks: Why They Are Important and How to Apply Them Effectively: Department of Electrical Engineering and Computer Science Vanderbilt University Nashville, TN 37203, ACM Queue magazine, 2004

[3] Django framework. https://docs.djangoproject.com, retrieved 14.1. 2016.
[4] Ruby on Rails framework, http://rubyonrails.org/documentation/, retrieved 14.1.2016.
[5] Play framework, https://www.playframework.com, retrieved 14.1.2016.
[6] Java Persistence API, http://www.oracle.com/technetwork/java/javaee/tech/persistence-jsp-140049.html, retrieved 14.1.2016.
[7] JAX-WS, https://jax-ws.java.net/, retrieved 14.1.2016.
[8] JSON, http://www.json.org/, retrieved 21.1.2016
[9] B Milosavljevic, M. Vidakovic, S.Komazec, G. Milosavljevic, User interface code generation for EJB-based data models using intermediate form representations,2nd International Symposium on Principles and Practice of Programming in Java, PPPJ 2003, Kilkenny City, Ireland, 2003
[10] T. Cerny, K. Cemus, M. J. Donahoo, and E. Song. Aspect-driven, Data-reflective and context-aware user interfaces design. Applied Computing Review, 13(4):53–65, 2013
[11] G. Milosavljevic, M. Filipovic, V. Marsenic, D. Pejakovic, I.Dejanovic, Kroki: A mockup-based tool for participatory development of business applications.. SoMeT (p./pp. 235-242), : IEEE. ISBN: 978-1-4799-0419-8, 2013.
[12] Kroki, www.kroki-mde.net
[13] Kroki demo, http://youtu.be/r2eQrl11bzA
[14] M. Filipović, S. Kaplar, R. Vaderna, Ž. Ivković, G. Milosavljević, I. Dejanović, Aspect-Oriented Engines for Kroki Models Execution, Intelligent Software Methodologies, Tools and Techniques (SoMeT), 2013
[15] http://trends.builtwith.com/framework, retreived 30.1.2016.

Java code generation based on OCL rules

Marijana Rackov *, Sebastijan Kaplar**, Milorad Filipović**, Gordana Milosavljević**
* Prozone, Novi Sad, Serbia
** Faculty of Technical Sciences, Novi Sad, Serbia
*marijanarackov@gmail.com, **{kaplar, mfili, grist}@uns.ac.rs

***Abstract*— The paper presents an overview of the underlying framework that provides Java code generation based on OCL constraints used in Kroki tool. Kroki is a mockup-driven and model-driven tool that enables development of enterprise applications based on participatory design. Unlike other similar tools, Kroki can generate a custom set of business rules along with standard CRUD operations. These rules are specified as OCL constraints and associated with classes and attributes of the developed application.**

I. Introduction

Software development on internet time dictates very fast and flexible approach to programming where traditional coding "from scratch" is often substituted with model-driven development. The main factors in this shift are market requirements imposed on the companies, forcing them to adapt to constant changes. A lack of communication between client and developers is another source of constant changes which affect delivery time even more [1].

Kroki [4] is mockup-driven and model-driven development tool developed at the Chair of Informatics of the Faculty of Technical Sciences from Novi Sad, Serbia that uses business application mockups and models in order to generate a fully functional prototype of the developed system. Resulting prototype is an operational three-tier business application which can be executed within seconds and used for hands-on evaluation by the customers immediately. Developed prototype offers not only basic CRUD operations that can be easily generated from the specification, but also, an implementation of the custom business rules specific for the currently developed system. These rules are represented as OCL [5] constraints.

In order to transform these OCL constraints into working programming code, a Dresden OCL parser is incorporated into Kroki tool and custom code generator for OCL is developed. This paper presents mechanisms and techniques used to implement the code generator and provide examples how it is used and what the resulting programming code looks like.

The paper is structured as follows: Section 2 reviews the related work. Section 3 covers the basic concepts of the solution. It covers introduction to Kroki elements that require OCL rules specification and an overview of the implemented components that enable OCL constraints specification and processing. Section 4 presents some techniques used to generate Java code from OCL expressions, while Section 5 gives an example of how it is incorporated into Kroki tool.

II. Related Work

Related work presented here focuses primarily on the OCL implementations and accompanying tools used to generate programming code from it. Most of the analyzed tools do not implement the complete OCL specification.

A. Dresden OCL

Dresden OCL [6] is an open-source tool for writing, parsing and implementation of the OCL code and the generation of Java, AspectJ and SQL code from it. It can be used as stand-alone Java library or as Eclipse plugin. If used as Eclipse plugin, it provides its own Eclipse perspective which simplifies constraint definition. Generated AspectJ code can be highly customized by specifying various verification and error handling options. SQL code can be generated as one of the following versions: Standard SQL, Postgre SQL, Oracle, and MySQL. It supports Ecore [7] models import and models specified from reverse-engineered Java programs.

Dresden OCL parser is used in our solution for parsing OCL expressions and constraints, but we have implemented our own code generator for Java.

B. Octopus

Octopus (OCL Tool for Precise UML Specifications) [8] is an open-source Eclipse plugin that provides an editor for writing and verification of OCL constraints and generation of Java programming code. It also supports model creation via textual UML specification or by importing XMl representation exported from external modeling tools. Octopus generates a Java method for each OCL constraint and provides numerous customization options for generated code. There is also an option for generating additional Java classes to store hand-written code which is integrated with the generated code.

C. OCLE

OCLE (OCL Environment) [9] is a stand-alone application that enables the creation of OCL models and constraints and code generation. It provides its own graphical modeling editor in which new models can be created, but it also supports XMI model import. OCL constraint editor enables code formatting and highlighting which is also very helpful. OCLE supports all of the OCL constraint and data types, but not all of the operators.

OCLE generates Java code which depends on some libraries contained in the OCLE tool which need to be imported separately. Also, a Java method is generated for every constraint, which makes some of the generated methods somewhat vast and cumbersome.

D. OCL2J

OCL2J [9] is another tool that generates AspectJ code from OCL constraints. It supports most of the OCL operations, but not all of the constraint types. Constraint code and the modeled application code can be generated

separately. If the constraint is broken, generated code throws an exception.

III. IMPLEMENTING THE OCL CONSTRAINTS

This section describes Java code generation process based on OCL constraints attached to Kroki derived fields (attributes), classes and class operations.

Kroki has three derived field types: Aggregated, Calculated and Lookup fields. These fields expand the *VisibleProperty* stereotype and add their own tags (Figure 1) [2].

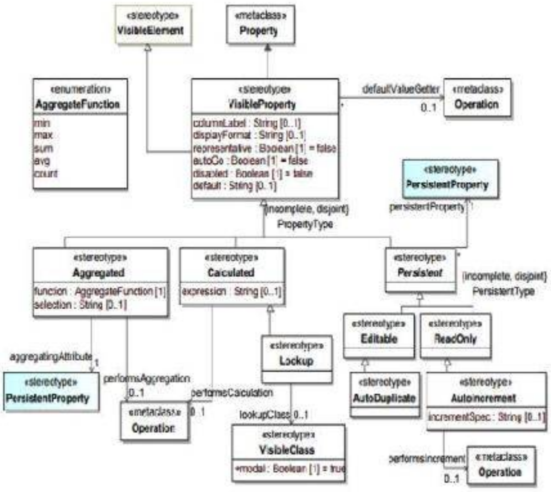


Figure 1. A part of a metamodel that defines form fields within Kroki [2]

Values of aggregated fields are calculated with the use of functions such as: Min, Max, Sum, Average, and Count. Field being aggregated can be of any kind (including derived fields).

Calculated fields are those whose values are calculated based on a formula. Formula is given with the OCL expression by filling the expression attribute in calculated field (please see Figures 5 and 7).

Values of a lookup fields are based on the value of OCL expressions that specify navigation to the class and its attribute we want to display.

A. Processing OCL constraints

Process started in `ProjectExporter`, Kroki's class for exporting modeled application to one of its engines (figure 2). `ProjectExporter` parses OCL expressions and prepares them for further analysis with `ConstraintAnalyzer` and `ConstraitBodyAnalyzer` classes. Finally, prepared set of rules is passed to `ConstraintGenerator` class which uses freemarker [11] templates to generate `EntityConstraint` classes. Figure 2 shows main steps of generating Java classes based on the given constraint. Blue color depicts classes that are used for generating Java code, while the yellow color shows helper classes that are used in the particular step.

`EJBClass` which instance is created for every defined class in the modeled application is extended with additional references to classes that model OCL constraint and its programming metadata. Those classes are `Constraint`, `Operation`, and `Parameter`. Figure 3 shows relationships of these classes.

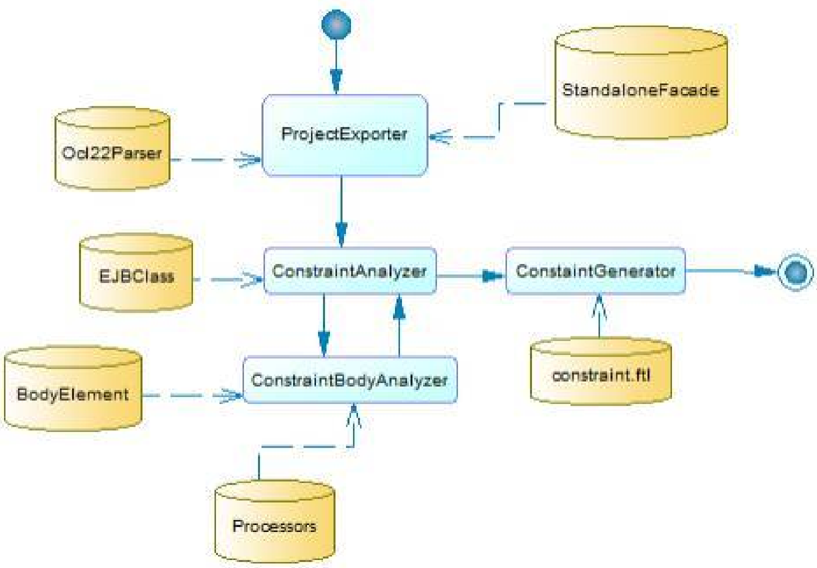


Figure 2. Constraint generation process

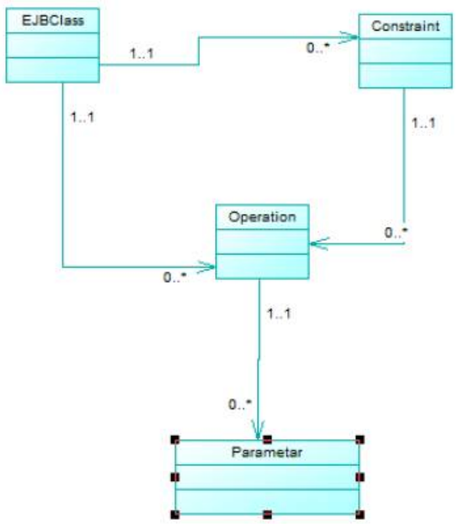


Figure 3. Kroki classes that support definition of OCL constraints

The `Constraint` class represents an OCL constraint while the `Operation` represents Java method that enforces given constraint. Method parameters are modeled with the `Parameter` class.

B. OCL expressions parsing

Parsing of the OCL constraints and generation of the corresponding Java methods is incorporated in the process of Kroki prototype execution. When analyzing each entity in the developed system, OCL expressions are extracted from derived fields and passed to Dresden OCL `OCL22Parser` parser [6], which assembles a list of parsed constraints for the given model (Kroki project). `ConstraintAnalyzer` associates OCL constraints from the parsed list with the corresponding `EJBClass`.

C. Java operation parsing

Another task for `ConstraintAnalyzer` class is building the `Operation` class instances for each `Constraint` instance. Data extraction starts with parameter definition. For ordinary (invariant) OCL constraints, parameters are built from the `Constraint` parameter list, while the definition constraint parameters are parsed from the OCL expression. Once the parameters have been prepared, the operation header can be constructed. The operation header is a string that represents future Java function signature. After that, OCL expressions from the constraint body are parsed by the `ConstraintBodyAnalyzer` class. If the expression contains an operation (data operation, mathematical or logical expression), it is delegated to one of the corresponding `Processor` classes for the further processing. The Figure 4 shows a list of all of the available `Processor` classes and the way they are connected to `ConstraintBodyAnalyzer` via the `Controller` class.

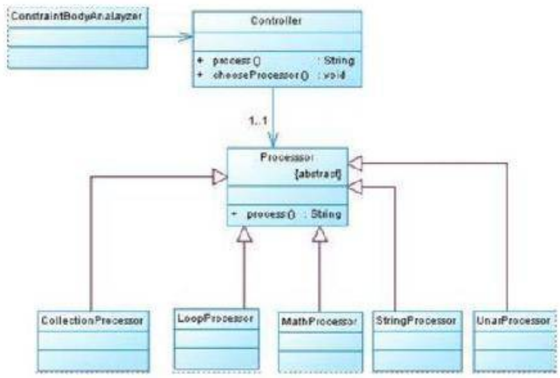


Figure 4. Operation processors

1) Binary operations

Supported binary logical and arithmetic operations are given in Table 1.

Since Java expressions are being built recursively, complexity of the operation is not an obstacle for parsing. The first step when processing the operations is checking whether the operator syntax is the same in Java and OCL, and altering the expression if it is not the case. An example of this are logical operators where OCL "and" “and” needs to be translated to Java "&&" and so on.

The binary operation that is not directly available in Java programming language is implication (implies). In order to support this operation, a custom function that checks whether one boolean parameter implies the other is integrated into each generated class. Another operator worth mentioning is the equality operator. When parsing equality check, if the operands (or return types of the functions used as operands) are integer or floating point numbers, the "==" operator is used, of all other types, "equals" function is generated.

Operation	Operator
Addition	+
Subtraction	−
Division	/
Multiplication	*
Modulo	%
Integer division	div
Equality	=
Less than	<
Greater than	>
Less than or equal	<=
Greater than or equal	>=
Not equal	<>
Logical and	and
Logical or	or
Exclusive or	xor
Implication	implies

Table 1. Supported binary operations

2) Other operations

For the sake of simplicity, a detailed overview of processing of the other operations is hereby skipped. This subsection covers some basic information on this topic

and provides an insight into some specific cases. All mathematic operations processed by the `MathProcessor` (`abs`, `min`, `max`, `round`, `floor`) are directly supported in Java. The same holds for the supported string operations (`concat`, `size`, `toLowerCase`, `toUpperCase`, `substring`), with the exception of the size function, which is used for calculating a collection size in Java. An additional check needs to be executed on the operand of the size function and, if its type is string, `length` is generated instead of a `size` function.

IV. JAVA CODE GENERATION

After all OCL expressions have been processed and the constraint lists of `EJBClass` objects have been populated, processed data is passed to `ConstraintGenerator` which generates corresponding Java classes. `ConstraintGenerator` is a generator class added to Kroki tool in order to support code generation based on OCL expressions. A full list of Kroki code generators is presented in [4].

`ConstraintGenerator` uses freemarker [11] template engine to generate files. For each entity of the modeled application, a Java class is generated in the same package with the constraints specified for it. The constraint class name is generated based on the `EJBClass` name, according to the pattern: `<Class name>Constraints.java`. The constraints class also extends the corresponding `EJBClass` in order to gain access to its methods and attributes. Also, during constraint processing, a list of Java import declaration is assembled for all of the used data types and classes that need explicit imports in Java. According to the list, a custom import section is generated for each constraint class. Basic constraint types (invariant, precondition, postcondition, definition, body) are generated in the constraint template while additional operations are defined in separate template files and imported into constraint template. That way, a dynamic support for OCL operations is enabled where additional functions can be added by specifying a template file and mapping it to a function name in the `ConstraintAnalyzer` class. As an example of one external template, the Listing 1 contains a freemarker template for the `exists` Java function that is generated for OCL function of the same name.

```
public Boolean ${method.name}{
  Iterator<${ method.iterType}>
    iter=${ method.forParam}.iterator();
  while(iter.hasNext()){
    ${ method.iterType} x =
       (${method.iterType})iter.next();

         if(${method.ifCondition})
           return true;
         }
         return false;
}
```

Listing 1. Freemarker template for *exist* OCL function

This function implements the OCL operation `exists`, that checks whether at least one element of the collection conforms to the specified condition. The `method` attribute in the template is `Operation` instance from the class operations list. Its attributes contain the information processed by the `ConstraintAnalyzer`. The `iterType` attribute holds the data type contained in the collection, and `ifCondtion` is Java `if` statement parsed from the OCL expression. The usage of iterators ensures that the generated method can be used on any Java collection type (if `get()` function is used, it could only work on `List` collections). If the method with the same name is already generated in the same class (e.g. an OCL defined method with the same name, but different functionality), a unique numeric suffix is added to the function name. If the operation is called upon another operation that returns a collection, the mentioned operation needs to be processed and generated before processing the current operation. An iterator is then created on the collection returned from the first operation.

V. OCL CONSTRAINTS IN KROKI

As mentioned before, Kroki supports special UI elements called *derived fields*, defined in the underlying DSL (EUIS DSL[12]) Those fields are used to hold the values that are not entered by the user, but instead, they are calculated from other values in the business system. Derived fields are located in the Actions section of the Kroki component palette (Figure 5).

Figure 5. Kroki derived fields

Aggregated fields hold the values that are calculated according to one of the supported functions (`min`, `max`, `sum`, `avg`), as shown in Figure 6.

Calculated fields can specify a custom expression for their values calculation. The settings panel for calculated field is shown in Figure 7 and it contains "Expression" text area which is used to specify OCL expression that defines rules imposed on the specified field value. The following subsections illustrate the usage of the calculated fields with two examples of the OCL rules and the accompanying generated code (Figure 7).

```
context Driver inv:
self.age >= 18
```

Listing 2. Driver age OCL constraint

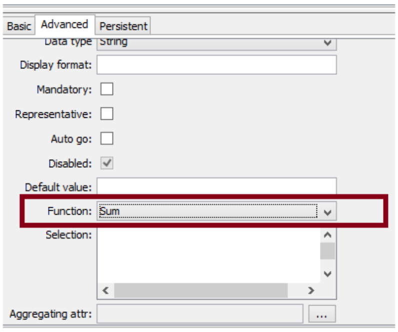


Figure 6. Aggregated field settings

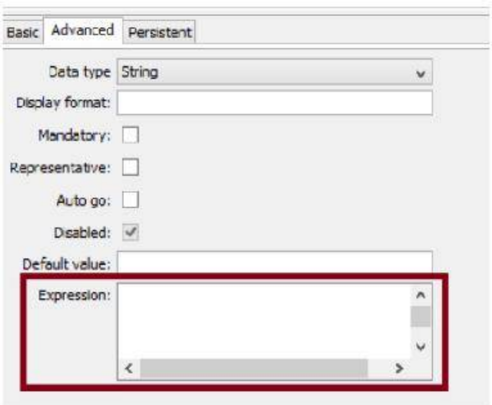


Figure 7. Calculated fields settings

A. Driver age restriction

Suppose that enterprise application we are developing contains a record about drivers and the vehicles used in client's company. The rule is that a driver must be at least 18 years old. In order to impose this restriction, a field that specifies driver age has to be a calculated field, and an OCL invariant constraint given in the Listing 2 has to be applied to it.

`ConstraintAnalyzer` output while processing specified expression is given in Listing 3, and the generated Java function is shown in Listing 4.

```
>=     OperationCallExpImpl       Boolean
age    PropertyCallExpImpl        Integer
18     IntegerLiteralExpImpl      Integer
```

Listing 3. OCL expression processing steps

```
public void checkdriverInvariant0()
  throws InvariantException {
  boolean result=false;
  try {
    result = (getA_age() >= 18);
  }catch (Exception e) {
      e.printStackTrace();
  }
  if(!result ) {
    String message = "invariant " +
    + "checkdriverInvariant0" +
    + "is broken in object '" +
    this.getObjectName() +
    + "' of type '" +
    this.getClass().getName() + "'";
    throw new InvariantException(message);
  }
}
```

Listing 4. Generated Java function that checks driver age restriction

As can be noticed from this example, all generated functions based on the boolean constraint are void Java functions, where `InvariantException` is thrown if constraint is broken.

B. Number of vehicles restriction

Another restriction for drivers in the observed system is that one driver can have up to three associated vehicles. The OCL constraint for this rule is shown in Listing 5, whereas the generated Java function is in Listing 6.

```
context Driver inv:
self.vehicles -> size <= 3
```

Listing 5. OCL constraint for vehicle number

```
public void checkDriverInvariant1()
throws InvariantException {
  boolean result=false;
  try {
     result = ( getA_vehicles().size() <= 3);
  }catch (Exception e) {
      e.printStackTrace();
  }
  if(!result) {
    String message = "invariant" +
      + "checkDriverInvariant1 " +
      + "is broken in object '" +
      + this.getObjectName() +
      + "' of type '" +
      + this.getClass().getName() + "'";
   throw new InvariantException(message);
  }
}
```

Listing 6. Generated Java function for vehicle number restriction

VI. CONCLUSIONS

The paper presented a process of implementing Java code generation functionality based on the OCL constraints specification in the Kroki tool. Implementing a custom rule specification is a very powerful feature of a development tool since most of the available software of that kind is able only to produce GUI skeletons or working prototypes that support only basic CRUD and navigation features. Implementing OCL constraints

specification combined with the already existing ability to export Kroki models as Eclipse projects and integrate hand-written code with the generated portion [13] makes Kroki prototypes one step closer to the final products.

The solution presented here requested a development of the custom OCL parsing framework based on Dresden OCL parser implementation and custom-made code generator for Kroki tool. The implemented OCL parser supports all of the OCL data types and most of the functions and operations. Currently `oclMessage` and `Tuple` data types are not supported as well as the functions that require AspectJ support in generated Java code. As a further enhancement, a syntax highlighting and a code completion support can be built in the Kroki OCL expression editor areas.

References

[1] A. Cockburn, J. Highsmith, “Agile Software Development: The Business Of Innovation”, IEEE Computer, pp. 120-122, Sept. 2001.

[2] B. Perišić, G. Milosavljević, I. Dejanović, B. Milosavljević, “UML Profile for Specifying User Interfaces of Business Applications”, Computer Science and Information Systems,Vol. 8, No. 2, pp. 405-426., 2011.

[3] A. Kleppe, J. Warmer, W. Bast, MDA Explained – The Model Driven Architecture: Practice and Promise, Addison-Wesley, Boston, 2003.

[4] G. Milosavljevic, M. Filipovic, V. Marsenic, D. Pejakovic, I. Dejanovic, Kroki: A mockup-based tool for participatory development of business applications. SoMeT (p./pp. 235-242), : IEEE. ISBN: 978-1-4799-0419-8, 2013.

[5] OMG. OCL 2.2 specification.

[6] B. Demuth, “The Dresden OCL toolkit and its role in information systems development”, Dresden University of Technology, http://dresden-ocl.sourceforge.net/downloads/pdfs/BirgitDemuth_TheDresdenOCLToolkit.pdf

[7] Eclipse Modeling Framework. http://www.eclipse.org/modeling/emf

[8] Octopus OCL. http://octopus.sourceforge.net/

[9] OCLE. http://lci.cs.ubbcluj.ro/ocle/

[10] A. Rensink, J. Warmer, "Model Driven Architecture – Foundations and Applications0", Second European Conference, ECMDA-FA 2006, Bilbao, Spain, July 10-13, 2006. Proceedings

[11] Freemarker Java template engine, http://freemarker.incubator.apache.org

[12] B. Perišić, G. Milosavljević, I. Dejanović, B. Milosavljević, “UML Profile for Specifying User Interfaces of Business Applications”, Computer Science and Information Systems,Vol. 8, No. 2, pp. 405-426., 2011

[13] M. Filipović, S. Kaplar, R.Vaderna, Ž.Ivković, G.Milosavljević, "Aspect-Oriented Engines for Kroki Models Execution", 5th International Conference on Information Society and Technology, Kopaonik, Serbia, March 8-11, 2015, Proceedings

[14] Kroki tool, http://kroki-mde.net/

Specification and Validation of the Referential Integrity Constraint in XML Databases

Jovana Vidaković[*], Ivan Luković[**], Slavica Kordić [**]
[*] University of Novi Sad, Faculty of Sciences, Department of Mathematics and Informatics, Novi Sad, Serbia
[**] University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
jovana@uns.ac.rs, ivan@uns.ac.rs, slavica@uns.ac.rs

***Abstract*—The referential integrity constraint (RIC) is one of the key constraints that exists in every data model. Current XML database management systems (XML DBMSs) do not completely support this type of constraint. The structure of an XML document can contain the information about the relationship between elements, but modern XML DBMSs do not perform automatic RIC verification during database operations execution. It is therefore necessary to implement and perform additional actions to provide a fully functional validation of the referential integrity constraint. In this paper we discuss different structures of XML documents, and their impact on the referential integrity constraint. We also present two types of its implementation in XML databases: using XQuery functions and triggers. Those techniques correspond to the characteristics of a selected XML DBMS.**

I. INTRODUCTION

Referential integrity constraint (RIC) is one of the most important and most common constraints in every type of database management system. While it is completely supported with both specification and implementation in relational database management systems (RDBMSs), it is only partially covered in XML database management systems (XML DBMSs).

XML databases allow in a certain way specification and implementation of several most used constraints like keys, primary keys, foreign keys, and unique constraints. Those are the building blocks for the specification of the RIC. However, modern XML DBMSs do not support either specification, or an automatic validation of this constraint. It is possible to insert, update, or delete elements in DML DBMSs in a way to violate the RIC.

It is possible to validate a database against basic constraints, e.g. primary keys or foreign keys, on demand. However, it cannot automatically maintain the database consistence, as the data has already been changed at the moment of a constraint validation.

A goal of this paper is to propose an approach to the specification and implementation of the referential integrity constraint (RIC) under XML DBMSs. XML DBMSs show some limitations caused by the structure of XML documents describing a database schema. Since an XML document can be structured following the well-known rules in different ways [19], we discuss how a selected document structure can support XML constraints, and particularly RICs.

Apart from Introduction and Conclusion, the paper is organized as follows. Section 2 describes related work. In Section 3, approaches for structuring XML documents are discussed. The notion of RIC, as well as its validation, are presented in Section 4.

II. RELATED WORK

RIC is one of the constraints that exists in every data model. This constraint is very well supported in relational data model. There are numbers of papers dealing with this constraint [1], [2], [3]. It also exists in XML data model, as it is presented in [4] and [5]. In [6] and [7], the notions of XML inclusion dependency (XID) and XML foreign key (XFK) are introduced over Document Type Definition DTD.

Due to hierarchical structure of an XML document, a scope of the uniqueness of an element is to be defined. The uniqueness scope can be a whole document or only a part of the document. A specification of XML constraints starts with declaring XML relative and absolute keys [8], [9], [10].

An XML document specifying a database schema can have a hierarchical or relational structure [11]. A hierarchical structure is the one where relationships between elements are modelled by nested structures. If the relationship among entities is “many to many”, it yields unnecessary repeating of some elements and leads to unexpected redundancy. If the structure is relational, relationships between elements are expressed by using keys and foreign keys. Each element has its own identification that is used as a key value to be referenced from other elements via foreign keys. There are pros and contras of both the approaches. In a hierarchical XML document, the normalization cannot be applied. RIC cannot be applied either, as stated in [11]. Since this paper deals with RIC, we cannot consider the hierarchical XML document structure in our approach. By this, we adopt a relational structure of XML documents, where all elements are direct descendants of the root element. As a rule, such documents are smaller and with no redundancy [11]. Instead of nesting, elements are linked using foreign keys. In our approach, RICs are used over such structure.

In [12], the authors introduce an approach to the constraint taxonomy in the XML data model, by means of an analogy to a similar taxonomy in the relational data model. In this paper we also use the notion of RIC according to the taxonomy proposed in [12].

There are many papers dealing with RICs. Most of them propose approaches to their specification only, while the implementation is not considered.

We propose the implementations of RICs based on XQuery functions or triggers, depending on the trigger support level in a selected XML DBMSs. A typical way of

constraint implementation in XML DBMSs is by means of triggers, as stated in [13], [14], and [15]. An XML DBMS supporting triggers in a way similar to relational databases is Sedna [16]. We have used Sedna for an implementation of the RIC by triggers. On the other hand, there are XML DBMSs with no appropriate support of triggers, such as eXist [17]. For those XML DBMSs, we propose a RIC implementation by XQuery functions [18].

III. Approaches to the Stuctures of an XML Document, Representing a Database Schema

One of the most important challenges in forming an XML document is defining relationships between elements. A relationship can be represented by physical positioning using nested structures. This is the hierarchical approach.

The relational approach of forming an XML document does not use nesting of elements. If two elements are related, keys and foreign keys are used to represent the relationship. Each element has identifier that can be used for referencing in another element using key and keyref elements.

Beside the data integrity, elements must be uniquely identified, the referential integrity constraint has to be satisfied, and there should not be update anomalies. If the hierarchical structure is used, a relationship between a parent element and a child element is represented only by physical positioning and the referential integrity constraint cannot be validated in that document.

If a relational structure is used, the key element is used to uniquely identify an element. The keyref element is used to connect a child element with its parent element. In this case, the referential integrity constraint is checked during validation, because each value in the keyref element must be an existing value in one key element.

The relational structure is better for use if the referential integrity constraint should be implemented using XML Schema document. Documents created using this kind of XML Schema document avoid redundancy, but they are larger. If the hierarchical structure is used, documents are more readable and clear, but foreign key cannot be declared and implemented in XML database. If an XML document is hierarchically structured and a child element is under its parent element, which represents “parent-child” relationship, then there is no foreign key in a child element to reference a parent element. Foreign key is necessary for referential integrity constraint.

The example of a hierarchical XML Schema document is given in Figure 1. An XML document created using this hierarchical schema is given in Figure 2. This schema describes the relationship between parents and children. A parent can have several children, but a child belongs to one parent. The element Child is dependent on the element Parent. The element Child is under the element Parent in the hierarchical structure. A position of the element Child represents its dependency on the element Parent. Each element Child has unique value of attribute child_id within its element Parent.

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified"
attributeFormDefault="unqualified">
 <xs:element name="Root">
  <xs:complexType>
   <xs:sequence>
    <xs:element name="Parent" maxOccurs="unbounded">
     <xs:complexType>
      <xs:sequence>
       <xs:element name="Child" maxOccurs="unbounded">
        <xs:complexType>
         <xs:attribute name="child_id" type="xs:string"
use="required"/>
         <xs:attribute name="child_name"
type="xs:string" use="required"/>
        </xs:complexType>
       </xs:element>
      </xs:sequence>
      <xs:attribute name="parent_id" type="xs:string"
use="required"/>
      <xs:attribute name="parent_name" type="xs:string"
use="required"/>
      <xs:attribute name="parent_lastname"
type="xs:string" use="required"/>
     </xs:complexType>
    </xs:element>
   </xs:sequence>
  </xs:complexType>
 <xs:key name="Parent_PK">
  <xs:selector xpath="Parent"/>
  <xs:field xpath="@parent_id"/>
 </xs:key>
</xs:element>
</xs:schema>
```

Figure 1. Hierarchical XML Schema document

```
<?xml version="1.0" encoding="UTF-8"?>
<Root xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
 xsi:noNamespaceSchemaLocation="ParentChild.xsd">
 <Parent parent_lastname="Peric" parent_id="R1"
parent_name="Pera">
  <Child child_id="D1" name="Maja"/>
  <Child child_id="D2" name="Ana"/>
 </Parent>
 <Parent parent_lastname="Maric" parent_id ="R2"
parent_name="Mara">
  <Child child_id="D1" name="Marko"/>
 </Parent>
 <Parent parent_lastname="Ilic" parent_id ="R3"
parent_name="Ilija">
  <Child child_id="D1" name="Ivan"/>
 </Parent>
 <Parent parent_lastname="Antic" parent_id ="R4"
parent_name="Ana">
  <Child child_id="D1" name="Aca"/>
  <Child child_id="D2" name="Maja"/>
 </Parent>
</Root>
```

Figure 2. Hierarchical XML document

The key of the element Child is a relative key, since the values of the attribute child_id are unique only in its parent element. This key can be defined in the XML Schema document like it is presented in Figure 3.

```
<xs:key name="Child_PK">
        <xs:selector xpath="Child"/>
        <xs:field xpath="@child_id"/>
</xs:key>
```

Figure 3. Relative key of the element Child

If the element Child does not depend on the element Parent, i.e. it has its own key, that key is now absolute key. It is presented in Figure 4. The XPath expression Parent/Child denotes that value of the attribute @child_id in each element Child is unique within the whole document.

```
<xs:key name="Child_PK">
 <xs:selector xpath="Parent/Child"/>
 <xs:field xpath="@child_id"/>
</xs:key>
```

Figure 4. Absolute key of the element Child

XML documents can be structured in different ways: all data are placed in elements, all data are placed in attributes, or some of data are placed in elements and some in attributes. In this paper we used the second rule, i.e. all data are placed in attributes. We also adopted that an XML document has a root element and all other elements are on the same level under the root element.

This is important for the following example. If an element has more than one foreign key from different elements, physical positioning in hierarchical structure is not a good way for representing this kind of relationship. Attributes that represent foreign keys have to exist in the element. The relational structure is better for representing this document.

A relational XML Schema document is given in Figure 5. This schema decribes relationship between workers and projects in a company. One worker can work on many projects. On one project there can work many workers. For each worker assigned to a project, there is number of hours spent on that project (attribute hrs). Between elements Worker and Project, there is the relationship with cardinality “many to many”. Element Engagement represents the engagement of one worker on one project. It has foreign keys wId and pId. Every element Engagement is unique since its primary key is union of attributes wId and pId, which are foreign keys.

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified"
attributeFormDefault="unqualified">
 <xs:element name="Work">
  <xs:complexType>
   <xs:sequence>
    <xs:element name="Worker" maxOccurs="unbounded">
     <xs:complexType>
      <xs:attribute name="wId" type="xs:string"
use="required"/>
      <xs:attribute name="firstName" type="xs:string"
use="required"/>
      <xs:attribute name="lastName" type="xs:string"
use="required"/>
     </xs:complexType>
    </xs:element>
    <xs:element name="Project" maxOccurs="unbounded">
     <xs:complexType>
      <xs:attribute name="pId" type="xs:string"
use="required"/>
      <xs:attribute name="name" type="xs:string"
use="required"/>
     </xs:complexType>
    </xs:element>
    <xs:element name="Engagement"
maxOccurs="unbounded">
     <xs:complexType>
      <xs:attribute name="wId" type="xs:string"
use="required"/>
      <xs:attribute name="pId" type="xs:string"
use="required"/>
      <xs:attribute name="hrs" type="xs:int"
use="required"/>
     </xs:complexType>
    </xs:element>
   </xs:sequence>
  </xs:complexType>
  <xs:key name="Worker_PK">
   <xs:selector xpath="Worker"/>
   <xs:field xpath="@wId"/>
  </xs:key>
  <xs:key name="Project_PK">
   <xs:selector xpath="Project"/>
   <xs:field xpath="@pId"/>
  </xs:key>
  <xs:key name="Engagement_PK">
   <xs:selector xpath="Engagement"/>
   <xs:field xpath="@wId"/>
   <xs:field xpath="@pId"/>
  </xs:key>
  <xs:keyref name="Engagement_FK1" refer="Worker_PK">
   <xs:selector xpath="Engagement"/>
   <xs:field xpath="@wId"/>
  </xs:keyref>
  <xs:keyref name="Engagement_FK2" refer="Project_PK">
   <xs:selector xpath="Engagement"/>
   <xs:field xpath="@pId"/>
  </xs:keyref>
 </xs:element>
</xs:schema>
```

Figure 5. Relational XML Schema document

The XML document valid to the schema given in Figure 6. There is no redundancy and every key in this document is an absolute key. If a hierarchical structure was used in this example, it would lead to repeating of either element Worker or element Project.

```
<?xml version="1.0" encoding="UTF-8"?>
<Work xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:noNamespaceSchemaLocation="Work.xsd">
 <Worker wId="R1" firstName="Petar" lastName="Petric"/>
 <Worker wId="R2" firstName="Marko" lastName="Maric"/>
 <Worker wId="R3" firstName="Ana" lastName="Antic"/>
 <Worker wId="R4" firstName="Nina" lastName="Ninic"/>
 <Project pId="P1" name="Plate"/>
 <Project pId="P2" name="PDV"/>
 <Engagement wId="R1" pId="P1" hrs="10"/>
 <Engagement wId="R2" pId="P1" hrs="20"/>
</Work>
```

Figure 6. Relational XML document

IV. The Referential Integrity Constraint in XML Databases

In the XML data model, a referential integrity constraint can be modeled with keyref element in XML Schema document. This constraint can be violated by update operations. XML database management systems do not implement the referential integrity constraint completely because validation is not done after the operation but only on demand. If the validation is not requested it will not be done.

The taxonomy of constraint in XML databases is given in [12]. According to that taxonomy, the referential integrity constraint in XML data model is defined in the following way:

RICon(2,
me,
E1@X⊆E2@Y,
e1@X⊆e2@Y, e1∈E1, e2∈E2, Key(E2, Y),
((e_1, referencing, {(insert, noAction), (update, noAction)}),
(e_2, referenced, {(delete, noAction, cascade), (update, noAction, Cascade)})))

The specification scope of the referential integrity constraint is bounded by two element types. It is interpreted over 2 elements, so the interpretation scope is ‘more elements’ (me). The parametrized formula pattern for this constraint is E1@X⊆E2@Y, where E1 is a referencing element type, E2 is a referenced element type, and the set of attributes Y is the key of the element type E2. A validation rule, used to validate RIC is e1@X ⊆ e2@Y, where the element e1 is of the type E1, and the element e2 is of the type E2. Critical operations, which RIC can cause, are insert or update of the referencing element, and delete and update of the referenced element.

The validation of RIC is done on insert, update or delete operations, over either referencing or referenced element. Delete or update of the referenced element is not allowed

(noAction) if that element is connected with at least one referencing element. Delete or update of the referenced element can be cascaded (cascade), which means that both referenced and referencing elements will be deleted or updated. Insert or update of referencing element is not allowed (noAction), if there is no corresponding referenced element.

This constraint cannot be implemented in the same way like in the relational databases. An XML document is validated against an XML Schema document and then the referential integrity constraint is checked. However, there is no mechanism to validate the constraint during update operations. XML database management systems, which we used, do not have support for cascade action or restrict action. Those actions should be implemented using procedural mechanisms. We used XQuery functions and triggers. Those techniques correspond to the characteristics of a selected XML DBMS. Since XML DBMS Sedna does not support XQuery functions, the referential integrity constraint is validated via triggers.

Let us validate the RIC over the XML document that is given in Figure 6. The RIC can be violated when updating or deleting a worker. The constraint validation can be implemented either by XQuery functions, or by triggers. XQuery functions are used in the eXist DBMS. The validation of an RIC constraint using XQuery functions is presented in Section IV.A. Triggers are used in Sedna DBMS. A validation based on triggers is presented in Section IV.B. We propose here the implementation methods for the RICs, using the mechanisms provided by the two representative XML DBMSs. We select eXist DBMS for the RIC implementation using XQuery functions, and Sedna DBMS for the implementation via triggers.

A. RIC implementation for eXist

eXist is one of the native XML database management systems [17]. eXist enables storing of XML documents and querying using XQuery language [18]. We used eXist DBMS to create functions that control the referential integrity constraint.

In this paper we describe deleting and updating of the referenced element, i.e. deleting and updating of the Worker element. The Worker element is part of the XML document, which is given in Figure 6. It can be implemented in two ways: by preventing (noAction) or by cascade action (Cascade). These actions have to be done procedurally because, to the best of our knowledge, XML databases do not support prevention or cascaded action.

XQuery function for deleting of the Worker element with preventing the operation, is given in Figure 7. Deleting of the Worker element can be forbidden if there is an Engagement element that referencing that Worker element. The Worker element can be deleted if there are no Engagement elements referencing to that Worker element. The function canDeleteWorker checks if the Worker element can be deleted. The function doDeleteWorker deletes the Worker element, if it is possible.

```
declare function local:canDeleteWorker($OLD as
element(Worker))
    as xs:boolean{
        let $r :=
not(exists(doc('Work.xml')/Work/Engagement[@wId=$OLD/@w
Id]) or not
(exists(doc('Work.xml')/Work/Worker[@wId=$OLD/@wId]))
        return ($r)
};

declare function local:doDeleteWorker($OLD as
element(Worker))
    as xs:boolean{
        let $i := update delete
doc('Work.xml')/Work/Worker[@wId=$OLD/@wId]
        return true()
};

declare function local:deleteWorker($OLD as
element(Worker))
    as xs:boolean{
        let $delWID := if(local:canDeleteWorker($OLD))
            then local:doDeleteWorker($OLD)
            else false()
    return $delWID
};
```

Figure 7. XQuery function for deleting a worker - NoAction

If the cascade deleting has to be done, all the Engagement elements that references to the Worker element, have to be deleted first. The function canDeleteWorker checks if there is the Worker element with the required value of the wId attribute. If it is found, the function doDeleteEngagement deletes all the Engagement elements that have the same value of the wId attribute. After that, the function doDeleteWorker deletes the Worker element. XQuery function for deleting the Worker element with cascaded action is given in the Figure 8.

```
declare function local:canDeleteWorker($OLD as
element(Worker))
    as xs:boolean{
        let $exists :=
exists(doc('Work.xml')/Work/Worker[@wId = $OLD/@wId])
        return ($exists)
};

declare function local:doDeleteWorker($OLD as
element(Worker))
    as xs:boolean{
        let $i := update delete
doc('Work.xml')/Work/Worker[@wId=$OLD/@wId]
        return true()
};

declare function local:doDeleteEngagement($OLD as
element(Worker))
    as xs:boolean{
        let $i := update delete
doc('Work.xml')/Work/Engagement[@wId=$OLD/@wId]
        return true()
};

declare function local:deleteWorker($OLD as
element(Worker))
    as xs:boolean{
        let $delWID := if(local:canDeleteWorker($OLD))
            then local:doDeleteEngagement($OLD) and
local:doDeleteWorker($OLD)
            else false()
    return $delWID
};
```

Figure 8. XQuery function for deleting a worker - Cascade

Updating of the Worker element is the operation that can violate RIC, too. If the value of the attribute wId, which is the key of the Worker element, does not exist in any of the child elements Engagement, update cannot be completed, as the referential integrity constraint will be violated. Beside that, the function canUpdateWorker checks if there is such Worker element with the same value of the wId attribute. If one of these conditions is satisfied, the function doUpdateWorker is called and the old Worker element is replaced with the new Worker

element. Updating of the Worker element by preventing action is given in Figure 9.

```
declare function local:canUpdateWorker($OLD as
element(Worker), $NEW as element(Worker))
    as xs:boolean{
        let $exists :=
not(exists(doc('Work.xml')/Work/Engagement
                           [@wId=$OLD/@wId]) or
exists(doc('Work.xml')/Work/Worker[@wId=$NEW/@wId]))
   return ($exists)
};

declare function local:doUpdateWorker($OLD as
element(Worker), $NEW as element(Worker))
    as xs:boolean{
   let $i := update replace
doc('Work.xml')/Work/Worker[@wId=$OLD/@wId] with $NEW
        return true()
};

declare function local:updateWorker($OLD as
element(Worker), $NEW as element(Worker))
     as xs:boolean{
   let $updWID := if(local:canUpdateWorker($OLD, $NEW))
            then local:doUpdateWorker($OLD, $NEW)
            else false()
    return $updWID
};
```

Figure 9. XQuery function for updating a worker - NoAction

If update is cascaded, after updating the Worker element, every child Engagement element is updated too. The whole Engagement element with the matching wId value is replaced with the new element with the new wId value. The values of the attributes pId and hrs are the same as in the old Engagement element. Updating with cascaded action is given in Figure 10.

```
declare function local:canUpdateWorker($OLD as
element(Worker), $NEW as element(Worker))
    as xs:boolean{
        let $exists :=
exists(doc('Work.xml')/Work/Worker[@wId = $OLD/@wId])
and not (exists(doc('Work.xml')/Work/Worker[@wId =
$NEW/@wId]))
   return ($exists)
};

declare function local:doUpdateWorker($OLD as
element(Worker), $NEW as element(Worker))
    as xs:boolean{
        let $i := update replace
doc('Work.xml')//Worker[@wId=$OLD/@wId] with $NEW
        return true()
};

declare function local:doUpdateEngaement($OLD as
element(Worker), $NEW as element(Worker))
    as xs:boolean {
        let $r :=
        for $i in doc('Work.xml')/Work/Engagement[@wId
= $OLD/@wId]
         let $pId:=$i//@pId
         let $hrs:=$i//@hrs
          let $r := update replace
doc('Work.xml')/Work/Engagement[@wId=$OLD/@wId
and @pId=$pId  and @hrs=$hrs ] with
<Engagement wId="{$NEW/@wId}"  pId={$pId}  hrs={$hrs}/>
            return <res/>
        return true()
};

declare function local:updateWorker($OLD as
element(Worker), $NEW as element(Worker))
    as xs:boolean{
        let $updWID := if(local:canUpdateWorker($OLD,
$NEW))
            then local:doUpdateEngagement($OLD, $NEW)
and local:doUpdateWorker($OLD, $NEW)
            else false()
    return $updWID
};
```

Figure 10. XQuery function for updating a worker - Cascade

B. RIC implementation for Sedna

We select Sedna XML DBMS because of its full support of triggers. Sedna [16] is a native XML DBMS. It also offers a large number of services for managing XML documents: querying, modifying, ACID transactions, query optimization, etc. It supports W3C XQuery implementation. Sedna allows the use of either XQuery functions or triggers.

Triggers in Sedna do not provide cascade deleting. In this paper, triggers for deleting and updating prevention are given.

Deleting of the Worker element is possible only if it is not referenced in any Engagement element. Otherwise, or if such worker does not exists, deleting is not allowed. The trigger for Worker deleting is given in Figure 11.

```
CREATE TRIGGER "RefIntWorkerBeforeDelete"
BEFORE DELETE
ON collection('RefInt')/Work/Worker
FOR EACH NODE
DO {
        if(exists(fn:doc('Work', 'RefInt')/
             Work/Engagement[@wId=$OLD/@wId]) or
 not (exists(fn:doc('Work', 'RefInt')/
               Work/Worker[@wId=$OLD/@wId])))
        then
        error(xs:QName("RefIntWorkerBeforeDelete"),"
Can’t delete worker because wId is used in Engagement
")
        else
                        ($OLD);
}
```

Figure 11. Trigger that does not allow deleting

Update of the Worker element is not allowed if there is referencing Engagement element or there is no matching value of the wId attribute in any Worker element. In other cases, the old Worker element is replaced with the new Worker element. The trigger for Worker updating is given in Figure 12.

```
CREATE TRIGGER "RefIntWorkerBeforeUpdate"
BEFORE REPLACE
ON collection('RefInt')/Work/Worker
FOR EACH NODE
DO {
        if (exists(fn:doc('Work', 'RefInt')/
              Work/Engagement[@wId=$OLD/@wId]) or
     exists(fn:doc('Work', 'RefInt')/
              Work/Worker[@wId=$NEW/@wId]))
        then
        error(xs:QName("RefIntWorkerBeforeUpdate"),"
Can’t update worker because wId is used in Engagement
")
        else
                        ($NEW);
}
```

Figure 12. Trigger that does not allow update

V. Conclusion and Future Work

In this paper we discuss different approaches to the structure of XML documents representing database schemas and their impact on the Referential Integrity Constraint (RIC). We find the relational XML structure as more suitable for specifying RICs.

Further on, we present two approaches to the RIC specification and implementation in XML databases. One is based on the usage of XQuery functions, while the other on triggers.

We find the approach based on triggers as better for the RIC implementation, if a selected XML DBMS offers the appropriate trigger support, as it is a case with the Sedna

DBMS. However, a support of triggers in the eXist DBMS is a quite basic, so it cannot be used for the RIC implementation. In this case, we have proposed a usage of the XQuery functions.

Future work will cover approaches to the specification and implementation of other types of constraints over various XML DBMSs. The process of generating XQuery functions and triggers for constraint validation could be fully automated and based on the paradigm of a model driven software engineering and the appropriate transformation specifications. A research work on a code generator development is in progress. This code generator would make database designer's and developer's job easier and free them from manual coding of XQuery functions and triggers for validating constraints.

Acknowledgment

The research presented in this paper was supported by Ministry of Education, Science and Technological Development of Republic of Serbia, Grant OI-174023.

References

[1] Luković I, Ristić S, Mogin P, "On The Formal Specification of Database Schema Constraints", I Serbian – Hungarian Joint Symposium on Intelligent Systems, September 19-20, 2003, Subotica, Serbia and Montenegro, Proceedings, Polytechnical Engineering College, Subotica, Serbia and Montenegro, and Budapest Polytechnic, Budapest, Hungary, pp. 125-136

[2] Mogin P., Lukovic I., Govedarica M., Database Design Principles, 2nd Edition, University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia and Montenegro, 2004

[3] Lukovic I., Ristic S., Mogin P., On The Formal Speci¯cation of Database Schema Constraints. 1st Serbian-Hungarian Joint Symposium on Intelligent System SISY 2003, September 19-20, 2003, Subotica, Serbia and Montenegro, Proceedings, pp. 125-136.

[4] Fan, W., XML Constraints: Specification, Analysis, and Applications, DEXA, 2005, pp.805-809.

[5] Fan, W., Simeon, J., Integrity constraints for XML, PODS, 2000, pp.23-34.

[6] Shahriar, M.S., Liu, J., On Defining Referential Integrity for XML, International Symposium on Computer Science and its Applications, 13-15 Oct 2008. ISBN 978-0-7695-3428-2, DOI 10.1109/CSA.2008.36, pp. 86-91.

[7] Shahriar, M.S., Liu, J., Towards a Definition of Referential Integrity Constraints for XML, International Journal of Software Engineering and Its Applications, Vol. 3, No. 1, January 2009. pp. 69-82.

[8] Buneman, P., Fan, W., Simeon, J., Weinstein, S., Constraints for Semistructured Data and XML, SIGMOD Record, 2001, pp. 47-54.

[9] Buneman, P., Davidson, S., Fan, W., Hara, C., Tan, W., Keys for XML. Computer Networks, 2002, 39(5), pp. 473-487.

[10] Buneman, P., Davidson, S., Fan, W., Hara, C., Tan, W., Reasoning about Keys for XML, Information Systems, 28(8):1037-1063, 2003.

[11] Chaudhuri, N., Glace, J., Wilson, G., Hierarchical vs. Relational XML Schema Designs, A Study for the environmental Council of States, Report ECO41T1, http://www.exchangenetwork.net/dev_schema/schemadesigntype.pdf, June 2006

[12] Vidaković, J., Luković, I., Kordić, S., Specification and Implementation of the Inverse Referential Integrity Constraint in XML Databases, 7th Balkan Conference in Informatics, ACM New York, USA, ISBN 978-1-4503-3335-1, DOI: 10.1145/2801081.2801111

[13] Tatarinov, I., Ives, Z., Halevy, A., Weld, D., Updating XML, Proccedings of the ACM SIGMOD International Conference on Managenent of Data 2001, pp. 413-424

[14] Bonifati, A., Braga, D., Campi, A., Ceri, S., Active XQuery, ACM SIGMOD 2001 May 21-24, Santa Barbara, California, USA

[15] Grinev, M., Rekouts, M, Introducing Trigger Support for XML Database Systems, Proceedings of the Spring Young Researcher's Colloquium on Database and Inforation Systems SYRCoDIS, St. Petersburg, Russia, 2005

[16] Sedna, Native XML Database System, www.sedna.org

[17] eXist, http://exist-db.org/exist/apps/homepage/index.html

[18] XQuery, http://www.w3.org/TR/xquery/

[19] Vidaković, J., Specification and Validation of Constraints in XML Data Model, Ph.D. Thesis, Faculty of Technical Sciences, University of Novi Sad, 2015. (in Serbian)

Benchmarking of Tools for Distributed Software Development and Project Management Support

Ljubica Kazi, Miodrag Ivkovic, Biljana Radulovic, Branko Markoski, Vesna Makitan
University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
ljubicakazi@hotmail.com, misa.ivkovic@gmail.com, biljana.radulovic66@gmail.com, markonins@yahoo.com, vesna@tfzr.uns.ac.rs

***Abstract*— Distributed software development as trend raises many issues. Some of these issues could be categorized as the need for proper software tools support, particularly in the area of distributed software development and project management. It is of a high importance to integrate software development (as data source) and project management (that uses such data for decision making). This paper presents results of benchmarking of most frequently used tools that support both distributed software development and project management.**

I. Introduction

Importance of project management and values that brings to the quality and success of projects and business processes in organizations has been explored within the Project Management Institute research, conducted with 418 projects, 60 reports, 447 interviews with project managers from 65 business organizations from all around the world [1]. Key results are described as elements that influence importance and value of project management:

1. Context – Economy status, Human factor, Culture, Project type and context, Organization characteristics, Strategic directions

2. Context influences implementation via: Training, Tools, People, Motivation and Organization

3. Implementation influences values: Satisfaction, Goals Alignment, Application Consistency, Process Results, Business Results and Implemented Advantages.

Obviously, it is of high importance to consider tools as important implementation factor.

Distributed software development (DSD) projects are very specific. Many research efforts have been made in systematization of issues of this specific type of projects, such as [2] and [3]. DSD issues could be categorized to the area of software engineering and project management. These two dimensions are used to systematize review papers results using specific SE-PM matrix [4] based on SWEBOK (Software Engineering Body of Knowledge) [5] and PMBOK (Project Management Body of Knowledge) as standards.

Integration of tools that support distributed software development and project management improve project management efficiency. Aim of this paper is to present part of results from Ljubica Kazi's PhD thesis [7] that is related to benchmarking model for evaluation of software tools that support distributed software development and project management. Applicability of the proposed model is illustrated with software tools that are extracted as most used in practice, as per previously conducted survey.

II. Theoretical Background

Benchmarking could be defined in the context of quality management in organizations, as per International Benchmarking Clearinghouse [8] as "systematic and continual process of measurement and comparison of organizational business processes in accordance to business processes of leaders in the same field in the world, in aim to get information that could help organization to take activities of its own performances improvements." According to [9], benchmarking could be defined as "test or set of tests that are used to compare performances of alternative tools and techniques". In [9], three important components of benchmarking are described:

- Motivated comparison – motivation for research and technical description of characteristics that are compared among tools and techniques that are compared
- Relevant set of tests – tests present part of tasks that tools/techniques should perform or solve in practice. It is important to choose relevant tasks.
- Performance measures – performance is link between technology and the way it is used, i.e. purpose. Measuring performance exšresses level of alignment with purpose.

According to [9], good benchmarking model should have characteristics such as: accessibility, applicability, economical sustainability, relevance, portability, scalability.

III. Related work

Even benchmarking is primarily focused on business excellence, it could be used in evaluation of tools and techniques. In [10], research on criteria for evaluation of IT project success is based on benchmarking within different business organizations, using surveys.

In [9], using benchmarking as a tool for research in software engineering area is examined. In [11], benchmarking is used to explore good practice elements application in the field of software project management from European countries.

Particular role of stakeholders in selection of project and portfolio management tools is examined in paper [12]. In this paper it has been proposed the approach of analytical hierarchical process. According to survey applied at National Documenting Center of Greece, criteria for evaluation and selection of project management tools are defined.

IV. THE PROPOSED BENCHMARKING MODEL

The key element of benchmarking is selection of characteristics that will be measured and compared. Selection of characteristics depends on the type of tools and its purpose, as well on the goals of the analysis.

In [7], key areas for selection of relevant characteristics are aligned with standards in software engineering (according to SWEBOK [5]), project management (according to PMBOK [6]) and strategic management methodology Balanced Scorecard (BSC) (introduced for the first time in [13]). The proposed benchmarking model is presented at Table I, as a three column table. First column represents standard source, second represents area within the standard source and third represents questions to be asked within the appropriate area.

TABLE I. THE PROPOSED BENCHMARKING MODEL

SOURCE	AREA	QUESTION (Yes, No answer) – does the tool support…?
PMBOK	Integration	Interoperability with other tools, especially development environments and different file formats? Data Export?
	Scope	Display of defined scope of software functions? Changes in scope of software functions? Display of implemented scope of needed software functions?
	Time	Estimation and planning of implementation time? Display of time characteristics of implementation flow?
	Quality	Measurement of human resources (project participants) quality? Team quality measurement within the project participation? Measurement of results quality? Measurement of results flow within the implementation process?
	Human resources	Project participants' records? Records on work finished? Personalization of functions for different user profiles according to different roles in project?
	Communications	File exchange among participants? Communication – exchange of messages, ideas, questions?
	Risks	Risk records? Risk measurement? Risk display?
	Procurement	Improvement of existing solution with additional modules?
	Stakeholders	Communication with stakeholders?
SWEBOK	Requirements	Requirements records? Requirements change records? Follow-up and display of requirements changes? Records on the level of alignment of solution with requirements?
	Design	Integration with tools for system design?
	Construction	Integration with tools for software development?
	Testing	Integration with tools for software testing?
	Maintenance	Not covered
	Configuration Management	Not covered
	Engineering Management	Covered with other questions
	Engineering Process	Display of implementation process flow? Measurement of phase success or success of implementation flow?
	Engineering models and methods	Application of different approaches and methodologies of project management? Application of different approaches and methodologies of software development?
	Quality	Results quality measurement? Measurement of quality of process results flow?
	Engineering economy	Financial indicators of phase or process success? Financial indicators of product success?
Balanced Scorecard	Finance	Financial indicators of phase or process success? Financial indicators of product success?
	Users	Communication with users? Creating reports? User satisfaction measurement? Records on needs and requirements from users?
	Internal Processes	Display of process flow? Planning of process flow? Measurement of process flow success?
	Learning and Growth	Records and display of experiences? Records and display on problems and answers?

The proposed benchmarking model is applied in analysis of selected software tools that support both distributed software development and project management. Selection of tools for analysis is explained in section V.

V. METHODOLOGY

A. Selection of data sample

The source of data for performing benchmarking analysis represent software tools that are practically tested for characteristics defined in the proposed benchmarking model.

Names of software tools for performing analysis are extracted from survey results, implemented among experienced software development employees as well among software project managers.

B. Survey participants characteristics

Prior to benchmarking, survey (Figure 1) has been conducted in aim to explore attitudes of experienced software development employees and software project managers regarding different aspects of distributed software development and project management, including software tools that they use in everyday practice.

3. The most frequent type of software project implementation:
a) (Teamwork) (single work)
b) (Collocated – at the common location) (distributed – at separate locations)

4. a) Did you use any tool for teamwork support, during your work on software projects?
YES NO

b) Which tools did you use?

5. a) Did you use any project management tool, during your work on software projects?
YES NO

b) Which tools did you use?

Figure 1. Part of survey questionnaire related to IT project conduction and tools usage [7]

Survey is conducted with questionnaire that was filled by 165 participants, where 32% are employed in foreign countries and 68% are employed in Serbia. Some of demographic data are represented as follows:

- 92 participants finished Bachelor level of studies, while 44 finished Master level and 16 are PhD.
- 72 participants are employed in IT companies, 36 in educational institution, 29 in business organization and 21 in government agency or institution.

It is important to emphasize examined experience of survey participants regarding teamwork, distributed work and using appropriate software tools:

- 112 participants had experiences in teamwork, while 26 mostly had experiences in single work engagements
- 51 participants worked collocated, 62 participants worked in distributed environment, while 6 participants worked equally in collocated and distributed work environments
- 87 participants used software tools for teamwork, while 61 participant never used teamwork supporting software tools
- 71 participants used project management software tools, while 73 participants never used project management tool.

VI. Results

A. *Survey results – tools selection*

According to previously presented questionnaire (question 3,4 and 5), survey participants entered answers related to their previous experiences in distributed teamwork, as well as their experiences with appropriate software tools that support teamwork and software tools that support project management. This way, two sets of software tools are extracted: software tools for software development teamwork support and software tools for project management support. By intersection of these two sets, the final set included names of software tools that are used for both purposes. These tools are then sorted by the number of occurrence in survey answers.

Most frequently tools from the intersection set were extracted and they are: Active Collab [14], Jira [20], Basecamp [15], RedMine [22] [23], dotProject [16], Microsoft Team Foundation Server [24] [25] [26], gForge [17] [18] [19], TeamViewer [26].

Analysis of each tool's description from appropriate tool's website led to the conclusion that Team Viewer could not be considered as project management tool, but as file sharing tool in distributed software development. Therefore, final selection of software tools for further analysis within benchmarking included:

- Active Collab
- Jira
- Basecamp
- RedMine
- dotProject
- Microsoft Team Foundation Server
- gForge

B. *Benchmarking results – tools ranking*

More detailed analysis of characteristics of each tool's website led to conclusion that these tools could be categorized as:

- General tools for teamwork and project management support: ActiveCollab, Basecamp, dotProject
- Tools for software development teamwork support and software project management support: GForge, JIRA, Redmine, Microsoft Team Foundation Server / Visual Studio Online.

In benchmarking analysis, each tool was analyzed for appropriate characteristic, according to previously presented benchmarking model, by answering appropriate question with Yes/No. Number of "Yes" answers was summarized and percentage is calculated, related to the maximum of possible "Yes" answers for each characteristics category.

In this paper, only tools from the second group (particularly related to software project management) are presented with summary results from benchmarking analysis (Table II).

TABLE II. Benchmarking Analysis – Summary Results

Category standard / Percent	GForge	Jira	RedMine	Microsoft Team Foundation Server
PMBOK	52	62	95	81
SWEBOK	62	69	77	92
BSC	40	50	90	70
TOTAL	52	61	89	82

According to results of benchmarking analysis, the most appropriate software tool that supports both teamwork in software development and software project management is RedMine. In the area of PMBOK characteristics this tool support is 95%, in SWEBOK 77%, BSC 90%.

VII. Conclusion

Distributed software development brings many challenges, where the most important are related to teamwork in software development, as well as related to software project management.

This paper presents a benchmarking model that could be used for analysis and evaluation of software tools that could support software development in both teamwork and project management dimension. The proposed benchmarking model is based on PMBOK, SWEBOK and BSC.

Applicability of the proposed model is illustrated with particular software tools, that are selected from the survey where participants were experienced software developers and software project managers.

Benchmarking analysis is performed upon the selected tools that support teamwork and project management for the particular area of software development: GForge, JIRA, Redmine, Microsoft Team Foundation Server / Visual Studio Online. Results of benchmarking analysis show that the most appropriate tool is RedMine, with 89% of needed characteristics, that are defined within the proposed benchmarking model.

Future research in this field could include research in the field of different areas of interest for defining relevant characteristics for benchmarking and extending number and quality of questions within the proposed model. Another direction for research could lead to new approaches and solutions for automated benchmarking of software tools.

VIII. Acknowledgment

This research is financially supported by Ministry of Education and Science of the Republic of Serbia under the project number TR32044 "The development of software tools for business process analysis and improvement", 2011-2014.

References

[1] Thomas J: Researching the value of project management, PMI research conference, Warshaw, Poland, 2008.

[2] Herbsleb J. D, Moitra D. (2001). "Global Software Development", IEEE Software (March/April 2001), pp. 16-20.

[3] Dhar S, Balakrishnan B (2006). Risks, Benefits, and Challenges in Global IT Outsourcing: Perspectives and Practices. Journal of Global Information Management, vol. 14, issue 3, Idea Publishing Group

[4] Kazi Lj, Radulovic B, Letic D, Radosav D, Berkovic I: „Applying SE-PM matrix in analysis of research in distributed software development", Global conference on Computer science, Software, Networks and Engineering, COMENG 2014, Antalya, Turska

[5] A Guide to the Software Engineering Body of Knowledge SWEBOK v.3 (2014). IEEE Computer Society

[6] A Guide to the Project Management Body of Knowledge PMBOK 5th edition (2013). International Project Management Association IPMA

[7] Ljubica Kazi: "Development of an Adaptable Distributed Information System for Software Project Management Support", University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, 2015, http://www.cris.uns.ac.rs/publicTheses.jsf

[8] Ajelabi I, Tang Y: "The Adoption of Benchmarking Principles for Project Management Performance Improvement", International Journal of Managing Public Sector Information and Communication Technologies (IJMPICT), Vol 1, No 2, December 2010

[9] Sim S.E, Easterbrook S, Holt R.C: "Using Benchmarking to Advance Research: A Challenge to Software Engineering", Software Engineering, 2003. Proceedings. 25th International Conference on, 3-10 May, 2003

[10] Say JAL International Consulting: "IT Project Management Benchmarking",November 12, 2008

[11] Dutta S, Kulandiswamy S, Van Wassenhove I.N: "Benchmarking European Software Management Best Practices", Communications of ACM, 1998, 41 (6): 77-86.

[12] Georgiannis V.C, Fitsilis P, Kakarontzas G, Tzikas A: "Involving Stakeholders in the Selection of Project and Portfolio Management Tool", (21ο Εθνικό Συνέδριο Ελληνικής Εταιρίας Επιχειρησιακών Ερευνών (ΕΕΕΕ), Αθήνα, 28-30 Μαϊου 2009) 21. National conference of Greek Operational Research (EUSR), Atina, 28-30 maj 2009

[13] Kaplan, R. S., Norton D. P., The balanced scorecard - Measures that drive performance, Harvard Business Review (January-February 1992.): 71-79.

[14] https://www.activecollab.com/

[15] https://basecamp.com/

[16] www.dotproject.net/

[17] https://gforge.com/

[18] https://gforge.inria.fr/

[19] https://gforgegroup.com/

[20] https://www.atlassian.com/software/jira

[21] http://blogs.atlassian.com/2013/06/using-watchers-and-mentions-effectively/

[22] www.redmine.org/

[23] www.easyredmine.com

[24] https://msdn.microsoft.com/en-us/vstudio/ff637362.aspx

[25] https://www.visualstudio.com/en-us/products/tfs-overview-vs.aspx

[26] https://www.teamviewer.com/sr/index.aspx

Real-time Biofeedback Systems: Architectures, Processing, and Communication

Anton Kos, Anton Umek, and Sašo Tomažič
Faculty of Electrical Engineering, University of Ljubljana, Slovenia
anton.kos@fe.uni-lj.si, anton.umek@fe.uni-lj.si

***Abstract*— Ubiquitous computing and wearable devices are offering new approaches in sports. Real-time biofeedback systems and applications can be used to speed-up the proper movement learning process. Movements must be processed in real time, what can pose a problem to small, light-weight, wearable devices with limited processing and communication capabilities, and battery power. We study two architectures of real-time biofeedback systems: local and distributed. The emphasis is on their properties when used in various biofeedback usage scenarios. Special focus is given on feedback loop delays and its real-time operation. A multi-user signal processing in football match is given as an example of high performance application that needs high speed communication and high performance computing. With growing number of biofeedback applications in sport, their complexity will grow, requiring new approaches in communication and motion processing.**

I. Introduction

Science and engineering offer new knowledge, expertise, and tools for achieving a competitive advantage in sports. The combination of wearable devices and ubiquitous computing are opening new dimensions in many areas of sport. One such example is the application of biomechanical *biofeedback systems*. One of the most common uses of biomechanical biofeedback is motor learning [2]-[3], which is based on repetition [1]. The primary focus of this paper is *real-time biomechanical biofeedback system* that can reduce the frequency of improper movement executions and speed up the process of learning the proper movement pattern.

The operation of biomechanical biofeedback systems largely depends on parameters of human motion and its analysis algorithms. Biomechanical biofeedback is based on sensing body rotation angles, posture orientation, body translations, and body speed. These parameters are generally calculated from raw data that represent measured physical quantities. Important parameters of human motion should therefore be adequately acquired by the chosen capture system (sensors). Sensors of the motion capture system should have: (a) sufficiently large dynamic ranges for the measured motion quantity, (b) sufficiently high sampling frequency that covers all frequencies contained in the motion, (c) sufficiently high accuracy and/or precision. The processing devices should have sufficient computational power for the chosen analysis algorithms. In biofeedback systems with real-time processing all computational operations must be done within one sampling period. When sampling frequencies are high, this demand can be quite restricting, especially for local processing devices attached to the user.

II. Challenges in Real-time Biofeedback

Real-time biofeedback can only be incorporated successfully when: (a) human reactions are performed in-movement, i.e., inside the time frame of the executed movement pattern, and (b) the biofeedback system operates in real-time with minimal delay. An ideal real-time biomechanical biofeedback system is an autonomous, wearable, lightweight system with large enough number of sensors to capture all the important motion parameters. Sensor signals exhibit high enough sampling frequency and accuracy. Processing is done instantly and the feedback modality is chosen in a way that it is not interfering with the principal modality of the motion.

The main challenges in real system implementation are often contradictory. For example, under the constraints of technology, the ideals of being wearable and lightweight contradict the ideals of autonomy and processing power.

One important challenge is related to the sampling frequency. Achieving high sampling frequency is generally not a problem, but it leads to large amounts of sensor data that needs to be first transferred to the processing device and then analyzed. Problems that may occur are available bandwidth of the communication channels and the computational power of the processing device. The later is especially a problem in real-time biofeedback systems. Here it should be noted that higher sampling frequency yields shorter sampling time T_s, thus allowing less time for the complete computation cycle needed for each sensor signal sample.

Communication channel bandwidth, range and delays are another set of potential problems. Low power wearable devices usually have low channel bandwidth, and limited communication range. In packet based technologies the delay is linearly proportional to packet length and inversely proportional to bandwidth. Longer packets and/or lower bandwidth cause higher delays.

III. Biomechanical Biofeedback Systems

The architecture of the biomechanical biofeedback system includes sensor(s), a processing device, actuator, and communication channels. Together with a user they form a biofeedback loop. The architecture can be shelled out from Figure 1 or Figure 2

Sensors are essential components of the system, which should be designed to work with one or multiple sensor devices. Sensors capture body movements and are attached to the user. They are the source of signals and data used by the processing device. Inertial sensor based systems are very common. It should be emphasized that inertial sensor based motion tracking systems are

generally mobile and have no limitation in covering space. Modern inertial sensors are miniature low-power devices integrated into wearable sensor devices. The *processing device* receives sensor signals, analyses them, and, when necessary, generates and sends feedback signals to the actuator (biofeedback device). Processing device is any device capable of performing computation on sensor signals. The computation can be performed in two basic modes: (a) during the movement; this mode requires processing in real time, (b) after the movement; this mode allows post-processing. The processing device can be located locally, on the user, performing *local processing* or remotely, away from the user, performing *remote processing*. The *actuator* uses human senses to communicate feedback information to the user. The most commonly used senses are hearing, sight, and touch. *Communication channels* enable communication between the independent biofeedback system elements. Although wireless communication technologies are most commonly used, wired technologies can also be used in practice.

A. Real-time biofeedback system groups

Real-time biofeedback systems can be divided into two basic groups on the grounds of processing device location. We denote a system with local processing as a personal biofeedback system, and a system with remote processing as a distributed biofeedback system.

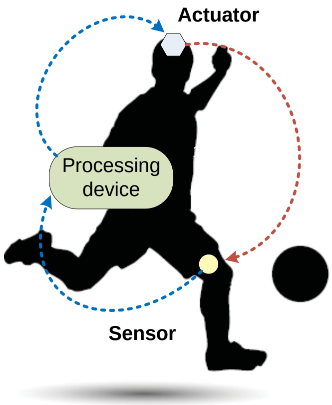


Figure 1. Personal biofeedback system. All devices of the system are attached to the user. Wearable processing device tends to be the most critical element of the system in terms of its computational power and/or battery time.

Personal biofeedback system is compact in the sense that all the system devices are attached to the user of the system and in close vicinity of each other, see Figure 1. Because the distance between devices is short, communication can be performed through low-latency wireless channels or over wired connections. The primary concern of personal biofeedback systems is the available computational power of the processing device. Personal version is completely autonomous. The user is not limited to confined spaces but free to use the system at any time and at any place.

In *distributed biofeedback system* sensor(s) and actuators are attached to the user's body, while the processing device is at some remote location, see Figure 2. The primary concern of distributed biofeedback systems are communication channel ranges, bandwidths, and increased loop delays. Distributed versions, especially the network one, have high computational power. With the local version of the system the user might be limited to a confined space if communication channel technology has short range.

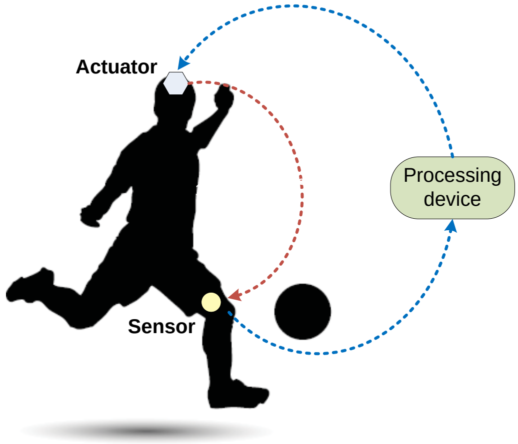


Figure 2. Distributed biofeedback system. Sensor(s) and actuators are attached to the user. The processing device is at the remote location, away from the user. Communication channels tend to be the most critical element of the system in terms of range, bandwidth, and delays or any combination of mentioned.

B. Local vs. remote processing

In wearable systems device energy consumption in the biofeedback loop is of the prime concern. One should consider choosing the system with the optimum energy consumption for the given task. According to [5] sensor devices consume many times more energy for radio transmission and memory storage than for local processing. This means that personal version with local processing could be more favorable option than distributed version with remote processing. Energy wise local processing at sensor device is very attractive, but there are some limitations that should be considered [5]:

(a) Algorithms developed in environments such as MATLAB are difficult to port to sensor devices.
(b) Sensor devices use fixed point microcontrollers for signal processing. Floating point operations must be simulated by using fixed point operation. This is slow and induces calculation errors.
(c) Total energy needed for all operations of one cycle could be higher than the energy needed for radio transmission of the raw data of the same cycle.
(d) Data from more than one sensor is processed by a single algorithm instance.
(e) Computational load of the algorithm could be too high to be handled by the sensor device.

When one or more of the abovementioned limitation apply, distributed system with remote processing is a better option. Its advantages are:

(a) The processing device has practically unlimited energy supply, high available processing power, and large amounts of memory storage.

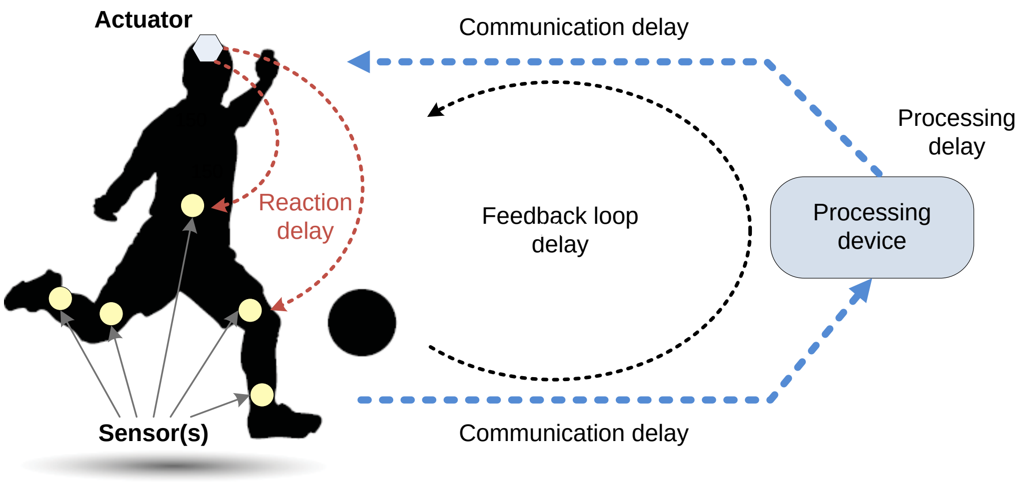


Figure 3. Real-time biofeedback system operation and delays. User movement (action) is captured by sensor(s) and their signals are sent to the processing device for analysis. Analysis results are sent to actuators, which use one of the human modalities to communicate the feedback to the user, who tries to react on it. The biofeedback system devices can control only the feedback loop delay, defined as a sum of all communication and processing delays of sensor(s), processing device(s), actuator(s) and communication paths.

(b) The processing is flexible in the terms of software environments usage, algorithm changes, algorithm complexity, choice of technology, choice of computing paradigm, etc.

(c) High performance computing (HPC) solutions can be used when the amounts of data increases and/or computational complexity rises.

C. *Delays and processing times in the biofeedback loop*

The delays in the real-time biofeedback system are illustrated in Figure 3. There are two basic points of view on delays in real-time biofeedback systems: (a) user's point of view and (b) system's point of view.

From the user's perspective the feedback delay occurs during the movement execution. This is the delay that occurs between the start of the user's action and the time the user reacts to the feedback signal. It is the sum of the feedback loop delay and the reaction delay and should be as low as possible. The feedback loop delay consists of system devices processing delays and communication delays between them. The processing delays of sensors and actuators are considered to be negligibly low comparing to communication delays and processing delay of the processing unit; therefore we consider them to be zero. The feedback loop delay should be a small portion of the human reaction delay, which depends on the modality (visual, auditory, haptic) used for the feedback. For example, reaction delays for visual feedback are around 250 ms for amateurs and around 150 ms for professional athletes [6].

Communication and processing delays within the feedback loop depend on the parameters of the equipment and technologies used. Some of the most important parameters are: sensor sampling frequency, processing unit computational power, communication channel throughput, and communication protocol delay. The feedback loop delay t_F is the sum of communication delays t_{c1} and t_{c2}, processing delay t_p, sensor sampling time t_s and actuator sampling time t_a:

$$t_F = t_s + t_{c1} + t_p + t_{c2} + t_a \quad (1)$$

In general the biofeedback system operates in cycles that are equal to sensor sampling time. If we want to ensure the real time operation of the system, processing time should not exceed sensor sampling time $t_p \leq t_s$.

IV. COMMUNICATION

Motion capture systems can produce large quantities of sensor data that are being transmitted through communication channels of a biofeedback system. When real-time transmission is required, the capture system forms a stream of data frames with sensor's sampling frequency. In real-time biofeedback systems two main transmission parameters are important; bit rate and delay. While bit rate depends on the technology used, delay t_{delay} depends on signal propagation time t_{prop}, frame transmission time t_{tran}, and link layer protocol t_{MAC} or medium access control protocol (MAC).

$$t_{delay} = t_{prop} + t_{tran} + t_{MAC} \quad (2)$$

At constant channel bit rate R, the transmission delay is linearly dependent on frame length L.

$$t_{tran} = \frac{L}{R} \quad (3)$$

Propagation time on different transmission media is 3.3 to 5 nanoseconds per meter and is sufficiently small to be neglected. MAC delays vary considerably with channel load, from a few tens microseconds to seconds. In lightly to moderately loaded channel MAC delays are below 1 ms. In most cases that leaves the transmission delay as the main delay factor.

Personal biofeedback systems can use wires or body sensor network (BSN) technologies that have bit rates from a few tens of kilobits per second up to 10 Mbit/s [5]. Considering the projected sampling frequency of 1000 Hz, that yields the maximal possible frame size in each sampling period in the range of a few tens of bits (a few bytes) up to 10,000 bits (1250 bytes). The range of BSN is typically a few meters.

Distributed biofeedback systems use various wireless technologies with bit rates from a few hundreds of kilobits per second up to few hundreds of Mbit/s [7]. Considering the projected sampling frequency of 1000 Hz, that yields the maximal possible frame size in each sampling period in the range of a few hundreds of bits up to 100,000 bits. The range of considered wireless technologies is between 100 m (WLAN) and a few kilometres (3G/4G).

V. Real-time processing

In real-time biofeedback systems the processing device is receiving a stream of data frames with inter-arrival times equal to sensors' sampling time. To assure real-time operation of the system, the processing time of any received frame should not exceed the sampling time. This depends on many factors: computational power of the processing device, sampling time, amount of data in one streamed frame, number of algorithms to be performed on the data frame, complexity of algorithms, etc.

In section III we have studied the trade-offs between the local and remote processing of biofeedback signals. While many examples of biofeedback applications, that do not require a lot of processing, exist, enough opposite examples can be found. One example is a real-time biofeedback system for a football match. Parameters in the capture side of the system are: 22 active players and 3 judges, 10 to 20 inertial sensors per player, 1000 Hz sampling rate, up to 13 DoF data. The data rate produced is between 92 Mbit/s. and 200 Mbit/s. Such data rates can presently be handled only by the most recent IEEE 802.11 technologies that promise bit rates in Gbit/s range.

The presented example clearly implies powerful processing device and high speed communication channels. Algorithms that are regularly performed on a streamed sensor signals are [8]-[10]: statistical analysis, temporal signal parameters extraction, correlation, convolution, spectrum analysis, orientation calculation, matrix multiplication, etc. Processes include: motion tracking, time-frequency analysis, identification, classification, clustering, etc. Algorithms and processes can be applied in parallel or consecutively.

Delay is the primary parameter defining the concurrency of a biofeedback system, as viewed from the user's perspective. The feedback delay, which is the sum of all delays of the technical part of the biofeedback system (sensors, processing device, actuator, communication channels), should not exceed a small portion of the user's reaction delay. To present two exemplary calculations for movements with high dynamics and movements with low dynamics. Let us set the maximal feedback delay at 20% of user's reaction delay. Considering that the reaction delay is around 150 ms [6], the maximal feedback delay is at most 30 ms. the two examples are:

High dynamic movement: biofeedback system sampling frequency should be high, for example 1000 Hz. Samples of captured motion are occurring every millisecond, accordingly the processing device must calculate one result every millisecond. The processing device receives a new frame of sensor data every millisecond and it has 1 ms to perform all the calculations, leaving 29 ms for the communication path delays. The implementation of high dynamic real-time biofeedback systems is feasible, if the processing device has enough processing power, the communication delays should not be a problem. The use of the *distributed version* of the biofeedback system is more likely.

Low dynamic movement: biofeedback system sampling frequency can be set relatively low, for example at 40 Hz. Samples of captured motion are occurring at 25 ms intervals, accordingly the processing device must calculate one result every 25 ms, leaving only 5 ms for the communication delays. The implementation of real-time biofeedback systems is feasible, if the communication delays are low, the processing power should not be a problem. The use of the *local version* of the biofeedback system is more likely.

VI. Conclusion

Science and advanced technology with connection to ubiquitous computing are opening new dimensions in many areas of sport. Real-time biofeedback systems are one such example. To assure the operation in real time, the technical equipment must be capable of real-time processing with low delay within the biofeedback loop. Challenges are present in all phases of real-time biomechanical biofeedback systems; at motion capture, at motion data transmission, and at processing. With growing number of biofeedback applications in sport and other areas, their complexity and computational demands will grow, possibly requiring new approaches in communication and processing paradigms.

References

[1] Oonagh M Giggins, Ulrik McCarthy Persson, Brian Caulfield, Biofeedback in rehabilitation. Journal of NeuroEngineering and Rehabilitation, 2013

[2] Baca, A., Dabnichki, P., Heller, M., & Kornfeind, P. (2009). Ubiquitous computing in sports: A review and analysis. *Journal of Sports Sciences*, 27(12), 1335-1346.

[3] Lauber, B., & Keller, M. (2014). Improving motor performance: Selected aspects of augmented feedback in exercise and health. European journal of sport science, 14(1), 36-43.

[4] Umek, Anton, Sašo Tomažič, and Anton Kos. "Wearable training system with real-time biofeedback and gesture user interface." Personal and Ubiquitous Computing, 2015, 10.1007/s00779-015-0886-4.

[5] Poon, C.C.Y.; Lo, B.P.L.; Yuce, M.R.; Alomainy, A.; Yang Hao, "Body Sensor Networks: In the Era of Big Data and Beyond," in Biomedical Engineering, IEEE Reviews in , vol.8, no., pp.4-16, 2015.

[6] Human Reaction Time, The Great Soviet Encyclopedia, 3rd Edition 1970-1979.

[7] IEEE 802.11 standards. Available online: http://standards.ieee.org/about/get/802/802.11.html. Accessed on 22.12.2015

[8] Lee, James B., Yuji Ohgi, and Daniel A. James. "Sensor fusion: Let's enhance the performance of performance enhancement." *Procedia Engineering* 34 (2012): 795-800.

[9] Wong, Charence, Zhi-Qiang Zhang, Benny Lo, and Guang-Zhong Yang. "Wearable Sensing for Solid Biomechanics: A Review." *Sensors Journal, IEEE* 15, no. 5 (2015): 2747-2760.

[10] Min, Jun-Ki, Bongwhan Choe, and Sung-Bae Cho. "A selective template matching algorithm for short and intuitive gesture UI of accelerometer-builtin mobile phones." In *Nature and Biologically Inspired Computing (NaBIC), 2010 Second World Congress on*, pp. 660-665. IEEE, 2010.

Healthcare Information Systems Supported by RFID and Big Data Technology

Martin Stufi*, Dragan Janković**, Leonid Stoimenov**
*Solutia, s.r.o., Prague, Czech Republic,
**Faculty of Electronic Engineering, Niš Serbia,
martin.stufi@solutia.cz, dragan.jankovic@elfak.ni.ac.rs, leonid.stoimenov@elfak.ni.ac.rs

Abstract **– This paper proposes a secured radio-frequency identification for smart healthcare environment with specialized system functionalities in order to enhance the quality of health service by improving healthcare, unique patient identification, specialized message delivery and activity related with patients rescue. Big Data part is presented for the analytic and data processing purposes in Health Care Information System, for massive data analysis and their visualization. In this paper we demonstrate how proposed prototype of information system can significantly improve daily emergency and healthcare operations and shows its benefits.**

I. INTRODUCTION

Radio Frequency Identification (RFID) [1,2] and Near-Field Communication (NFC) [2] is the fastest incoming technology in the recent years used to identify items, assets, products, humans or animals. RFID is the technology, which allows unique identification by using radio waves. NFC technology is specialized subset of that technology within the family of RFID technology. NFC is a branch of High-Frequency (HF) RFID and it has been designed to be a secure form of data exchange and NFC device, and is capable of working as NFC reader and an NFC tag. Any system based on RFID or NFC technology [2] can be successfully deployed in many sectors where the emphasis is on the fastest and accurate information processing and immediate transfer of the loaded data for subsequent processing.

Big Data [3, 4] definition is relative today, very attractive and popular. Big Data can be defined as all data that is not a fit for a traditional Relational Data Base Management System (RDBMS), regardless of its use, for online transaction processing (OLTP) or analytic purpose. Big data are not only associated with the size of the data, but most important with the format of the data.

Medical technologies are used by healthcare organizations and their workers to enhance operational efficiency and reduce workload on their professional side. The level of information technology used in medical institutes is already quite high, and based on that the amount of data are growing in line with Big Data trends.

Most researches of the applications of RFID technology in medicine are focused on emergency to automatically identify: people – locating hospital staff and patients tagging for error prevention, objects as blood products, medication and locating assets. RFID is technology that saves lives, prevents errors, saves costs and increases security.

To support a standard medical and specialized hospital practices we are moving from relatively ad-hoc and subjective decision making to the new challenges in healthcare based on Big Data technology.

As an example of taking care of a patient through specialized conceptual Hospital Information System we demonstrate usage of RFID and NFC elements through specialized mobile applications running on smart phones, tablets, and also by using standard PC mobile clients. An Android/iOS based client is used to identify, collect and exchange patient healthcare information with the backend Hospital Information System based on Big Data principles. Within this utilization it is advantageous to use options of RFID/NFC and establish control of taking medical patient care through the Specialized Hospital IS and its life cycle.

This paper is organized as follows: Section 2 gives details about RFID and NFC and their physical characteristics. Section 3 introduces use case scenario of emergency care and smart care hospital principles and its processes. Section 4 presents the system architecture of proposed solution. Section 5 and Section 6 describe principle of Smart Health Care Hospital IS and its functionalities and also new Big Data challenge in Healthcare Information Systems.

II. SMARTER HEALTH

Nowadays healthcare is becoming smarter and it revolutionizes our life. The smart healthcare hospital offers a number of advantages:

- Provides a beneficial strategy for the better healthcare services
- It helps to manage and integrate complex healthcare functionality
- It helps to integrate ICT technology, their products and services
- It supports developing and building new educational models and learning/teaching strategies at no risk to patients
- Continual emphasis on patient and welfare environment and satisfaction

Deployment of smart healthcare hospital information system in particular hospital settings will involve development and implementation of modern technologies and their integration with existing information systems.

A. Big data challenges in healthcare

Big data is a broad term for data sets, so large or complex that traditional data processing applications are inadequate. One way to describe characteristics of Big Data and help us to differentiate data categorized as “Big” from other forms

of data is through the five letters V - “5Vs” [3]. They are Volume, Velocity, Variety, Veracity and Value.

Data volume attributes provide representation of the large volume of data being created on daily bases. Velocity attribute of big data is about fast speed of data, which arrive, and are accumulated within the short period of time. Variety is about the multiple formats and data types that need to be supported by information systems. Veracity refers to the quality or fidelity of data and Value is dependent on how long data processing takes.

The amount of healthcare data that exists in Healthcare information system is growing faster than we expected based on the large amount of historical data, data driven by record keeping, based on legal, regulatory and compliance requirements. Rapid digitization in healthcare is one of these massive quantities of data, which are in line with big data described in primary attributes, “5Vs”.

Big Data Challenges in Healthcare are as follows:

- Deducing knowledge from complex and heterogeneous sources for the patients.
- Understanding unstructured clinical notes in the right context.
- Efficiently handling large amount of medical data.
- Analyzing medical data.
- Capturing the patient’s behavioral data through several sensors.

B. *RFID and NFC sensing technology*

RFID and NFC are wireless sensing technologies based on electromagnetic signal detection with automatic identification. Radio frequencies are used for communication with other appliance or items called RFID tag, by using RFID reader. RFID tag is a small object like adhesive sticker that can be attached or incorporated into a product. RFID tag has antenna connected to an electronic module (chip). RFID reader is device which can read unique information stored in RFID tag and transceiver to respond by sending back information to backend information system.

Passive RFID tags are priced significantly cheaper, have different special reading distance from 0.5m to 10m, long lifetime tag using method. Tags that operate at the highest frequency UHF, have radius - about 3-10 m. On the opposite side, the lowest frequency of 125 kHz LF have a range of only about 0.5 m. Therefore passive RFID tags are widespread used in many hi-tech solutions. Due to a several applications of RFID technology in medical identification application, we had a passive RFID wristband bracelet in the 13.56 MHz, which is also NFC enabled bracelet.

C. *Use Case scenario in emergency care and smart care hospital*

RFID or NFC technology can contribute to create Smart Hospital Healthcare Information System [5]. We can distinguish two use case scenarios:

- Simplified Emergency Care
- Simplified Hospital Care

In Simplified Emergency Care scenario, on emergency arrival, patient receives a wristband with an embedded RFID tag storing all the information related with emergency occurrence. Rescuer collects patient's information as patient’s identity, patient's position, insurance company, photos, as well as rescuer identity, and transmits the information to the back end system. Rescuer and other staff members wear a smart RFID badge storing their Id number, rescue car identification etc.

In Simplified Hospital Care scenario, doctors, nurses, caregivers and other staff members also wear a smart RFID badge storing their Id number, Name, Specialization, etc. On arrival, each patient has stored unique identifier, and other personal information as Patient Id, Name, Surname, Gender, Date of Birth, Blood type, Insurance Company Id etc.). All the patients' medical procedures, medical histories and other important information are stored in Smart Healthcare Information System based on RFID label.

III. SYSTEM ARCHITECTURE

Proposed system encompasses three-tier architecture combined with front-end devices (RFID tag, RFID reader). As a client, Smart Health IS uses mobile device (Smartphone, Tablet, RFID reader), wireless communication technologies to integrate and simplify human to application and human to human interaction.

This concept shows secured communication based on modern and new technologies (RDBMS and Big Data) for emergency service and their communication for Internet and intranet users. Smart Heath care Information System architecture is shown on the Figure 1.

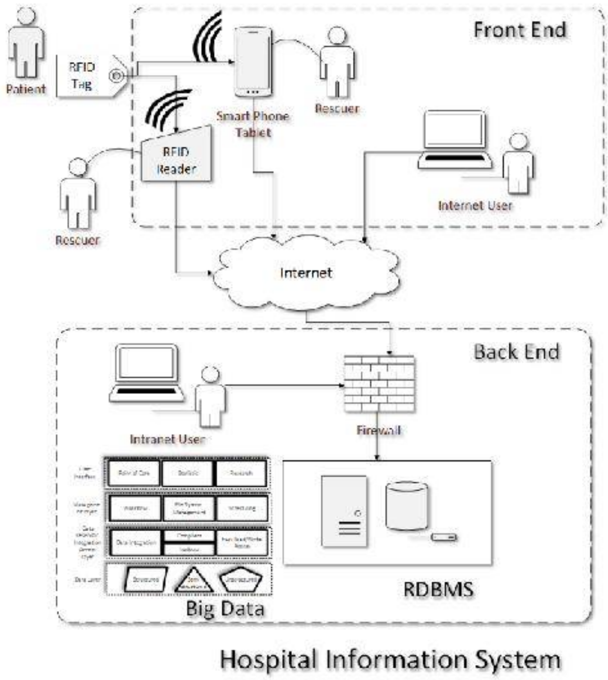


Figure 1. Smart Healthcare High Level Architecture

The architecture of a system for Smart Health Care should have the Front-End part and Back-End part of Hospital Information System.

Front-End part of Hospital Information System has following modules:

- Smart phone/tablet with an Android, iOS or Microsoft OS based smart platform and specialized mobile application shown in Figure 2, Figure 3 or Figure 4. This is the entry point of data which are collected

based on the information from “RFID Tag”. In this stage Rescuer collects the data as follows: Patient’s Id, Name, Surname, Gender, date of Birth, Country, City, Social Security number, Insurance Company etc. The data collection is fully automated based on RFID tag (Wristband) and the information a patient has given to the Rescuer. Afterwards, collected data are transferred to the Back-End Information System via the Internet. After transfer and collection of data in Back-End part, data are being processed on daily bases in “Smart HCIS”. Hospital Information System generates Datasets [3], which are being transferred to the Big Data part of “Smart HCIS” for data analysis, processing and visualization.

- Mobile device - this is also Front-End part of the information system which is Microsoft OS, Linux or OS X based connected with RFID Reader via Bluetooth or WiFi. The Data collection is organized via RFID Reader to the Mobile Client and also transferred as in previous scenario to the Back-End system via Internet.

In Back-End part of the Hospital Information System we are distinguishing two functional parts:

- Data processing part – this module gathers patients data and also collects all the data concerning examination and therapy phases of a patient healthcare life cycle.
- Big Data part – this module gathers structured, unstructured and semi-structured healthcare and operational data about patients, diseases, provided medical services, financial and operational data, clinical data, patient’s treatments, risks for diseases. Structured data are data that can be easily stored, queried, analyzed and this data are stored in Relational DBMS. Unstructured and semi-structured data are representing the information that typically requires a human touch to read, capture and interpret properly by analytic tools in this Big data part of Healthcare Information System. This module includes medical imagining, laboratory, pharmacy, insurance and other administrative patient’s data.

IV. FRONT-END ANDROID SMARTPHONE CLIENT – “SMART-HCLIENT”

This section presents front-end client, named “Smart-HClient”. This application is developed for smart devices with Android OS based on contact-less technology (near-field communication) [6]. Application improves connectivity and coordination between medical team members, helping clinicians respond faster to critical patient events. Smart client (smartphone and RFID reader) allows rescuing worker to send and receive emergency messages at any time and any place. The worker also collects patients’ information as Name, Surname, Gender, Location Coordinates, Blood type, Insurance Company Id etc.

Smartphone application as “Smart HClient” for emergency and hospital message delivery is registered with Smart healthcare information system within account name and password. Once the registration is completed, the system checks whether the user operating this device is sending or receiving relevant emergency messages correctly.

The smartphone operation is described as follows:

- Healthcare employee activates the smart healthcare notification system with their smartphone or RFID reader.
- The employee inputs his username and password.
- Smart health care information system verifies the worker’s identity.
- Once the logon procedure is completed, the smartphone, RFID client shows the system menu (see Fig. 2, Fig. 3, Fig. 4). The system then starts the message collection and transceiving.
- Upon reaching the patient, the healthcare worker starts the rescue. First, the healthcare worker confirms patient’s identity. Second, the healthcare worker checks and know what’s the correct rescue that patient really need.
- Smartphone application client has its local database and independently collects the information. Client is communicating with the backend Smart Healthcare information system based on asynchronous principle through web services published on the backend information system and unifies collected information.
- Once the patient is registered by the Smart Healthcare IS and accepted by a hospital medical staff responsible for his examination, the lifecycle continues inside the Smart Healthcare IS through the smartphone application or web browser as a light client.

Figure 2. “Smart-HClient” Patient information

In the Therapy part of “Smart-HClient” application, values as Temperature, Blood pressure, Time, Date, Patient’s Id, Location, are being measured on daily basis. This data are structured and transfered to the local database of Healthcare Information System. This data is at the same

time being transferred to the Big Data part of information System for processing and analysis by Big Data Analysts [3].

Figure 3 - “Smart-HClient” Patient therapy

On the patient examination part of Front-End application, there are collected information about diagnosis, medication, Time, Data, Description of disease, shown on Figure 4.

Figure 4 - “Smart-HClient” Patient examination

Patient dashboard part shows the graphical presentation of tracked and followed patient’s parameters from previous parts as Examination and Therapy. It allows medical staff to make analysis of patient’s health checks in predefined time period, and to make decision for further examination based on previous results, shown in Figure 5.

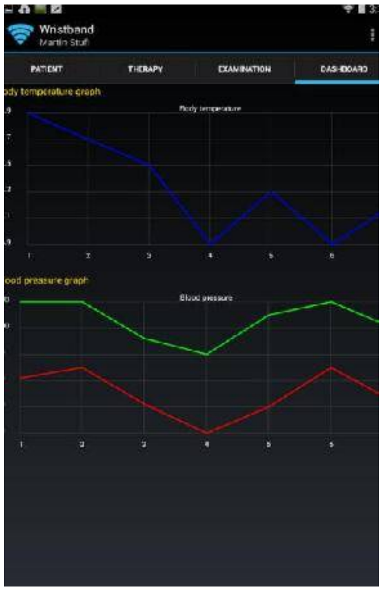


Figure 5 - “Smart-HClient” Patient dashboard

V. BACK-END SMART HEALTHCARE IS – “SMART HCIS”

Backend Smart Healthcare IS can be developed on the best fitting technology, which supports a variety of ways to build Business services, Web services, Java objects, Business Components. In our use case demands, Back-End application was developed such as Service Oriented Architecture [7] with minimal integration effort based on Oracle GlassFish application server for J2EE enterprise application. Our framework follows Model-View-Controller [8] design paradigm. The layered architecture of the framework simplifies maintenance, decouples implementations from interfaces, and improves reusability of the components across applications.

Key functionalities like Create, Edit, Refresh and Manage of Back-End system are shown in Figure 6 – “Smart HCIS” Patient record and Figure 7 - “Smart HCIS” Patient examination.

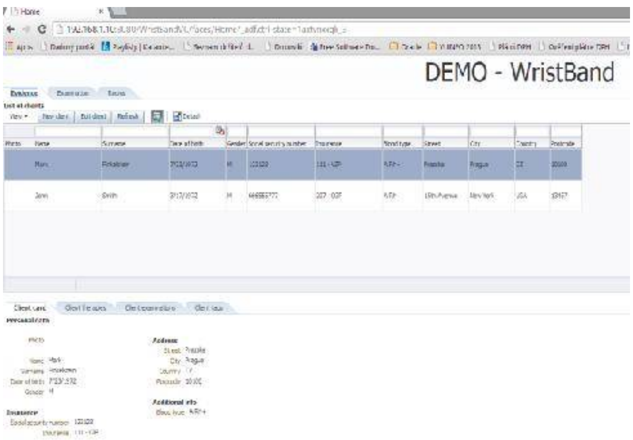


Figure 6 – “Smart HCIS” Patient record

Information related to the patients’ healthcare is in the Evidence part, shown in Figure 6. Evidence part collects the data as follows:

- Patient's records (Patient Photo, Patient Cards, Insurance Company, Blood Type Information, Address, Examinations, Therapies, Tags, etc.)

Examination part, shown in Figure 6 collects the patient data as follows:

- Examination (Temperatures, Pressures, Medications, etc.)
- Rooms (Patient location information, Hospitalizations, etc.)
- Medical staff administration

Administration part, shown in Figure 7 is related with application administration of:

- Web Services address settings
- Dashboard settings
- Administrator's console

Web Services Administration part allows administration tasks related with configuring the enterprise application that contains Web Service, starting and stopping deployed application, configuring the Enterprise application, such as the deployment order, session time-out for Web application or transaction type for Enterprise Java Beans (EJBs). Also, this part of administration is needed for creating, updating, monitoring and testing the Enterprise application.

Dashboard Settings is a part of Hospital Health Care Information System related with the application administration. Based on the set of parameters via dashboard, users are able to see visualized [9] and fit data related with a patient as Temperature indicator, Blood pressure and another performance measures which have to be monitored.

Administrator console is a Web browser-based, graphical user interface which is being used to manage GlassFish server domain and their instances, including Web-Services, that are deployed to the server or cluster.

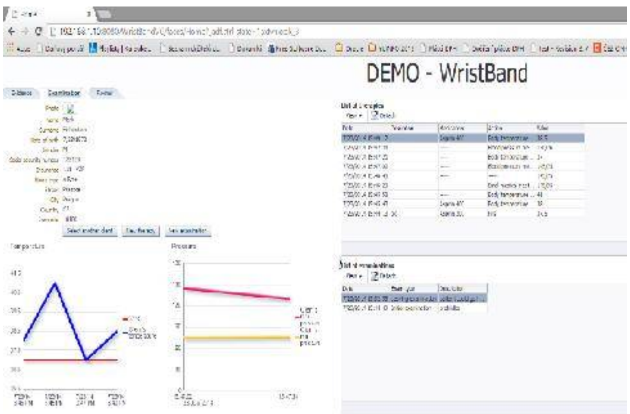


Figure 7 - "Smart HCIS" Patient examination

There are analytic tools in Big Data part of Healthcare Information Systems and their purpose is to derive values from big data and also to help Data Analysts and Data Scientist to prepare valuable information in order to make smart and right decisions by management, which brings competitive advantage for the hospital.

VI. CONCLUSION

This article provides an overview and functioning example of RFID/NFC based mobile medical patient tracking and diagnosis Smart Healthcare Information System and its relationship to Big Data architecture for patient's care, analytic, statistic and research purposes.

Main benefits of smart devices like smartphones, tablets, RFID/NFC readers and other new technologies improve connectivity, responsibility and coordination between medical or hospital team members, helping clinicians respond faster to usual and critical patient events inside of smart healthcare information system. Smart devices provide an important delivery mechanism for combing alert messages with graphic presentation resulting in positive impact on workflow.

Big Data has huge potential to transform and improve the way for a healthcare information system by using sophisticated technology like RFID/NFC. Big Data therefore becomes mainstream in information system governance.

REFERENCES

[1] Igoe T., *"Getting Started with RFID"*, O'Reilly Media, Inc., 2012, pp. 1-4.

[2] Finkenzeller, K. *"RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication, Third Edition"*, John Wiley & Sons, Ltd., 2010, pp. 6, pp. 57–59, pp. 339-346.

[3] Erl T., Wajid K., *"Big data Fundamentals: Concepts, Drivers & Techniques"*, Prentice Hall, 2015, pp. 19, pp. 21, pp. 29.

[4] Iafrate F. *"From Big Data to Smart Data"*, John Wiley & Sons, Inc. 2015, pp. 18-19.

[5] Winter A., Haux R., *"Health Information Systems: Architectures and Strategies"*, ©Springer-Verlag London Limited 2011, pp. 1.

[6] Mier R., *„Professional Android™ 4 Application Development"*, John Wiley & Sons, Inc., 2012, pp. 28, pp. 693.

[7] Erl T., "*SOA Design Patterns*", 1st ed., Prentice Hall, 2009, 35.

[8] Phillips B, Hardy B., "*Android Programming: The Big Nerd Ranch Guide"*, Big Nerd Ranch, Inc., pp. 2013, pp. 35-37

[9] Marr B., *"Big Data: Using smart big data, analytics and metrics to make better decisions and improve performance"*, John Wiley & Sons Ltd., 2015, pp. 97.

Survey of Internet-of-Things platforms

Milan Zdravković*, Miroslav Trajanović*, João Sarraipa**, Ricardo Jardim-Gonçalves**, Mario Lezoche***, Alexis Aubry***, Hervé Panetto***
* Laboratory for Intelligent Production Systems (LIPS),
Faculty of Mechanical Engineering, University of Niš, Niš, Serbia
** CTS, Uninova, Dep. de Eng. Electrotécnica, Faculdade de Ciências e Tecnologia, FCT ,
Universidade Nova de Lisboa , Caparica , Portugal
*** Research Center for Automatic Control of Nancy (CRAN), Université de Lorraine, UMR 7039, Boulevard des Aiguillettes B.P.70239, 54506 Vandoeuvre-lès-Nancy, France.

milan.zdravkovic@gmail.com, miroslav.trajanovic@masfak.ni.ac.rs
{jfss, rg}@uninova.pt
{mario.lezoche, alexis.aubry, herve.panetto}@univ-lorraine.fr

***Abstract* – Internet-of-Things (IoT) platform (often referred to as IoT middleware) is a software that enables connecting the machines and devices and then acquisition, processing, transformation, organization and storing machine and sensor data. The objective of the research behind this paper is to establish a state of the art in the development of IoT platforms. In specific, here we present the main conclusions regarding the functional and design perspective to current IoT platforms and related research. The focus was made on cloud-based IoT platforms with significant user base and successful record of exploitation. We also consider the relevant theoretical foundations of IoT platform research, mostly by taking into account the results of European research. The conclusions and respective discussion are meant to be used in the development of a formal runtime model-driven IoT platform, which conceptual design decisions are shortly presented.**

I. Introduction

Internet of Things (IoT) has attracted attention of major players in industrial landscape and it is currently one of the most expected emerging technologies. According to the 2015's Gartner's Hype Cycle for Internet of Things [1], IoT is currently at the "peak of inflated expectations". Although some of its technologies and enablers are getting closer to the so-called Plateau of Productivity (Wireless Healthcare Asset Management, ZigBee, RFID for logistics and transportation), IoT platforms are still at inception phase of development. The potential market impacts are considered as quite significant.

The development of IoT platforms is driven by the need to facilitate machine-to-machine (M2M) connectivity, which is emerging at unprecedented rate. Machina Research [2] predicts that M2M connections will rise from two billion in 2012 to 12 billion in 2020. Cisco [3] values IoT market to 19 trillion USD. According to the same source, only 0.6% of the physical objects the potential candidates for IoT are currently connected. Different sources refer to estimated 50 billion objects online on 2020.

The previously mentioned Gartner's study [1] suggests that the need for IoT platforms is most precisely articulated in the sector of consumer-centered enterprises, which are currently classified into early adopters of IoT. According to Gartner, in 5 years, 2-3 of 10 homes will be connected homes, with about 500 connectable devices. In service sector, the increase in overall effectiveness of employees and workplaces has been sought as the key expectation from IoT. Intelligent business operations, machine learning and RFID for logistics and transportations are foreseen as the earliest facilitators (within 5 years). Industrial sectors expect a significant impact of IoT on factory floors; this impact will be driven in a short term by enterprise manufacturing intelligence and facilities energy management (within 5 years).

Obviously, all these devices and technologies need a platform which will act as a command centre in homes, workplaces and factory floors. Google Nest and Apple Homekit are some of the examples of "home" aggregators, IoT platforms capable to implement home automation functionality. IoT platform (often referred to as IoT middleware) is thus defined as a software that enables connecting the machines and devices and then acquisition, processing, transformation, organization and storing machine and sensor data.

There exists a strong need for quite diverse sets of methodologies and tools for effective deployment of IoT solutions on different scales. In fact, large-scale deployment of IoT projects under realistic conditions has been considered only recently [4]. The set of methodologies and tools deployed in the specific domain, as the specific IoT solution, is typically referred to as an integrated IoT platform. IoT platform will be built within the complex ecosystem of machines, software and people, dealing with different relevant issues, spanning from M2M connectivity, to data analytics and visualization, IoT application development and others. The most important common features of the different domains are massive scaling and security.

Main objective of this paper is to define the key directions in the development of a formal model-based, generic IoT platform. These directions are presented in the conclusions of the paper (Section 5). They are made based on the discussion (Section 4) of the theoretical foundations for IoT platforms (presented in Section 2); and existing IoT platforms (Section 3).

II. Theoretical Foundations for IoT Platforms

Stankovic [5] highlighted the following research directions for IoT:

- massive scaling (addressing, discovery, architectural models that can support the expected heterogeneity),
- architecture and dependencies (IoT apps, deployment, resolving interference problem in using the utility device from different apps by some kind of multiplexing, dependencies across applications especially for safety critical apps or when actuators can cause harm),
- creating knowledge and big data (real-time data interpretation, knowledge formation, new inference techniques, trusting data by using confidence levels, reliable data associations),
- robustness and openness,
- security (detection and diagnosis of attack and deployment of countermeasures),
- privacy (evaluate requests against policies, reconciliation of the different policies) and
- humans in the loop (modeling human behaviors, human use and control).

Obviously, each of the listed directions is highly relevant for the IoT platform. In fact, the design of one IoT platform must take into account all these directions and its development approach must be defined so IoT platform becomes an enabler of each of the above factors.

As the technologies and standards used for IoT device manufacturing and communication are still in very early phase of development and adoption, IoT platforms should embrace a role of IoT experimentation facilities. Gluhak et al [4] identified the requirements for a next generation of experimental research facilities for the IoT:

- scale (supporting thousands of nodes: minimized human intervention, maximized plug-and-play configuration, automatic fault management),
- heterogeneity (management of devices, easy programmability of heterogeneous devices),
- repeatability (across different test beds: agreements on standards),
- federation (with other test beds, or other experiments: common framework for authentication, interoperability),
- concurrency (virtualization of devices, multiple experiments for one device),
- experimental environment (robustness towards the environmental conditions),
- mobility (handling system dynamics, movement of devices) and
- user involvement and impact (multi-modal mechanisms for user feedback, automated detection of situations where user behavior influences the data validity).

III. Cloud-based IoT platforms and services

In this section 16 different cloud-based IoT platforms and services are presented, with short overview of their distinguishing architectural designs and functionality. Initial selection of platforms was done by a Google search. Then, based on the analysis of the website content, the selection is filtered to the platforms with a significant customer base and partnerships with device manufacturers and system integrators.

Some platforms without M2M connectivity features are included, considered as IoT support platforms. Domain-specific platforms are not presented in this paper. The platforms are listed in alphabetical order.

Arrayent. The IoT platform is composed of four components. Connect Agent is a firmware, a lightweight agent deployed in devices (bulk firmware updates enabled). It exchanges data with the Connect Cloud by using 128-bit AES encryption. Each of these devices has its own digital copy in Connect Cloud which hosts the virtual devices to which mobile apps can connect. Mobile Framework is used for development of apps which manage connected devices. It uses also engine for managing and sending triggered alerts that could also then trigger response actions in the product which generated the alert. Finally, Insights provides secure access to data via dashboards, batch exports, data streaming and data connectors.

Axeda. Connectivity middleware facilitate connecting machines and devices to the cloud. Application enablement platform simplifies development of IoT apps, with capabilities such as data management, scripting engine, integration framework, SDKs and web services for accessing data and apps in the cloud. Connected machine management applications facilitate remote monitoring, management, service and control of remote devices. Capabilities also include software (client, firmware) distribution and configuration management.

Bugswarm is a lightweight platform that can acquire data from and control devices using JavaScript or plain HTTP. It defines a “swarm” – system of resources which can communicate to other resources within the system, according to the defined access policy. A resource is considered as anything that can communicate through HTTP, not only devices but also web or mobile applications. Device-specific, client-side applications, device connectors are available for use, to connect existing device as a resource to a swarm. When the specific device is connected, it sends the private message to all swarm members, with the list of its capabilities or services that the device can provide (feeds). Other resources interested in these services could send a feed request to a device, which then responds with a feed response (typically, with sensed data).

Carriots. The platform is an aggregator; it enables connecting any type of device with web connectivity which can send a stream of data, by using MQTT, cURL, hURL or Poster, to Carriots REST API. For each of the protocols, a client installation is needed on the device. Then, Listener or Trigger component can be developed and deployed on platform to perform operations with data. Device control and maintenance is enabled (checking status, managing configurations, updating firmware). All development is being done by using Java Carriots SDK, by putting a code to the specific fields in Carriots Control Panel web application (Java code interpretation/execution). Free use is enabled with limited functionality (up to 10 devices).

Evrythng is natively a digital identity management platform, often referred to as “Product Relationship Management” (PRM) platform. Semantic data store is

used to customize dynamic data profiles – digital identities of the products, so they can exchange data with authorized applications.

Exosite is cloud-based IoT platform offering M2M connectivity and data visualization tools and services. Open API is available for advanced data processing and integration with enterprise applications.

GrooveStreams is data analytics cloud platform, allowing data collection from multiple platforms, including IoT devices. Open API can be used to send data streams at a fixed (up to 10 second) or random interval or as a point stream of a fixed value. Data analytics tools are offered with near real-time performance. Data can be redistributed as Derived Steams, or visualized with customizable charts and graphs. The platform is open access. Premium features are available, related to number of organizations, users and increased (scalable) data I/O rate.

Ifttt (If This Then That) is not a native IoT platform; it is an interoperability-as-a-service platform which allows users to create chains of simple conditional statements, called “recipes”, which are executed upon the particular events recorded from the different services. It is the platform which enables users to create their own recipes, which can also include events from the different devices. Some of the existing examples of IoT related recipes are: “delay watering your garden if it’s going to rain tomorrow”, “receive and emergency call if smoke is detected”, and others. There also exist alternatives to this service, such as Zapier and Yubnub.

Kaaproject is open-source IoT middleware platform. It enables management and maintenance (firmware updates distribution) of device inventory and near real-time communication between the devices. It is transport-agnostic and promotes use of structured data. It provides endpoint SDKs that can be embedded into devices. Complete solutions already exist for Android, IoS, Raspberry Pi and other platforms. It is pre-integrated with existing data processing solutions, such as mongoDB, Hadoop, Oracle and others.

LinkSmart open source middleware platform is a framework and a service infrastructure for creation of IoT applications, originally developed by Hydra EU project. The project is hosted by Fraunhofer FIT. It includes Device Connector for integrating devices (with different implementations for specific devices), Resource Catalog for managing devices and resources they expose, Service Catalog (services used to access devices and resources) and GlobalConnect tunneling service that enables access to devices beyond the boundaries of a private (routable) network.

Mbed platform aims at even tighter integration, by treating all its connected devices as embedded devices. They all have in-house mbed open-source Operating System, event-driven single-threaded architecture (scales down to the simplest, lowest cost, lowest power consumption devices). Mbed supports only devices based on ARM Cortex-M microcontroller. Key principles are security, connectivity and manageability (uses OMA Lightweight M2M, a popular protocol for monitoring and managing embedded devices). The architecture consists of mbed OS, Device connector (works with REST APIs), TLS (includes cryptographic and SSL/TLS capabilities in embedded products), client (library that connects devices to device connector service and mbed server – free, high-level C++ API) and server (essentially, a middleware, also hosted as a cloud service, connects IoT devices to web applications).

Nimbits is data logging service and rule engine platform for M2M connectivity. It provides nimbits.io open source Java library for developing Java, web and Android solutions that use Nimbits Server as a backend platform. Backend platform collects geo and time-stamped data and executes rules on this data, such as calculations, email alerts, xmpp messages, push notifications and others. Free and Enterprise editions of the server are available.

Particle.io (former Spark.io) offers hardware development kits for building the firmware for the devices, by using web-based IDE and deploying this firmware over the air. Then, ParticleJS and Mobile SDK libraries can be used to build web and mobile apps, based on the collected data.

Autodesk SeeControl is IoT cloud service which virtualizes machines, links them with reporting devices and use analytics to unlock their data. No-coding, drag-and-drop approach is implemented. Platform is focused to the needs of the manufacturing industries, in specific to generating product performance data, predicting a product failure, performing maintenance and optimizing supply chain and material replenishment costs. It provides a large library of existing protocol/vendor device adapters. Service also includes light ERP modules and business management tools.

SensorCloud is a cloud IoT platform for acquisition, visualization and analysis of data. The platform natively supports connectivity with LORD MicroStrain’s wireless and wired sensors. Visualization tools are available. It is possible to setup simple alerts, triggered by the data threshold values. MathEngine analysis tools are provided, with a simple interface which facilitates common operations such as FFTs, smoothing, filtering and interpolation.

PTC ThingWorx. In the platform, each device is represented by so-called Thing Template. Template defines properties (for example, temperature), services (for example, posting to Facebook) and events (for example, malfunction). Devices use agents to connect to IoT platform; different agents are used for the different types of devices. Composer application is used to model the things, business logic, visualization, data storage, collaboration and security required for IoT application. Mashup can be assembled by using different thing templates, namely, UI widgets which are pre-wired to the thing templates. The mashups are then used as interactive IoT applications, real-time dashboards, collaborative workspaces and mobile interfaces. BPM component is included to enable definition and execution of the processes, starting with an alert or event from a remote connected device. Device asset management tool is also included to facilitate remote diagnostics, control and scheduled software update of things. Free use is possible, with limited functionality.

ThingSpeak is another IoT platform, with features very similar to SensorCloud. It features open channels with available data from different devices, published by the users. Platform enables actuation, namely talking back to the device, which is done over HTTP.

IV. Discussion

Today's IoT platforms are typically cloud based, delivering PaaS (Platform-as-a-Service). Communication with devices is made by installing or developing light clients, which only purpose is to facilitate connectivity to the central platform. Pervasive (over any communication channel, including cellular, WiFi and satellite) and reliable (where connection quality is considered as a primary criteria) connectivity is a key feature.

Typical features of IoT platforms are: connectivity as a service, monitoring and maintenance of devices (including firmware updates), data visualization, data analytics, basic application logics through alerts and triggers.

Connectivity as a service is achieved by enabling the unconditional (with installed client) access to devices, no matter if they are located behind the firewall, a NAT or mobile network router. The service should work with any device that provides a TCP socket.

The following categories of existing IoT platforms were identified:

- Domain-specific platforms are the IoT platforms which facilitate specific domain scenarios. Often, these platforms are built on the top of the generic M2M connectivity providers. Examples of such platforms are rachio for smart irrigation, nest for home automation, getcleverpet, fishbed and others.
- Technology-specific platforms are the platforms which take into account only specific set of devices. These platforms are sometimes closed, in the sense that they are based on the devices with proprietary technology. Examples of such platforms are Mbed, which supports only devices based on ARM Cortex-M microcontroller, Zatar, Nest and others.
- M2M connectivity providers offer connectivity as a service as a core service, with only a few other features, mostly related to data analytics. Their primary objective is data acquisition and analysis.
- Full scale generic IoT middlewares (for example, ThingWorx) provide full range of connectivity services, but they also facilitate the application development, based on data collected by the devices and transformed by analytical tools. Such development is possible by using integrated development environments (IDE), API's or even language interpreters.
- Some platforms offering supporting services were considered as important. They do not offer M2M connectivity services, so they are not IoT platforms. However, they offer functionality which can be useful for IoT scenarios. Examples of such platforms are GrooveStreams, a data analytics cloud platform, and ifttt, interoperability-as-a-service platform.

With the rise of IoT platforms, cross-platform interoperability and reuse is starting to emerge. There are cases where the domain-specific IoT platforms are made by using M2M connectivity providers. For example, getcleverpet, groovelabs.io and fishbit are implemented by using particle.io. Similarly, first cases of collaboration between platforms appear, with interoperability solutions. For example, ThingSpeak platform is connected to realtime.io. Further collaboration may be facilitated by increasing number of stable open-source solutions. In fact, they already exist on market (Kaaproject, LinkSmart) offering significant opportunities for development of complex systems over existing core communication management and communication platform. Finally, IoT ecosystem will certainly benefit from the further development of supporting services (interoperability-as-a-service, storage-as-a-service, data analytics and visualization, etc.).

HTTP and REST will probably be the way to communicate with devices, as well as between platforms. Guinard et al [4] found that RESTful Web Services are more suitable for programming access to IoT devices, than WS-* service architecture. However, it is highlighted that the latter are better choice when complex requirements related to advanced security and QoS are considered.

IoT at scale is a crucial technical requirement for further development, scalable performance (concurrency), storage and connectivity being the most important topics. Scaling at a client side is considered with least priority; devices are still the most sensitive and weakest components of IoT architecture in terms of reliability (power consumption).

A. Issues and challenges of current cloud IoT platforms

Centralized approach to managing IoT ecosystem may pose challenges to the devices' reusability in multiple contexts/applications, due to potential conflicts between clients. It may also affect the future of IoT ecosystem architecture and pose an approach characterized by exclusive ownership over a device, where services and data of this device would be offered through a central platform that controls that device, forming a network of networks. This may lead to application silos, with potential risk of interoperability issues. Is P2P communication between devices worth considering as an alternative, even through their digital identities stored in a cloud (like in Arrayent platform)? Obviously, current devices' energy consumption issue makes the case for centralized approach, but for how long?

Though convenient, turnkey solutions, such as the group of full scale generic IoT platforms may affect the development of IoT market. Namely, when considering current cloud computing applications, the only feature that distinguishes IoT platforms from the other cloud services platforms is M2M connectivity as a service. All the other services, such as analytics, visualization and application development are value-added, non-IoT core services and they may be provided by the third-parties. However, the use of non-core services by the M2M connectivity providers is not at the significant rate.

Finally, the third major issue is lack of support to complex data structures and business logics (beyond the level of simple triggers or rules), to be used for the development of applications, based on collected and/or transformed data.

V. Conclusions

Although Gartner's analysis of the emerging technologies positions IoT platforms at the very early phase of development, experiences from this survey show that cloud-based M2M connectivity services offer is well established. Some characteristic market niches are already recognizable, namely, M2M connectivity, data storage and

analysis, data visualization, Interoperability-as-a-Service and others. What is clearly missing at this point is IoT ecosystem application building environment.

While the objective of this paper was to identify the gap in the current state of art of IoT platforms, comparing to the theoretical foundations and vision of IoT, its motivation was to setup the novel design of IoT platform which core feature will be exactly application development.

Based on the survey, following main principles for the development of formal model-driven IoT software execution platform (InoTEP) are defined:

- InoTEP is web application for devices in IoT which enables composition and realization of IoT scenarios, by using peer-to-peer approach (multiple InoTEP instances installed on multiple devices, communicating over REST).
- InoTEP provides Application-as-a-Service service which will interpret any formal model (RDF/RDFS/OWL ontology) in a runtime and deliver CRUD (create/read/update/delete data) application.
- InoTEP enables formal definition of the device's capability to sense and/or actuate, by using Capabilities ontology.
- InoTEP uses RDF as a transport protocol for communication between devices (over REST).
- InoTEP tries to match any data received through its own REST interface (external data), with domain and capabilities ontologies.

The above listed principles are further used in selection of the enablers of the key components of InoTEP. Application-as-a-Service component will be implemented by using OntoApp system [6]; W3C Sensor ontology [7] is being extended to develop a Capabilities Ontology; Active Semantic Model [8] approach will be used for a matching engine.

Acknowledgment

This work presented in this paper has been co-funded by the projects VIHOS (III41017, supported by the Government of Republic of Serbia), H2020-644715 "AquaSmart", supported by the European Commission and "Towards Context-aware Smart Cyber-Physical Ecosystems" (451-03-01765/2014-09/23, supported by Bilateral scientific co-operation programme - Serbia and Portugal).

References

[1] Stankovic, J.A. (2014) Research Directions for the Internet of Things. Internet of Things Journal, 1(1)3-9

[2] Machina Research press release, Machine-to-Machine connections to hit 12 billion in 2020, generating EUR714 billion revenue. https://machinaresearch.com/static/media/uploads/machina_research_press_release_m2m_global_forecast_analysis_2010_20.pdf

[3] Nair, S., Why is "Internet of Everything" a future growth driver for Cisco? http://finance.yahoo.com/news/why-internet-everything-future-growth-130020843.html

[4] Gluhak, A., Krco, S., Nati, M., Pfisterer, D., Mitton, N. et al. A Survey on Facilities for Experimental Internet of Things Research. IEEE Communications Magazine, Institute of Electrical and Electronics Engineers, 2011, 49 (11), pp.58-67

[5] Stankovic, J.A. (2014) Research Directions for the Internet of Things. Internet of Things Journal, 1(1)3-9

[6] Zdravković, M., Trajanović, M. (2015) On the Runtime Models for Complex, Distributed and Aware Systems. 5th International Conference on Information Society and Technology (ICIST 2015), March, 8-11 2015, Kopaonik, Serbia. In: Zdravkovic, M., Trajanovic, M., Konjovic, Z. (Eds.): ICIST 2015 Proceedings, pp.235-240, 2015

[7] Compton, M., Barnaghi, P., Bermudez, L., Garcia-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowitz, K., Kelsey, D., Le Phuoc, D., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., Taylor, K.: The SSN Ontology of the W3C Semantic Sensor Network Incubator Group, Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 17, 25-32. (2012)

[8] Stojkovic, M., Trifunovic, M., Misic, D., Manic, M. (2015) Towards Analogy-Based Reasoning in Semantic Network. Computer Science and Information Systems, 12(3) 979–1008

Concepts for Agriculture and Tourism Cyber-Physical Ecosystems

Joao Sarraipa*, Milan Zdravković**, Ioan Stefan Sacala***, Elsa Marcelino-Jesus*, Miroslav Trajanović**, Ricardo Jardim-Goncalves*
* UNINOVA, DEE, FCT, Universidade NOVA de Lisboa, Caparica, Portugal
** Faculty of Mechanical Engineering, University of Niš, Niš, Serbia
*** University Politehnica Bucharest, Faculty of Automatics and Computer Science, Bucharest, Romania
jfss@uninova.pt, milan.zdravkovic@gmail.com, sacalaioan@yahoo.com, ej@uninova.pt, miroslav.trajanovic@gmail.com, rg@uninova.pt

*Abstract***— Nowadays is quite common the usage of robotics and automation combined with smart electronics and embedded systems, in manufacturing or enterprise processes. Such elements work together to augment systems intelligence in decision making for actuation representing the so-called Cyber Physical Systems concept. Due to the amount of data generated by enterprise processes tools, which go either uncollected or unprocessed, as well as, the need for greater coordination of those processes for reducing inefficiency and downtime among other factors, Cyber-Physical Systems have been introduced to support such activities. Thus, and apart from the level of automation or the industrial sector, their dependence in these types of systems have been increased. The objective of this paper is to describe some concepts; approaches and analysis to facilitate an efficient and structured implementation and use of technologies related to Cyber-Physical Systems specifically in agriculture and tourism industries. Authors introduce the concept and the idea that enterprises systems have four states of working, which Cyber Physical Systems act as the catalyst for the transitions between the solid, the liquid, the gas and the plasma states, establishing by this way a parallelism to the states of matter.**

I. Introduction

The agriculture and tourism sectors are highly important for the European Union's social and economic welfare. Agriculture represents around 10% of Europe's GDP, while tourism is the third largest socio-economic activity in the EU. Agriculture employs around 25 million people, while the tourism industry provides work for another 10 million. However, European Agriculture is faced with three main challenges, namely, changes in climate which affect agricultural production rates, as well as, increasing energy costs (electricity and fuel) and reduced water sources, due to increased demand for this scarce resource. The EU's tourism industry faces similar challenges. Although Europe is a culturally rich and interesting tourist destination, the emergence of new destinations in other regions, has decreased some of the demand for European tourism spots. Additionally, the impact of rising transportation costs, due to higher fuel prices, can also deter attracting tourists from outside Europe. There is therefore a need for solutions which can assist both agriculture and tourism enterprises to become more efficient and cost effective in their operations.

This paper intends to contribute towards helping the EU's agricultural and tourism enterprises in remaining competitive, providing a solution and tools which will help them in better managing their resources and to more effectively adapt themselves to changes in their operating environment. The paper provides a new paradigm, implementing a new concept based on state of matter (solid, liquid, gas and plasma), developing a novel platform and toolbox, which will assist agricultural and tourism enterprises in effectively manage their business and environmental context. With appropriate smart sensorial-based tools, as well as, pro-active event processing, enterprises can react and adapt themselves in real-time to events which can influence their business, such as weather, market or social phenomena. It is propose the necessary components, from embedded systems and sensing devices, to a framework and software platform, which will enable companies to deploy smart Cyber-Physical Systems (CPS) solutions in their surrounding environment. The solution will be demonstrated in agricultural and in a ski resort settings, performing proper real-time sensorial data acquisition, knowledge analytics and management, and taking pro-active actions which will assist them in reducing costs, in being more efficient and in providing better service to their customers.

In this paper, the authors intend to draw attention to the use of CPS technology in agriculture and tourism scenarios with the goal of showing the potential benefits and respective cost savings provided by the implementation of these tools and demonstrate the importance of these technologies nowadays in industry. In section II the author`s proposed, a study of CPS, where is described it`s definition tools, objectives and applications. In section III, it is presented various enterprises' types (e.g. sensing enterprise) and its four states (the solid, liquid, plasma and gaseous states). Section IV identifies the positioning in the agriculture and tourism industry challenges and respective scenarios, one for each industry. In section V it is presented the modular architecture to enable the scenarios' application development. Conclusions and prospective future work are described in section VI.

II. Cyber-Physical Systems

CPS or "smart" systems can be defined as co-engineered interaction networks of physical and computational components [1]. These systems are at the heart of our critical infrastructure and form the basis of our future smart services. These promise increased efficiency and interaction between computer networks and the physical world enabling advances that improve the quality of life, including advances such as in personalized health care, emergency response, traffic flow management, and the electric power generation and delivery [2]. Other tools or technologies related to CPS

include: Internet of Things (IoT); Industrial Internet; Smart Cities; Smart Grid and "Smart" Anything (e.g., Cars, Buildings, Homes, Manufacturing, Hospitals, Appliances) [1]. Key stakeholders in CPS have identified the need to develop a consensus definition, reference architecture, and a common lexicon and taxonomy. These will facilitate interoperability between elements and systems, and promote communication across the breadth of CPS stakeholders. As these concepts are developed, it is critical to ensure that timing, dependability, and security are considered as first order design principles [2].

With the emergence of high speed broadband and the IoT, the embedded world is meeting the Internet world and the physical world meets the cyber world. In the future world of CPS, a huge number of devices connected to the physical world will be able to exchange data with each another, access web services, and interact with people [3]. One example of CPS is an intelligent manufacturing line, where the machine can perform many work processes by communicating with the components [4].

In conclusion, CPS are physical and engineered systems whose operations are monitored, coordinated, controlled and integrated by a computing and communication core. Just as the Internet transformed how humans interact with one another, CPS will transform how we interact with the physical world around us [5]. Therefore, using sensors, the embedded systems monitor and collect data from physical processes, like steering of a vehicle; energy consumption or human health functions are networked systems that make the data globally available. CPS make it possible for software applications to directly interact with events in the physical world, for example to measure peaks in energy consumption [4].

III. Enterprise Systems and their Four States

This paper foresees the development of a modular CPS architecture, as presented in section V, which could facilitate the establishment of collaborations between enterprises through the sharing of resources and assets among them with the objective of increasing autonomous systems able of satisfying multiple critical constraints including safety, security, power efficiency, high performance, size and cost. The main idea is to use CPS features to act as a catalyst of business collaborations generation and decision making, which underpins the generation of dynamic enterprises processes that are able to smoothly change their state of working. This means that such dynamic enterprises continually transform the multitude of changes occurring around it into coordinated strategic actions by its people or systems to further the development of its products and services [6].

Sensing Enterprise was introduced in 2011 by the Future Internet Enterprise Systems (FiNES) initiative to describe a business environment in which all assets are endowed with sensing capabilities and connected to networks to 'treat' information, generate new knowledge, help quick and effective decision making and, in so doing, make the Future Enterprise [7]. This concept intends to go beyond of such enterprise handling sensing features, accomplishing on its top a new generation of modular architectures elements accomplished by embedded systems characteristics and acting as software agents, which will enable a high degree of autonomy ensuring adaptability, scalability, complexity management, security and safety, and providing trust to humans in the loop. These new generation of novel architectures are represented as 'cyber assets', which as defined by Tow [8] in include in its component, the key information and knowledge resources, including the data, policies, reports, IP, algorithms and applications, programs and operational procedures that a modern society in the 21st century relies on to operate and manage its business.

The proposed main concept resides in the idea that enterprises have four states of working: the solid, the liquid, the gas and the plasma states, establishing a parallelism to the state of matter (Figure 1.). In the state of matter, the enthalpy represents the thermodynamic potential of the matter to change its state to another. Thus, the change depends mainly to the heat added or released in the matter; the other characteristics are the volume and the pressure. On the enterprises case, such potential is characterized mainly by the capability of absorbing or releasing knowledge between the states by the system; the other characteristics that have influence in such process are the interoperability and the automation abilities. As stated before are the CPS features the main catalysts of such dynamism, which would potentiate the changing of the state.

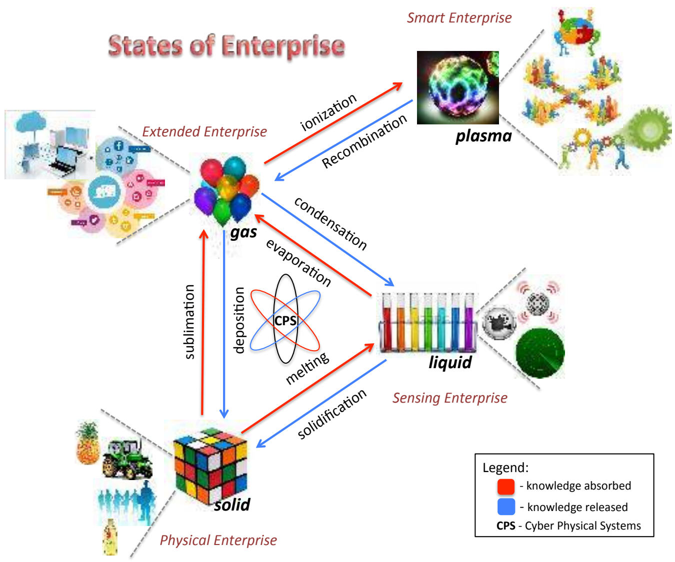


Figure 1. States of Enterprise

A. Solid State or Physical Enterprise

An enterprise is a socio-economic unit, integrated by human, material and technical elements, which aims the goal to obtain utilities through their participation in the market of goods and services. In this sense, makes use of production factors (labour, land and capital). A physical enterprise is characterized by do not use technological systems to establish business networks and to handle knowledge about its domain. In additional, it is similar to a "solid", meaning that structural rigidity and resistance to changes of shape or volume characterize it [9]. In the proposed concept case, enterprise's processes will interact with the physical or solid assets of an enterprise managing them with indications and receiving (human) feedback.

B. *Liquid State or Sensing Enterprise*

The Sensing Enterprise state relies upon the collection and processing of several types of data, from a wide variety of sources, including sensor networks. If the data collected is to be useful in the support and decentralization of its decision-making capabilities, the data gathered must be transformed into actionable information and knowledge. Knowledge management principles can and should be applied as they can assist in the fulfilment of the Sensing Enterprise vision. It states that enterprises with sensing capabilities will be able to anticipate future decisions by the understanding of the information handled, which enables specific context situations awareness. Thus, the Sensing-Enterprise interconnects physical (e.g. sensors, actuators), digital (e.g. sensors & web data, explicit knowledge) and virtual (e.g. simulation & prediction information) worlds in the same way as a semi-permeable membrane permits the flow of liquid particles through itself. This represents the Liquid State, like in the states of matter, the enterprise has a definite volume but no fixed shape, it has the capability of handling the knowledge associated to a predetermined set of resources, physical or not, that are interconnected to the system. It can be manageable and adapted to all kind of circumstances or objectives as liquid do when in contact to other shapes it simply takes its shape form.

C. *Gas State or Networked Enterprise*

The state of matter distinguished from the solid and liquid states by: relatively low density and viscosity; relatively great expansion and contraction with changes in pressure and temperature; the ability to diffuse readily; and the spontaneous tendency to become distributed uniformly throughout any container [10]. A networked enterprise is any undertaking that involves two or more interacting parties. A virtual enterprise is a temporary network of autonomous firms dynamically connecting themselves stimulated and driven by a business opportunity arising on market. Every member makes available some proprietary sub processes and part of their own knowledge.

D. *Plasma State or Smart Enterprise*

Plasma is one of the four fundamental states of matter (the others being solid, liquid, and gas). It comprises the major component of the Sun. Heating a gas may ionize its molecules or atoms (reducing or increasing the number of electrons in them), thus turning it into a plasma, which contains charged particles: positive ions and negative electrons or ions [11]. Smart Enterprise is about transforming our organizations to take advantage of the capabilities of smarter ecosystems, so enterprises can make more informed decisions, build deeper relationships and work with more agile and efficient business processes. The new enterprise organizations, such as virtual enterprise and smart enterprise, require the usage of collaborative automation approaches, addressing the flexibility and dynamic re-configurability [12].

E. *States of Enterprise and CPS*

This concept intends to address the processes required to change of state of working by an enterprise. It intends to answer the necessary characteristics which enable enterprises to have systems able to handle knowledge and collaborations underpinning the idea of a dynamism, where enterprises can change their state concerning specific situations as matter do in nature. It depends on the knowledge shared, automation, interoperability and intelligence. CPS is the fuel to introduce new knowledge to the system. CPS is able to acquire knowledge in some situations that will enable enterprise systems to actuate and originate the creation of specific approaches or processes that in relation to the presented concept will push the enterprise to another state of working.

IV. Current Agriculture and Tourism Industry Challenges and Scenarios

A. *Positioning in the Agriculture Industry*

One of the most important challenges of our century is to ensure the proper framework that will allow feeding the entire population of the planet, in the context of severe climate, different soil quality and population growth. The quality and quantity of agricultural products as primary elements of the food requirements for the population are representing since long ago a major preoccupation of both agriculture specialists and, in particular, of governments. The increase in the production and, implicitly, of the productivity in agriculture are subject to severe constraints with respect to inherently limited resources, less qualified workforce and agricultural terrains that are decreasing in size and productivity. According to “Food and Agriculture Organization of the United Nations” [13] predictions show that in 2050 the agricultural production will double as a result of the increasing level of automation, but also of the scientific and technological progresses with respect to the soil improvement and development of new and more efficient species of plants and animals.

One of the major resources for the increase of the productivity is potable water, which is now, in a continuous decrease, as it is estimated that 2/3 of the Earth population suffers, more or less, from its insufficiency. Consequently, the management of irrigation systems needs to be re-designed and soil humidity control and monitoring systems shall be improved in accordance with the evolution of integrated, embedded and CPS, sensing networks and of interconnected actuators. Mechanical and automated systems, successfully implemented in agriculture will lead to increased productivity, a decrease of the human resources involved and, globally to a more efficient exploitation of the soil. Also, the progresses in ICT will lead to a paradigm change, where the concept of CPS could find a rich domain of implementation.

Smart Plantation Scenario

In the current globalised context, enterprises and especially manufacturing enterprises survival depends on their ability to produce goods in a competitive way and adapt their organization, processes and products in a quick and suitable way to the market and social changes.

The agro food supply chain can be considered as a Collaborative Networked Organization (CNO) with a federated governance model, with the responsibilities shared among players and stages all along the supply,

from farm to the point of sale, reaching the consumer. To have the ability to efficiently produce highly customized and unique products, the agro food supply chain need to work as a single ecosystem where devices, machines, information systems, external data are orchestrated and correlated each other so as to make the knowledge behind the manufacturing line tacit and available for making decisions at tactical and operational level.

The integration of CPS in agriculture can be defined in three different stages. The first implementation stage is to ensure the parameterisation of the customer request in terms of different groups: filed, harvest, elaboration, quality measures, and sensory properties. A wireless sensory network (WSN) can monitor permanently the different parameters along the manufacturing process from the field to the manufacturing. Collecting data by such WSNs, their correlation with the parameters of field control and monitoring systems as well as with external factors (customer outputs, market evolution) can impact agricultural productivity and support strategic decisions with respect to the development of plantations. Thus, at field level can be found: weather stations to monitor the most relevant environmental parameters, such as temperature, relative humidity, solar radiation, wind speed and rainfall; soil sensors to estimate moisture, temperature and electrical conductivity (indicator of nutrients), plant sensors to measure size and thickness of stems, leaf wetness and fruits sizes; and specific visual sensors to obtain remote images of some quality parameters (aspect, diseases, damages).

A second stage consists of the acquisition of production data from the information systems (e.g. sensors) spread all over the production places. Information about traceability, quality control, stock, production planning can be acquired and compiled so as to serve as a knowledge base of the intelligent control at the physical level of the agricultural production and manufacturing, able to ensure a constant quality of products. In order to ensure the adaptability of the production to the external conditions – weather, market, food and feed alerts, regulations – it appears as a necessity the existence of a web intelligence network features (crawling, scraping), which represents an example of virtual sensors able to be used. Web services architectures, semantics and interoperability technologies to allow the orchestration of data coming from heterogeneous legacy systems sources are required to then integrate CPS at the next stage.

The final stage implies a CPS approach and results in an emergent pro-active behaviour of the agro food supply as a system, putting the previous stages all together. By correlating the real-time monitoring system and the existent knowledge embedded with the production planning and status, integrated with an intelligent control system and with estimation of the external world evolution (with respect to the agro food supply chain), the CPS system will be able to interpret different information received from channels as sensors, humans, devices, systems, even social media channels or web public sources, in order to identify and even predict relevant contextual changes at tactical and operational level, to which the company should adjust to align with the customer request.

As a scenario example using the presented enterprises states concept, if a possible disease in a plantation field is detected by a specific sensor it could require further confirmation from other sensors nearby, which may belong to another neighbour enterprise or field. This process represents the move from a sensing enterprise situation (liquid state) to another state (gas state), where through specific network of knowledge and services the system can check other environment variables that have a direct influence in plants diseases propagation as the weather. This enables more efficient actuation as in the request to a specific treatment. In such mentioned actuation process the system supplies additional information or knowledge to the “physical” part of the enterprise in a kind of “deposition” process (see Figure 1). Through such information, the “physical” part of the enterprise (e.g. farm employees) could react accordingly and conduct the appropriate treatment.

B. Positioning in the Tourism Challenges

As the third largest socio-economic activity in the EU, tourism is important for growth and employment. Despite the depth of the economic crisis, the tourism industry in the EU has proved resilient with numbers of tourist trips remaining high. However, long-term trends suggest Europe is losing position in the global marketplace, with new destinations gaining ever-growing market share. The Lisbon Treaty provides for faster and easier decision-making on EU measures in the field of tourism, allowing decisions to be taken through the ordinary legislative procedure, even though it has not substantially increased the scope of EU powers in the area. Drawing on the new Treaty provisions, the European Commission has prepared a new policy framework, whose main objective is to make European tourism more competitive, modern, sustainable and responsible. The strategy, entitled “Europe, the world’s No 1 destination – a new political framework for tourism in Europe”, was welcomed by the European Parliament, which nonetheless underlined the need to better coordinate tourism-related issues within the Commission and to clearly signpost financial support for tourism-related projects. The Parliament has also recently underscored the importance of tourism-related activities in different policy fields such as in rural, maritime and coastal areas [14].

Smart Mountain Resort Scenario

Winter resort is a complex organization which involves the coordinated deployment of the multiple and diverse services, such as: customer services, hospitality services, lift management, snow making, slope grooming, slope inspection, technical maintenance and repair, procurement, mountain rescue services, forest services, road maintenance, medical, parking, ski equipment rental, weather monitoring, ski school and fire rescue. At this moment, Bergfex.com lists a total of 1668 winter ski resorts in 12 European countries. Such ski resorts can be also considered as a CNO [15] with a federated governance model, with the responsibilities shared among

winter resort management, hotel property owners, mountain rescue, regional roads' maintenance company, municipality and even state-level organizations (e.g. national parks), and others. Smooth operations of the winter resort management and their coordination with other stakeholders have significant impact on the development of the local region, in terms of revenue, employment and sustainable development.

However, due to complexity of the federated governance model, the services coordination is often far below the optimum level, leading to the issues, such as lower customer satisfaction, inefficient energy consumption, increased total cost of operation and sometimes even higher probability of the critical events, such as injuries, avalanches and forest fires. The problems above are multiplied by the fact that, due to a climate change, the resorts are also forced to continuously work on adapting the services and even on natural hazards management [16]. Figure 2. presents the illustration of the governance (solid lines) and dependences (dashed lines) of selected services in Winter Resort Management. It also demonstrates the inefficiencies of the current operation practices, mostly related to the coordination of the different services.

For example, slope inspection is the service provided by the specific department of the winter resort enterprise that collect and share information about the current snow and slope conditions, but also about the slope congestion. In case of the bad conditions, a work orders are issues to slope grooming and/or snow making departments, which then start the respective processes. These processes can be launched only in the case of favourable meteor-conditions. Also, in a case of congestion on specific slopes, inspectors are informing mountain rescue personnel about the increased risk of injuries. Only this routine needs the coordination of 5 services and three different, independent actors. Given that it is performed continuously, during the day (even nightly, excluding the mountain rescue services) and given that it is highly dependent on the human involvement, high rate of idle time and subjective judgments, it involves great cost and risk of waste of resources in case of inaccurate decisions.

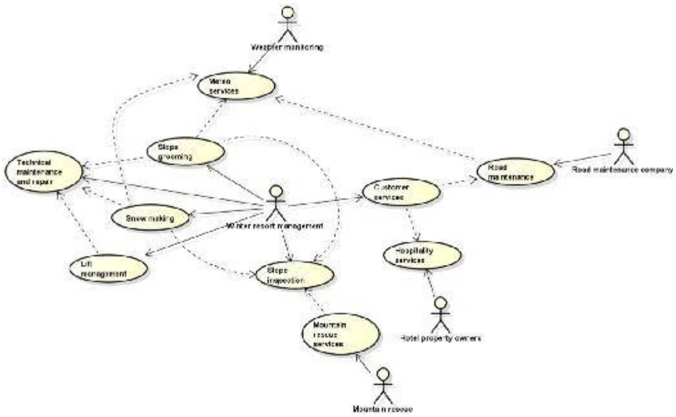


Figure 2. Illustration of the governance and dependence of selected services in Winter Resort Management

This scenario aims the CPS implementation to reduce (if not eliminate) specific human effort in the coordination of services, hence automating the dependencies between them. The following example takes in consideration the enterprises states concept presented at section III.

Sometimes there are enterprises in a winter resort area that need to share some specific management services through the cloud/internet to establish a more collaborative stage in order to answer to common goals as in slope inspection or snow making. To answer such necessity, systems from various enterprises as winter resort hotels, security inspectors and snow making suppliers could rearrange all their available services and resources (e.g. sensors, actuators) to answer to more complex situations where these resources can be shared in a kind of an only one brand to the customer. When this kind of collaborations is established, an advanced stage is reached representing a re-shoring situation that represents the "plasma state". This represents smart enterprises moments that together fulfil specific complex contexts in the best to their business/customers demands. If these collaborations are not necessary anymore there is a kind of "recombination" of each enterprise's systems, where the knowledge related to such collaborations are absorbed to a knowledge base to improve future similar situations. In this case, enterprises come again to a "gas state", which the services of each enterprise keep online ready for new services orchestration if needed.

V. CPS SUPPORTED BY A MODULAR ARCHITECTURE APPROACH

All the interactions or processes described before should be accomplished through a modular architecture, which each atomic element (e.g. aggregated to a sensor or actuator or a network of those) is the "cyber asset" that is configured for a particular purpose or objective and will be active as a software agent till accomplishes everything to what it was generated for. Each element type of the overall enterprise system has the capability to clone the "cyber assets" as necessary. In a first draft composition of a modular part of a CPS architecture is represented in the Figure 3.

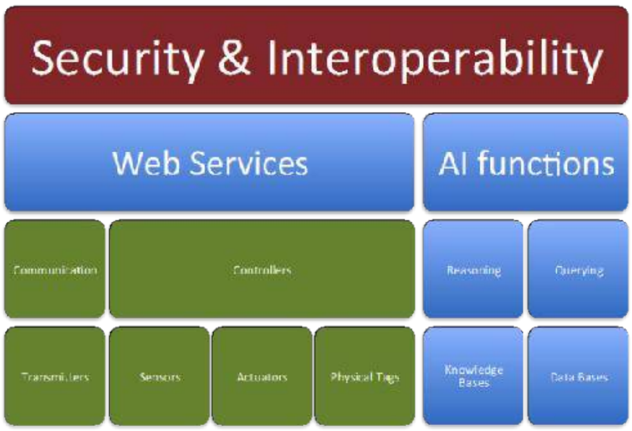


Figure 3. Main elements of a modular architecture (cyber asset element)

Each module (cyber asset) may integrate sensors, physical tags and other smart devices to accomplish sensing information about a determined ecosystem. The

data that comes from these various devices is managed by built-in controllers, which then transmit them to particular knowledge bases accomplishing a digital representation of the world. Or on the other hand, such controllers autonomously actuate over other devices that will execute specific tasks accordingly to specific formalised and accessible contexts, representing the shared situational awareness of the physical world (could be physical assets of the enterprise). To accomplish some predictions through simulations, Artificial Intelligence (AI) algorithms will be integrated on such elements as machine learning features able to be installed in function. These features would supply the system of dynamic capacities. All of these interconnectivities and features require a well security, privacy and interoperability policy defined & adequate technical solutions integrated.

VI. Conclusions and Future Work

The greatest innovation potential of this paper is the integration of various advanced technologies and concepts, such as CPS, sensing enterprise, knowledge management, proactive event-driven computing and smart electronics or embedded systems to offer advanced sensing and decision-making capabilities for enterprises. This paper presents the state of CPS by enabling them with new autonomous decision-making capabilities, moving CPS from sensor and actuator-based systems, to systems with greater intelligence and computational capabilities. This study relies heavily on advanced developments in embedded systems and smart electronics, as enabling CPS with greater capabilities, as well as, pushes the envelope of existing hardware.

These tools increases an enterprise´s market value, by providing a mean of predicting behavioural dynamics of the domain of interest and actuate accordingly. This is achieved by aggregating information and knowledge from sensors, databases and contextual knowledge bases, analysing it and use the output knowledge to support and optimize the decision-making process. An architecture able to self-sense enterprises environment and their own assets is also proposed. In conclusion, an enterprise should be able, not only to acquire information and knowledge in a static way, but it should be, able to re-configure and reprioritize targets of interest, as humans do. This architecture should be able to integrate/disintegrate new components (sensors, actuators, controllers, databases, knowledge bases) in order to fulfil enterprises assets. In other words, enterprises should be able to be reconfigurable to achieve a determined asset, and not limited to the fact of the developments are designed to deal with a limited set of components.

As a future work the authors intend to apply these concepts to the aquaculture domain. In this case the various sensors will be integrated in systems able to potentiate the sharing of enterprises knowledge and services in such way that could enable some dynamicity and intelligence in establishing business collaborations.

Acknowledgments

The authors acknowledge the European Commission for its support and partial funding and the partners of the research project Horizon2020 - AquaSmart – Aquaculture Smart and Open Data Analytics as a Service, project ID nr. 644715, (http:// www.aquasmartdata.eu/). It has also received funding from the Portuguese-Serbian bilateral research initiative specifically related to the project entitled: Context-aware Smart Cyber–Physical Ecosystems.

References

[1] National Institute of Standards and Technology (NIST), 2015. Cyber-Physical Systems. Retrieved from the web in January 2016 at: http://www.nist.gov/cps/.

[2] National Institute of Standards and Technology (NIST), 2014. NIST Cyber-Physical Systems Public Working Group. Retrieved from the web in January 2016 at: http://www.nist.gov/cps/cpspwg.cfm.

[3] Digital Agenda for Europe, 2015. Cyber-Physical Systems. Retrieved from the web in January 2016 at: https://ec.europa.eu/digital-agenda/en/cyberphysical-systems-0.

[4] Stéphane Amarger, 2010. Cyber-Physical Systems. Retrieved from the web in January 2016 at: https://www.eitdigital.eu/innovation-entrepreneurship/cyber-physical-systems/.

[5] Rajkumar R., Lee I., Sha L., and Stankovic J., 2010. Cyber-Physical Systems: The Next Computing Revolution. Retrieved from the web in January 2016 at: https://www.cs.virginia.edu/~stankovic/psfiles/Rajkumar-DAC2010-Final.pdf.

[6] Lisa Friedman and Herman Gyr, 2004. Creating The Dynamic Enterprise. Retrieved from the web in January 2016 at: http://enterprisedevelop.com/wp-content/uploads/2015/05/ED_Primer_2004.pdf.

[7] Gérald Santucci, Cristina Martinez and Diana Vlad-Câlcic, 2013. The Sensing Enterprise. Retrieved from the web in January 2016 at: www.theinternetofthings.eu/g%C3%A9rald-santucci-cristina-martinez-diana-vlad-c%C3%A2lcic-sensing-enterprise.

[8] David Tow, 2014. FutureWorld: The Future of Life Civilisation and the Planet. Retrieved from the web in January 2016 at: https://books.google.pt/books?id=7cSKvGsf6G0C&hl=pt-PT.

[9] Wikipedia, the free encyclopedia. (2014). Solid. Retrieved from the web in January 2016 at: http://en.wikipedia.org/wiki/Solid.

[10] Free Online Dictionary, Thesaurus and Encyclopedia. (n.d.). Gaseous State definition. Retrieved from the web in January 2016 at: http://www.thefreedictionary.com/gaseous+state.

[11] Luo, Q.-Z., D'Angelo, N., & Merlino, R. L. (1998). Shock formation in a negative ion plasma. Physics of Plasmas, 5(8), 2868. doi:10.1063/1.873007.

[12] Colombo, A. W., Schoop, R., Leitão, P., & Restivo, F. (2004). A collaborative automation approach to distributed production systems. In 2nd IEEE International Conference on Industrial Informatics, INDIN'04 (pp. 27–32). Faculty of Engineering, University of Porto, Porto, Portugal.

[13] FAO, (2014). FAO - News Article: 2050: A third more mouths to feed. Retrieved from the web in April 2014 at: http://www.fao.org/news/story/en/item/35571/.

[14] European Parliamentary Research Service. (2014). The European Union and tourism: challenges and policy responses. Retrieved from the web in April 2014 at: http://epthinktank.eu/2014/03/14/the-european-union-and-tourism-challenges-and-policy-responses/.

[15] Camarinha-Matos, L. M., Afsarmanesh, H., Galeano, N., & Molina, A. (2009). Collaborative networked organizations - Concepts and practice in manufacturing enterprises. Computers & Industrial Engineering, 57(1), 46–60. Retrieved from the web in April 2014 at: http://dblp.uni-trier.de/db/journals/candie/candie57.html#Camarinha-MatosAGM09.

[16] OECD - edited by Shardul Agrawala, 2007. Climate Change in the European Alps: Adapting Winter Tourism and Natural Hazards Managemen. Retrieved from the web in January 2016 at: http://www.oecd.org/env/cc/climatechangeintheeuropeanalpsadaptingwintertourismandnaturalhazardsmanagement.htm. ISBN: 9789264031692.

Aquaculture Knowledge Framework

Joao Sarraipa*, Elsa Marcelino-Jesus*, Pedro Oliveira*, Pedro Amaral*, Marcio Ponte**, Ruben Costa* and Milan Zdravković***,
* UNINOVA, DEE, FCT, Universidade NOVA de Lisboa, Caparica, Portugal
** UFOPA - Universidade Federal do Oeste do Pará, Santarém, Pará, Brasil
*** Faculty of Mechanical Engineering, University of Niš, Niš, Serbia
jfss@uninova.pt, ej@uninova.pt, pgo@uninova.pt, mjmoutinho@yahoo.com.br, rddc@uninova.pt, milan.zdravkovic@gmail.com

Abstract— **Despite to seem infinite, oceans have fragile ecosystems and limited resources. This way, aquaculture is an activity extremely important nowadays namely in the preservation and in the creation of a good alternative to fisheries. Moreover, aquaculture is extremely important to supply the population with food at a global level since fishing is not enough. In line with this, the authors propose an aquaculture knowledge framework to facilitate the development of information technologies solutions to improve the efficiency of the traditional aquaculture industry processes.**

I. Introduction

As vast as the world's oceans may seem, their resources are limited and their ecosystems fragile. Aquatic resources, although renewable, are not infinite and need to be properly managed, if their contribution to the nutritional, economic and social well being of the growing world's population is to be sustained [1].

Aquaculture, exists since ever, however it has had a greater impact in the last 50 years namely due to the fact of population growth. Thus, aquaculture is accomplished of others benefits such as: health benefits because eat fish are healthier and help fight cardiovascular disease, cancer, alzheimer's and many other major illnesses; economic benefits due to the increase of aquaculture farmers which origins new ways of develop and raise local/national economies moving the involved business chain such: researchers, breeders, fish food manufacturers, equipment manufacturers, marinas, storage facilities, processors, transportation and marketing companies as well as restaurants; and environmental benefits there are real advancements in all types of aquaculture systems. Especially for offshore systems, there are bio-security systems, cameras and surveillance infrastructure, as well as trained inspectors who ensure that farms are complying by environmentally safe practices. This helps to reduce diseases transfer in the waters and so on [2]. Accordingly to the population growing and to the reducing of natural and aquatic resources aquaculture is one of the solutions that contribute to the nutritional, economic and social well being of the growing world's population.

Thus, in this paper, the authors present a Knowledge Framework (KF) to be used by the aquaculture industry (farmers) and to demonstrate that such KF is essential to organize the semantics in the system to enable an harmonised communication between the business actors and as well to facilitate an effective knowledge transfer. Thus, in section II a study about aquaculture domain (e.g., definition, types, advantages and disadvantages) is presented. In section III, the proposed Aquaculture KF is introduced. Section IV presents the AquaSmart semantic referential and section V specifies the AquaSmart Knowledge Mechanisms. Conclusions and prospective work are analysed in section VI.

II. Aquaculture Domain

The word aquaculture it has its origin in mid of 19th century and its divided in two words "aqua" and "culture". The word "aqua" derives from the Latin and means "water" while the word "culture" derives from English on the pattern of words such as agriculture [2].

The term aquaculture refers to the cultivation of both marine and freshwater species and can range from land-based to open-ocean production [3]. Aquaculture also includes the production of ornamental fish for the aquarium trade, and growing plant species used in a range of food, pharmaceutical, nutritional, and biotechnology products [4].

Aquaculture is also mentioned as the farming of aquatic organisms such as fish, molluscs, crustaceans, aquatic plants, crocodiles, alligators, turtles, and amphibians [5]. Farming is the process that implies an intervention in the growth process to increase production, such as feeding, protection from predators, regular stocking, individual or corporate ownership of the stock being cultivated, etc. For statistical purposes, aquatic organisms which are harvested by an individual or corporate body which has owned them throughout their rearing period contribute to aquaculture, while aquatic organisms which are exploitable by the public as a common property resource, with or without appropriate licences, are the harvest of capture fisheries [6].

Aquaculture has any cultivating science has advantages and disadvantages not only to the aquaculture but also to the consumers and farmers. Thus, some of its advantages are presented in the following sentences.

Accordingly to the Food and Agriculture Organization (FAO), aquaculture has the highest production rate per area (hectare) in comparison with other cultivations, thus one cultivated hectare can produce more fish than any other animal. So, aquaculture is of great importance in the panorama of world food supply, mainly because it is possible to maintain a balanced and adequate diet to ensure that species develop in a healthy way, without altering its nutritional value [7].

Through aquaculture is achieved significantly increase the amount of fish produced compared to fishing, which contributes for a sustainable fish supply business. The explanation relates to artificially fertilizing eggs that lead to higher chance of fertilization, as in the wild only 10% of eggs are fertilized [8].

However, some of the aquaculture disadvantages are related to the rations and products used for this practice can harm the ecosystem if they are released into the

environment without proper treatment. Other point is related to the fact that farmers use large quantities of low cost proteins for animal feed, producing high-cost products (e.g. shrimp) instead of bet in the production of other fish population, less costly, and sometimes use high doses of antibiotics leading to bioaccumulation in humans that consume it.

Additionally, farmers' management behaviours sometimes results in bad environmental impact as in the increase of the spread of invasive species [9], or due to low spatial equality and organisation leads to large wastes concentrated in one area, which can facilitate disease to spread, leading easy transmission from host to host [8].

To better understand the relevance of aquaculture, it is relevant to distinguish the three types of aquaculture. These correspond to the terms used for terrestrial agricultural production [9]. The three types of aquaculture are the following:

- Extensive aquaculture it uses craft techniques, the species are raised in tanks next to their natural habitat where they receive nutrients and the renewal of the waters is made depending on the tides. Low rearing density and little or no food input accomplished with low production levels also characterizes this type of aquaculture.
- Semi-intensive aquaculture uses little primitive techniques, it is made in tanks on shore or land, it uses industrial rations and the created species reach up to 1 year old. It represents an intermediate level of production.
- Intensive aquaculture uses technologically advanced techniques. The species are raised of in tanks where the water is renewed every hour through pumps. High density and total food input. The species are totally fed through rations. High production levels characterize this type of systems [9].

In any type of aquaculture it is necessary to improve the existent tools in a way that these ones can contribute to increase the production of the number of high quality of fishes in aquaculture and reducing the production costs at the same time. Thus, the authors will present in the next chapter a KF that can support the development of systems to help in such objective. The defined framework proposes the use of ontologies, which related features and mechanisms could be deployed as a tool in organizations for advanced aquaculture industry systems establishment that would facilitate the increase of production management efficiency.

III. Aquaculture Knowledge Framework

Knowledge is considered the key asset of modern organizations and industry. The aquaculture domain has a proper nomenclature and the knowledge associated with that economic activity that needs a proper type of knowledge structuring and management. That king of knowledge organization can be achieved by the development of a specific ontology-based framework aiming to support Aquaculture knowledge, research and operational activities. The authors propose a framework to be the foundation for the aquaculture knowledge organisation and representation (Figure 1). It specifies the aquaculture knowledge model in four main parts: the aquaculture glossary or thesaurus; the aquaculture domain ontology; the aquaculture training ontology; and the IT infrastructures ontology. The framework also establishes the principles for the knowledge use and management services establishment. It encloses three main parts: searching and reasoning mechanisms; semantic enrichment mechanisms; knowledge and lexicon management mechanisms.

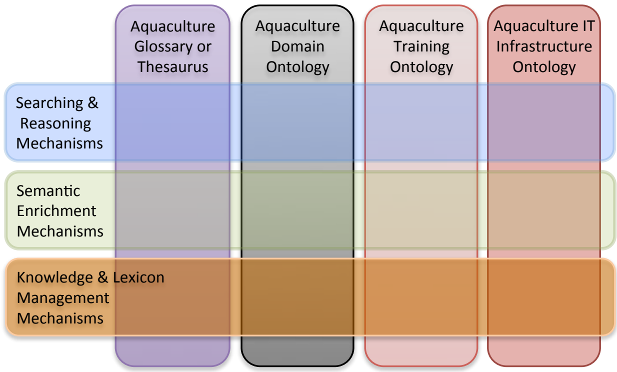


Figure 1. Aquaculture KF

When an information system intends to represent a domain knowledge needs to be aligned to the community that it represents. Consequently it is required to have a solution where community members could present their view on the domain and discuss it with their peers. Additionally, such knowledge must be available and maintained by all the involved actors.

Fundamentally, ontologies are used to improve communication between people and/or computers. By describing the intended meaning of "things" in a formal and unambiguous way, ontologies enhance the ability of both humans and computers to interoperate seamlessly and consequently facilitate the development of knowledge-based (and more intelligent) software applications.

A. *Aquaculture Glossaries or Thesaurus*

The main objective of a glossary or thesauri is to be a lexicon reference for a particular community. Thus, an aquaculture glossary or thesauri is such reference but for the aquaculture domain. This domain lexicon integrates terms and concepts with shared definitions (semantics) defined by domain experts. Due to such characteristics, these lexicon elements facilitate the semantic alignment between actors (systems or people) enabling interoperable communications. Additionally, a multi-language glossary that has mappings between the various languages concepts and synonyms outreaches a bigger community.

B. *Aquaculture Domain Ontology*

Ontologies allow key concepts and terms relevant to a given domain to be identified and defined in a structure able to express the knowledge of an organisation [10]. A good ontology model of any particular domain knowledge facilitate its understanding [11]. Additionally, Its recognised capacity to formally represent knowledge, to facilitate use and maintenance through semantic searching and reasoning, if integrated in a system could be handled for problem solving [12] contributing to such system computational intelligence increasing. Aquaculture domain ontology represents the knowledge in the domain in such way that if defined by domain experts with the support of knowledge engineers, will provide the

necessary insights towards the improvement of the efficiency of the aquaculture production processes. Thus, it can enclose knowledge for representing fish diseases, aquaculture production equipment, water quality, etc.

C. Aquaculture Training Ontology

The aquaculture training ontology will be used to represent the training knowledge base facilitating the categorization of its elements and subsequently reasoning over it. It comprises the model to represent any training curriculum and it is composed by generic training elements as courses, modules, competences, skills, etc. Its main objective is to specify a training curriculum which, addressed by appropriate reasoning mechanisms, will be able to generate customizable training programmes. It should contribute to the skills and competencies development of the trainees as required for specific understanding and exploitation.

D. Aquaculture IT Infrastructure Ontology

In the context of any project a set of use cases are normally identified to describe required functionalities that can be provided through particular services. Thus, these services' can accomplishes or support particular business processes and applications. In order to allow future reuse or sharing of these services, an ontology to formalise such IT services or infrastructures in a kind of services UDDI are necessary. This framework will be supported through Semantic Web technologies by providing tools: (i) to define an information model (as an ontology), (ii) to semantically enrich and relate the modelled data and (iii) to query this information. This framework will essentially provide:

1. An information model that allows users to instantiate and catalogue information that describes the functionality and interface of modularized services;
2. A query interface, providing service filtering capabilities and access to the descriptions of individual services.

IV. The AquaSmart Semantic Referential

The AquaSmart is an Innovation Action project that received support by the European Commission under the H2020 program. The purpose of AquaSmart project is through the usage of open and big data analysis along with a cloud infrastructure to promote the aquaculture industry. Thus, it intends to solve one of the main problems that aquaculture companies are facing nowadays, which is related to the fact that they cannot interpret the data they capture neither use others companies data. If they were able to do so, they would be able to dramatically improve the production in terms of feed conversion rate, cost, mortality, diseases, environment impact, etc.

In line to this objective, the AquaSmart consortium is developing a cloud based platform with a backend based on machine learning and data mining techniques to provide assistance to aquaculture managers in the decision making process. Such platform encloses big and open data analytics as a service to make more accurate estimations of the growth of the fish enabling a better view of the living inventory (biomass) that exist in a farm.

However, to accomplish such platform features authors also developed a set of mechanisms that instantiates the KF presented in the previous section. Such instantiation establishes the so-called AquaSmart Semantic Referential. It represents the foundation of both content-based information access and semantic interoperability over AquaSmart platform.

The AquaSmart semantic referential is the backbone for the technical developments which handle the semantic interoperability dimension. Such technical developments fit in the application areas presented in Figure 2. and can be described as follows:

- Aquaculture Glossary of terms - a glossary that contains terms and definitions of aquaculture domain, in a multilingual manner to accomplish a reference lexicon in the domain.

- Semantic Reasoning Mechanisms - services that make use of the knowledge contained in this semantic referential to apply reasoning techniques able of infer logical consequences from a set of asserted facts.

- Semantic Enrichment - services that use the semantic referential to enrich knowledge sources as documents or training courses.

- Knowledge Management - services that use appropriate semantic queries to retrieve or formalise knowledge from or to the semantic referential ontologies.

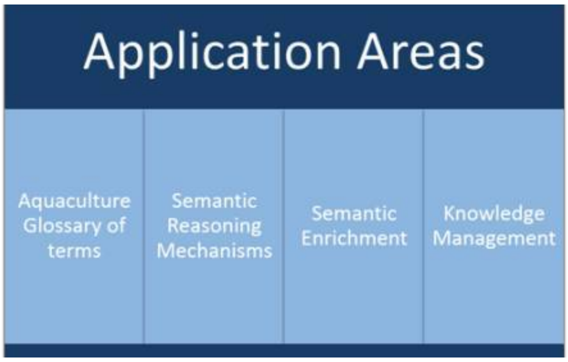


Figure 2. Application areas of the semantic referential

The creation of a semantic referential followed a method for designing and developing a domain ontology with inputs from knowledge experts, providing the necessary insights towards the improvement of the efficiency of the aquaculture production processes. Such experts, contributed with their knowledge about the aquaculture production, the actors involved, and the data generated during the production process [13].

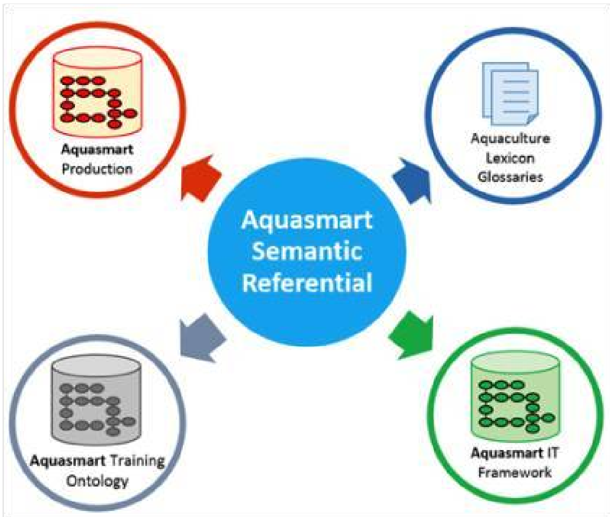


Figure 3. AquaSmart Knowledge Framework

In Figure 3. is presented the semantic referential with the four different modules as proposed by the Aquaculture KF: the AquaSmart Training Ontology; the AquaSmart Glossary; the AquaSmart IT Ontology; and the AquaSmart Production Ontology.

A. AquaSmart Production

In the AquaSmart context, knowledge experts are the end users (mainly fish farmers). The purpose of evolving such experts in the process is not only to provide input to the semantic referential, but to perform a quality review of the AquaSmart training courses. With the help of these experts, the main structure of the AquaSmart ontology was developed to accommodate all the important and necessary information that will support all the project services and functionalities. The Figure 4 shows the ontology with the structure proposed by the experts (i.e. Collector, Manager, Analyst, Administrator, Feeder, Diver, Vet, Biologist).

As depicted in Figure 4. the ontology is mainly separated in two concepts, the "Aquaculture Production Entities" and the "Grow Out Data Analysis". The first contains the all the aquaculture related entities, while the second one contains the key performance, the process and production related data.

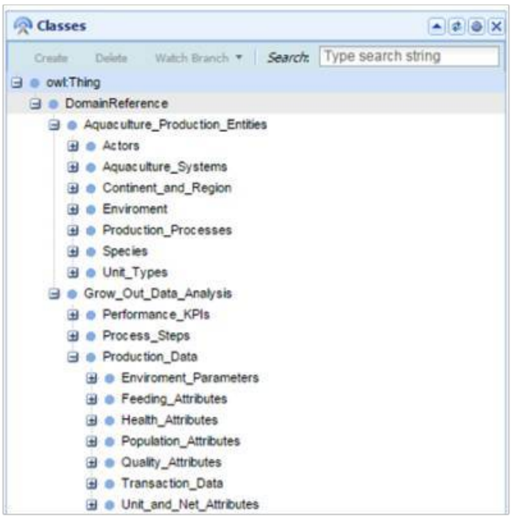


Figure 4. Domain Ontology

To make available fully interoperable multi-lingual data products and services this ontology establishes links with an aquaculture glossary that contains specific domain terms and their definitions. Such glossary integrates contents from the FAO glossary for aquaculture [14].

B. Aquaculture Lexicon Glossaries

From the AquaSmart perspective, the multilingual data generated within the aquaculture domain, can be exploited as a layer of services and resources by seamlessly adding (i) linguistic information for data and vocabularies in different languages, (ii) mappings between data with labels in different languages, and (iii) services to dynamically access and traverse linked data across different languages.

We envisage a multilingual aquaculture system where an end-user would query the "Aquaculture Open Data Cloud" in his/her own language, and would get the relevant data in that language. The glossary includes English, French, Greek, Hebraic, Spanish and Portuguese terms. Each term has properties, which defines it (Name, definition, related term, synonyms, subject area, translation, and image). Figure 5. shows an example of a glossary term.

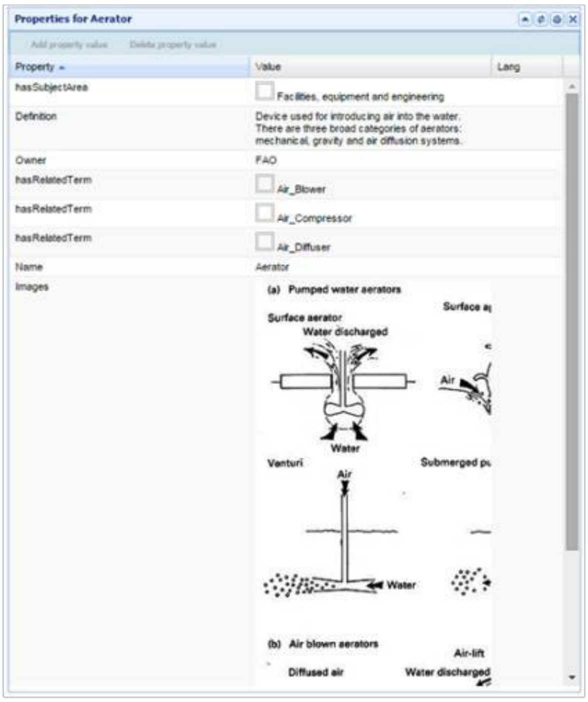


Figure 5. Glossary term

C. AquaSmart Training Ontology

The AquaSmart training ontology reflects the need to develop, organise and run courses to train "future users" about how to use the AquaSmart software, for instance, in extracting relevant patterns from aquaculture production data.

As an example, in this ontology model each learning *Module* has several concepts associated. The *Sources* concept contains information about the training materials sources used in the *Module*. *Contact* includes the contact information of the author of a *Module* or *Course* and *Keywords* contain a list of all relevant concepts needed for describing their contents. A *Course* belongs to a *Curriculum Main Area* that is divided by *Content Areas* and *Learning Levels*. Each *Module* and *Course* has a *Target Audience*, which are the ones that are recommended to attend to it. Such set of recommendations materials (*Courses* and *Modules*) can compose instances of pre-defined training *Programmes*.

D. AquaSmart Information Technology Framework

In the context of any project a set of use cases are normally identified to describe required functionalities that can be provided through particular services. Thus, such services can accomplish or support particular business processes and applications. In order to allow future reuse or sharing of such services, an ontology to formalise such IT services or infrastructures in a kind of services UDDI is necessary. This ontology component provides that. It encloses:

1. An information model that allows users to instantiate and catalogue information that describes the functionality and interface of modularized services;

2. A query interface, providing service filtering capabilities and access to the descriptions of individual services.

V. AquaSmart Knowledge Mechanisms

Associated to the AquaSmart Semantic Referential presented it was developed a set of ontology related services that represent the various mechanisms defined in the also proposed Aquaculture KF. Thus, in the overall such services supply the system with: 1) searching & reasoning: 2) semantic enrichment; and 3) knowledge & lexicon management mechanisms.

These mechanisms provide users with the ability to query the semantic referential knowledge base for terms, potential partners and public results as new knowledge resulted from big data analysis.

There is a partner search functionality, which enables users of a searching functionality able to find companies in the aquaculture domain by determined criteria such as water temperature, type of fish produced, size of production, country, etc.

Since an ontology is not a static entity, this set o proposed mechanisms also include specific ontology management services to update its represented knowledge. Additionally, due to the multi-language nature of the stakeholders of AquaSmart the ontology interface integrates a translation service to translate newly added terms.

As a consequence of all these features, an ontology API was developed to allow the users to handle such provided services to interact with the proposed Semantic Referential. A User Interface (UI) will be also provided to ease these interactions, allowing the user to explore the full provided services functionalities. This same UI will communicate with the developed ontology API, and consequently such ontology API at other level will communicate with a FUSEKI API via HTTP. FUSEKI is a SPARQL server that can be used to communicate with the ontology in order to do particular reasoning by the mean of TDB RDF Datasets.

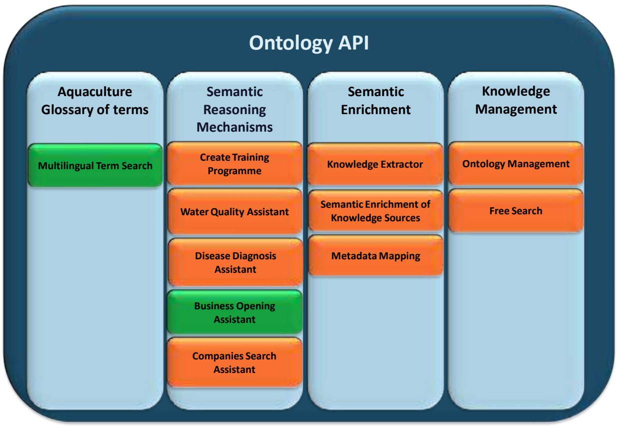


Figure 6. AquaSmart Knowledge Mechanisms

The proposed ontology API contains all the services categorized the four different modules presented on the semantic referential; the aquaculture glossary of terms, the semantic reasoning mechanisms, the semantic enrichment and the knowledge management (Figure 6.). Each of the services provided are described in the following sub-sections.

1) *Aquaculture Glossary of terms*

The introduction of an innovative multilingual knowledge base capacity suitable for the Aquaculture sector, which would enable an end-user query the "Aquaculture Open Data Cloud" in his/her own language, and would get the relevant data in that language, could significantly improve semantic interoperability solutions and the formal knowledge representation in the sector, which could contribute for the competitiveness increase. Thus, one of the main components of the semantic referential is the Aquaculture glossary of terms. As explained before, the glossary contains information regarding terms related to the aquaculture domain.

a) *Multilingual Term Search*

The aquaculture glossary of terms UI allows the user to search for information regarding aquaculture terms making use of this glossary and the amount of information contained within. This UI uses the multilingual services, which extends the usability of the semantic referential component. This multilingual service searches locally at the AquaSmart semantic referential for information related to a specific aquaculture term in three different languages (English, French and Spanish). Additionally the multilingual service also integrates an extra translation functionality to integrate other languages supported by other external API (e.g. BING Translator, Google Translate) (Figure 7).

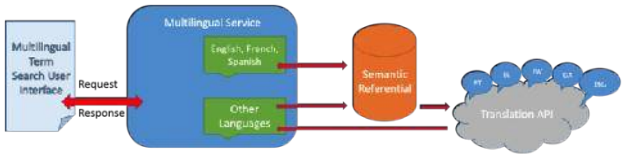


Figure 7. Multilingual service structure

Service description:

- Roles: user
- Parameters:
  - data – term to be searched
  - languages – original language and desired translation language
  - Returns: Translated term and all information (e.g. definition, images, etc.) regarding this term.

2) *Semantic Reasoning Mechanisms*

The semantic reasoning related services are services with the objective to provide specific information or knowledge to the user in order to facilitate and improve his work in relation to aquaculture business. In relation to this mechanism there are five proposed different services. However these can be more in the future. From these five there is one related to the training program and other four more specific to the particular business processes. The training service is a service that allows the creation of a training programme that meets the needs of the user. The

other four services: water quality control assistant, disease diagnose assistant, business opening assistant and search for company assistant are services from which the user can ask for assistance. Based in the provided characteristics the services will then do some reasoning and narrow down the problem.

a) Create Training Programme

The system could establish orchestrations using the existent training modules and courses materials for the creation of customized training programmes.

Service description:

- Roles: user
- Parameters:
    - data – topics; keywords; trainee’s profile; target audience; Roles and Competences; skills.
    - Returns: Customized Training Programme accordingly to the input data

b) Water Quality Assistant

The water quality includes all physical, chemical and biological characteristics that interact individually or collectively influence the performance of the fish growth. The Water Quality Control Parameters in aquaculture are usually the ones presented in the following table.

Table 1 Water Quality Parameters

PHYSICAL	***CHEMICAL***	***EXTERNAL STRUCTURE***
Temperature	*Ph*	*Algae*
Color	*alkalinity*	*sediments*
turbidity	*Toughness*	*Oils*
Visibility and transparency	*Dissolved oxygen*	
	Ammonia	
	Salinity	

There are other factors that also affect water quality and the result of fish farming. However the factors listed in Table 1 are the most critical and easier to monitor, and this can and should be done by the producer. This way the water quality control system aims to help the fish farmer to monitor the characteristics presented in Table 1, indicating possible adjustments parameter in order to maintain an aquatic environment suitable for productive development of aquaculture.

Service description:

- Roles: user
- Parameters:
    - data – Characteristics seen in the water (e.g. Table 1 values)
    - Returns: Possible problem regarding the input characteristics and a suggested solution.

c) Disease Diagnose Assistant

With the consolidation of aquaculture, new emerging technologies for intensive production and more diversity of fish species with potential for cultivation appear, however health problems or disease transmission problems may present obstacles at different stages of production. Thus, early diagnosis of diseases and the appropriate management constitutes a key factor for the success of the activity. In this sense the disease diagnostic assistant it is a computer system that aims to identify diseases in fish species based on physical and behavioural characteristics observed in the species and the environment in which are inserted, this includes water quality, presence of foreign matter to production environment as sediment or algae amongst other features. This service allows early diagnosis of disease and treatment suggestions. However the service does not aim to replace the veterinarian, but propose practical for immediate treatment and mortality containment, aiming to reduce the evolution and disease dissemination.

Service description:

- Roles: user
- Parameters:
    - data – Characteristics seen in the fishes (these are defined from existent concepts in the semantic referential)
    - Returns: Possible problem regarding the input characteristics and a suggested solution.

d) Business Opening Assistant

The opening of a new business within the aquaculture, involves criteria and requirements that could help to determine if it is an economically viable activity. These criteria are related to the location, climate, topography, water quality (salinity), distribution logistics and capture inputs. Through them, it is possible to determine what kind of fish are best suited to these conditions, what kind of equipment the entrepreneur will need to purchase and other valuable information. The service "Business opening assistant" aims to identify the best business approach taking into account specific criteria and requirements.

Service description:

- Roles: user
- Parameters:
    - data – Characteristics of the desired business (e.g. farm location)
    - Returns: Description of the possible business type and infrastructures needed

e) Companies Search Assistant

The service "Companies search assistant" aims to search and list companies that are active in the market. Such research is done by specific characteristics of the aquaculture sector, as fish species marketed, production volume, production site, among other features. This service will use the knowledge gathered by the semantic referential and then provide the user the companies that

fulfil the characteristics chosen. This will help the user to access to open data that could help in his business.
Service description:

- Roles: user
- Parameters:
    - data – Characteristics of the desired company (e.g. fish species)
    - Returns: Full description of the company regarding the input characteristics.

3) Semantic Enrichment of Knowledge Sources

There are three kinds of services related to semantic enrichment; the first intends to formalise new knowledge resulted from data analytics executions; then one that establishes semantic alignments with the raw datasets concepts, and finally other that enriches the existent knowledge base (semantic referential) with the knowledge sources as documents.

a) Knowledge Extractor

The service that make use of the data analytics results will allow users not only to get a better understanding of the results obtained as also will allow for extract knowledge in order to enrich the ontology.
Service description:

- Roles: user
- Parameters:
    - data – dataset analysis results
    - Returns: Description of the results.

b) Metadata Mapping

This service uses raw datasets. These datasets are composed by a set of standardized columns along with some company-specific columns. These company-specific columns may not be properly identified if they are meant to be kept as a company private information.
The purpose of this service is to:

1. Allow the disambiguation of the role of each column in the dataset;
2. Check the consistency of the values used in each column (i.e. temperature values above 50ºC or inconsistent temporal values).

These functionalities will be supported by:

1. A set of unique semantic concepts defined in a semantic referential, that describe precisely the meaning of each possible column;
2. A set of rules that define the range of allowed values for each column.

This service exposes operations that allow:

1. The mapping of the information in each column header of a dataset to a specific semantic concept;
2. The creation of new semantic concept, if needed to describe a non-standardized column and/or a private column;
3. The checking of the value set used in a column and, consequently, the reporting of errors found in the data (outliers).

The concrete implementation of this service will be achieved through a domain specific ontology (which will include aquaculture specific terms) backed by an RDF Triple Store, for data persistence, management and interrogation. The following diagram (Figure 8) contains a high level architectural description of this service, which encompasses three main operations:

1. Search concepts allows searching through the available concepts in the ontology;
2. Select concept enables the association between a dataset column and a concept, allowing its identification and the verification of its values;
3. Create concept exposes the functionality of adding previously undefined concepts to the ontology.

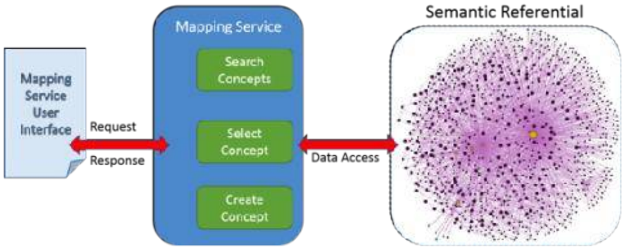


Figure 8. An high level description of the architecture of the metadata mapping service

Service description:

- Roles: user
- Parameters:
    - data – dataset upload
    - Returns: Full description of the dataset uploaded.

c) Semantic Enrichment of Knowledge Sources

This service supports the management (storing, indexing and retrieval) of aquaculture knowledge sources, using a formal representation method based on enriched Semantic Vectors. The method explores how traditional knowledge representations can be enriched through incorporation of implicit information derived from the complex relationships (semantic associations) modelled by domain ontologies with the addition of information presented in documents.

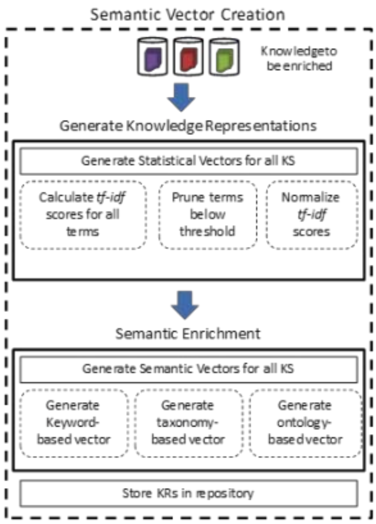


Figure 9. Overall approach for the semantic enrichment process

This service is responsible for enriching sources of non-structured information, combining with background knowledge available in the AquaSmart ontologies. The service can be composed into several steps as presented in the Figure 9. In the Document Analysis step, the tf-idf scores for all sources are calculated, next a procedure reduces the size of the statistical vector according to a certain relevance degree defined by the knowledge expert (Prune terms below threshold), after, another procedure normalizes the statistical vector after pruning the terms. Next, the semantic enrichment is performed by three different methods, responsible for the generation of the keyword, taxonomy and ontology-based vectors, respectively.
Service description:

- Roles: user
- Parameters:
    - data sources
    - Returns: semantic representations

4) Knowledge Management

Regarding knowledge management two services are being developed, one that allows ontology management, create/update/delete concepts; and other that allow free search within the ontology.

a) Ontology Management

Request an update to the ontology relating a concept or a relation.
Service description:

- Roles: ontology manager
- Parameters:
    - type – concept or relation
    - attributes – relation name, concept name
    - Returns: void (ontology is updated)

b) Free search

Submit a query to the ontology service. Incoming queries will be validated and subjected to limitations to prevent excessive resource utilisation. The ontology is queried and the response wrapped as a JSON message.
Service description:

- Roles: user
- Parameters:
    - data – search query
    - Returns: Information presented in the ontology related to the input query.

VI. CONCLUSIONS

The greatest innovation potential of this paper is the development of a KF for the aquaculture industry using a semantic layer in the form of the ontology service which will provide an open data service allowing interoperability between aquaculture companies. Openness of certain data was discussed while maintaining privacy and security for the owners of that data.

In conclusion, with an appropriate KF, aquaculture actors could benefit of advanced services to facilitate semantic interoperability and new knowledge formalization.

For future work it is intended to accomplish the others aquaculture knowledge mechanisms, represented in orange in Figure 6. during the second and last year of the project e.g., until end of January of 2017.

ACKNOWLEDGMENTS

The authors acknowledge the European Commission for its support and partial funding and the partners of the research project Horizon2020 - AquaSmart – Aquaculture Smart and Open Data Analytics as a Service, project ID nr. 644715, (http:// www.AquaSmartdata.eu/).

It has also received funding from the Portuguese-Serbian bilateral research initiative specifically related to the project entitled: Context-aware Smart Cyber –Physical Ecosystems.

REFERENCES

[1] FAO and MCS, 2013. "YUNGA International Drawing Competition: Protecting our fisheries - inheriting a healthier world". Retrieved from the web in February 2016 at: http://www.fao.org/climatechange/35157-0edd1699112a5dba2ea70240699c2c98e.doc.

[2] eSchooltoday, 2015. "Importance of Aquaculture". Retrieved from the web in February 2016 at: http://www.eschooltoday.com/aquaculture/importance-of-aquaculture.html.

[3] Department of Marine Resources State of Maine, 2006. " What is Aquaculture?". Retrieved from the web in February 2016 at: http://www.maine.gov/dmr/aquaculture/what_is_aquaculture.htm.

[4] NOAA Fisheries, 2015. " What is Aquaculture?". Retrieved from the web in February 2016 at: http://www.nmfs.noaa.gov/aquaculture/what_is_aquaculture.html.

[5] FAO and CWP, N/A ." CWP Handbook of Fishery Statistical Standards Section J: AQUACULTURE". Retrieved from the web at: http://www.fao.org/fishery/cwp/handbook/j/en.

[6] FAO, 2016. " Definition of Aquaculture". Retrieved from the web in February 2016 at: http://www.fao.org/fishery/cwp/handbook/J/en.

[7] AQUA, 2010. "Vantagens e Desvantagens da Aquacultura". Retrieved from the web in February 2016 at: http://aquacultura-aqua.blogspot.pt/2010/06/desvantagens-e-vantagens-da-aquacultura.html.

[8] Answers, 2016. " What are the advantages and disadvantages of aquaculture?". Retrieved from the web in February 2016 at: http://www.answers.com/Q/What_are_the_advantages_and_disadvantages_of_aquaculture.

[9] Aquaculture, 2011. " Different types of aquaculture". Retrieved from the web in February 2016 at: http://en.aquaculture.ifremer.fr/World-statistics/General-Introduction/Different-types-of-aquaculture.

[10] Sarraipa, J.; Jardim-Goncalves, R., and Steiger-Garcao, A. (2010). 'Mentor: An Enabler for Interoperable Intelligent Systems', International Journal of General Systems, 39: 5, 557 — 573, First Published on: 13 May 2010

[11] Camarinha-Matos, L. M., & Afsarmanesh, H. (2007). A comprehensive modeling framework for collaborative networked organizations. Journal of Intelligent Manufacturing, 18(5), 529–542. doi:10.1007/s10845-007-0063-3.

[12] Karayel, D., Sancak, S., & Keles, R. (2004). General framework for distributed knowledge management in mechatronic systems. Journal of IntelligentManufacturing, 15(4), 511–515. Retrieved from http://dblp.uni-trier.de/db/journals/jim/jim15.html.

[13] P. Oliveira, R. Costa, J. Lima, F. Ferreira and J. Sarraipa, "A Knowledge-Based Approach for Supporting Aquaculture Data Analysis Proficiency," in ASME International Mechanical Engineering Congress and Exposition, Houston, Texas, USA, 2015.

[14] FAO, "FAO Glossary,". Retrieved from the web in February 2016 at: www.fao.org/faoterm/collection/aquaculture/en/.

Simulation of a railway mainline junction using High level Petri nets

Dušan Jeremić*, Milan Milosavljević**, Sanjin Milinković***, Slavko Vesković****, Zoran Bundalo*****

*Railway College of Vocational Studies, Belgrade, Serbia, jeremicd@gmail.com
**Railway College of Vocational Studies, Belgrade, Serbia, mimilan89@gmail.com
***Faculty of Transport and Traffic Engineering, Belgrade, Serbia, sanjin@sf.bg.ac.rs
****Faculty of Transport and Traffic Engineering, Belgrade, Serbia, veskos@sf.bg.ac.rs
*****Railway College of Vocational Studies, Belgrade, Serbia, cheminot2@gmail.com

Abstract — *Railway transport is one of the most complex kind of transport. Solving railway infrastructure problems and conflict situations is very difficult task. This paper presents a simulation model of mainline railway junction using High level Petri nets. The aim of the paper is to detect the conflict situations in station and analyses impact of infrastructure elements malfunction on train delays. Model was tested on mainline station Mjölby in southeastern Sweden.*

I. Introduction

Modelling and simulating railway infrastructure is often challenging and demanding task which mostly depends on desired accuracy of the model. For microscopic models which are most precise and accurate large number of elements and connections need to be included. There are number of tools and software packages developed for this purpose, with different levels of complexity and accuracy. Petri nets as a modelling tool are mostly used in manufacturing processes and communication technology. Although it's not their primary usage, railway application of Petri nets could be found in literature in past 20 years. Van der Aalst and Odijk (1995) proposed the interval timed coloured Petri Nets (ITCPN) for modelling and analysis of railway stations, where train delay is specified by an upper and lower bound, i.e., an interval. Fanta, Giua and Seatzu (2006) used a coloured Petri Nets (CPN) model of a dynamic rail system for determining deadlock situations. Daamen, Goverde, and Hansen (2009) developed a CPN tool for route conflict identification and estimation of knock-on delay [1].

Main goal of this research was creation of a model of mainline railway junction using High level Petri nets. For this task station Mjölby in southeastern Sweden was chosen. Reasons for this decision lay in complex infrastructure layout of station Mjölby as in its busy timetable. Model was created using Exspect, a software tool for modelling discrete processes. Modelling was performed by creating basic modules like module for block section and switch, entering them in order as in real system and then connecting them to each other. Number of other elements is then added to establish dispatching logic and provide functionality of model.

II. Petri nets

A. Petri net theory

Petri net theory is based on mathematical graph theory, more precisely to the two-part (bipartite) graphs. Bipartite graph is a graph whose nodes can be divided into two different sets V1 and V2 so that each branch connects a node in V1 and node in V2 therefore there are no two branches in the same set. Extension of the theory of graphs has become possible with computer development and from graph theory Petri nets were created.

Petri nets are mathematical modelling tool that is used in the analysis and simulation of concurrent systems. The system is modelled as a two-part directed graph with two sets of nodes: a set of places that are representative of the state or system objects and the set of events or transitions that determine dynamics of the system. The distributed system is modelled as a bipartite directed graph with two sets of nodes: the set of places that represent state or system objects and the set of events or transitions that determine the dynamics of the system [1].

B. High level Petri nets

The term High level Petri nets is used for many Petri net formalisms that extend the basic Petri net formalism; this includes colored Petri nets, hierarchical Petri nets and others.

Colored Petri net (CPN) has its each token attached with a color, indicating the identity of the token. Moreover, each place and each transition has attached a set of colors. A transition can fire with respect to each of its colors. By firing a transition, tokens are removed from the input places and added to the output places in the same way as that in original Petri nets, except that a functional dependency is specified between the color of the transition firing and the colors of the involved tokens. The color attached to a token may be changed by a transition firing and it often represents a complex data-value [2].

III. Modelling railway infrastructure using Petri nets

A. Railway infrastructure simulations

Simulating a railway infrastructure is a complex and demanding task since there are many variables and dependencies which must be taken into account to

provide accurate model. The infrastructure can be modeled on a macroscopic or microscopic level depending on the effects a specific study aims to capture. A microscopic model consists of a node-link system and can contain all necessary characteristics and parameters of the real infrastructure, such as switches, signals, and speed and gradient profiles. The track layout can be defined with high accuracy. Macroscopic models represent data much more aggregated, stations can for example be defined by single nodes with attributes regarding the handling capacity. Links contain information on number of tracks, average speeds and other relevant information [3].

There are number of software packages developed for the simulation of railway infrastructure, most notable OpenTrack and RailSys but also Dons, Peter, Fasta, Railcap and many others [4]. On the other hand, there are number of software packages which were not designed primarily for railway simulations, but nevertheless can be used effectively. Best example is Matlab with its component Simulink.

B. Simulations using Petri nets

Petri nets as a simulation tool are mainly used in manufacturing and communication systems. They are suitable for use in concurrent systems but they are applicable to other uses like biological and chemical processes. In general, all systems which can be shown as set of places and have clear transition rules can be modeled with Petri nets.

In Petri nets trains can be represented as tokens while all other infrastructure elements (block sections, switches, station tracks…) can be represented as places. Since tokens are colored, they can carry number of information about train attributes like train number, length, train type which are constant trough model but also some information which can be changed as train (token) passes through model. Those information include arrival, departure and waiting times at each place, train path and other relevant information concerning train movement through system. Information which include time be carried only on timed Petri nets.

Transitions in Petri nets are responsible for applying rules for token movements through model or in other words for token firing from input to output places. There are many rules with various complexity which can be used in transitions. Simpler rules usually do not include token color or other properties as a decision criterion, for example, rule that forbids transition (firing) if output place is busy. Simpler rules are usually embedded in simulation software. More advanced rules use token color as a decision input, for example, train type or train length are used to determine if transition would occur or not. Or in case of multiple output places rules which include train properties could be used to determine to which output place token is going to be transferred (fired).

IV. MODELLING ELEMENTS OF RAILWAY INFRASTRUCTURE

A. Moddeling software

For this research model of railway infrastructure was created using software ExSpect [5]. ExSpect is a software tool for discrete process modelling. It is suitable for business process modelling, production chain modelling, and use case modelling, etc. ExSpect embodies a colored Petri nets approach to process modelling. An ExSpect model describes a process in terms of a collection of subtasks that communicate by message passing. The network of subtasks is displayed and edited graphically. A subtask can itself have a process model; this allows large models to be decomposed into manageable parts. Models can be analyzed for structural correctness properties, and their behavior can be observed through simulation, either step-by-step, continuous or with breakpoints.

Components of the program include predefined blocks which can be used to create objects. Basic blocks are places (channels), processors (transitions) and connections (arcs) which are basic Petri net elements but also some other blocks. Those include store, which are used to store and export data during simulations, time generator and input and output pins, which are used for creation of modules inside a model.

ExSpect as a software package is easy to learn, uses intuitive programming language, allows creation of complex models through creation of smaller modules and supports exporting data to number of formats. Unfortunately, its development stopped in 2000 due to number of bugs and compatibility issues with newer versions of Windows.

B. Modelling elements of railway infrastructure

There are two ways how system can be modeled in ExSpect. The first way is to create a model by entering the places and transitions for each element of the model, and then link them together and define the transition conditions for crossing (firing). This method requires that each element or object it the system is specifically defined and connected. For small and simple systems this method is efficient, accurate and reliable as the programmer can easily analyze all the connections and put all data into the model.

The second method involves the creation of subsystems or modules. These modules are stored in the program database and can be accessed and easily incorporated into the model. This feature is very useful for creating large models where some elements of the model appear more than once. For simulation of railway lines, modules correspond to isolated sections of track. The model is created by entering modules in order as in the real system, connect them to one another and with stores in the model. Such a process of modeling takes more time during the initial programming, but enables that these created modules could be used to model any system where there are similar processes, and by creating additional modules it is possible to model large number of railway systems.

C. Modules in Petri net simulations

Basic modules which are used in modeling any railway line are block section and switch. Block sections can be either line block sections or track sections in stations. Basic principle is the same for both kinds of sections, each module has two input and two output pins for both directions. They are used as an external connection to the rest of the model. Aside from those, there are three more external connections for stores time, info and stanje. Time is connected to main time store which provides time reference, stanje is connection to a store which provides information about occupancy of the section and info is connection to a store which provides information about

section length and maximum speed for all train types. When token enters the module, data is written in store stanje so section is marked as occupied. Occupancy time is calculated by dividing section length with section speed, which is provided from external store info. That way module is universal and can be used for any bidirectional block section. For block with only one possible direction module is simpler and includes only one subsystem. Module for bidirectional section is shown in Figure 1.

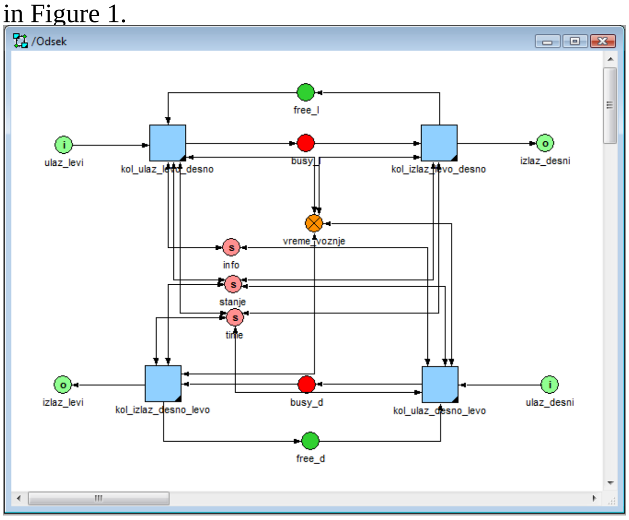


Figure 1. Bidirectional block section module

Module for switch is a bit different since it has a one entrance (input) and two exits (output). It has two branches, one for normal position of the switch and other for reverse. As in block section it has three stores, stanje, info and time which have same function as in block section, but also one more store put. Purpose of this additional store is to set the position of a switch to normal or reverse. There are two types of modules, one for diverging and one for converging switch. Diverging switch when observed from left to right has one input and two outputs while converging switch has two inputs and one output while they are similar in other aspects. (Figure 2)

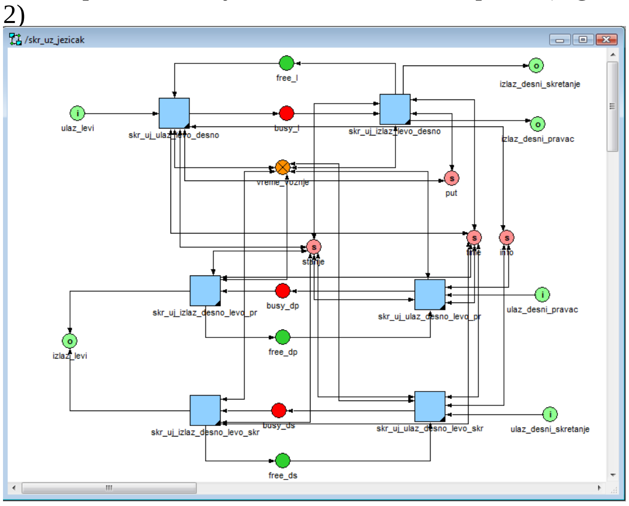


Figure 2. Module for diverging switch

These modules represent basic building blocks for railway infrastructure model. They can be further modified to fulfil any specific need.

D. *Application of Petri net models*

Petri nets can be used to model number of infrastructure elements. Simplest task is to model open line or block sections. Model is created just by adding section modules and connecting them with one another. More advanced application is modelling of a railway junction and adjoining block sections [1] [6]. Besides using modules for block sections and switches it requires much more elements for functions such as route setting, route releasing, etc. Most advanced task is modeling a railway station. Depending of the size of the station number of elements can be significantly different. Main factor in model size and complexity is number of possible routes in station. Smaller stations in that aspect are not much different from junctions since number of possible routes is generally small. Larger stations provide much more complexity, especially stations with many connecting lines.

V. Simulation of Station Mjölby

As an example of Petri net application for complex railway junction, Petri net model of station Mjölby was created.

A. *Location and importance of station Mjölby*

Station Mjölby is located in southeastern Sweden on Southern main line (Södra stambanan) which connects Stockholm with Malmö. It is one of most important railway lines in Sweden, connecting central and northern Sweden to southern Sweden, Denmark and rest of the Europe. It is a double track line, congested with many types of trains such as fast X2000, local commuter trains such as Östgötapendeln and many freight trains. Station Mjölby is also a junction from which line to Hallsberg, largest marshaling yard in Sweden diverges. On top of it, station Mjölby is terminal station for many local passenger trains from Linköping and Hallsberg. This adds to more than 600 scheduled trains daily from which around 300 are in service regularly. Track diagram of station Mjölby as seen on display in dispatching center Norrköping is shown in Figure 3 [7].

Figure 3: Track diagram of station Mjölby

Station was mostly reconstructed as a part of a double track project from Mjölby to Hallsberg in preparation of large increase in freight traffic from that line. Project was very soon stopped, freight traffic did not increase and as a result Mjölby is a station with lots of capacity and possibilities for parallel movements.

B. *Petri net model of station Mjölby*

Modules which were described in previous section and in [6] were used as a basis for constructing a model of station Mjölby. Some modifications were performed, mainly adding characteristic of types of trains (X2000, pendel_tranås, pendel_mjölby and godståg). Modules are placed and connected to each other in the order in which

tracks are located in station. They are also connected to stores time, info, stanje and put for switches. All stores except time need predefined initial value. For store stanje initial condition is 0 (free) and for store put it depends of switch position, it is usually 0 (normal position). Definition of store info a bit more complicated. In each store type info is necessary to define the name of section, length in meters and speed for all four types of trains (X2000, pendel_tranås, pendel_mjölby and godståg) in meters per second. After connecting basic elements it is necessary to define dispatching logic.

Since Mjölby is a terminating station for number of local passenger trains (pendel_mjölby), special section module was created for tracks on which these trains terminate. For all other train types this section acts as a normal block section but for terminating trains, instead of passing them trough, it returns those trains (tokens) to the point of entry.

Dispatching logic in station Mjölby is not limited to one or two processors, but is incorporated in almost every main signal (signal processor). This allows on the one hand simplification of the code and on the other hand much more flexibility in creating a model. There's the fact that there are a number of different possible train routes in Mjölby and often only entrance routes are used for passenger trains, therefore was necessary to disperse dispatching logic on multiple processors. This is done so that all entrance and most intermediate signals are presented with a separate processor [7].

There are number of other elements which are necessary for proper functioning of this model. Those include section clearance (release) processors and some additional processors. Due to the large number of crossovers, short-track parts and complex track diagram on the one hand, and longer trains than some main tracks on the other hand, section release logic need to be solved differently. It is defined in modules that that train (token) entrance is registered in the section store as a condition 2 (occupancy), but the clearance (release) of sections, due to factors mentioned above is not performed in the processor inside the module, but in special release processors outside the module. They are variously arranged in the model and carry out route release function of various numbers of module sections. In general, section release processors are arranged so that they can approximate the length of the train factor in the best way. Therefore an approximation needs to be performed where some points are grouped so sections are cleared when train (token) enters last section (point), some short sections are not cleared immediately upon exiting the train from that section but only when train leaves next section, etc.

Additional processors mostly include platform processors and train number changing processors. Platforms processors are used to simulate stopping time of certain passenger trains in the station, and only on the tracks that have a platform next to it. Number changing processors are used next to terminating track modules and they change number of terminating trains before their return towards destination station. There are also few additional stores like sig which is used to store information relating to the signal aspect and prolaz which is used to transmit information to intermediate signals if the route is already formed over that signal by another (usually the entrance) signal or it needs to be set when train enters a section in front of the intermediate signal.

Upon completion of all definitions, it is necessary to integrate all elements. In addition to linking all the elements, order of connection must be considered, for example signal processor with a section stores or processor for release with the section stores. The result is a completed model of station Mjölby (Figure 4).

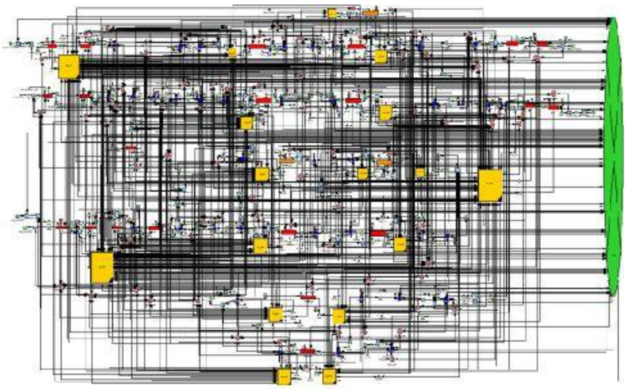


Figure 4. Completed model of station Mjölby

After developing model, to launch the simulation it is necessary to define the initial value, i.e. trains. The program allows input train data from external sources (txt file) which must be written in necessary form with a specific sequence of data. Trains must be sorted by time of entry into the system and other information must be sorted in alphabetical order.

VI. Simulation results and discussion

A. *Simulation parameteres*

Simulation of station Mjölby was performed with two sets of input data. In first set train arrival was according to schedule and in other set train arrival was randomly generated according to proper distribution for different train types. For random arrivals ten combinations were generated and accordingly ten simulations were performed. For scheduled arrivals ten simulation were also performed but with same results since only one set of input data was used.

Since ExSpect has the ability to export data into an Excel file, in this model it is defined that the data is entered into an Excel file from each channel between the sections (or groups of sections in case of switches) and from each section store stanje. In this case Excel was chosen between other formats due to suitable processing and well known file format. Information from each channel is written into one xls file (izlaz.xls), which serves mostly to review the functioning of the whole system, of delays, the most used routes, while data from each store is entered into separate file which can be later merged into single file for comparison and better view of section occupancy time. Same principle applies to stores put which export data about switch positions. Every change in store or place is recorded along with time of change. For every place which token (train) passes, all train properties (token color) are exported to xls file. Every entry includes data about train that usually do not change like train length, train type, destination and data which changes at each place that train passes like current train path and current position. All entries include simulation time at which they were recorded.

B. *Simulation results*

Goal of this simulation was to determine critical places or points of possible conflict in station but also impact of track element malfunctions on train delay. Same methodology was used both for scheduled and random train arrivals. Both tasks were performed using data from export xls file.

There are several criteria used to determine points of possible conflict. First was to compare frequently used routes and determine which are conflicting. Second, to determine which switches were most frequently used as part of different routes. Next was to determine occupancy time of those switches, both with train and train route. Another criterion was to determine number of train routes per switch. And as a last criterion, minimal time between consecutive passes of different trains over switch was measured.

Data about frequently used routes was gathered from main xls file (izlaz.xls) which recorded all data about trains through system. All data is written in one sheet chronologically using simulation time. Since every route has predefined switches which are included in that route, simply by comparing conflicting routes it is possible to determine which switches are most used by conflicting routes. Therefore, switches which were most used by conflicting routes were likely to be places of possible conflict. (Figure 5)

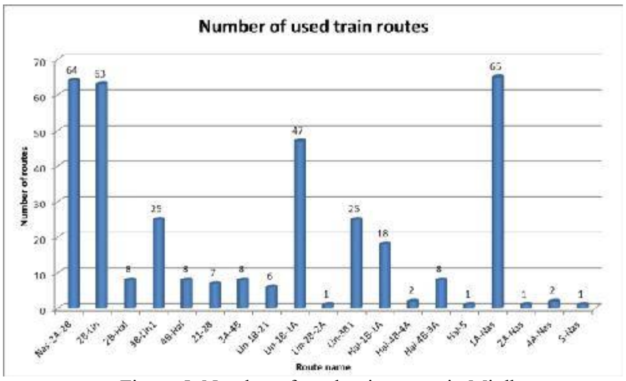


Figure 5. Number of used train routes in Mjölby

Switch occupation time is determined from each switch store. There are three possible conditions of a switch, and each condition associated with numeric value. Those conditions are free (value 0), occupied by route (value 1) and occupied by train (value 2). All values are written in separate xls file for each switch, together with time of condition change in simulation. Overall occupancy time is obtained by adding time difference between each change of condition. Switches with longest occupancy time are likely candidates for place of possible conflict. Another way to determine which switch could be place of possible conflict is to compare number of train routes per switch or in other words number of trains that cross each switch. That information is obtained from izlaz.xls simply by counting how many times each switch appears in file (Figure 6).

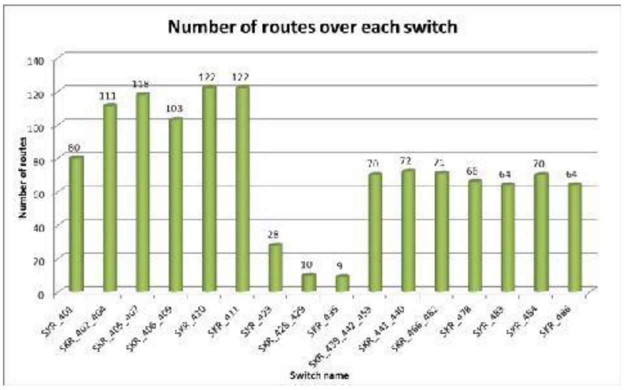


Figure 6. Number of train routes per switch

Minimal time between consecutive passes of different trains over switches was performed only on switches which were by previous methods determined to be likely places of possible conflict. It was also measured from izlaz.xls by comparing time difference between each passing over certain switches. All time differences were grouped into time intervals. For each switch and each interval, number of occurrences was shown. (Figure 7).

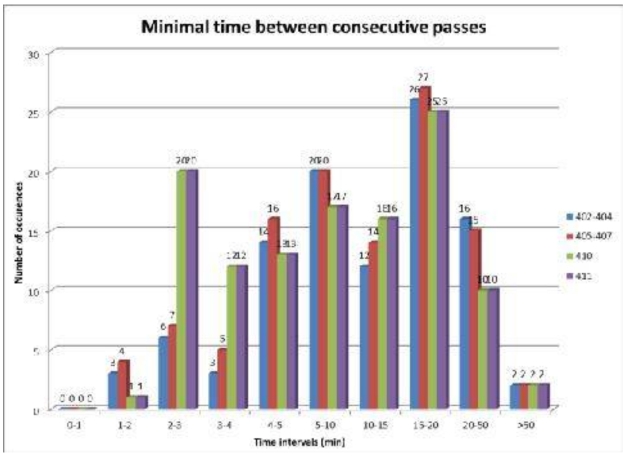


Figure 7. Number of minimal times between consecutive passes

Same methodology was used in case of random train arrivals. With ten different sets of input data, ten different simulation results were obtained. Difference between them was not significant, all results were with approximately 10% variation. For example, number of routes over switches for first five simulations with random train arrivals is shown in Figure 8.

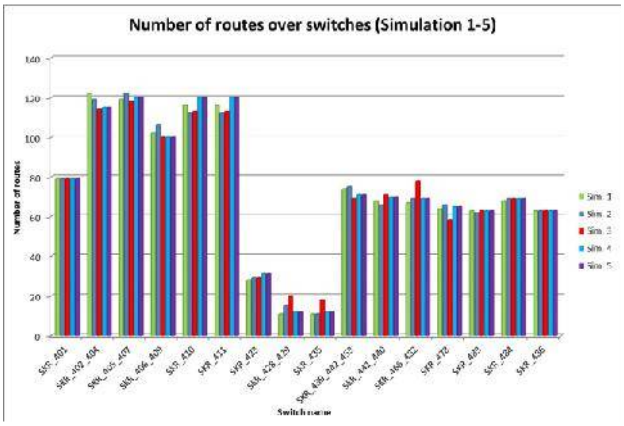


Figure 8. Number of routes over switches in case of random train arrivals

As it can be seen results are only marginally different then in case of scheduled train arrivals. Similar results were obtained for frequently used routes, switch occupation time and minimal time between consecutive passes of different trains over switches.

Second goal was determining impact of track element malfunctions on train delay. Track elements included both track sections and switches. Two kinds of malfunctions were considered, one was malfunction of isolated section which can occur both in blocks and switches, and other was malfunction of switch mechanism. Reasons and probabilities of isolated section malfunctions were not considered since that requires much broader analysis of interlocking device in station Mjölby and its reliability. Analysis of possibility of switch mechanism malfunction was performed by analyzing their usage, or number of throws per switch. Number of throws per switch is obtained from each switch store put, which holds information about point position. As switch can be in two positions, normal and reverse, store put can also be in two conditions, normal (value 0) and reverse (value 1). Number of throws is determined by counting number of changes from 0 to 1 and vice versa. Total number of throws is shown in Figure 9.

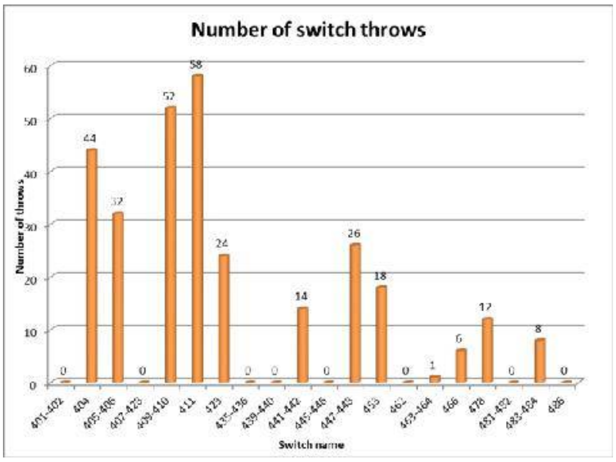


Figure 9. Number of throws per switch

Impact of track element malfunction on train delay was tested by changing the state of observed element from free into busy at the start of simulation. Tracks and switches whose malfunction would block entire station were not analyzed. Analysis was performed by comparison of delays in case of malfunction and delays in normal operating conditions. Results showed small delays, around 15-20 seconds when malfunction was on through main tracks and up to one hour when malfunction was on one of the terminating tracks, namely 3B. Those high delays were only observed on one type of train, local terminating trains from Hallsberg in case they could not use their designated terminating track 3B.

VII. Conclusion

High level petri nets proved to be powerful tool for creating models of railway infrastructure where greatest advantage is their flexibility. Most important advantage is possibility of creating modules for any part of railway infrastructure. Mostly used are modules of repeating or frequently used elements like track section or switch, but even complete station could be used as a module. Since modules are not predefined but created by user, they can also be modified for specific purposes. Block section for terminating trains is a good example, where standard block section module is modified to return certain type of train i.e. terminating train to a point of entry instead of passing them trough.

Another flexibility is dispatching logic, which could be defined as centralized in one processor or as in this paper decentralized to several processors. Route setting rules can be defined in number of ways, using practically any train attribute as a route setting criterion. Train attributes are also custom category defined by user and they can be defined in number of formats (integer, real, string…). Number of other processors and transition rules can be easily defined and incorporated into model.

As for station Mjölby simulation showed that it is well projected with lots of possible routes and possibilities for parallel movements. Number of places of possible conflict is low, both with scheduled and random train arrivals. Impact of track circuits and switches malfunctions produces only small delays. In general, it is hard to find any part of the station which could be source of conflict situations. Reasons may lay in fact that data for this research is from 2008, and in that period there were not so many train using line to and from Hallsberg. Therefore, functions of this station as a junction were not so pronounced and it mostly served as a terminating mainline station. Today is a bit different situation, there are much more passenger train which use line to Hallsberg. And for a future research it would be interesting to test a model with current train schedule and at the same time determine if Mjölby is really well projected as this research showed.

Acknowledgment

This paper is supported by The Ministry of Educations and Science of the Republic of Serbia, within the research projects No. 36012.

References

[1] S. Milinković, M. Marković, S. Vesković, M. Ivić, N. Pavlović, A fuzzy Petri net model to estimate train delays, Simulation Modelling Practice and Theory 33, 143-157

[2] Petri Nets for Dynamic Event-Driven System Modeling, Jiacun Wang

[3] Simulation of rail traffic - Methods for timetable construction, delay modeling and infrastructure evaluation, Hans Sipilä, Ph. D. thesis, Stockholm 2015

[4] Barber, F., Abril, M., Salido, M. A., Ingolotti, L., Tormos, P., Lova, A. 2007. Survey of automated Systems for Railway Management. *Technical Report DSIC-II/01/07.*Technical University of Valencia.

[5] http://www.exspect.com/

[6] Milinković, S., Vesković, S., Marković, M., Ivić, M., Pavlović, N. (2010). Simulation model of a railway junction based on Petri Nets and fuzzy logic. *Selected Proceedings of the 12th World Conference on Transport Research Society*, 11-15 July Lisbon, Portugal.

[7] Jeremić, D. Utvrđivanje konfliktnih tačaka i parametara rada stanice Mjölby simulacionim modelom Petri mreža, Diplomski rad, Saobraćajni fakultet, Beograd 2010

The Application of the Topic Modeling to Question Answer Retrieval

Jelica Vasiljevic*, Tom Lampert**, Milos Ivanovic*
* Faculty of Science, University of Kragujevac, Serbia
** Laboratoire Quantup, Strasbourg, France
jeca.kg@hotmail.com
mivanovic@kg.ac.rs
t.lampert@laboquantup.eu

Abstract*—Topic modeling (TM) is used for the extraction of the information from unstructured documents. The aim of this study is to investigate the application of the Latent Dirichlet Allocation Topic modeling algorithm to question answer retrieval. The most appropriate answer is automatically selected from a database of answers based on a combination of several similarity measures. The primary hypothesis assumed in this study is that a question and its correct answer are thematically similar. All TM results were compared to a simple word count approach, employed as the reference model. Results show that the topic modeling approach performs better than the reference model as the number of the documents increase. It is also proved that the difference in results is statistically significant. Nevertheless, basic LDA turned out to be insufficient for efficient question answering. It is therefore hypothesized that additional expert knowledge would greatly improve its performance.

I. Introduction

Community Question Answering [1] sites, such as StackExchange and Yahoo! Answers, have become very important sources of information on the internet. They enable users to exchange knowledge by posing questions and offering answers, and as these exchanges are stored, such sites have created vast stores of valuable knowledge. A significant portion of these databases can be used to answer new question if they are related to the information that already exist in the database.

Question Answer retrieval refers to the selection of one or more correct answers to a given question. An answer is selected from a set of potential answers that already exist. Typically, a question is considered to be lexically similar to the correct answer. Thus, the question is processed and important lexical characteristics are extracted which are then used in order to retrieve the correct answer. Commonly used preprocessing methods in the lexical processing chain are: lowercase transformation, stop word removal, stemming, lemmatization, and so on. However, it is possible that a question and its correct answer do not have any words in common. Thus, lexical similarity may not always be sufficient to select the correct answer to a given question. Because of this, it is important to encode a question's semantics when inferring the correct answer.

Using Latent Dirichlet Allocation (LDA) topic modeling [1] it is possible to encode the semantics of a given document. In [2], the authors propose a number of LDA based similarity measures which could be applied to the question answering problem. Furthermore, a novel statistical topic model for the question answering problem in community archives is proposed in [3].

In this paper, we explore the limits of the classic LDA topic modeling algorithm. We combine several similarity measures in order to rank answers according to their semantic similarity. We also test the influence of synonyms, stemming, and lemmatization on the final result.

This remainder of this manuscript is organized as follows: Section II presents the methods used within the presented study, Section III the experimental setup and results, and finally our conclusions are drawn in Section IV.

II. Method

A. Topic Modeling Approach

Latent Dirichlet Allocation is described in [1]. It is an unsupervised, statistical approach to document modeling that discovers latent semantic topics in a large collection of text documents [1]. Each document is represented as a distribution over a fixed number of topics, while each topic is represented as a distribution over words. In the presented work, these two probability distributions are used to calculate the similarity between a question and all possible answers. The answer which is the most similar to the question is then proposed to be the correct answer.

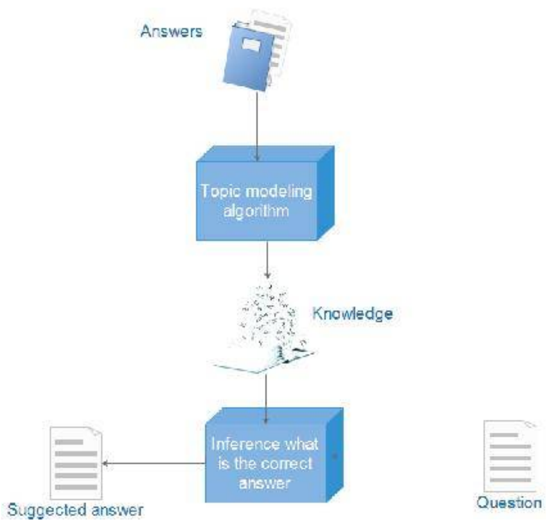


Figure 1. Conceptual overview of the LDA Topic Modeling approach.

Figure 1 presents a conceptual overview of the proposed solution. An LDA topic model is constructed using a training set composed of all the answers that currently exist in a database. Therefore, topic distributions over all possible answers are known, and word distributions over all topics are known. This model can then be used to infer the distribution over topics and distributions over words of a new, user defined question, and the most similar answer that exists in the database can be selected.

B. Similarity Measures

The output of LDA is a multinomial distribution over topics for each document in the (answer) database and a multinomial distribution over words for each topic. The nature of the algorithm also allows for the inference of topic distributions for unseen documents. In the presented experimental setup, questions are the unseen documents for which a topic distribution is determined using LDA, according to the topic model learnt using the database of answers. A similarity measure is therefore required to infer the answer with the most similar topic distribution to the question, and in this work several are evaluated.

B.1 Cosine Similarity

The cosine similarity measures the cosine of the angle between two vectors. As the measure tends towards 1, two vectors are more similar. The cosine similarity between two vectors a and b is measured as follows

$$\cos(\theta) = \frac{\vec{a} \cdot \vec{b}}{\|a\| \cdot \|b\|}. \tag{1}$$

In topic modeling, the distribution over topics is discrete, so it can be represented as a vector. The first coordinate of that vector is the probability of the first topic in the document, the second coordinate is probability of the second topic in the same document and so on. Since the distribution over topics for each document D is known, it is possible to measure the similarity of two documents using (1).

Given a question, the answer that results in a cosine similarity closest to 1 is selected as the correct answer.

B.2 Similarity Measure Based on Query Likelihood Probability

It has already been stated that the distribution over topics is known for each document. Furthermore, for each topic the distribution over words is known. Let K be total number of topics, w a particular word, and θ_d the topic distribution in document D. The probability of word w appearing in document D can be expressed as [5]

$$P_{lda}(w|D) = \sum_{z=1}^{K} P(w|z) P(z|\theta_D), \tag{2}$$

where:

- $P(w|z)$ is the probability of word w in topic z,
- $P(z|\theta_D)$ is the probability of topic z in document D.

Equation (2) can be interpreted as the probability of generating word w given document D.

Using (2), the probability of a set of words Q belonging to document D can be defined as

$$P(Q|D) = \prod_{w \in Q} P_{lda}(w|D). \tag{3}$$

Equation (3) can be interpreted as the probability of generating the set of words Q given document D.

Specifically, if we take Q to be a question and D a particular answer, (3) gives the probability of generating the question from the answer. As the probability increases and approaches 1, it is more likely that the answer is correct. Therefore, (3) can also be used to measure the similarity between a question and an answer.

Besides (2), the probability of word w appearing in document D can be expressed in terms of classical probability, as follows

$$P(w|D) = \frac{f_{w,D}}{|D|}, \tag{4}$$

where:

- $f_{w,D}$ is number of occurrences of the word w in document D,
- $|D|$ is total number of words in document D.

It is not guaranteed that word w belongs to document D, and therefore it is possible that (4) is zero. This would problematically cause (3) to also be zero. In the proposed application, this is not logical. For example, two documents could have many words in common and one that is not. According to (3), the similarity between two such documents would, incorrectly, be zero.

The solution to this problem is to use pseudo-counts. A pseudo-count is the default number of occurrences of words that do not exist in a document [5]. By extending (4) pseudo-counts are introduced as follows

$$P(w|D) = \frac{f_{w,D} + \mu \frac{c_w}{|C|}}{|D| + \mu}, \tag{5}$$

where:

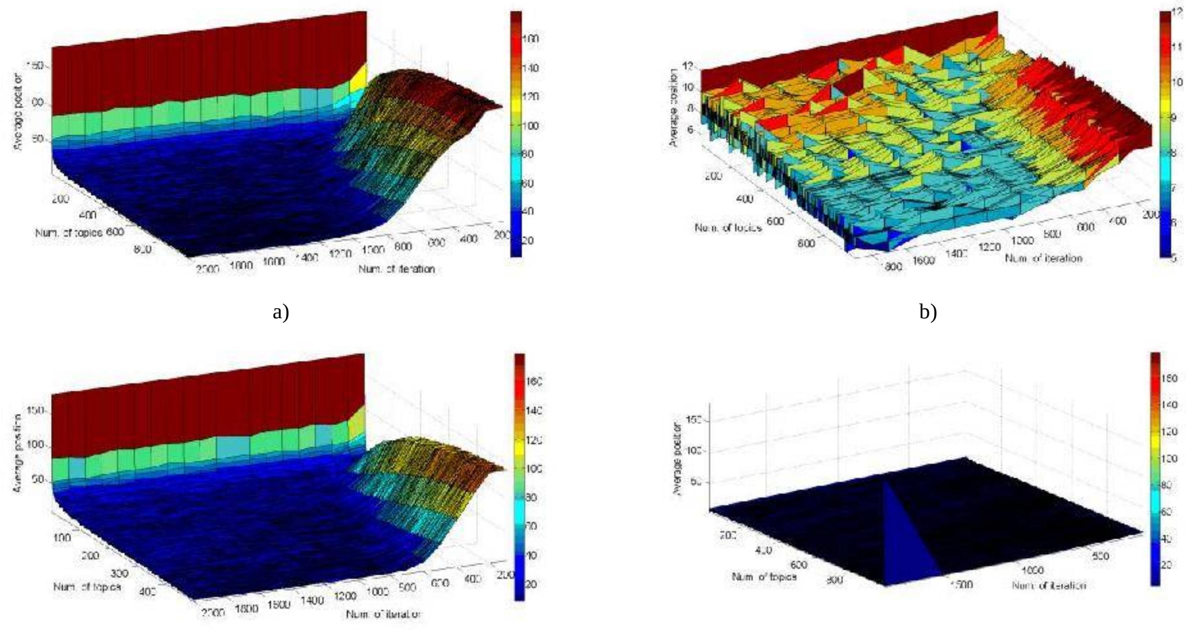
- μ is the pseudo count and is determined experimentally,
- C is set of all possible answers,
- c_w is the number of occurrences of word w in C.

It is still possible that (5) results in zero, which is the case when word w from a question does not occur in any answer, i.e. $c_w = 0$. In that case, we sample a hyper-parameter β from a discrete Dirichlet distribution, as the default probability [1].

Equations (2) and (5) both define the probability of a word appearing in a document, but with different physical meanings. Equation (2) exploits topic similarity, while (5) exploits lexical similarity. Both similarities are important for correct answer selection, but not necessarily with the same importance, and therefore they can be combined as follows

$$P_{lda}(w|D) = \lambda \left(\frac{f_{w,D} + \mu \frac{c_w}{|C|}}{|D| + \mu} \right) + (1 - \lambda) \sum_{z=1}^{K} P(w \mid z) P(z|\theta_D) \tag{6}$$

The influence of each term is controlled by the parameter λ which takes values from the range $[0,1]$. By increasing λ, lexical similarity becomes more important.



a) b) c) d)

Figure 2. Average position of the correct answer (height) as a function of the number of iterations (x-axis) and the number of topics (y-axis): a) when using the cosine similarity measure and preprocessing steps 1—4 (see subsection II.E of the text); b) when using the query likelihood probability similarity measure and preprocessing steps 1—4; c) when using the cosine similarity measure and preprocessing steps 1—5; d) when using the query likelihood probability similarity measure and preprocessing steps 1—5.

In the presented experiments the value of μ is fixed to 200, while the value of λ is fixed to 0.2. The similarity between two documents is measured using (3).

C. Mallet

The topic modeling algorithm is implemented in the software framework *Mallet* [6]. Mallet is an open source Java-based package for topic modeling, natural language processing, and document clustering. More information regarding Mallet can be found in [6] and [7].

In the presented work, the *ParallelTopicModel* class is primarily used, which is an implementation of the Gibbs Sampling LDA topic modeling algorithm [8]. Also, the *TopicInferencer* class was used for inferring the topic distribution of each question.

D. Experimental Data

Experiments are conducted using question-answer pairs taken from two publically available real-world online collaborative question answering platforms: StackExchange[1] and Yahoo! Answers[2]. These portals enable users to collaborate in the form of asking question and proposing answers. They use a collaborative voting mechanism, which allows members of the community to vote up or down the questions and answers that they think are appropriate or not. Furthermore, a person who asks a question can mark one of the answers as the best answer.

Due to the nature of these sources, each question has multiple answers associated with it. We choose the best answer (marked by the person who asked the question) as the correct answer, resulting in one answer for each question (a question-answer, or Q-A, pair). If no best answer exists then the answer that received the most votes is used instead.

In the StackExchange dataset, 120 question-answer pairs from the health, fitness, and engineering categories are selected at random (the dataset therefore contains 360 Q-A pairs). Various stages of the Topic Modeling approach are optimized using this dataset (preprocessing steps and the similarity measure, see Section III.A), which is referred to in this manuscript as the training set (this is distinct from the training set used to train the topic modeling algorithm, which is formed from the answers that exist in the dataset being evaluated, and therefore the questions form the same dataset form the test set).

The final Topic Modeling approach is then compared to the reference approach (described in subsection F below) using test datasets extracted from the *health* category of the Yahoo! Answers website. A number of test sets are constructed, containing 100, 400, 700, 5 000, 10 000 and 20 000 randomly selected question-answer pairs.

E. Preprocessing

As is common in text based analysis, it is necessary to preprocess the raw data to make it suitable for automatic analysis. The following preprocessing steps are used (in order of application):

1. HTML tag removal;
2. Lowercase transformation;
3. Removing all non alphanumeric characters (smileys, special symbols etc.);
4. Stop word removal;
5. Lemmatization, using the Stanford Lemmatizer [9].

[1] Available from https://archive.org/details/stackexchange.

[2] Available from http://webscope.sandbox.yahoo.com/catalog.php?datatype=l.

Steps 1—4 are mandatory for effective application of the topic modeling algorithm as they remove irrelevant data (Steps 1, 3, and 4), and remove typographic variations that prevent automatic matching of syntactically equivalent text (Steps 2).

It is often not known *a priori* whether lemmatization or stemming should be used in document retrieval tasks [10] and we therefore conducted preliminary experiments to determine the best for this application. As such, the training set was used to determine whether stemming and augmenting documents with word synonyms improves performance. In addition to this, all combinations of stemming, lemmatization and synonym augmentation were tested. The Porter stemmer [11] was used in this part of the experimentation. It was experimentally found that the combination of steps listed above gave the best results.

F. Reference Model

The commonly applied Word Count approach is used as the reference model, in which the similarity of two documents is proportional to the number of words that they have in common. The TF-IDF measure [12] is used as the similarity measure and preprocessed answers and questions act as the input. Documents are preprocessed in exactly the same way as in the Topic Modeling approach to enable a fair comparison.

III. Results and Discussion

The first part of this section is dedicated to the choice of preprocessing stages and the selection of parameter values in the Topic Modeling approach, and the second part compares the final Topic Modeling algorithm to the Word Count approach.

A. Preliminary Experiments

This section presents the relevant results obtained during preliminary experimentation using the StackExchange training set. In Mallet, the LDA topic modeling algorithm is implemented using Gibbs sampling. Therefore, the performance of the algorithm depends upon the number of iterations used in the Gibbs sampling process.

The average position of the correct answer in the list of answers ranked by similarity is used to evaluate performance. These are presented as functions of the number of topics and iterations when using several variations of the algorithm in Fig. 2. The minimum values

Table I.
Preprocessing and parameter values that result in the minimum average position of the correct answer in each of the experimental setups, see Figure 2.

Min Avg. Position	Preprocessing Stages	Similarity Measure	# Iterations	# Topics
8	1—4	cosine	1200	693
5	1—4	query likelihood	300	51
8	1—5	cosine	1300	481
4	1—5	query likelihood	900	417

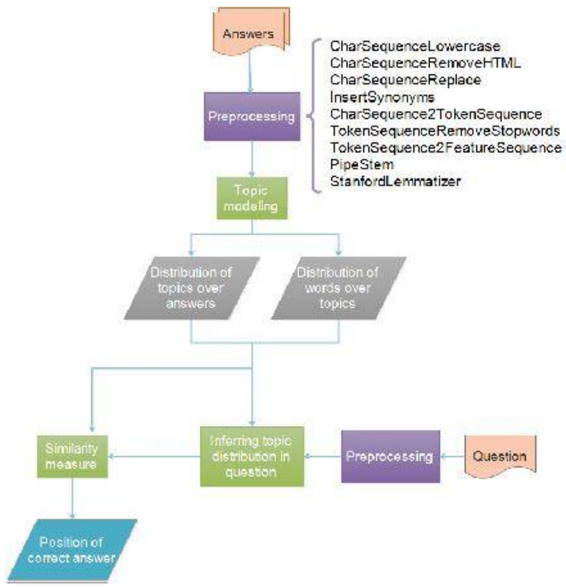


Figure 3. Detailed diagram of the proposed question-answer matching algorithm.

obtained in each of these experimental setups are presented in *Table* I.

It can be observed that within each of the two different preprocessing chains (Steps 1—4, and Steps 1—5) the query likelihood similarity measure gives the best results. Moreover, it achieves these results with fewer topics and iterations when compared to the cosine similarity measure. Indeed, it can be observed in Figure 2b that the performance is relatively constant when using this similarity measure in comparison to the cosine similarity measure results presented in Figures 2a and 2c (which performs badly when few iterations or topics are used). This indicates that exploiting information inherent to the topic modeling algrithm in the similarity measure leads to a simpler and more accurate model. It can also be observed that lemmatization (Stage 5) increases the number of topics needed to achieve a similar minimum average position when using the query likelihood measure, this can be explained by the fact that lemmatization reduces the diversity of words in the corpus, and therefore more topics are needed to distinguish between two documents' semantics.

The large peak in Fig. 2d appears when a very large number of topics is used. When the number of topics is greater than some specific number (which depends on the data, preprocessing steps, and similarity measure), then all answers have approximately equal topic distributions and all topics have approximately equal word distributions. As such all answers are equally probable, which causes the average position to be $N/2$, where N is the total number of answers. If this peak were to be excluded the landscape would look similar to that in Fig. 2b.

Following this optimization stage, the design of the algorithm is fixed for comparison to the Word Count approach. Figure 3 presents a detailed view of the proposed approach. The preprocessing classes are listed in order of their application during subsequent experimentation. All preprocessing classes, except

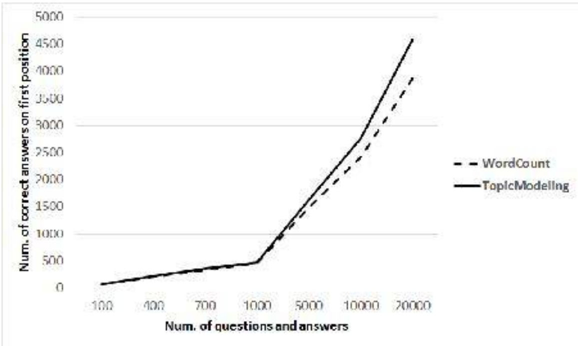


Figure 5. Number of correct answers in the first position for the Topic Modeling and Word Count approaches.

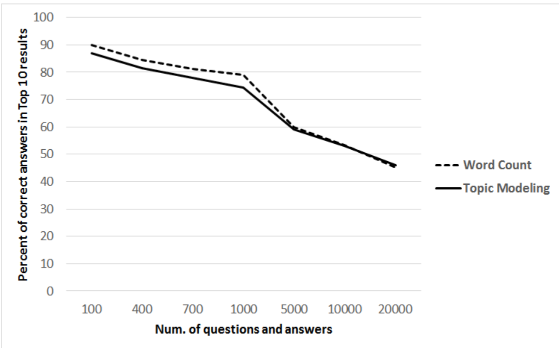


Figure 4. Percentage of correct answers in the top 10 most similar answers for the Topic Modeling and Word Count approaches.

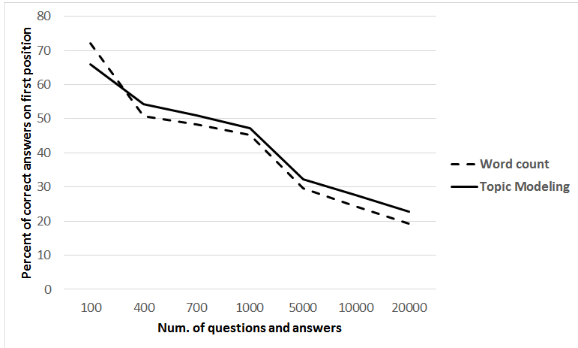


Figure 7. Percentage of correct answers in the first position for the Topic Modeling and Word Count approaches.

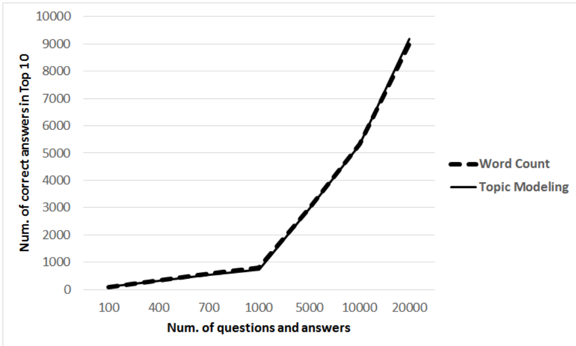


Figure 6. Number of correct answers in the top 10 most similar answers for the Topic Modeling and Word Count approaches.

InsertSynonyms, *PipeStem*, and *StanfordLemmatizer*, were taken from *Mallet*.

B. Main Results

It has been shown that the best results are obtained when using preprocessing steps 1—5 (as described in subsection II.E) and the query likelihood probability similarity measure.

In order to avoid overfitting and to present a fair comparison between the Topic Modeling and Word Count approaches, the algorithms are compared using the unseen Yahoo! Answers test datasets (containing 100, 400, 700, 1 000, 5 000, 10 000, 20 000 question-answer pairs).

The number of correct answers in first position, i.e. those with the highest similarity to the question according to (6) for the Topic Modeling approach and those with the highest TF-IDF measure for the Word Count approach, are presented in Fig. 5. When the number of Q-A pairs is less than 1000, both methods result in approximately equal performance (note that this does not mean that both methods give the same answers for each question). As the number of question-answer pairs increases, however, the difference between the two methods become more pronounced.

An additional important characteristic of each solution is the percent of correct answers in first position. These results are presented in Fig. 7 and it can be observed that as the number of documents increase, performance (according to this measure) decreases. Nevertheless, as the number of documents increases past 400, the Topic Modeling approach results in a slower performance decrease when compared to the Word Count approach.

Besides the number of the correct answers in first position, another important evaluation criteria is the number of times that the correct answer appears in the top 10 most similar results. This evaluation is presented in Fig. 6. Both methods result in almost equal performance. The differences in the results range from $\pm 0.03\%$ to $\pm 0.01\%$, depending on the number of documents (when using a smaller number of documents, the difference is greater).

The percentage of questions in which the correct answer appears in the top 10 most similar results is presented in Fig. 4. This confirms what was previously discussed: when the number of documents is greater than 5000, the performance of each model is approximately equal.

All of the results presented in this section demonstrate that the Topic Modeling approach gives better performance when evaluating according to stricter criteria, such as the number of correct answers in first position. When the criteria is weaker, such as evaluating the top 10 results, the lexical part of the similarity measure becomes more significant and the model starts to act in a similar way to the Word Count approach. This is because of the model's inability to find semantically similar answers, therefore the contribution from the topic similarity term in (6) becomes almost zero. Nevertheless, it can be concluded that the Topic Modeling approach ranks correct answers higher in the list of results when compared to the Word Count approach.

Finally, the statistical significance of the differences between the first position results of each approach were tested using the Wilcoxon matched pairs test. The results of these statistical tests, using a significance level of 0.05, are presented in Table II.

These results show that the Topic Modeling approach to question answering gives significantly better results than the reference model.

IV. Conclusions

It has been shown that a question's topic structure, as well as its lexical structure is important for correct answer selection. An interesting question that arises when dealing with online, collaborative question-answer data sources is: What exactly is the correct answer? In this work, the best answer or the best rated answer was assumed to be correct, but there is no guarantee that this is true. Without any expert knowledge to judge which answer is truly correct, any system validated using this data cannot be completely trusted and should instead be used as an aid to find correct answers.

Experimentation has been conducted using two distinct, real-world datasets, giving weight to the generality of the findings presented in this manuscript. Design decisions for the Topic Modeling approach (preprocessing and parameter values) were made using one dataset and applied to another. It was shown to outperform the Word Count approach, however, the differences may be more pronounced when optimizing the parameter values using data derived from the same source.

The most appropriate application of this work would be in interactive systems, which contain a lot of similar and frequently asked questions. In this case the user can be presented with a list of possible answers to their question (say the top 10 most similar), as it has been shown that the discussed approach performs well in this setting. It has been shown that using the Topic Modeling approach, the user would find the correct answer more quickly (as it will be located higher in the ranked list of results) when compared to the Word Count approach.

To further improve the method, future work should be directed towards integrating and encoding expert knowledge within the discussed Topic Modeling approach.

TABLE II.
RESULTS OF THE WILCOXON MATCHED PAIRS TEST FOR STATISTICAL DIFFERENCE BETWEEN THE FIRST POSTION RESULTS FOR THE TOPIC MODELING AND WORD COUNT APPROACHES, STATISTICALLY SIGNIFICANT RESULTS ARE IN BOLD, STATISTICAL SIGNIFICANCE IS TAKEN TO BE 0.05.

# of docs	WordCount	TopicModeling	Significance
100	72	66	0.710
400	**203**	**217**	**0.031**
700	**338**	**357**	**0.021**
1000	**453**	**472**	**0.002**
5000	1484	1614	0.075
10000	**2422**	**2766**	**0.027**
20000	**3866**	**4576**	**0.000**

References

[1] David M. Blei, et al., "Latent Dirichlet Allocation," Journal of Machine Learning Research,, vol. 3, pp. 993-1022, 2003.

[2] A. Celikyilma, et al., "LDA Based Similarity Modeling for Question Answering," In Proceedings of the NAACL HLT 2010 Workshop on Semantic Search, pp. 1-9., Los Angeles,2010.

[3] Z. Zolaktaf, et al., "Modeling community question-answering archives," presented at the Second Workshop on Computational Social Science and the Wisdom of Crowds NIPS, 2011.

[4] Guangyou Zhou, et al., "Improving Question Retrieval in Community Question Answering Using World Knowledge," Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, pp. 2239 - 2245, 2013.

[5] B. Croft, et al., *Search Engines, Information Retrieval in Practice*, Addison-Wesley, 2010.

[6] A. McCallum. (2002)., *MALLET: A Machine Learning for Language Toolkit* [online]. Available http://mallet.cs.umass.edu

[7] D. Mimno, "Machine Learning with MALLET," Department of Information Science, Cornell University.

[8] T. Griffiths and M. Steyvers, "Finding scientific topics," Proceedings of the National Academy of Sciences, vol. 101, pp. 5228-5235, 2004.

[9] C. Manning, et al., "The Stanford CoreNLP natural language processing toolkit," in In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, 2014, pp. 55-60.

[10] V. Hollink et al., "Monolingual document retrieval for European languages," Information Retrieval, vol. 7, no. 1, pp. 33-52, 2004.

[11] M. Porter, "An algorithm for suffix stripping," Program, vol. 14, no. 3, pp. 130-137, 1980.

[12] H. Wu et al., "Interpreting TF-IDF term weights as making relevance decisions," vol. 26, no. 3, pp. 13:1-13:37, 2008.

[13] J. Ramos, "Using TF-IDF to determine word relevance in document queries," in Proceedings of the 1st Instructional Conference on Machine Learning,, Piscataway, 2003.

Application of adaptive neuro fuzzy systems for grinding process modeling

Pavel Kovač*, Dragan Rodić*, Marin Gostimirović*, Borislav Savković*, Dušan Ješić**
*Univeristy of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia
** Medjunarodna Tehnoloska-Menadzerska Akademija – MTMA, Trg Dositeja Obradovica 7, 21000 Novi Sad, Serbia
pkovac@uns.ac.rs; rodicdr@uns.ac.rs; maring@uns.ac.rs; savkovic@uns.ac.rs; dusanjesic@hotmail.com

***Abstract*—Modeling and prediction of cutting forces by intelligent techniques in grinding operations play an important role in the manufacturing industry. This paper proposes a method using an adaptive neuro-fuzzy inference system (ANFIS) to currently establish the relationship between the machining conditions and cutting force, and consequently can effectively predict cutting force using cutting parameters work speed, feed rate and depth of cut. The results indicate that the ANFIS modeling technique can be effectively used for the prediction of cutting force in grinding process.**

I. Introduction

The grinding process has been shown to be a good method as a final operation for heat-treated materials. The machining parameters of grinding process, such as feed-rate, cutting speed and depth of cut are usually selected before machining according to standard manuals or experience [1]. In setting the machining parameters, the main goal is the minimum cutting forces [2]. The main problem is how to currently obtain the actual cutting forces using various parameters of cutting operations. It is difficult to utilize the optimal functions of a machine owing to there being too many adjustable machining parameters. As a result, many machining systems are inefficient and run under the operating conditions that are far from optimal criteria [3]. To know the optimal criteria, it is necessary to employ intelligent models making it feasible to do predictions in function of response parameters [4]. In order to achieve the minimal cutting forces, machining parameters must be optimally set.

Artificial intelligent techniques, such as adaptive neuro fuzzy systems, has been successful applied to machining processes through recent years. A broad literature survey has been conducted on the application of artificial intelligence systems to machining processes [5-7]. As all these researches paid high attentions on the fuzzy correlation between the experimental results and influential factors. A lot of ANFIS methods were also developed and used for predicting cutting forces [8-10]. From the review of literature, it is observed that neuro-fuzzy systems have found wide applications in modeling of process parameters.

In solving problems of modeling and prediction of cutting forces, this study uses a more powerful learning tool known as a combination of fuzzy logic and neural networks. It is known that the adaptive neuro-fuzzy inference system (ANFIS) is efficient for non-linear mapping [11]. ANFIS is a fuzzy inference system implemented in the framework of an adaptive neural network. The main problem with fuzzy logic is that there is no systematic procedure to define the membership function parameters. ANFIS eliminates the basic problem in fuzzy system design, defining the membership function parameters and design of fuzzy if-then rules, by effectively using the learning capability of neural network for automatic fuzzy rule generation and parameter optimization.

In this paper an adaptive neuro-fuzzy inference system (ANFIS) is used to correlate the machining parameters to cutting force in grinding process using the data generated based on experimental observations. The results of the experiments were also compared with those of the ANFIS predictions [12].

II. Experimentals procedures

Cutting tests were performed on a 4 kW cylindrical grinder machine. Tool was cylindrical grinding wheel Ø 350x40x127 mm, type B60L6V. The working material was a cylindrical shaped of Ø 60 x 150 mm of steel EN 34Cr4 and was fixed on grinding machine table.

The experiment was carried out for different combinations of work speed, feed rate and depth of cut according to the planning of experiment. The workpiece was mounted on a three-component piezo-electric dynamometer. Output parameter was Tangential force F_t (N). Other parameters were kept constant: tool geometry, tool wear, cooling and lubricating fluid, dynamical system machine-tool-workpiece.

III. ANFIS modeling of grinding

The architecture of the ANFIS used in the proposed method is shown in Fig. 1. The process followed in this study is illustrated in Fig. 2. There are three input parameters (v, f, a) and one output value, the predicted cutting force (F_t). Denote the output node i in layer l as $O_{l,i}$. The used five-layer ANFIS is described as follows:

Layer 1:

Every node in this layer is an adaptive node with a node output defined:

$$\begin{aligned} O_{1,i+1m} &= m_{vi}(v),\ i=1,\dots,m; \\ O_{1,i+2m} &= m_{fi}(f),\ i=1,\dots,m; \\ O_{1,i+3m} &= m_{ai}(a),\ i=1,\dots,m; \end{aligned} \tag{1}$$

Where $\boldsymbol{v}$, $\boldsymbol{f}$, $\boldsymbol{a}$ *are the inputs to the nodes,* $\boldsymbol{v_i}$, $\boldsymbol{f_i}$, $\boldsymbol{a_i}$ are the i^{th} fuzzy sets associate with the membership functions

of this nodes, and ***m*** is the number of fuzzy sets for each input parameter. In other words, $O_{1,i}$ which are outputs of this layer, are the membership values of the premise part v_i, f_i, a_i. Here, we choose the membership function to be *Gaussian* shaped with maximum equal to 1 and minimum equal to 0.

$$O_i^1 = \mu_{Ai}(x) = e^{\frac{-(x-c)^2}{2\sigma^2}} \tag{2}$$

where x, c and σ is the parameter set.

Layer 2:

Every node in this layer is a fixed node labeled Π, which multiplies the incoming signals and outputs the product. For instance:

$$O_i^2 = w_i = \mu_{Ai}(x)\mu_{Bi}(y) \quad i = 1,\, 2 \tag{3}$$

Each node output represents the firing strength Π of a rule.

Layer 3:

Every node *i* in this layer is a fixed node labeled *N*. The i^{th} node calculates the ratio of the i^{th} rule's firing strength to the sum of all rule's firing strengths:

$$O_i^3 = \overline{w}_i = \frac{w_i}{\sum_i w_i} = \frac{w_i}{w_1 + w_2 + w_3} \quad i = 1, 2, 3 \tag{4}$$

For convenience, outputs of this layer are called *normalized firing strengths.*

Layer 4:

Every node *i* in this layer is an adaptive node with a node function:

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i(p_i v + q_i f + rs_i) \quad i = 1, 2 \tag{5}$$

Where $\overline{w}_i$ is the output of layer 3 and {*pi*, *qi*, *ri*} is the parameter set. Parameters in this layer are referred to as *consequent parameters.*

Layer 5:

This layer is called as the output nodes. The single node in this layer computes the overall output as the summation of contributions from each rule.

$$O_i^5 = f(x.y) = \sum_i \overline{w}_i \cdot f_i = \overline{w}_i f_1 + \overline{w}_i f_2 = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{6}$$

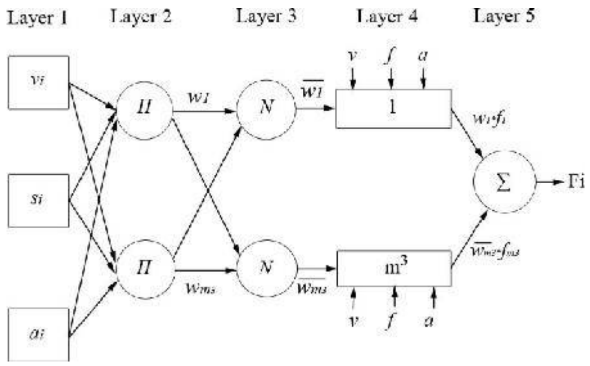


Figure 1. ANFIS architecture

TABLE I.
EXPERIMENTAL DATA

No.	Machining factor			Experimentally measured values
	v	*f*	*a*	F_t
	[m/min]	[mm/o]	[mm]	[N]
1	18,4	20	0,01	10,9
2	36,8	20	0,01	11,8
3	18,4	30	0,01	13
4	36,8	30	0,01	13,2
5	18,4	20	0,02	17,4
6	36,8	20	0,02	17,9
7	18,4	30	0,02	18,8
8	36,8	30	0,02	20
9	26	25	0,014	12,3
10	26	25	0,014	13,5
11	26	25	0,014	12,2
12	26	25	0,014	14
13	16	25	0,014	12,6
14	42,4	25	0,014	13,6
15	26	18,4	0,014	12,4
16	26	32,6	0,014	14,1
17	26	25	0,0086	10,8
18	26	25	0,023	18,1
19	16	25	0,014	12,1
20	42,4	25	0,014	13,9
21	26	18,4	0,014	12,6
22	26	32,6	0,014	14,6
23	26	25	0,0086	11
24	26	25	0,023	18,5

IV. RESULTS

In this study, an ANFIS model based on both ANNs and FL has been developed to predict cutting force in grinding process. Four machining parameters namely work speed, feed rate and depth of cut were taken as input features.

A full factorial experimental design was adopted to study to collect cutting force values. The data set was used as inputs of ANFIS in training and testing stage. The experiments were divided into two group for training (the first 16 experiment) and testing (remaining) of ANFIS. According to the experimental results, the proposed method is efficient for estimating of the cutting force in grinding process. The average deviation of the testing data for cutting force is 7.81 % (Table 2).

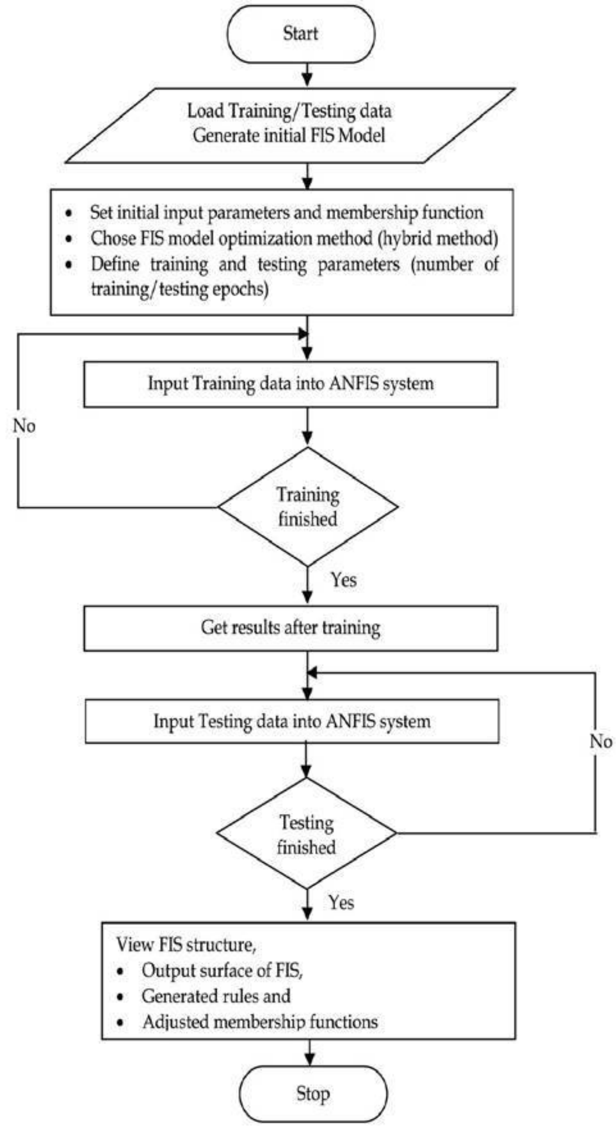


Figure 2. Comparison Flowchart of cutting force prediction of ANFIS system [13]

TABLE II.
TEST DATA

No.	Machining factor			Tangential force		
	v	*f*	*a*	F_t	*Fanfis*	*Error*
	[m/min]	[mm/o]	[mm]	[N]	[N]	[%]
1.	36.8	20	0.02	16.6	15.8855	4.4978
2.	18.4	30	0.02	17.2	15.9767	7.6568
3.	26	25	0.014	12.6	14.5713	13.5286
4.	26	25	0.014	12.8	14.5713	12.1561
5.	26	25	0.0086	11.0	11.1825	1.6320
6.	26	25	0.023	17.1	15.6066	9.5690
7.	26	25	0.0086	11.1	11.1825	0.7378
8.	26	25	0.023	17.6	15.6066	12.7728
Average error:						**7.81**

In solving problems of modeling and prediction of cutting force in grinding process, this study uses a more powerful learning tool known as a combination of fuzzy logic and neural networks. On Fig. 3. is shown ANFIS system for prediction cutting force in grinding process. Figs. 4 depict the comparison of experimental and ANFIS results for the cutting force, respectively. It proved that the method used in this paper is feasible and could be used to predict the Fc in an acceptable error rate for grinding process. The compared lines seem to be close to each other indicating with good agreement.

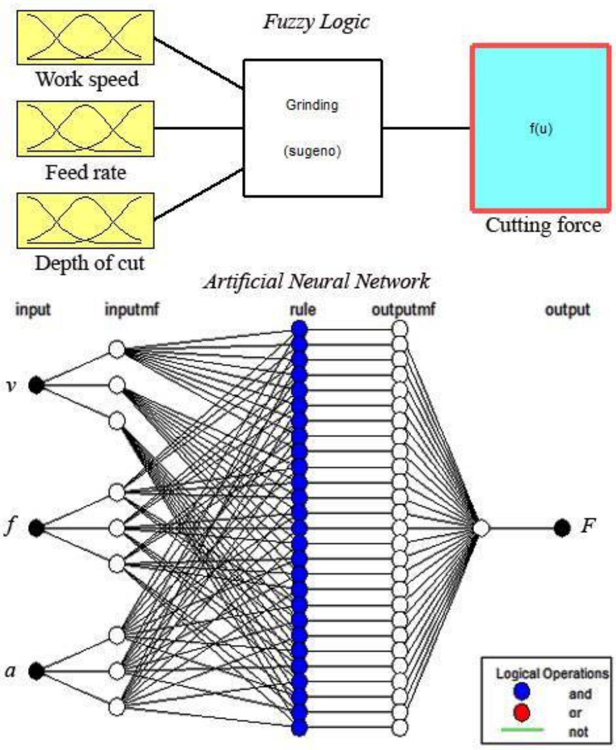


Figure 3. ANFIS architecture for prediction cutting force in grinding process

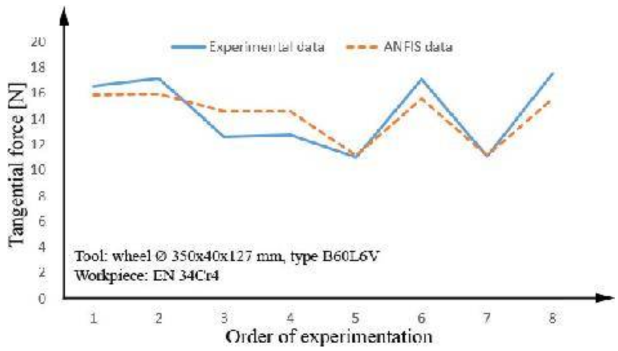


Figure 4. Comparison of predicted results with actual values for Tangential force.

Fuzzy 3D surface viewer give a good overview of the process behavior, displaying how the response vary with two different parameters, where remaining three of the inputs must be held constant. The form of the surface viewer depends of the membership functions and their parameters. The selection of most influential parameters was based on fuzzy logic.

At medium and high feed rate levels from 26 to 32 mm/o and depth of cut is constant 0.014 mm, the cutting force decrease with decreasing work speed for all cutting speed ranges, Fig. 5. At work speed ranging from 15 to 40 m/min, the cutting force increase with increasing depth of cut for ranging from 0.01 to 0.02 mm, Fig 6.

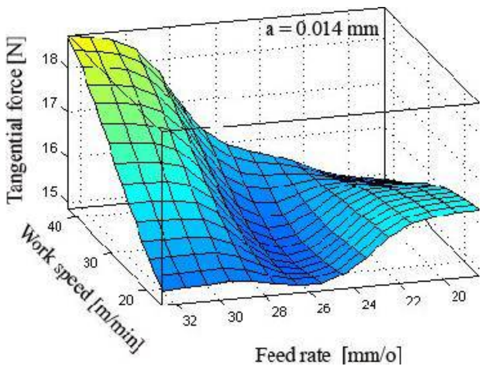


Figure 5. Fuzzy 3D surface viewer, effect of work speed and feed rate on the tangential cutting force, where a=const.

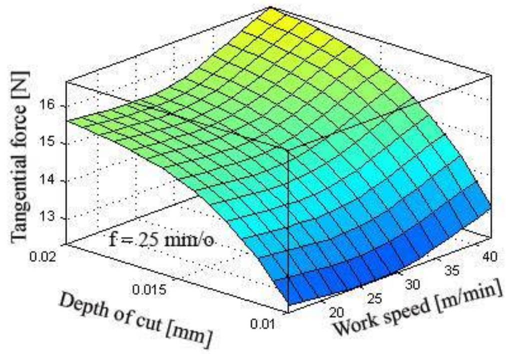


Figure 6. Fuzzy 3D surface viewer, effect of work speed and depth of cut on the tangential cutting force, where f=const.

V. Conclusion

This paper proposes a method using an adaptive neuro-fuzzy inference system (ANFIS) to currently establish the relationship between machining conditions and cutting force, and consequently can effectively predict cutting force using cutting parameters work speed, feed rate and depth of cut. The advantages of the proposed method is obtain the actual cutting force using various parameters of cutting operations. Experimental results have shown that the proposed ANFIS based method outperforms the existing polynomial network based method in terms of modeling and prediction accuracy.

Acknowledgment

The Ceepus mobility program and The Tehnological Development program of Republic Serbia, supported this project. For their support authors show great appreciation.

References

[1] Zuperl, U., F. Cus and M. Milfelner. "Fuzzy control strategy for an adaptive force control in end-milling." Journal of Materials Processing Technology Vol. 164-165, pp. 1472-1478, 2005.

[2] Gostimirović, M., D. Rodić, P. Kovač, D. Ješić and N. Kulundžić. "Investigation of the cutting forces in creep-feed surface grinding process." Journal of Production Engineering Vol. 18 (2), pp. 21-24, 2015.

[3] Tönshoff, H. K., J. Peters, I. Inasaki and T. Paul. "Modelling and Simulation of Grinding Processes." CIRP Annals - Manufacturing Technology Vol. 41 (2), pp. 677-688, 1992.

[4] Kovac, P., D. Rodic, V. Pucovsky, B. Savkovic and M. Gostimirovic. "Application of fuzzy logic and regression analysis for modeling surface roughness in face milliing." Journal of Intelligent Manufacturing Vol. 24 (4), pp. 755-762, 2012.

[5] Çaydaş, U., A. Hasçalık and S. Ekici. "An adaptive neuro-fuzzy inference system (ANFIS) model for wire-EDM." Expert Systems with Applications Vol. 36 (3), pp. 6135-6139, 2009.

[6] Adeel H. Suhail, N. Ismail, S. V. Wong and N. A. A. Jalil. "Cutting parameters identification using multi adaptive network based Fuzzy inference system: An artificial intelligence approach." Scientific Research and Essays Vol. 6 (1), pp. 187-195,, 2011.

[7] Liang, Z., B. Xie, S. Liao and J. Zhou. "Concentration degree prediction of AWJ grinding effectiveness based on turbulence characteristics and the improved ANFIS." The International Journal of Advanced Manufacturing Technology Vol. 80 (5-8), pp. 887-905, 2015.

[8] Iqbal, A., N. He, L. Li and N. Dar. "A fuzzy expert system for optimizing parameters and predicting performance measures in hard-milling process." Expert Systems with Applications Vol. 32 (4), pp. 1020-1027, 2007.

[9] Maher, I., M. E. H. Eltaib, A. A. D. Sarhan and R. M. El-Zahry. "Cutting force-based adaptive neuro-fuzzy approach for accurate surface roughness prediction in end milling operation for intelligent machining." The International Journal of Advanced Manufacturing Technology Vol. 76 (5-8), pp. 1459-1467, 2014.

[10] Ramanujam, R., K. Venkatesan, V. Saxena, R. Pandey, T. Harsha and G. Kumar. "Optimization of Machining Parameters Using Fuzzy Based Principal Component Analysis during dry turning operation of Inconel 625 – A hybrid approach." Procedia Engineering Vol. 97, pp. 668-676, 2014.

[11] Mohd Adnan, M. R. H., A. Sarkheyli, A. Mohd Zain and H. Haron. "Fuzzy logic for modeling machining process: a review." Artificial Intelligence Review Vol. 43 (3), pp. 345-379, 2013.

[12] Prabhu, S., M. Uma and B. K. Vinayagam. "Adaptive neuro-fuzzy interference system modelling of carbon nanotube-based electrical discharge machining process." Journal of the Brazilian Society of Mechanical Sciences and Engineering Vol. 35 (4), pp. 505-516, 2013.

[13] Zuperl, U. and F. Cus. "Optimization of cutting conditions during cutting by using neural networks." Robotics and Computer-Integrated Manufacturing Vol. 19 (1-2), pp. 189-199, 2003.

Free-hand human-machine interaction in vehicles

Tomaž Čegovnik and Jaka Sodnik
University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia
tomaz.cegovnik@fe.uni-lj.si
jaka.sodnik@fe.uni-lj.si

***Abstract* - This paper focuses on alternative interaction techniques and user interfaces for in-vehicle information systems (IVIS). Human-machine interaction in vehicles needs to introduce new intuitive and more natural interaction approaches, which would reduce driver's distraction and increase safety. We present a prototype of a gesture recognition system intended for in-vehicle free-hand interaction. Our solution is based on a Leap Motion Controller. We report also on a short user study with the proposed system. Test subjects performed a set of tasks and reported on their experience with the system through a user experience questionnaire. The study reveals that free-hand interaction is attractive and stimulating, but it still suffers from various technological issues, particularly with the efficiency and robustness of free-hand gestures recognition techniques.**

I. Introduction

Today's vehicles provide an increasing number of new functionalities that enhance the safety and driving performance of drivers or raise their level of comfort. In addition to a variety of passive and active safety systems, information systems have also become increasingly common, enabling modern communication mechanism and luxury facilities [1]. The most common examples are navigation systems, multimedia devices and connectivity services. With every generation of vehicles, the range of these functionalities increases and requires new techniques for human-machine interaction (HMI). Due to the growing functional complexity and mostly restriction to tactile input and visual output, many user interfaces show very poor usability. In addition these systems require long learning periods, which often increases the potential of errors and user frustration [2][3].

However, the primary task in vehicles remains the driving process itself, which demands a certain amount of visual and cognitive attention. Secondary tasks, such as controlling a complex infotainment system, can distract the driver from controlling the car and focusing on the traffic. Since inattention proved to be a major cause of many car accidents, it is reasonable to search for interaction options, which cause less driver distraction in both cognitive and visual domains. Interaction with IVISs is most commonly enabled through buttons and rotary knobs attached to different parts of the vehicles' steering wheel or dashboard [4]. Other input possibilities are speech and interaction via touch interfaces. All interactions, except speech, require eye contact and lead to a visual distraction of the driver.

HMI in vehicles needs to introduce intuitive and natural interaction approaches such as multimodal interaction interfaces. Multimodality defines an interaction form in HMI in which more modalities (i.e. communication channels) are used simultaneously. Input and output processes are combined in a coordinated manner [5]. Currently available conventional interfaces can be upgraded by simultaneous use of speech, gestures and touch for input and displays, speech non-speech sounds and haptic feedback for output. Theoretical foundation for this principle is the unique characteristic of human working memory which seems to be working in fully functional separable components for different human senses [6]. Should a component be overwhelmed (e.g. visual working memory while driving a vehicle) other components (e.g. gestures and speech) can be used instead. By using more components simultaneously, the processing capacity of the working memory is optimized and the visual working load is relived, which leads to less visual distraction and errors.

Such intuitive and natural interaction systems address also the problem of increasing amount of complex functionalities since this approach shortens drivers' learning and adaptation period. Namely, the learning process represents even higher distraction for the primary task as it derives a huge amount of visual and mental attention to a new user interface. During the learning period, driver's brain builds a mental representation of the interaction system which is a mental mirror image of the real system. We call a system intuitive, if the mental representations are already present and no or very little learning period is needed. Speech, movements and natural gestures are such examples since they are widely used in common every day's situations and interactions.

In this paper, we are focusing on free-hand gesture interaction as a big potential technique to enhance

intuitiveness of interaction with in-vehicle devices [7]. It poses numerous advantages over tactile and touch modalities, since it requires lower amount of visual load, reduces driving errors and increases the level of user acceptability. It follows the intuitive interface approach as it uses natural hand gestures where no learning effort is needed. As a part of a multimodal interaction system (e.g. in addition to visual and auditory modalities), it allows the driver to keep his eyes on the road and improve his or her safety.

II. Related work

A lot of in-vehicle interaction research focuses on minimizing driver workload and reducing distraction caused through visual attention of such systems [8]. Different interaction techniques were compared focusing on their effects on driver performance and eye glance behavior. The results showed that gesture interaction could reduce eye glances on simple secondary tasks although gesture interaction is not fully attention free [9].

A research of the BMW Group shows that using a gesture vocabulary reduced to meaningful gestures can be recognized using simple state of the art hardware. Their system is able to distinguish 17 different hand gestures, which can be used to provide skipping between radio stations and navigating. It also distinguishes six different head gestures, which can be used for simple yes/no decisions (e.g. accepting and denying incoming calls) [3].

Other research states that two techniques provide “low attention” interaction (i.e. speech recognition and gesture-based interaction) opposed to touch-based and tactile-based techniques [9]. The latter suffer from a number of inherent limitations in terms of reducing visual attention [10]. Speech recognition on the other hand, can also be very cognitive demanding and therefore impractical and flawed for in-vehicle interaction [11][12] although it seems very useful as it provides hands-free and visual-free interaction.

Since gesture-based interaction could provide a suitable alternative to other types of interaction, research focuses on developing gesture recognition systems. Almost all developed interfaces for gesture recognition are vision based [13][14] since there are numerous advantages of such approach. Firstly, the camera could serve a multi-purpose analyzer of also other activities and not just hand-gestures. Secondly, it offers flexibility where the gestures can be performed and allows location customization. Thirdly, there are advantages in terms of cost and simplicity of installation. On the other hand, the major challenge of vision-based gesture recognition systems is how to generalize a great number of different users and possible variation of gestures. The algorithms must be resistant also to varying in-vehicle illumination conditions [1].

III. Gestures

Firstly, we need to define gestures since they have numerous definitions, depending on a specific research field. Physically, gestures are movements of individual limbs. Often the term refers also to facial expressions, gaze tracking, head movements and whole body postures [4]. The primary goal of various movements of body parts is augmenting the basic verbal interpersonal communication by exchanging also expressions. Gestures, as a mean of communication between persons, are used and identified subconsciously and their semantics is often hidden or unknown.

In this paper, we primarily refer to gestures as conscious and intentional movements of a selected hand and arm to communicate information with a system. Gestures can be identified and described by two well-defined attributes: temporal seclusion and movement trajectory [3].

In order to recognize various gestures, we have to understand their primary structure. Although gestures look like one continuous movement, they consist of three phases [4].

- Preparation: The hand is brought to the position to start the stroke.
- Stroke: Main part of the gesture which defines its meaning.
- Retraction: The hand is brought back to the resting position.

The stroke phase is essential to the gesture, but in order to identify more gestures merged into one continuous sequence, we need to take into account also the first and last phase.

According to Geiger gestures in HMI can be divided into seven categories [10].

- Mimic gestures: imitating an object (e.g. pick up a phone),
- Schematic gestures: special kind of mimic, standardized symbols witch need to be learned,
- Kinemimic gestures: imitating a direction,
- Symbolic gestures: imitating an abstract feature like an emotion, feeling or thought, e.g. thumb up for yes.
- Deictic gestures: pointing towards the intended destination.
- Technical gestures: used by experts in their working field where no other communication is possible (e.g. diving).
- Encoded gestures: language of technical gestures.

A literature review and analysis of a research published in [15] led to an overall classification of hand

gestures based second controls (see Figure 1).

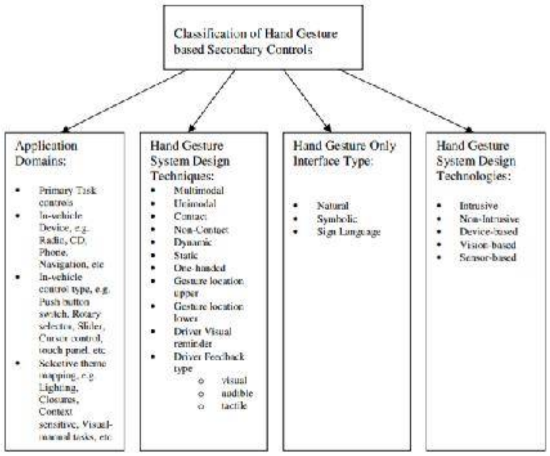


Figure 1: The diagram shows the organization and the categories used for classifying gestures [15]

IV. Gesture recognition

There are two different technology bases for gesture recognition (i.e. video-based systems and sensor-based systems). Video-based recognition is realized by cameras and requires constant illumination. The various lighting conditions in vehicles present a problem for robust image processing, therefore a constant near-infrared lightning source is needed and a daylight filter to compensate the varying conditions. Sensor-based gesture recognition depends on distant measurement sensors, and works better than the video-based gesture recognition since the background is farther away from the sensor and can be blocked out [4].

However, gesture recognition requires light independent techniques. A motion based entropy technique has to be applied to the near infrared imaging. Human skin has the characteristic of high reflectance of infrared radiation and in majority of cases the hand is the brightest object in the scene [3]. An effective segmentation algorithm is required to segment a hand shape from the background, such as entropy motion segmentation or approaches based on restricted coulomb energy [14].

In-vehicle HMI requires a hand to be displayed in a 2D image field visible to the camera and restricted by the camera's view cone. If the human body is beyond the effective interaction space, the hands may not be fully displayed in the image and the gesture recognition would not be complete. In order to address this problem and to design better gesture interfaces, several researches have defined the regions within the vehicle's cockpit for optimal performance of gesture recognition systems [16].

V. Leap Motion Controller

The Leap Motion Controller is a small cost effective and accessible USB device, which recognizes all types of hand and finger movements and measures their position and velocity [17]. It illuminates the space over the controller with three infra-red LED light sources and captures the hands with two cameras (see Figure 2). The captured stereo-image is processed with a segmentation algorithm resulting in a data structure, which contains precise position of each finger at every moment. The Leap Motion's API processes this data and gives precise velocities of each finger and hand, and also combines movement patterns into gesture frames. It recognizes four different gesture types:

- Circle - A finger drawing a circle.
- Swipe - A long, linear movement of a hand and its fingers.
- Key Tap - A tapping movement by a finger as if tapping a keyboard key.
- Screen Tap - A tapping movement by the finger as if tapping a vertical computer screen.

The controller's field of view is an inverted pyramid centered on the device [18]. The effective range of the controller extends from approximately 25 to 600 millimeters above the device. The main limitation of the controller's performance is the low and inconsistent sampling frequency (i.e. its mean value is less than 40Hz).

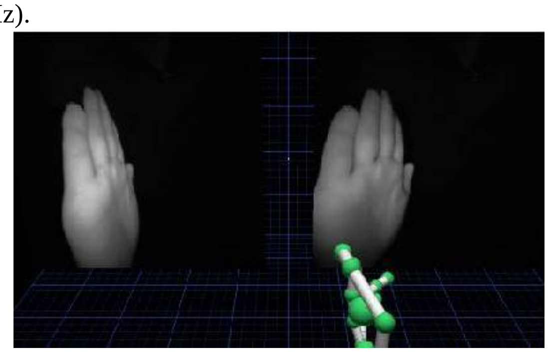


Figure 2: Hand illuminated by infrared LED diodes and captured by the two cameras of the Leap Motion controller (i.e. the bottom right part of the figure [21]).

VI. Free-hand gesture interaction with IVIS

We used the Leap controller's gesture models to develop a simple free-hand gesture interface for IVIS. Our IVIS simulated the majority of common functionalities related to navigation (e.g., traffic reports, navigation assistance), entertainment (e.g., audio, video, communication) and vehicle control (e.g., air conditioning, system information, cruise control) [19]. All features were accessible through a hierarchical menu structure. The top-most level of the structure was called the "Main menu" and each level of the individual sub-menu consisted of up to eight options. At each level, the user could freely navigate between all the available

options, select one of them and enter the corresponding sub-menu or exit and return to the previous menu.

The output of the interface was a simple graphical menu displayed on a small dashboard screen (Figure 3). Four different commands were required to control this menu: up, down, confirm and return. Up and down were moving the selecting cursor up and down on the list while confirm and return entered or left a submenu. Four free-hand gestures were bound to these four commands (see Figure 4). A "circle" gestures drawn with a finger were used for up and down commands – a clockwise circle for moving down and counterclockwise for moving up. A finger "tap" gesture was used for selecting an item and a "swipe to the left" gesture was used for returning to the previous menu.

Figure 3: The study setup: the Leap motion controller is located below the hand and the IVIS display is located next to the screen of the simulator.

Figure 4: Gestures controlling the IVIS.

VII. Preliminary user study with the proposed system

We performed a short user study to evaluate the usability of the proposed free-hand gesture interaction system. The goal was primarily to get some user feedback on gesture interaction and to assess its acceptance and efficiency.

16 subjects (9 female and 7 male) participated in the user experience evaluation. They had to perform three different tasks with the system:

- set a temperature to the selected value,
- set a radio receiver to the specific station or
- check the vehicle's status (e.g. battery state, fuel level or errors).

The experiment was performed in provisionary driving simulator with stable light conditions. After performing the task, each subject filled out the User Experience Questionnaire (UEQ) [20]. The UEQ is intended to be a user-driven assessment of a system quality and usability. It consists of 26 bipolar items rated on a seven-point Likert scale. The UEQ allows the experience to be rated using the following six subscales: attractiveness, perspicuity, efficiency, dependability, stimulation and novelty of the display technique.

VIII. Results and discussion

Results of our user study show high interest of test subject for this new interaction technique. Figure 5 compares mean values for all six UEQ categories. The gesture interface has very positive values in the categories novelty, stimulation and attractiveness. We believe since the technology is new for the subjects, they are attracted to it and highly stimulated to use it.

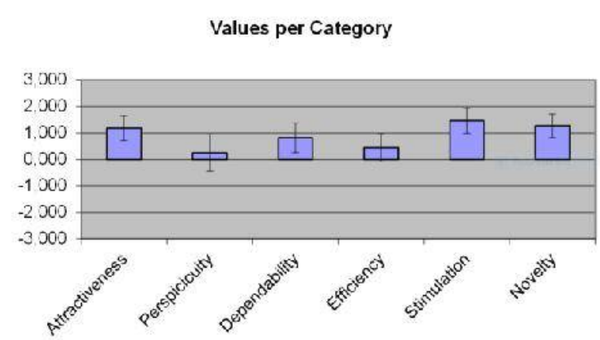


Figure 5: Result values per category of the UEQ [21].

On the other hand, mid-value results in the category dependability indicate that test subjects are not entirely convinced they could rely on this type of interface in real vehicles. The poorest results, but still positive, are in the

categories perspicuity and efficiency. The latter reflects the problems of the proposed interface to clearly and robustly detect gestures performed by variety of test subjects. Despite using only simple and intuitive gestures, their correct execution depended strongly on individual's performance and varied a lot between different users [21].

In the course of the study, we also noticed that several users did not clearly understand the exact form for performing individual gestures. The biggest problem seemed to present the execution of a circular gesture, which should be only a simple circular movement of one finger. Instead, some users performed big circles with their entire arm which often escaped the controller's field of view and could not be successfully detected. We believe this is the main contributor to the poor result in the UEQ efficiency category.

IX. Conclusion

Nowadays the increasing range of functionalities included in in-vehicle interaction requires new interaction techniques. Conventional techniques often result in decreased driving performance since they require a huge amount of driver's visual attention and visual workload. Increased visual workload or even overload causes high distraction and great deficit of focus on the road and traffic situations. In this paper, we proposed a possibility to lower the visual workload in vehicles by using new and very intuitive interaction technique. We explored the opportunity of using gestures as interface input - they can be used intuitively because they are already part of interpersonal communication.

Our experimental results show that there is an attraction and interest among people for using this type of interaction. However, there is still a clear need to improve the detection performance of such systems. Majority of the development and research in this area try to recognize gestures through vision-based technology, i.e. capturing the hand with cameras and illuminating the vision field with an infra-red light source, which tries to nullify the restrictions due to variable lighting condition in vehicles. The appropriate mathematical segmentation algorithms need to be applied to extract usable data from the captured images.

An example of such technology is already in production in the BMW 7 series for triggering some basic phone functions and controlling the volume. Other car manufacturers (Jaguar, Mercedes, VW, Nissan) announced to implement gesture control systems at latest in 2018.

Although the free-hand gesture interface is designed to be intuitive and use intuitive gestures, each person executes gestures slightly different and the interface needs to adapt to these differences. Furthermore, additional studies need to take into consideration other modalities and other combinations of interaction techniques in order to develop highly usable, adaptable and robust configurations.

References

[1] Ohn-Bar, E.; Trivedi, M.M., "Hand Gesture Recognition in Real Time for Automotive Interfaces: A Multimodal Vision-Based Approach and Evaluations," in *Intelligent Transportation Systems, IEEE* Transactions on, vol.15, no.6, pp.2368-2377, Dec. 2014.

[2] Tönnis, M.; Broy, V.; Klinker, G., "A Survey of Challenges Related to the Design of 3D User Interfaces for Car Drivers" in *Proceedings of the 2006*

[3] Althoff, F.; Lindl, R.; Walchshäusl, L., "Robust Multimodal Hand- and Head Gesture Recognition for Controlling Automotive Infotainment Systems", In: VDI-Tagung: Der Fahrer im 21. Jahrhundert, 2005.

[4] Zeiser, J., "Use of Freehand Gestures for Interactions in Automobiles: An Application-Based View", In: *Ubiquitous Computing, University of Munich*, 2011.

[5] Sharon O., "Ten myths of multimodal interaction" In *Commun. ACM* 42, 11, pp. 74-81 1999.

[6] Ablaßmeier, M., "Multimodales, kontextadaptives Informations-management im Automobil", Dissertation, Technische Universität München, 2009.

[7] Loehmann, S.; Diwischek, L.; Schröer, B., „The User Experience of Freehand Gestures", In: *User Experience in Cars Workshop at INTERACT 2011 - 13th IFIP TC13 Conference on Human-Computer Interaction*, pp. 54-58, 2011.

[8] M. Geiger, M. Zobl, K. Bengler in M. Lang, "Intermodal differences in distraction effects while controlling automotive user interfaces," 2001.

[9] Majlund Bach, K.; Gregers Jaeger, M.; Skov, M. B., „You Can Touch, but You Can't Look: Interacting with In-Vechicle Systems" in *CHI 2008 Proceedings*, pp.1139-1148, 2008.

[10] Geiger, M., „Berührungslose Bedienung von Infotainment-Systemen im Fahrzeug", Dr.-Ing. Dissertation, Technische Universität München.

[11] Gellatly, A. W., „The Use of Speech Recognition Technology in Automotive Applications", Faculty of the Virginia Polytechnic Institute and State, 1997.

[12] Sezgin, T. M.; Davies, I.; Robinson, P., "Multimodal Inference for Driver-Vecicle Interaction", In: *Proceedings of the 2009 Inter-national Conference on Multimodal Interfaces*, 2009.

[13] McAllister, G.; McKenna, S. J.; Ricketts I. W., "Towards a non-contact driver-vehicle interface", in *IEEE Intelligent Transportation Systems*, pp.58-63, 2000.

[14] Dong, G.; Yan Y., Xie, M., "Vision-Based Hand Gesture Recognition for Human-Vehicle Interaction" in *5th Int. Conf. Control, Automation, Robotics and Vision*, Singapore, 1998, pp. 151–155.

[15] Niedermaier, F. B., "Entwicklung und Beertung eines Rapid-Prototyping Ansatzes zur multimodalen Mensch-Maschine-Interaktion im Kraftfahrzeug", Dr.-Ing. Dissertation, Technische Universität München.

[16] Riener, A.; Ferscha, A.; Bachmair, F.; Hagmüller, P.; Lemme, A.; Muttenthaler, D.; Pühringer, D.; Rogner, H.; Tappe, A. & Weger, F. (2013), Standardization of the in-car gesture

interaction space., *in* Jacques M. B. Terken, ed., 'AutomotiveUI' , ACM, , pp. 14-21.

[17] Leap Motion Controller https://www.leapmotion.com/, 1/2016.

[18] J. Guna, G. Jakus, M. Pogačnik, S. Tomažič in J. Sodnik, “An analysis ofthe precision and reliability of the leap motion sensor and its suitability forstatic and dynamic tracking,”, In: *Sensors*, vol. 14, no. 2, pp. 3702, 2014.

[19] Jakus, G., Dicke, C., Sodnik, J., “A user study of auditory, head-up and multi-modal displays in vehicles”. In: Applied ergonomics, vol. 46, 2015, pp. 184-192.

[20] B. Laugwitz, T. Held and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire", *Lecture Notes in Computer Science*, pp. 63-76.

[21] Čegovnik, T., “Prostoročna interakcija človek-stroj v vozilih”, In Master Thesis, University of Ljubljana, Faculty of Electrical engineering, 2015.

Co-training based algorithm for gender detection from emotional speech

Jelena Slivka*, Aleksandar Kovačević*
* University of Novi Sad/Faculty of Technical Sciences/Computing and Control Department, Novi Sad, Serbia
slivkaje@uns.ac.rs, kocha78@uns.ac.rs

***Abstract*—Automatic recognition of gender from spoken data is an important prerequisite for many practical applications. In this paper we consider the problem of automatic gender recognition from emotional speech. Analyzing emotional speech adds to the complexity of the problem due to the variability of the speech signal which is driven by the speakers' emotional state. The ever-present and dominating bottleneck in automatic analysis of spoken language is the scarcity of labeled data. Here we investigate the possibility of applying a co-training style algorithm in order to reduce the human effort needed for data labeling. We apply several co-training settings: random split of features, natural split of features, majority vote of several co-training classifiers and Random Split Statistic Algorithm (RSSalg) we have developed earlier in order to boost co-training performance and enable its application on single-view datasets. We test the performance of the competing settings on IEMOCAP, a large, publicly available emotional speech database. In order to test the robustness of the compared method, in the conducted experiments we vary the amount of available labeled data. In our experiments a random feature split yielded a better performance of co-training than the feature split that can be considered as natural. The best performing settings proved to be the majority vote of co-training classifiers and RSSalg. These settings also proved to be the most robust ones considering the amount of the available labeled data.**

Keywords:* *semi-supervised learning; co-training; computational paralinguistics; gender recognition; emotional speech; SVM;

I. INTRODUCTION

Automatic recognition of gender from speech plays an important role in a broad range of practical applications. Gender-dependent models achieve better performance than gender independent models in various Automatic Speech Recognition tasks as they help reduce the inter-speaker variability [1]. Gender information has been shown to aid the detection of the speakers' emotional state [2]. It was used to leverage recommender systems [3]. Studies suggest that the acoustic analysis of pathological voices should be conveyed using gender-dependent feature subsets [4]. Gender detection module is one of the main parts of audio segmentation system presented in [5].

In [6] it was stated that one of the ever-present and dominating bottlenecks in automatic analysis of spoken language is data scarcity. Manual annotation of data instances by human experts is very tedious, time-consuming and expensive. This problem can be alleviated by application of semi-supervised learning (SSL). SSL techniques incorporate both labeled and unlabeled data in the training process, resulting with the models that have comparable or even better performance than traditional supervised models trained on much larger portions of labeled data. Successful application of SSL techniques highly reduces the human effort needed for annotation of the appropriate training set.

One of the major semi-supervised techniques is co-training [7]. Co-training assumes that the features of the dataset can be naturally separated in two disjunctive subsets called views. The two views should be based on two independent sources describing the same data, a trait not so commonly found in real-world problems.

In this paper, we investigate the possibility of applying a co-training style algorithm to the problem of automatic gender recognition from emotional speech in order to reduce the human effort needed for data labeling. The inherited difficulties of analyzing emotional speech add to the complexity of the problem [8]. For example, pitch information (commonly used for gender recognition) varies over time, driven by the speakers' emotional state [9]. It is proven that using pitch as criterion for gender classification of emotional speech deteriorates the accuracy by 21% compared to the processing of spontaneous speech [10].

Previous work has shown that co-training can be successfully applied to the task of gender recognition from spoken data [11]. In order to apply co-training authors have identified a reasonable feature split, motivated partly by the independence of the features and partly by the size of the resulting views in order to ensure the sufficiency condition. Although these views are not completely conditionally independent [12], the experimental results demonstrate the validity and effectivity of such feature split for co-training [11][12]. In this paper we test several co-training settings: co-training run using a "natural" feature split (defined in [11] and [12]), co-training run with a random feature split, majority vote of the ensemble consisting of multiple co-training classifiers run with different random splits (MV), and *Random Split Statistic algorithm* (*RSSalg*) we have developed earlier in [13] with the goal of eliminating the need for the "natural" feature split in co-training, as well as boosting the performance of co-training.

We have tested the algorithms on interactive emotional dyadic motion capture (IEMOCAP) database [14]. We have decided to use IEMOCAP database as it is relatively large (approximately 12 hours of audiovisual data), contains emotional speech and is publicly available and free of charge.

The IEMOCAP database consists of 5 sessions, each recorded by two mixed gender actors. Sessions were

manually segmented at the dialog turn level (speaker turn), defined as the continuous segments in which one of the actors was actively speaking. Here we consider the problem of segment-level gender annotation and consider each segment independently. We evaluate our solution using a five-fold-cross evaluation scheme where in each turn one session is used as test data (1 male and 1 female actor) while the remaining 4 sessions are used as training data. In order to test the robustness of compared methods we vary the available amount of initially labeled data in our experiments.

Our results confirm that the feature split used in [11][12] is effective for co-training, as this setting has improved the initial classifier. However, in our experiments, co-training applied with random feature split has outperformed the performance of co-training applied with the defined natural split for all sizes of initial labeled set *L* except for the largest one. The best performing settings in our experiments were *RSSalg* and *MV* which achieved better improvement in accuracy than the competing solutions and displayed more robustness to the size of the initial training set *L* than the competing settings.

The rest of the paper is organized as follows. Work related to this paper is presented in Section 2. Section 3 describes our methodology. Section 4 presents the experiments conducted in this paper and discusses the achieved results. Finally, section 5 concludes the paper and gives directions for future work.

II. Related work

In this section, we present the previous effort of using semi-supervised techniques in order to leverage the problem of labeled data scarcity in gender detection from audio.

In [15] authors present an Expectation-maximization (EM) based algorithm for semi supervised learning in audio classification. In their setting, each audio class is modeled with a Gaussian mixture model, the parameters of which are estimated using EM. They apply their setting to gender and speaker identification tasks and show that adding unlabeled data in such a fashion may reduce the classification error rate by more than half.

Co-training was shown to be beneficial in paralinguistic tasks [11-12][16-18]. However, with the exception of [11], these studies focus on emotion recognition.

In [11] authors apply co-training on several representative tasks of paralinguistic phenomena. They chose to investigate three tasks officially studied in INTERSPEECH challenges from 2009-2011: recognizing emotion, sleepiness, and gender of speakers. Authors have used the standard set of acoustic features from INTERSPEECH 2010 (IS10) challenge [19] which they separated in three partitions suitable for the assumption of independence. The three partitions were rearranged into two views by agglomerating the two smaller groups of features in one view. A similar feature split was used in other co-training studies applied on computational paralinguistics task [12][16].

Authors in [11] have evaluated the task of gender recognition on Agender database, an official corpus of INTERSPEECH 2010 Paralinguistic Challenge (PC) Gender Sub-Challenge [19]. They report 2.1% improvement of unweighted average recall (UAR) when using co-training compared to the performance of the supervised classifier. However, for the gender detection task, co-training was not significantly better than self-training (improvement of 0.1% of UAR). In this paper, we follow the work presented in [11] and further evaluate the applicability of co-training style algorithms on the gender detection task. We note that we focus on the task of identifying gender form emotional speech as opposed to spontaneous speech used in [11].

III. Methodology

A. Acoustic features and the feature split for co-training

We have selected the standard set of acoustic features used in the INTERSPEECH 2010 (IS10) challenge [19]. The feature set consists of 1582 acoustic features which result from a base of 34 low-level descriptors (LLD) with 34 corresponding delta coefficients appended, and 21 functionals applied to each of these 68 LLD contours (1428 features). In addition, 19 functionals are applied to the 4 pitch-based LLD and their four delta coefficient contours (152 features). Finally, the number of pitch onsets (pseudo syllables) and the total duration of the input are appended (2 features). Features were extracted using the openSMILE framework [20].

Authors in [11] and [12] suggest the feature split for co-training which should be suitable for both sufficiency and independence assumption of co-training: the first view comprises of MFCCs (Mel-Frequency Cepstral Coefficients) while the second view comprises of LLDs (Low-Level Descriptors). In this paper, we adopt this feature split and refer to it as „natural“ feature split. The numbers of features in the resulting views are 630 and 952 for the first view and the second view, respectively.

B. Base learner

As base learners in co-training, we employ Support Vector Machines (SVMs) trained with the Sequential Minimal Optimization (SMO) algorithm available in the WEKA toolkit [21]. The classifier setup was: SVMs with linear kernel and a complexity constant of 0.2. This is the same base learner that was employed in [11] and used as a set-up for INTERSPEECH 2010 (IS10) challenge [19] with the exception of the complexity constant which we empirically determined to be 0.2 as it yielded better results for IEMOCAP database. As in [12] we employed a parametric method of logistic regression in order to transform the output distances of SVM into (pseudo) probabilistic values [22].

C. The applied co-training settings

Originally, co-training was designed for the datasets that have the natural separation of the features in two feature subsets, called views. In order to guarantee the successful application of co-training, each view should be sufficient for learning (i.e. given enough labeled data, the features from each view separately should be sufficient to train an accurate model) and conditionally independent of the other view given the class label [7]. Co-training uses the feature split in order to train two different classifiers – both classifiers are trained using the same set of labeled data, but using different views of the data. Classifiers are then applied on unlabeled data and each is allowed to select a certain amount of most confidently labeled

instances to label and add to initial training set. Both classifiers are then retrained on the enlarged training set and the process iteratively continues for the predefined number of iterations.

In this paper we experiment with several different co-training settings:

- ***Natural***: co-training algorithm [7] applied with "natural" feature split,
- ***Random***: co-training applied with random feature split (obtained by randomly splitting features in two equal-sized feature sets)
- ***Majority Vote (MV)***: an ensemble of diverse co-training classifiers is obtained by generating a number of different random feature splits. By using different feature splits, we are able to train different co-training classifiers as co-training is sensitive to the used feature split [23]. *MV* algorithm combines the predictions of the obtained ensemble in a simple majority vote fashion [13].
- ***Random Split Statistic Algorithm (RSSalg)***: In [13] we have developed *RSSalg* in order to boost the performance of co-training and enable its application to single-view datasets. As in *MV*, in *RSSalg*, an ensemble of diverse co-training classifiers is obtained by training multiple co-training classifiers using different random splits of features. The enlarged training set resulting from each co-training process consists of initially labeled examples and examples labeled during co-training. Each enlarged training set is different as different views will cause the resulting classifiers to select different instances as most confident and added instances may be assigned different labels depending on the accuracy of the classifier that classified them. Enlarged training sets obtained by co-training are processed by selection of most reliably labeled instances. An instance is considered reliable if it appears in the majority of resulting training sets and if most of the resulting co-training classifiers agree on its label. The selected reliable instances form the final training set used for learning a final RSSalg model.
- ***RSSalg$_{best}$*** will denote *RSSalg* optimized on the test data [13]. This is only considered as the upper bound of *RSSalg*, i.e. the performance it could achieve if we had additional labeled data for optimization. It is not considered a competing co-training setting but as a test of performance of the threshold optimizing technique used in *RSSalg*.

IV. Experimental results

A. Experimental setup

In order to evaluate our solution we use a five-fold-cross evaluation scheme: in each iteration, a different session is used as test data, while the 4 remaining sessions are used as training data. Each session in IEMOCAP database is recorded by a different pair of mixed-gender actors. This way we ensure that the test data contains segments spoken by both genders and avoid the situation where the model would be trained and tested on segments produced by the same speaker. In this way, in each round of five-fold-cross validation we use in average 3827 instances for training and 957 instances for testing.

In order to test the robustness of the compared methods we vary the size of the initial labeled set *L*. We test the situations where we have 20, 50, 100, 200, 400 and 800 labeled instances available which make around 0.5%, 1.3%, 2.6%, 5.2% and 20.9% of the training set, respectively. All instances that don't belong to the initial labeled set *L* are used as unlabeled data (i.e. the information about the class label is discarded for those instances).

Initial labeled set *L* is randomly chosen from the training data in the following way: in each round of five-fold-cross validation a different session (out of 4 used for training) is chosen to supply the labeled data; an equal number of randomly chosen 'male' and 'female' instances from that session are used as the initial labeled set *L*. Thus, the initial set *L* contains segments recorded by one male and one female actor. This setting makes the recording and labeling of the segments much easier as we employ just one speaker in order to record several segments of speech. All of these segments are automatically labeled with the gender of that speaker, without the need for latter expert annotation. Also, employing only two professional actors is much cheaper than obtaining data from a group of different actors.

As the measure of performance, we use the accuracy measure as IEMOCAP database is relatively balanced according to the gender label (2362 and 2422 segments are spoken by female and male actors, respectively).

The rest of the co-training parameters are fixed to the following values; the number of co-training iterations is 30; the size of the unlabeled pool *U*' is 100; in each iteration of co-training 20 instances most confidently labeled as 'male' and 20 instances most confidently labeled as 'female' are added to the initial training set; the number of different random splits we use in *Random*, *MV*, and *RSSalg* setting is 50. All of these parameters were empirically chosen.

The performance of the supervised classifier obtained using the whole training set (labeled and unlabeled data with correct labels assigned) was evaluated to be 94.12%. This accuracy will be referred to as All_{acc} in the remainder of this paper. The obtained accuracy justifies the features used for the task – it is reported that most gender classifications tested for clean speech demonstrate the accuracy below 95% [24]. We consider that IEMOCAP database contains only clean speech as it was recorded in the controlled environment.

In order to test the redundancy of the feature set as in [25], we test the performance of the supervised classifier trained using the whole training set (labeled and unlabeled data with correct labels assigned) when using different feature sets - "natural" views defined in section III.A, as well as random split of features in two equal-sized partitions. The results are shown in table I.

TABLE I.
THE PERCENT ACCURACY AND STANDARD DEVIATION OBTAINED BY A SUPERVISED CLASSIFIER TRAINED USING DIFFERENT VIEWS OF THE WHOLE TRAINING SET

View 1 (MFCCs)	View 2 (LLDs)	Random halves	All features
90.18 ± 6.59	93.03 ± 3.87	94.11 ± 0.51	94.12 ± 3.26

We can see from the results presented in table I that a significant amount of redundancy exists within the feature set of the constructed dataset - if we randomly select just half of the available features, the resulting performance is still very close to the performance obtained by using all available features. This suggests that co-training with a random split of features can yield an improved classifier performance, as it was shown that this setting is beneficial in the case where there is enough redundancy in the data [26].

Accuracies and standard deviations obtained by applying the compared settings using the described five-fold-cross validation process and different sizes of initial training set *L* are shown in table II. Alongside with the performance of the settings described in sub-section III.C we report the performance of the supervised classifier trained on the initial set of labeled examples *L*, denoted as L_{acc}. The highest accuracies for each setting of *L* are bolded.

TABLE II.
THE PERCENT ACCURACY AND STANDARD DEVIATION OBTAINED BY APPLYING TESTED ALGORITHMS ON DIFFERENT SIZES OF INITIAL TRAINING SET *L*. COLUMNS: NUMBER OF INSTANCES IN THE INITIAL TRAINING SET *L* ('MALE' INSTANCES/ 'FEMALE' INSTANCES). THE HIGHEST ACCURACIES OF THE COMPARED SETTINGS ARE BOLDED ($RSSALG_{BEST}$ IS NOT A COMPETING SETTING AS EXPLAINED IN SECTION III.C).

	10/10	25/25	50/50	100/100	200/200
L_{acc}	73.1±4.6	80.8±7.1	81.3±7.2	83.6±6.6	83.3±6.4
Rand.	74.0±5.0	85.5±3.5	**88.3±5.3**	88.2±2.4	87.2±5.3
Natural	72.3±8.2	83.2±5.2	85.0±5.9	87.1±4.2	88.2±4.6
MV	76.0±7.4	87.7±3.8	87.2±6.6	88.2±5.5	89.3±3.7
RSSalg	**76.8±7.0**	**88.2±3.2**	86.9±7.2	**89.3±4.3**	**89.6±3.5**
RSSalg best	81.0±6.0	89.3±3.3	88.5±6.0	90.2±4.4	90.8±3.1

In order to better visualize the results, we plot the percentage increase of accuracy achieved by each setting. We consider the difference All_{acc}-L_{acc} as 100% of possible increase in accuracy, i.e. we consider a 100% possible increase in accuracy if the algorithm was able to assign the correct label to all available unlabeled instances and add them to the initial training set for training a supervised classifier. The percentage increase of accuracy for each setting is calculated as (*obtained_accuracy*/All_{acc})*100%. The percentage increases in accuracy are plotted in figure 1.

Experimental results presented in table II and figure 1 show that all compared co-training settings succeeded in improving the accuracy of the initial classifier for all chosen sizes of the initial training set *L*, with the exception of *Natural* setting applied in the extreme case where *L* comprises of only 10 labeled instances from each class.

Furthermore, we observe that all settings except for *Random* seem to benefit from adding more examples in the initial training set *L*. However, it should be noted that if the initial classifier is too strong (i.e. the initial training set L is large enough to apply supervised training), co-training is unable to further improve its performance [27].

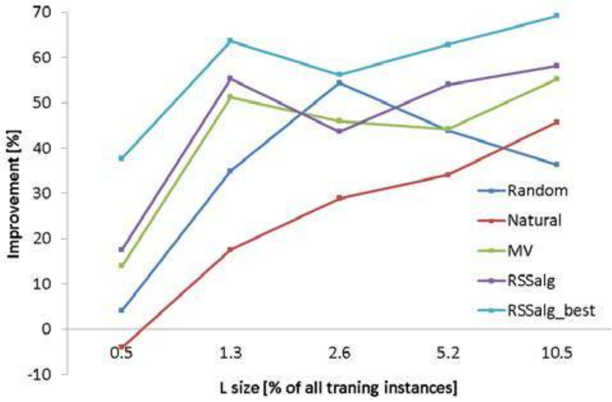


Figure 1. The percentage increase of accuracy for each setting (y-axes, improvement [%]) plotted against different sizes of initial labeled set *L* (x-axes, *L* size).

From table II and figure 1 we observe that *Random* yields with a higher increase in accuracy than *Natural* for all sizes of *L* except for the largest one (10.5% of the whole training set). This result could be attributed to the fact that *Random* feature split produced stronger views in average than MFCC and LLD views used in *Natural* (table I). However, *Random* setting is not optimal as we cannot guarantee the success of co-training with a particular random split of features.

The *Natural* setting was able to improve the initial classifier for all cases except for the extreme case where there are only 0.5% labeled instances in which case it degraded the performance of the initial classifier. From these results, we can further confirm that the split suggested in [11] is effective for co-training. The *Natural* setting yields maximal improvement when applied on largest initial labeled set L.

The best performing settings in our experiment were *RSSalg* and *MV*. In each setting, they succeeded in improving the accuracy of the initial classifier and outperformed the *Natural* setting. They outperformed the *Random* setting for each setting of *L* except for the size of 2.6% of the training data.

We can observe that *RSSalg* and *MV* settings are more robust to the size of the initial training set than *Natural* and *Random.* Although adding more instances does improve the performance of these settings, the increase of performance is relatively small, i.e., increasing the size of L from 50 to 400 increases the performance of RSSalg and MV by 1.4% and 1.6%, respectively, while the same increase of L results with a more significant increase of 5% for the *Natural* setting.

It should be noted that *RSSalg* and *MV* settings require significantly more resources than *Natural* and *Random* setting as they require multiple runs of the co-training algorithm. However, *RSSalg* and *MV* run completely off-line, without any human interaction which makes the time complexity less of a problem. Moreover, these solutions are highly parallelizable as multiple runs of co-training needed in *RSSalg* and *MV* are completely independent.

Finally, we analyze how close the performance of RSSalg is to its upper bound performance $RSSalg_{best}$. We can conclude that $RSSalg_{best}$ does outperform *RSSalg*, however their performances are correlated and relatively close with the exception of the extremely small initial labeled set *L*. This result indicates that the process of

optimizing *RSSalg* parameters without using any labeled data is relatively successful for this case; however there is space for further improvement: given the optimal parameters $RSSalg_{best}$ has significantly outperformed all compared settings. It is also the most robust setting – given only 1.3% labeled instances it was able to achieve the performance of 89.3%. Further adding of data to the initial training set *L* yields with little improvement of $RSSalg_{best}$ (the best accuracy for this setting was 90.8% when 10.5% of the training set is labeled).

In our experiments, we have tried using more than 400 labeled instances as the initial training set *L* (more than 20.9% of the training data). However, this did not result with better performance of supervised algorithm trained using *L* or any of the compared settings. These results are omitted here as they are very similar to the results presented in the 5^{th} column of table II (denoted as 200/200). This is probably due to using the initial labeled data recorded by only one male and one female speaker. It might be helpful to add examples recorded by different speakers to the dataset. However, it should be noted that the results obtained by applying *RSSalg* on the initial data recorded by only one female and one male speaker is reasonably high (89.6% compared to the performance of 94.12% obtained by using all labeled data from different speakers).

V. Conclusion

In this paper we have considered the problem of automatic gender recognition from emotional speech. In order to alleviate the problem of scarcity of labeled data, an ever-present bottleneck in automatic analysis of spoken language, we have investigated the possibility of applying a co-training style algorithm in order to reduce the human effort needed for data labeling. With this goal in mind we have tested the performance of several co-training settings applied to the task of gender recognition from emotional speech. In our experiments we have used co-training with random split of features, natural split of features, a majority vote of several co-training classifiers (*MV*) and Random Split Statistic Algorithm (*RSSalg*) developed with the goal of boosting co-training performance and enabling its application on single-view datasets. We have tested the performance of these settings using IEMOCAP, a gender-annotated emotional speech database.

Our experiments suggested that for this task, a random feature split yields a better performance of co training than the feature split obtained by the separation of MFCC (Mel-Frequency Cepstral Coefficient) features and LLD (Low-Level Descriptor) features which might be considered as the “natural” feature split. We attributed the high performance of Random split to the fact that the significant amount of redundancy exists among the constructed features and to the fact that the random feature split produced stronger views than LLD and MFCC views. The best performing settings in our experiments proved to be MV and RSSalg settings.

We have also tested the robustness of the considered solutions to the size of the initial labeled set L. Both MV and RSSalg proved to be most robust to this aspect.

In the future we plan to conduct more experiments using other emotional speech databases in order to account for speakers of different age groups and different speech quality.

One other branch of experiments would be to track the performance of individual random splits in order to discover the feature split best suitable for this problem.

Also, in the experiments conducted in this paper we have only tested the robustness of the compared solutions to the size of the initial data set L. In the future, we plan to experiment with other co-training parameters such as the growth size and number of iterations of co-training.

Finally, we plan to parallelize RSSalg in order to reduce the problem of computational complexity of this solution.

Acknowledgment

Results presented in this paper are part of the research conducted within the Grant No. III-47003 financed by the Ministry of Education and Science of the Republic of Serbia

References

[1] W.H. Abdulla, N.K. Kasabov, and D.N. Zealand. "Improving speech recognition performance through gender separation." *changes* 9 (2001): 10.

[2] I.M.A. Shahin. "Gender-dependent emotion recognition based on HMMs and SPHMMs." *International Journal of Speech Technology* 16.2 (2013): 133-141.

[3] S.E. Shepstone, Z.H. Tan and S.H. Jensen. "Demographic recommendation by means of group profile elicitation using speaker age and gender recognition." *INTERSPEECH*. 2013.

[4] A. Tsanas and P. Gómez-Vilda. "Novel robust decision support tool assisting early diagnosis of pathological voices using acoustic analysis of sustained vowels." *Multidisciplinary Conference of Users of Voice, Speech and Singing*. 2013.

[5] H. Meinedo, “Audio Pre-Processing and Speech Recognition for Broadcast News,” Ph.D. dissertation, IST, Lisbon, Portugal, 2008.

[6] B. Schuller, “The computational paralinguistics challenge,” Signal Processing Magazine, IEEE, vol. 29, no. 4, pp. 97–101, 2012.

[7] A. Blum and T. Mitchell, “Combining labeled and unlabeled data with co-training,” in Proc. 11th annual conference on Computational Learning Theory, Madison, WI, 1998, pp. 92–100.

[8] M. Kotti and C. Kotropoulos. "Gender classification in two emotional speech databases." *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008.

[9] P. Castellano, S. Slomka89, and P. Barger. Gender gates for telephone-based automatic speaker recognition. Digital Signal Processing, 7(2):65–79, 1997

[10] T. Vogt and E. Andr`e. Improving automatic emotion recognition from speech via gender differentiation. In Proc. Language Resources and Evaluation Conf., page 11231126. Genoa, Italy, May 2006.

[11] Z. Zhang, J. Deng, and B. Schuller. "Co-training succeeds in computational paralinguistics." *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013.

[12] Z. Zhang, E. Coutinho, J. Deng, and B. Schuller. "Cooperative learning and its application to emotion recognition from speech." *Audio, Speech, and Language Processing, IEEE/ACM Transactions on* 23.1 (2015): 115-126.

[13] J. Slivka, A. Kovačević, and Z. Konjović, "Combining co-training with ensemble learning for application on single-view natural language datasets," *Acta Polytechnica Hungarica*, Vol. 10, No 2, pp. 133-152, 2012.

[14] C. Busso, M. Bulut, C.C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, and S.S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," Journal of Language

Resources and Evaluation, vol. 42, no. 4, pp. 335-359, December 2008.

[15] P.J. Moreno and S. Agarwal. "An experimental study of EM-based algorithms for semi-supervised learning in audio classification." *Proc. of ICML-2003 Workshop on Continuum from Labeled to Unlabeled Data*. 2003.

[16] J. Liu, C. Chen, J. Bu, M. You, and J. Tao, “Speech emotion recognition using an enhanced co-training algorithm,” in Proc. IEEE International Conference on Multimedia and Expo (ICME), Beijing, China, 2007, pp. 999–1002.

[17] B. Maeireizo, D. Litman, and R. Hwa, “Co-training for predicting emotions with spoken dialogue data,” in Proc. 42nd Annual Meeting of the Association for Computational Linguistics (ACL), Stroudsburg, PA, 2004, pp. 203–206.

[18] A. Mahdhaoui and M. Chetouani, “Emotional speech classification based on multi view characterization,” in Proc. 20th International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, 2010, pp. 4488–4491.

[19] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. M¨uller, and S. Narayanan, “The INTERSPEECH 2010 Paralinguistic Challenge,” in Proc. INTERSPEECH 2010, Makuhari, Japan, 2010, pp. 2794–2797.

[20] F. Eyben, M. Wöllmer, and B. Schuller, “openSMILE–the munich versatile and fast open-source audio feature extractor,” in *Proc. 18th ACM Int. Conf. Multimedia (MM)*, Florence, Italy, 2010, pp. 1459–1462.

[21] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H.Witten, “The WEKA data mining software: an update,” ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10–18, 2009.

[22] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. ed. New York, NY, USA: Wiley, 2001.

[23] I. Muslea, S. Minton and C. Knoblock: “Active + Semi-Supervised Learning = Robust Multi-View Learning”, *Proc. ICML*, 2002.

[24] Y. Zeng, Z. Wu, T. Falk, and W. Y. Chan. Robust GMM based gender classification using pitch and RASTA-PLP parameters of speech. In Proc. 5th. IEEE Int. Conf. Machine Learning and Cybernetics, pages 3376–3379. Dalian, China, 2006.

[25] C. Jason, I. Koprinska and J. Poon. "Co-training with a single natural feature set applied to email classification". *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*. IEEE Computer Society, 2004.

[26] K. Nigam and R. Ghani: “Understanding the behavior of co-training”, *Proc. Workshop on Text Mining at KDD*, 2000

[27] P. David, and C. Cardie. "Limitations of co-training for natural language learning from large datasets." *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*. 2001.

The Minimal Covering Location Problem with single and multiple location coverage

Darko Drakulić*, Aleksandar Takači**, Miroslav Marić ***
* Faculty of Philosophy, University of East Sarajevo, East Sarajevo, Bosnia and Herzegovina
** Faculty of Technology, University of Novi Sad, Novi Sad, Serbia
*** Faculty of Mathematics, University of Belgrade, Belgrade, Serbia

*ddrakulic@ffuis.edu.ba **atakaci@uns.ac.rs ***maricm@matf.bg.ac.rs

Abstract— **The covering of Location Problems represents very important class of combinatorial problems. One of the most known problem from this class is Maximal Covering Location Problem (MCLP). Its objective is to cover as more as possible location with given facilities. The opposite problem is called Minimal (or Minimum) Covering Location Problem and its aim is to find places for given facilities such as they cover as few as possible location. This paper will present two models of MinCLP derived from the model of classical MCLP.**

I. Introduction

The main objective of covering location problems is in finding optimal positions for facilities with the satisfaction of given conditions. The main parameter of these problems is radius of coverage and it determines an impact of facilities to locations.

One of the most popular covering location problem is Maximal Covering Location Problem (MCLP) and it is introduced by Church and ReVelle [1]. The objective of MCLP is to find positions for the fixed number of facilities in given set of locations so that the total coverage is maximized. This problem is largely applicable in everyday life, such as finding optimal locations of emergency services (ambulance and fire stations), shops, schools etc. MCLP was studied well in past decades and a lot of results have been published. A list of different types of MCLP models follows, but detailed survey from this field could be found in [2]. Moore and ReVelle are introduced a hierarchical service location problem [3], Berman and Krass described gradual covering problem [4] and Qu and Weng described hub allocation maximal covering problem [5]. Capacitated MCLP is defined by Current and Storbeck [6], probabilistic MCLP by ReVelle and Hogan [7] and implicit MCLP is introduces by Murray et al. [8].

On the other hand, Minimal (in literature also known as Minimum) Covering Location Problem is an attempt to minimize the impact of facilities to locations. It is applicable on finding optimal location for undesired facilities, as jails, nuclear and power plants, pollutants etc. This problem has not been much studied in the past, only several papers in this field have been published. Drezner and Wesolowsky [9] presented the Minimum Covering Location problem on the plane with the Euclidean distance between locations. Some researchers studied similar problems, but with different name, like expropriation location problem (ELP) [10]. Exhaustive study of this problem is given by Berman and Huang in [11] and they described five models of Minimum Covering Location Problem with distance constraints (MCLPDC). This paper describes the problem of locating a fixed number of facilities on the network with pre-specified minimal distance between facilities. The objective of MCLPDC is minimization of covered customers.

The aim of this paper is presenting two mathematical models for Minimal Covering Location Problem with single and multiple location coverage. Both models are derived from mathematical model of MCLP presented in [1]. This paper presents theoretical approaches in modelling these problems, without practical methods for finding solutions.

This paper is organized as follows. Mathematical models of MCLP and MCLPDC from [1] and [11] are presented in the section 2. The new approach to modelling Minimal Covering Location Problem are presented in Section 3. Finally, Section 4 briefly summarizes the paper and describes plans for further researches.

II. Mathematical Models of MCLP and MCLPDC

Both problems are defined on given set of locations with pre-defined distances between them. Other constraints in this problems are number of facilities and radius of coverage. The objectives of these problems are locating facilities such as they cover maximum (in MCLP) / minimum (in MCLPDC) locations. Coverage function is determined by radius of coverage.

A. Maximal Covering Location Problem

The original mathematical model of MCLP is described in [1]:

$$\text{maximize} \quad g = \sum_{i \in I} a_i y_i \tag{1}$$

$$\text{subject to} \quad \sum_{j \in N_i} x_j \geq y_i, \forall i \in I \tag{2}$$

$$\sum_{j \in J} x_j = P \tag{3}$$

$$x_j \in \{0,1\}, \forall j \in J \tag{4}$$

$$y_i \in \{0,1\}, \forall i \in I \tag{5}$$

where

- I – set of locations (indexed by i)
- J – set of eligible facility sites (indexed by j)

- S – radius of coverage
- d_{ij} – distance from location *i* to location *j*
- $x_j = \begin{cases} 1, \text{if facility is located at location } j \\ 0, \text{otherwise} \end{cases}$
- a_i – population in node *i*
- P – number of facilities
- $N_i = \{j \in J | d_{ij} \leq S\}$ – set of all facilities *j* which cover location *i*

This model maximizes the sum of covered locations (1) with given conditions (2)-(5). Constraint (2) calculates the coverage of each location y_i - sum on the left side summarizes a number of facilities which cover *i*-th location and determines upper bounds for decision variable y_i. Constraint (3) provides placement of the exact number of facilities and constraints (4) and (5) restrict values of decision variables x_j and y_i.

B. *Minimum Covering Location Problem with distance constraints*

As mentioned before, Berman and Huang [11] described five mathematical models of Minimum Covering Location Problem with distance constraints. We will describe the first model from their research (MCLPDC1) because our model has a similar formulation. Other models in their paper require advanced concepts from graph theory and it is beyond the scope of this paper. Motivation for introduction of distance constraint they described with sensitivity and safety reasons ("if several nuclear reactors are clustered in the same region, they may all be attacked by an aggressor").

The model MCLPDC1 described in [11] is

$$\text{minimize} \quad g = \sum_{i \in I} a_i y_i \tag{6}$$

$$\text{subject to} \quad x_j + x_k \leq 1, \forall j,k \in J \tag{7}$$

$$y_i \geq x_j, i \in I, j \in N_i \tag{8}$$

$$\sum_{j \in J} x_j = P$$

$$x_j \in \{0,1\}, \forall j \in J \tag{9}$$

$$y_i \in \{0,1\}, \forall i \in I \tag{10}$$

where

- S – radius of coverage
- d – minimal distance between facilities
- d_{ij} – distance from location *i* to location *j*
- $x_j = \begin{cases} 1, \text{if facility is located at location } j \\ 0, \text{otherwise} \end{cases}$
- a_i – population in node *i*
- P – number of facilities
- I – set of locations (indexed by *i*)
- J – set of eligible facility sites (sites with minimal distance greater that *d*) (indexed by *j*)
- $N_i = \{j \in J | d_{ij} \leq S\}$ – set of all facilities *j* which cover location *i*

Function g (6) of this problem is the same as in the previous problem (1), but objective of this problem is its minimization. Other constraints determine the fulfillment of all conditions. Constraint (7) state that two locations cannot be located on sites with distance less that *d*. Constraint (8) is similar as constraint (2) and it calculates a coverage of each location y_i. Constraints (8), (9) and (10) are the same as constraints (3), (4) and (5).

III. The New Approach to Modelling Minimal Covering Location Problem

In this paper, Minimal Covering Location Problem will be denote as MinCLP. The main idea of this research is to change model of MCLP with the aim of obtaining the model of MinCLP. This approach, together with introduction of minimal distance constraints, will give modles of two different types of MinCLP. Solutions of both models will be graphic illustrated on the instance with 300 locations, 15 facilities and radius of coverage 4. All locations are generated as points in square with dimensions 30x30 and distances between them are Euclidean distance. Following models will not consider the node population a_i, but this does not reduce the generality of the problem. Population variables are removed order to simplify of the graphic representation of the solution.

Mathematical model of MinCLP obtained from classic MCLP model is follows:

$$\text{minimize} \quad g = \sum_{i \in I} y_i \tag{11}$$

$$\text{subject to} \quad \sum_{j \in N_i} x_j \leq y_i, \forall i \in I \tag{12}$$

$$\sum_{j \in J} x_j = P \tag{13}$$

$$x_j \in \{0,1\}, \forall j \in J \tag{14}$$

$$y_i \in \{0,1\}, \forall i \in I \tag{15}$$

The differences between this model and the MCLP model is in minimization of function g (11) and changing bounds of decision variables y_i (12). Bounds (12) and constraint (15) provide that each location y_i can be covered with at most one facility. It gives an important property of this model – a single coverage for locations. The name of this problem is derived from described properties – Minimal Covering Location Problem with single coverage (MinCLP-SC). Figure 1 illustrates a solution of one instance of MinCLP-SC and it is obvious that each location is covered with at most one facility.

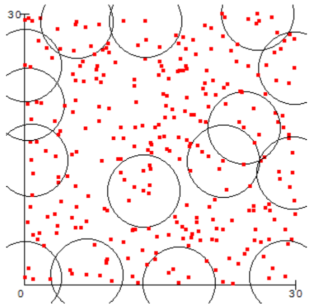


Figure 1. The solution of MinCLP-SC instance of 300 location, 15 facilities and radius of coverage 4

Single coverage is important in modelling problems with this requirement, but the main disadvantage of this model is low solution space. A lot of instances of MinCLP-SC do not have a solution because there is no enough space

for placing all facilities with condition of single coverage. In most real-life problems it is necessary to place all facilities, even if they cover more than one location.

The overcoming of this problem can be resolved if the

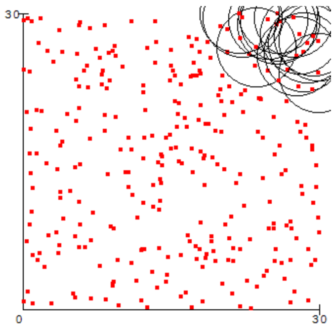


Figure 2. The solution of MinCLP-SC instance of 300 location, 15 facilities and radius of coverage 4

sum from constraint (12) is bounded with value 1. This new constraint allows the covering of each location with more facilities and it gives a Minimal Covering Location Problem with multiple coverage (MinCLP-MC).

Mathematical model of MinCLP-MC is:

$$\text{minimize} \quad g = \sum_{i \in I} y_i \tag{16}$$

$$\text{subject to} \quad \min(1, \sum_{j \in N_i} x_j) \le y_i, \forall i \in I \tag{17}$$

$$\sum_{j \in J} x_j = P \tag{18}$$

$$x_j \in \{0,1\}, \forall j \in J \tag{19}$$

$$y_i \in \{0,1\}, \forall i \in I \tag{20}$$

As mentioned before the difference between MinCLP-SC and MinCLP-MC is in the conditions (11) and (17) and description of this model is not necessary. This type of MinCLP does not have a constraints for a minimal distance between facilities. Figure 2 illustrates a solution of one instance of this type of MinCLP-MC. It is obvious that all facilities are located in near sites and that is not good solution in many real-life problems. Berman and Huang illustrated it on the sensitivity and safety reasons, but this approach is also not good in case of pollutants. Pollution in multiple covered area will be huge and this solution will not be acceptable.

If we add a condition for minimal distance between facilities, we will give a generalization of MinCLP-MC and its mathematical model follows.

$$\text{minimize} \quad g = \sum_{i \in I} y_i \tag{21}$$

$$\text{subject to} \quad \min(1, \sum_{j \in N_i} x_j) \le y_i, \forall i \in I \tag{22}$$

$$\sum_{j \in J} x_j = P \tag{23}$$

$$d_{ij} \ge d, \forall j_1, j_2 \in J, j_1 < j_2 \wedge \tag{24}$$

$$x_{j_1} \cdot x_{j_2} = 1$$

$$x_j \in \{0,1\}, \forall j \in J \tag{25}$$

$$y_i \in \{0,1\}, \forall i \in I \tag{26}$$

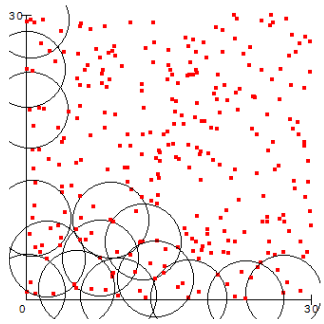


Figure 3. The solution of MinCLP-SC instance of 300 location, 15 facilities and radius of coverage 4 and minimal distance between facilities 4

As mentioned before, this model is generalization of the previous – previous can be obtained with $d = 0$.

The Figure 3 illustrates the solution of the generalized MinCLP-SC.

IV. CONCLUSION

This paper presents a new approach to modelling Minimal Covering Location Problem. This approach is based on conditions of single and multiple location coverage. Two different models are described and graphic representations of their solutions are presented. In the first model, each location is covered with at most one facility and in the second model, location can be covered with more facilities.

In our previous research, we have shown that using fuzzy logic in modelling MCLP significantly improves the quality of models. That will be the main direction for our further work – how to improve MinCLP models with fuzzy conditions.

ACKNOWLEDGMENT

This research was supported the Ministry of Science and Technological Development of Republic of Serbia, project 174009 and the project “Mathematical models of intelligent systems and their applications” of the Academy of Sciences and Arts of Vojvodina supported by Provincial Secretariat for Science and Technological Development of Vojvodina.

REFERENCES

[1] R. Church and C. ReVelle, “The maximal covering location problem”, *Papers of the Regional Science Association* 32, pp 101–118, 1974.

[2] R. Z. Farahani, N. Asgari, N. Heidari, M. Hosseininia and M. Goh: “Covering problems in facility location: A review”, *Computers and Industrial Engineering* 62, pp. 368-407, 2012.

[3] G. Moore and C. ReVelle, “The hierarchical service location problem”, *Management Science* 28, pp. 775-780, 1982.

[4] O. Berman and D. Krass, “The generalized maximal covering location problem”, *Computers and Operations Research* 29 (6), pp. 563-581, 2002

[5] B. Qu and K. Weng, “Path relinking approach for multiple allocation maximal covering problem”, *Computers and Mathematics with Applications*, pp. 1890-1894, 2009

[6] J. R. Current and J. E. Storbeck: "Capacitated covering models", *Environment and Planning B: Planning and Design* 15(2), pp.153-163, 1988.

[7] C. ReVelle and K. Hogan: "The maximum availability location problem", *Transportation Science* 23, 192-200, 1989.

[8] A. T. Murray, D. Tong and K. Kim: "Enhancing classic coverage location models", *International Regional Science Review* 33(2), pp. 115-133, 2010.

[9] Z. Drezner, G. Wesolowsky. "Finding the circle or rectangle containing the minimum weight of points". *Location Science* 2, pp. 83–90, 1994.

[10] O. Berman, Z. Drezner, G. Wesolowsky, "The expropriation location problem", *Journal of Operational Research Society* 54, pp. 769–76, 2003.

[11] O. Berman and R. Huang, "The minimum weighted covering location problem with distance constraints", *Computers and Operations Research* 35, pp. 356 – 372, 2008.

Volume 2

Comparative quality inspection of Moodle and Mooc courses: an action research

Marija Blagojević*, Danijela Milošević*
*Faculty of Technical Sciences Čačak, University of Kragujevac, Serbia
marija.blagojevic@ftn.kg.ac.rs
danijela.milosevic@ftn.kg.ac.rs

Abstract **— This paper shows comparative analysis of Moodle and Mooc courses. In the domain of Mooc courses, the edX platform was used for course creation and analyses. A survey was applied and participants were students from the Faculty of Technical Sciences in Cacak. Results point to possible differences betwen two chosen platforms. Future work relates to continuous improvement of courses and quality assurance of created courses.**

I. INTRODUCTION

The term Open Educational Resources (OER) was first mentioned in 2000 during the UNESCO conference. According to OECD [1], resources are not related only on content but on three different categories:

- Learning content - Relates to the content of learning, learning objects, collections and journals;
- Tools - Software that enables the development and distribution of content as well as searching and organization of the content;
- Implementation resources - Relates to the intellectual property licenses

The term "Open" in OER refers to a number of aspects, according to [2]:

- Openness in open source
- Openness in the social domain
- Openness in the technical domain

Massive open online courses (MOOC) have an unbreakable bond with Open educational resources. The term MOOC (Massive Open Online Course) was first used in 2008 and it was related to an online course "Connectivism and Connected Knowledge", designed by George Siemens and Stephen Downes. According to [3], MOOC integrates the advantages of social networking, a collection of open educational resources and the support of experts in the relevant field. "Massive" refers to the number of the course's participants, as well as the capacities of the course in terms of allowing access to a large number of activities.

E-learning is being standardized by the International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) - ISO / IEC, [4]. Standardization of e-learning is preceded by development and research results from other institutions, such as, for example: AICC, IMS, DCMI, ADL-SCORM, ALIC, IEEE LTSC, ADRIADNE, CEN/ISSS WS-LT, CEN/ISSS CDFS, CEN/ISSS WS on Privacy, W3C etc ... [5]. JTC 1 / SC 36 subcommittee for e-learning functions within the First unified technical committee (JTC1 ISO / IEC).

Along with development of electronic courses, a question arises about their quality. According to [6] “quality issues often manifest as dissemination on teaching effectiveness, faculty to student rations, attrition rates, student satisfaction, and institutional resources invested in online delivery”.

There are several approaches to evaluation of quality on electronic courses. Khan listed some possible approaches in [7] that are shown in Table I.

TABLE I.
KHAN'S QUALITY FACTOR

Quality factor	Concern
Pedagogical	Issues related to teaching and learning such as course contents, how to design it, how to offer it to target audience and how the learning outcomes will be achieved.
Technological	Issues related to hardware, software and infrastructure. e-learning environment, LMS, server capacity, bandwidth, security and backup are also covered in this dimension.
Interface design	The overall look and feel of an e/learning program. Interface design encompasses Web and content design, navigation, Web accessibility and usability testing.
Evaluation	The evaluation of e-learning at institutional level, evaluation learning assessment.
Management	The maintenance and modification of the learning environment, it also addresses issues related to quality control, staffing and scheduling.
Resource support	All technical and human resources support to create meaningful online environment which includes Web based digital libraries, journals, and online tutorials.
Ethical	Issues related to social and political influence, diversity and legal issues such as plagiarism and copy rights.
Institutional	Including three sub dimensions: issues of administrative affairs; issues of academic affairs; issues of student services.

However, according to [8] a consensus was not achieved when it comes to standards and methods of evaluation that would relate to electronic materials. Namely, it is necessary to determine different methods and criteria of evaluation for traditional and online learning materials.

By reviewing related research in the field of evaluation the quality of electronic course presented in [9-12] it was determined that most often a combination of various factors of quality was used in order to get relevant information.

II. Methodology

A. *Purpose, tasks and goals*

The purpose of the research is comparative analysis of quality of Moodle and Mooc courses. Research is of action type because the goals were to promote created courses and reach socially useful results in the domain of electronic learning.

The goal of the research relates to determining potential differences in the quality of courses on Moodle and Mooc platform that arise from different possibilities that these platforms offer. Regarding Mooc platform, in this research is analyzed edX platform (https://www.edx.org).

Research methodology:

- Creating the courses on Mooc and Moodle platform
- Enrolling beginners to appropriate Moodle and Mooc courses
- Creating a survey
- Applying the survey
- Analysis and evaluation of results

B. *Research framework*

Courses that were evaluated are created within the Moodle LMs at the Faculty of technical sciences in Cacak (itlab.ftn.kg.ac.rs/moodle) and edX courses, created within Tempus Baektel project (www.baektel.eu).

Within Moodle LMS over 100 courses was created from various fields within basic and master studies. There are around 2000 active users in the system.

Apart from Moodle course, courses developed within edX system were also evaluated. The edX platform, together with developed courses were established within Tempus Baektel project (www.baektel.eu). The main project objective is to "establish an OER framework for fostering technology enhanced learning (TEL) within HE institutions and life-long learning within enterprises in WB countries, develop and implement guidelines and procedures for quality assurance of OER according to EU practices at national level in WB beneficiary countries" [www.baektel.eu].

In the initial phase of the project, two courses on edX platform were created: introduction to programming and practical course on static electricity at gas stations. The comparative review on both systems edX and Moodle enviroment are described in detail in [13].

C. *Participants and survey*

Students from Faculty of Technical sciences in Cacak took part in the research. Selected students attended at least two courses at Moodle LMS and both courses on edX system. Table II shows data about the participants.

The easiest and fastest way to assess quality is self-evaluation that can be done by passing through a simple and concise evaluation. Survey is given in table III [14]. Survey contains the following categories:

- Course overview and Introduction
- Learning objectives (Competencies)
- Assessment and measurement
- Instructional materials
- Learner interaction and engagement
- Course technology
- Learner support
- Accessibility

TABLE II.
INFORMATION ABOUT PARTICIPANTS

Gender	Frequency	Percentage	Valid percentage
Male	15	68,18	68,18
Female	7	31,82	31,82
Total	22	100,00	100,00

TABLE III.
INFORMATION ABOUT PARTICIPANTS

		Moodle courses grade	edX courses grade
1	**Course overview and Introduction**		
1.1	Instructions make it clear how to get started and where to find various course components.		
1.2	Students are introduced to the purpose and structure of the course.		
1.3	Etiquette expectations (sometimes called "netiquette") for online discussions, email, and other forms of communication are stated clearly.		
1.4	Course and/or institutional policies with which the student is expected to comply are clearly stated, or a link to current policies is provided.		
1.5	Prerequisite knowledge in the discipline and/or any required competencies are clearly stated.		
1.6	Minimum technical skills expected of the student are clearly stated.		
1.7	The self-introduction by the instructor is appropriate and available online.		
1.8	Students are asked to introduce themselves to the class.		
2	**Learning objectives (Competencies)**		
2.1	The course learning objectives describe outcomes that are measurable.		
2.2	The module/unit learning objectives describe outcomes that are measurable and consistent with the course -level objectives.		
2.3	All learning objectives are stated clearly and written from the students' perspective.		
2.4	Instructions to students on how to meet the learning objectives are adequate and stated clearly.		
2.5	The learning objectives are appropriately designed for the level of the course.		
3	**Assessment and measurement**		
3.1	The types of assessments selected measure the stated learning objectives and are consistent with course activities and resources.		
3.2	The course grading policy is stated clearly.		
3.3	Specific and descriptive criteria are provided for the evaluation of students' work and participation and are tied to the course grading policy.		
3.4	Students have multiple opportunities to measure their own learning progress.		
4	**Instructional materials**		
4.1	The instructional materials contribute to the achievement of the stated course and module/unit learning objectives.		
4.2	The purpose of instructional materials and how the materials are to be used for learning activities are clearly explained.		
4.3	All resources and materials used in the course are appropriately cited.		
4.4	The instructional materials present a variety of perspectives on the course content.		
4.5	The distinction between required and optional materials is clearly explained.		
5	**Learner interaction and engagement**		
5.1	The learning activities promote the achievement of the stated learning objectives.		
5.2	Learning activities provide opportunities for interaction that support active learning.		
5.3	The instructor's plan for classroom response time and feedback on assignments is clearly stated.		
5.4	The requirements for student interaction are clearly articulated		
6	**Course technology**		
6.1	The tools and media support the course learning objectives.		

6.2	Course tools and media support student engagement and guide the student to become an active learner.		
6.3	Navigation throughout the online components of the course is logical, consistent, and efficient.		
6.4	Students can readily access the technologies required in the course.		
6.5	The course technologies are current.		
7	**Learner support**		
7.1	The course instructions articulate or link to a clear description of the technical support offered and how to access it.		
7.2	Course instructions articulate or link to the institution's accessibility policies and services.		
7.3	Course instructions articulate or link to an explanation of how the institution's academic support services and resources can help students succeed in the course and how students can access the services.		
7.4	Course instructions articulate or link to an explanation of how the institution's academic support services can help students succeed and how students can access the services.		
8	**Accessibility**		
8.1	The course employs accessible technologies and provides guidance on how to obtain accommodation.		
8.2	The course contains equivalent alternatives to auditory and visual content.		
8.3	The course design facilitates readability and minimizes distractions.		
8.4	The course design accommodates the use of assistive technologies.		

III. Results and discussion

During result analysis, a T-test of paired samples was used (or repeated measurements).

Research question was posted as following: Are the results of user satisfaction with courses created within Moodle LMS and edX considerably different?

The following was applied:

- One category independent variable (in this case that was the type of the system: 1-Moodle, 2-edX)
- One uninterrupted, dependable variable (platform rating)

T-test of paired samples shows us whether there is a statistically meaningful difference in mean values received for two selected systems. Received results are shown in Table IV.

In the above mentioned table, starting from the second to the last column named Sig., it can be noted that there isn’t any couple with Sig values less than 0.05, which points to the conclusion that there is no meaningful difference between two tested systems regarding the categories.

TABLE IV
PAIRED SAMPLE TEST

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	1 - 1	,600	1,647	,521	-,578	1,778	1,152	9	,279
Pair 2	2 - 2	,900	1,595	,504	-,241	2,041	1,784	9	,108
Pair 3	3 - 3	,300	1,059	,335	-,458	1,058	,896	9	,394
Pair 4	4 - 4	1,100	1,853	,586	-,226	2,426	1,877	9	,093
Pair 5	5 - 5	,800	1,814	,573	-,497	2,097	1,395	9	,196
Pair 6	6 - 6	1,000	2,160	,683	-,545	2,545	1,464	9	,177
Pair 7	7 - 7	,900	1,853	,586	-,426	2,226	1,536	9	,159
Pair 8	8 - 8	,900	1,101	,348	,113	1,687	2,586	9	,029
Pair 9	9 - 9	1,000	1,633	,516	-,168	2,168	1,936	9	,085
Pair 10	10 - 10	,600	1,430	,452	-,423	1,623	1,327	9	,217
Pair 11	11 - 11	,200	1,317	,416	-,742	1,142	,480	9	,642
Pair 12	12 - 12	-,200	,919	,291	-,857	,457	-,688	9	,509
Pair 13	13 - 13	1,100	1,595	,504	-,041	2,241	2,181	9	,057
Pair 14	14 - 14	,800	1,619	,512	-,358	1,958	1,562	9	,153
Pair 15	15 - 15	,100	1,595	,504	-1,041	1,241	,198	9	,847
Pair 16	16 - 16	,400	1,713	,542	-,825	1,625	,739	9	,479
Pair 17	17 - 17	,300	2,003	,633	-1,133	1,733	,474	9	,647
Pair 18	18 - 18	,600	1,647	,521	-,578	1,778	1,152	9	,279
Pair 19	19 - 19	-,400	1,578	,499	-1,529	,729	-,802	9	,443
Pair 20	20 - 20	,000	1,700	,537	-1,216	1,216	,000	9	1,000
Pair 21	21 - 21	,500	1,716	,543	-,728	1,728	,921	9	,381
Pair 22	22 - 22	,500	1,958	,619	-,901	1,901	,808	9	,440
Pair 23	23 - 23	-,200	1,398	,442	-1,200	,800	-,452	9	,662
Pair 24	24 - 24	,200	1,619	,512	-,958	1,358	,391	9	,705
Pair 25	25 - 25	-,300	1,889	,597	-1,651	1,051	-,502	9	,627
Pair 26	26 - 26	,200	1,317	,416	-,742	1,142	,480	9	,642
Pair 27	27 - 27	,400	1,350	,427	-,566	1,366	,937	9	,373
Pair 28	28 - 28	,300	2,003	,633	-1,133	1,733	,474	9	,647
Pair 29	29 - 29	,700	2,312	,731	-,954	2,354	,958	9	,363
Pair 30	30 - 30	,500	1,780	,563	-,773	1,773	,889	9	,397
Pair 31	31 - 31	,600	1,897	,600	-,757	1,957	1,000	9	,343
Pair 32	32 - 32	,300	1,947	,616	-1,092	1,692	,487	9	,638
Pair 33	33 - 33	-,100	1,524	,482	-1,190	,990	-,208	9	,840
Pair 34	34 - 34	,300	1,494	,473	-,769	1,369	,635	9	,541
Pair 35	35 - 35	,100	2,132	,674	-1,425	1,625	,148	9	,885
Pair 36	36 - 36	,700	1,494	,473	-,369	1,769	1,481	9	,173
Pair 37	37 - 37	-,400	1,506	,476	-1,477	,677	-,840	9	,423
Pair 38	38 - 38	-,700	1,767	,559	-1,964	,564	-1,253	9	,242
Pair 39	39 - 39	-,900	,994	,314	-1,611	-,189	-2,862	9	,019

IV. CONCLUSION

According to the obtained results two conclusions can be made:

- The current state in regards to differences between two analyzed systems: based on the conducted T-test, it was determined that there are no meaningful differences.
- Further research directions: having in mind that the research is active and the work in both systems is still conducted, by improving and developing new courses and functionalities within those courses goals will be achieved.

Limitation of the research is apparent in inadequate number of test subjects (22).

Future work relates to the development of courses and continued surveying of both systems, analysis of comparative results and further improvements.

REFERENCES

[1] OECD (2007), Giving Knowledge for Free: the Emergence of Open Educational Resources, http://tinyurl.com/62hjx6

[2] I. Tuomi, "Open Educational Resources: What they are and why do they matter", retrieved from: http://www.meaningprocessing.com/personalPages/tuomi/articles/OpenEducationalResources_OECDreport.pdf, last access: August 27th 2015.

[3] A. McAuley, B. Stewart, G. Siemens G. and D. Cormie., "Massive open online courses: Digital ways of knowing and learning", retrieved from: http://www.elearnspace.org/Articles/MOOC_Final.pdf, Last access: November 23 rd 2015.

[4] ISO/IEC (2015). International Standards for Business, Government and Society, 35: Information technology, List of ICS fields, URL: http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_ics_browse.htm?ICS1=35, Last update 15. 01. 2012.

[5] Micić, Ž. Modeling of education/e- process based on standardization and PDCA platform, retrieved from: http://www.ftn.kg.ac.rs/konferencije/rppo13/Zbornik%20RPPO13.pdf, last access August 31st 2015.

[6] J. Bourne and J.Moore, "Elements of quality online ducation", The Sloan consortium 2004.

[7] B. Khan, "A framework for web-based learning" pp. 75-98, 2001, Educational Technology, Englewood Cliffs, NJ

[8] M. Alkhattabi, D. Neagu, A. Cullen, "Accessing information quality of e-learning system: a web mining approach, Computers un human behavior, retrieved from: http://scim.brad.ac.uk/staff/pdf/ajcullen/DOI_version.pdf, Last access December 10th 2015.

[9] N.E.koch, "E-learning: Didactical recommendations and quality assurance", retrieved from: https://www.uni-hohenheim.de/fileadmin/einrichtungen/ells/QA-documents/ELLS-eLearning-Didactical-Recommendation-final-version.pdf, Last access December 10th 2015.

[10] P. Kawachi, "Quality assurance for OER: Current state of the art and the TIPS framework", retrieved from: http://www.openeducationeuropa.eu/en/article/Assessment-certification-and-quality-assurance-in-open-learning_In-Depth_40_1, Last access December 10th 2015.

[11] W.Zhang and Y.L.Cheng, "Quality assurance in E-learning:PDPP evaluation model and its application", retrieved from: http://files.eric.ed.gov/fulltext/EJ1001012.pdf, Last access December 10th 2015.

[12] P. Russo, T. Heenatigela, E.Gomez, l. Strubble, "Peer-review platform for astronomy education activities", retrieved from: http://www.openeducationeuropa.eu/en/article/Assessment-certification-and-quality-assurance-in-open-learning_From-field_40_2, Last access December 10th 2015.

[13] Marija Blagojević, Danijela Milošević, "Massive open online courses: EdX vs Moodle MOOC", 5th international Conference on Information Society and technology, Kopaonik, 8-11 March 2015. pp. 346-351. (ISBN: 978-86-85525-16-2)

[14] QM criteria, retrieved from: www.qmprogram.org, last access December 14, 2015.

Exploring the influence of ubiquitous workplaces on individual learning

Sonia JEDDI [*],
[*]Doctor, University Tunis El Manar, Tunis, Tunisia
jeddisonia@yahoo.fr

Samia KAROUI ZOUAOUI[**]
[**]Professor, University Tunis El Manar, Tunis, Tunisia
samiazouaoui.fsegt@gmail.com

***Abstract*— Nowadays, the professional use of ubiquitous computing devices for data transfer and communication is becoming increasingly important. Workers become constantly connected and available. Moreover, they are constantly asked to be accessible. Falling in line with the abductive approach, this study has developed a model to investigate mobile and ubiquitous work consequences on user's learning. The theoretical framework for understanding learning is based on connectivism theory, as a learning theory for the digital age. An adapted survey was administered to 214 managers operating in IT and Telecoms sectors. Results show that social interactions generate an overload by providing redundant and useless information. Consequently, and contrary to what was said in previous researches, social interactions are an ongoing problem that can affect learning by enhancing the information load. This empirical evidence makes a valuable contribution for both managers and organizations. Future researches must include other workers categories and offer possible countermeasures.**

Key words: Mobile work, information overload, social interactions, individual learning

I. Introduction

Technological advancement and the promised rewards of mobile working have led to an explosion in mobile computing and telecommunications technologies in recent years [1]. Thanks to their unique characteristics, mobile devices facilitate value creation anywhere anytime. Although new forms of flexible works have emerged, we will interest on mobile work_as one of teleworking form_which is increasingly observed and practiced in business. It consists on activities leading mobile operator to perform his job, in whole or in part, outside the company, using mobile devices. While benefits are always expected [2], [3], we consider that it is interesting to enrich debates conducted around mobile work through studying its effect on individual learning.

Mobile work is not only a remote work mode, more important, it also generates mobilization of human interaction in the workplace in terms of spatiality, temporality and contextuality [4]. Learning is also a contextual activity since we can learn anywhere anytime [5]. The challenge for users is, therefore, to understand how best they might use mobile devices to support their learning. Whereas the literature is still unclear about mobile workers learning, this study will try to explore capabilities of mobile work to enhance or inhibit learning.

Our review of the literature reveals three major theories that provide an effect view of learning in many environments: behaviourism, cognitivism, and constructivism. Those theories fall short, however, when learning is considering as rich, informal, networked experience, whether in office, home, travelling or wandering. In this context, a blended approach to enabling learning with mobile technologies was necessary as successful and engaging activities draw on a number of different theories and practices [5].

With the changes that have occurred on learning possibilities, George Siemens has introduced the theory of learning called “connectivism” [6]. In this connectivist model, a learning community is described as a *node*, which is always part of a larger network. A network is comprised of two or more nodes linked in order to share resources. Nodes may have varied size and strength, depending on the concentration of information and the number of individuals who are navigating through the network [7]. In this way, "Know-how" and "know-what" are completed by "know-where" which refers to the question "where can we find the necessary information?" [8].

Two important elements are indispensable in understanding learning models in a digital era: Social network and information flow. In fact, new information is constantly being acquired through connection and interaction between nodes. Learning is thus becoming a process of connecting specialized nodes or information sources [6]. More important for this theory, that the ability to draw distinctions between important and unimportant information is vital. Consequently, connectivism stresses that two important skills that contribute to learning are the ability to seek out current information, and the ability to filter secondary and extraneous information [9].

While this conception is often associated with and proposes a perspective similar to Vygotsky's 'zone of proximal development' (ZPD) and Engeström's Activity theory (2001), we consider it as the most appropriate to approach the learning of mobile workers. In fact, as a result of the recent rapid advances made in information and communication technology, mobile work is consequently associated with an increasing amount of

information [10] and with development of a large networks [1], [11].

In this context, it will be useful to enrich the current debates conducted around mobile work through the study of its consequences at the individual level. This paper investigates how mobile work could influence mobile worker's learning. The specific objective of this study is to verify the significant correlation, if any, exists among learning and its determining factors identified through our exploratory study. Data were collected through observations and semi-structured face-to-face interviews with Tunisian managers operating in IT and Telecoms sectors. This inductive phase helped us to understand human thoughts and actions in organizational context [12].

A total of 17 people participated in our exploratory study. Posts targeted were very active and dynamic requiring movement and a frequent use of mobile communication technologies. All those interviews were transcripted using qualitative coding. Two emerging codes were identified: information overload and social interactions are an ongoing problem that can affect mobile worker's learning.

This finding joins basic idea of cognitive load theory of Sweller (1988) [13]. According to this theory, cognitive capacity in working memory is limited, so that if a learning task requires too much capacity, learning will be hampered [14]. Nevertheless, cognitive load will not be considered in this research. We will limit our investigation to both concepts identified by our inductive research.

II. Conceptual Framework and Research Hypotheses

The proposed model is a predictive model of mobile's work consequences generated by our qualitative exploratory study "Fig. 1". In following sections, we will discuss hypothesis

A. Mobile worker's social interaction:

In this study, we focus on the structural dimension of social capital and more specifically social interactions of mobile workers. Social interactions promote access to different resources [15]. According to our interviews, the most important resource is informational. Information can be received anywhere anytime. Previous researches have highlighted that only information benefits have interested researchers [15], [16]. However, the information exchanged can also represent a risk factor. The 17 interviewers have considered social interactions as source of information overload. This consequence is a result of two phenomena as identified by our exploratory research: Too much solicitation and especially redundant and useless information received by different mobile devices.

Therefore deducing from the foregoing discussion, it is hypothesized that:

H1: *mobile worker's social interaction have a positive impact on information overload*

According to Vygotsky (1978), learning is a process of internalization. It is generated by interaction with the other in different contexts [17]. According to Chiu et al. (2006), social interaction between members of a virtual community is an effective way of sharing knowledge and learning [18]. The more these interactions exist, the more important the intensity, frequency, and the exchange of knowledge and skills [19]. We then propose that:

H2 : *mobile worker's social interaction have a positive impact on individual learning.*

B. Mobile worker's information overload:

Information overload has been exacerbated by the recent rapid advances made in information and communication technology [17]. The use of mobile technology is associated with an increasing amount of information at the user's disposal [20]. This phenomenon can be defined as a situation in which the amount of information an individual receives exceeds an individual's information processing capacity [21].

Our empirical investigation has showed that learning may be disturbed by an information overload. While we could not be able to identify a direct link between these two concepts, a moderation relationship was identified. For this study, the positive relationship demonstrated between social interaction and learning (*see H1*) can, beyond a certain threshold, impede learning. This threshold will depend on the level of the informational load, which is also caused by the social interaction. Indeed, as the information overlap with a phone that keeps ringing, emails to read and transmit, meetings and face to face conversations, interactions become threatening and also sources of redundancy. It will disturb operator promoting additional informational load. The operator will in this case adopt a partial review of his tasks. Therefore this study proposes that:

H3: *mobile worker's information overload moderates relation between social interactions and individual learning.*

Figure 1 presents the research model. The dependent variable is "individual learning" [IL] that can be influenced positively and directly by social interactions [SI], negatively and indirectly by information overload. The moderator variable information overload [IO] will alter the strength of the causal relationship. Although classically, moderation implies a weakening of a causal effect, a moderator can amplify or even reverse that effect [22]. Barron and Kenny (1986) assumed that the moderation can be captured by an XZ product term. For our case, moderation will be tested as IOxSI product term.

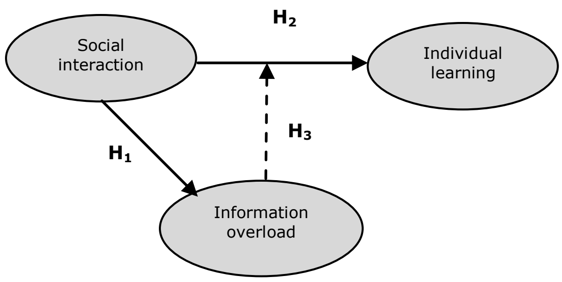


Fig1. Research model

III. Methodology

A random sample of 305 mobile middle or senior managers operating in IT and Telecoms sectors were selected. Survey questionnaires were mailed to them. A total of 214 responses were retained. All the respondents are males. The median age group of the respondent was that of less than 38 years (65.9%). About 81.4% of the respondents have university level of education. Consequently, results analysis will not be discussed according to gender, age and level of education as they can raise any interest in our case.

A. Measures:

All the scale items were measured on a five-point Likert-type scales that was anchored by 1= strongly disagree to 5= strongly agree to express the degree of agreement. The three constructs were modelled using reflective indicators. Scales are resumed in table I.

TABLE I.

SCALE MEASUREMENT

Constructs	Items	Sources
Information overload	12	Scale construction
Social interaction	4	Chiu and al. (2006).
Individual learning	11	Haueter and (2003).

Research scales were operationalized on the basis of previous work and also newly designed. Some adaptations were essential in order to fit the current research context and purpose. “Social interactions” used a four-item scale measure of Chiu and al. (2006), “Individual learning” used eleven-item scale measure from Haueter and al. (2003) [23]. “Information overload” was measured by twelve items developed for this study.

New measure of IO was created by using a psychometric approach based on Churchill’s “scale development” procedure and on the C-OARSE method of Rossiter. While Churchill defines the construct in terms of the attribute only [24], C-OAR-SE construct definition requires specification of (1) the construct, (2) the object and (3) the attribute.

Three dimensions were retained. Informational one will be measured by five items, communicational one by three items and temporal dimension by four items.

B. Data analysis and measurement model:

To analyse both measurement and structural models, we used Smart PLS software for Structural Equation Modeling (SEM) technique [25] which can support at the same time exploratory and confirmatory research. It will be the appropriate software to this study since our model is a prediction of an impact not yet approved by researchers. It is also recommended for small sample size.

Construct reliability was assessed using Composite Reliabilities and Cronbach’s Alpha values. All these values were above 0.68 as recommended by Hulland [26]. We can then admit that the scales are reliable. Convergent validity was assessed using the average variance extracted (AVE) measure and Item loading values. According to Fornell and Lacker [27], Variance Extracted (AVE) by each construct should excee d the variance due to measurement error for that construct, i.e., AVE should be greater than 0.5. For our case, all the items loadings and AVE values are satisfactory and confirm the existence of convergent validity (see table II).

TABLE II.
CONSTRUCT RELIABILITY AND CONVERGENT VALIDITY

Research construct	Cronbach’s alpha And CR values	AVE value
SI	α= 0.825 CR=0.919	0,765
IO3*	-	-
IO2	α= 0.437 CR=0.780	0,616
IO1	α= 0.772 CR=0.869	0,683
ISxIO	α= 0.898 CR=0.952	0,906
IL	α= 0.969 CR=0.973	0,734

*mono-item construct after purification process

Furthermore, to evaluate discriminant validity, the AVE of each construct should be greater than the shared variance between the construct and the other model constructs [28]. In Table IV we can remark that the diagonal elements are greater than the off-diagonal elements in the corresponding rows and columns, therefore confirming that discriminant validity is verified (table III.).

TABLE III.

DISCRIMINANT VALIDITY

	SI	IO3	IO2	IO1		IL	AVE
SI	1						**0,765**
IO3	0,020	1					
IO2	0,001	0,074	1				**0,616**
IO1	0,024	0,242	0,400	1			**0,683**
ISxIO	0,001	0,004	0,042	0,057	1		**0,906**
IL	0,009	0,004	0,002	0,001	0,008	1	**0,734**

C. Structural Modeling Results:

The structural model was tested using the R^2. It is the amount of variance explained by independent variables. The R^2 values for the dependent variables – information overload (IO) and individual learning (IL) are respectively 0.213 and 0.17. According to our results, Information overload explains 21.3% of social interactions. Moderating variable explains 35.6% of individual learning while social interactions explain only 17 % of individual learning.

Based on R^2, the global goodness-of-fit (GoF) will be calculated. This value will assess the quality of the measurement and the structural models [29]. Our GoF is 0.721 which can be satisfactory (between 0.3 and 0.9). Thus, this study concludes that the research model provides an overall goodness of fit.

IV. Hypothesis tests and discussion of results

Table IV. summarizes the hypothesis tests.

The purpose of this study was to investigate and verify the significant correlation between individual learning and two of its determining factors associated to mobile work. Hypothesis test showed that there is a significant and positive relationship between social interactions and information overload of mobile workers (t= 0.682, β = 7.532). A positive association exists between SI and IL. A significant and negative relationship was also founded between SIxIO and IL.

TABLE IV.
HYPOTHESIS TEST

hypothesis	Path coefficient	Statistics	Results
H1: SI → IO	0.682	7.532	Supported
H2: SI → IL	0.359	5.621	Supported
H3: IOxSI →IL	0.532	3.394	Supported

The findings have highlighted that social interactions influence positively the information overload of managers. While previous researchers have assumed that social interactions, as a structural dimension of social capital, generate benefits [15] [16], our results assume that social interactions is risk factor that can increase amount of information. Technological proximity is becoming a source of interruption and redundancy information.

Otherwise, in accordance with previous research, individual learning happens through social interaction. Based on our findings, information overload can moderate this positive effect. Our empirical investigation has showed that the more information overlap with mobile devices, the more interactions become threatening and sources of redundancy. Social interactions can thus disturb the operator by promoting additional informational load. Learning will consequently be disturbed.

V. Conclusion and Future research

Ubiquitous computing, through the social interaction that they mobilize, has been shown to be beneficial for learning. For this research, social interactions are source of overload and can disturb individual learning. An interesting area for future research is to examine other consequences of social interaction that can disturb learning (task interruption as an example). It will be then necessary for further research to verify the generalization of our findings in other sectors.

References

[1] Green, N. On the move: technology, mobility, and the mediation of social time and space. *Information Society*, vol. **18** (4), 2002, pp. 281-292.

[2] Chen, L. and Nath, R. A socio-technical perspective of mobile work, Information Knowledge Systems Management, vol. 7, 2008, pp.41-60.

[3] Allen, D.K, Shoard, M. Spreading the load: mobile information and communications technologies and their effect on information overload. *Information Research*,Vol. 10 (2). 2005. Available on http://informationr.net/ir/10-2/paper227.html.

[4] Kakihara, M., Sørensen, C., Wiberg, M. Fluid interaction in mobile work practices.Tokyo Mobile Roundtable (May) Tokyo, Japan. 2002.

Naismith ;;;;;; siemns 2004

Siemens, G. (2004). A learning theory for the digital age. Retrieved from http://www.elearnspace.org/articles/connectivism.htm

[5] Ruff, J. Information Overload: Causes, Symptoms and Solutions. Harvard Graduate School of Education's Learning Innovations Laboratory (LILA). Information Overload: Causes, Symptoms and Solutions.pdf [*PDF now hosted on WorkplacePsychology.Net for convenience.], 2002.

[6] Siemens, George: Connectivism: A Learning Theory for the Digital Age. 2005. http://www.elearnspace.org/Articles/connectivism.html

[7] Downes S. New technology supporting informal learning. Journal of Emerging Technologies in Web Intelligence, vol. 2 (1), 2010, pp.27–33.

[8] Zanden V. D. The facilitating university. Uitgeverij Eburon. 2009.

[9] Kop, R. and Hill, A. Connectivism: Learning Theory of the Future or Vestige of the Past?in *The International Review of Research in Open and Distance Learning*. 2008. Vol 9, No 3

[10] Eppler, M.J., Mengis, J., A Framework for Information Overload Research in Organizations. *The Information Society*, 2004.informaworld.com.

[11] Ljungberg F., Sørensen C. Overload: From transaction to interaction. *Planet Internet K. Braa, C. Sørensen, and B. Dahlbom(ed)*. 2000. pp:113-136

[12] Klein, H.K. Myers, M.D. A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, vol. **23** (1), 1999, pp.67-69.

[13] Sweller, J. Cognitive load during problem solving: Effects on learning. *Cognitive Sciences.* Vol. 12, 1988, pp. 257-285.

[14] DE Jong, T., Van Joolingen, W., Giemza, A., Girault, I., HOPPE, U., Kindermann, J. Learning by creating and exchanging objects: The SCY experience. *British Journal of Educational Technology*, vol. 41 (6), 2010, pp. 909-921.

[15] Burt, R. S. The contingent value of social capital. Administrative Science Quarterly, vol. 42, 1997, pp. 339-365.

[16] Adler, P. and Kwon, S. Social capital: Prospects for a new concept. Academy of Management Review vol. 27(1), 2002, pp.17-40.

[17] Koschmann, T. Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration. In C. Hoadley & J. Roschelle (Eds.), Computer Support for Collaborative Learning (pp. 308– 313). Mahwah, NJ: Lawrence Erlbaum Associates. 1999.

[18] Chiu, C.-M., M.-H. Hsu, E. T. G. Wang. Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. Decision Support Systems vol. 42(3), 2006, pp. 1872–1888

[19] Edmunds, A., Morris, A., The problem of information overload in business organizations: A review on the literature. *International Journal of Information Management*, vol 20 (1), 2000, pp17-28.

[20] Lang, J.C., Managing in knowledge-based competition. *Journal of Organizational Change Management*, vol. **14**(6), 2001, pp. 539-553.

[21] Jackson T. W., Hooff B. V. D. Understanding the Factors that Effect Information Overload and Miscommunication within the Workplace. Journal of Emerging Trends in Computing and Information Sciences Vol. 3(8), 2012.

[22] Baron R.M. et Kenny D.A. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, Vol. 51 (6), 1986, pp. 1173-1182.

[23] Haueter, J. A., Macan, T. H., Winter, J. Measurement of newcomer socialization: construct validation of a multidimensional scale. *Journal of Vocational Behavior*, vol. 63. 2003, pp. 20–39.

[24] Rossiter, J. R. The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, vol. 19(4), 2002, pp. 305–335.

[25] Ringle, C. M., Wende, S., Will, A. SmartPLS 2.0 M3. 2005, Available on http:// www.smartpls.de

[26] Hulland, J. Use of partial least squares (PLS) in strategic management research: A review of four recent studies, *Strategic Management Journal*, vol. 20, 1999, pp.195-203.

[27] Fornell, C., Lacker, D. F. Evaluating structural equations with unobservable variables and measurement error. *Journal of Marketing Research*, vol. 18, 1981, pp.39–50.

[28] Chin, W. W. The Partial Least Squares Approach to Structural Equation Modeling," in G. A. Marcoulides (Ed.) Modern Methods for Business Research, London, 1998, pp. 295-336.

[29] Tenenhaus M., Esposito Vinzi V., Chatelin Y., Lauro C. PLS path modeling », *Computational Statistics & Data Analysis*, Vol. 48 (1), 2005, pp. 159–205.

Reverse auction bidding in transport of goods - interCLEAN case

Vladimir Despotović
interCLEAN d.o.o., Belgrade, Republic of Serbia
vladimir.despotovic@interclean.rs , despotovic1@gmail.com

Abstract* - *In order to achieve business requests such as shorten the time of booking the transport of goods at a fair price with best delivery dates specific methodologies need to be applied.

This paper presents an application of the reverse auction bidding methodology in booking international transport of goods. It is the real interCLEAN Serbia case described from an idea, proof of concept and concluding remarks. This may be used as an example in other businesses and companies in order to improve transport procurement.

I. INTRODUCTION

Small and medium enterprises (SME) usually use a few ,,reliable,, carriers in a process of booking of transport of goods. In the best case scenario SME will pay market price for good service (delivery time, prompt communication etc.). For unregular full truck loading and less than full truck loadings situation is even worst, because ,,reliable,, carriers have to forward requirement to their own ,,reliable,, circle with additional margins included. One way to approach to this problem is to increase number of reliable carriers, but for SME it is, even more, time consuming process.

Online, so called freight exchanges have emerged (Timocom.com, Cargoagent.net), but there are no price, so SME (shipper) has to do all process of phone calls/chat conversations to find the ,,market,, prices manually. On the other side there is a trend of transport portals, more user friendly for SME (shippers-owners of the goods) with pre-determined prices (uship.com, ucandeliver.ru, cargoduck.com) and with/without transparent bidding between carriers (uship.com , nestcargo.com).

The main goal in this paper was to report a successful use of on-line reverse auction bidding in a company that has not previously used it and, as a result, and cost savings in its business operations.

The main goal in this paper was to report a successful use of on-line reverse auction bidding in a company that has not previously used it and, as a result, and cost savings in its business operations.

II. RELATED WORK

In the modern world, USA mostly, reverse auction bidding is greatly used. Since its deregulation in 1980s, the US freight transportation market has grown significantly and become very competitive and advancement of information technology leads to the online marketplaces for transportation services where the transportation capacities (carriers' offer) are dynamically matched with loads (shippers' demand) through auction mechanisms [1].

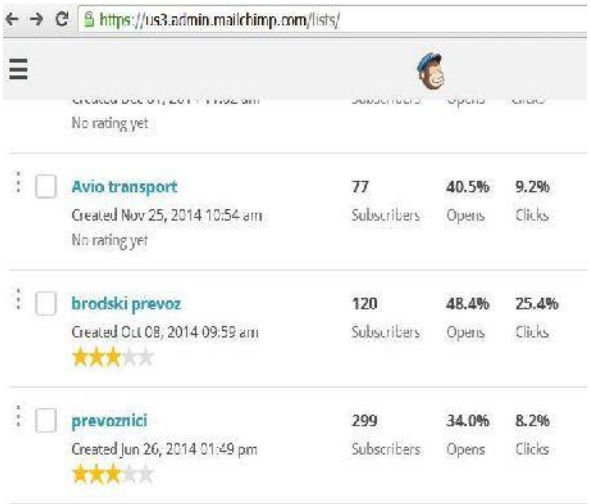
However, there are no sources, about reverse bidding auction application experiences in Serbia. This emphasized importance of this paper, which by described experience may positively influence on bigger reverse bidding auction application in our country, as well as on increasing company competitiveness, which uses this methodology.

There are several types of reverse auctions. For example it is not uncommon for buyers to use the terms a bundled bid, cherry picked or a scorecard auction [2]. In the case of a bundled auction the buyer usually bundles together his/her requirements into a single lot. Suppliers then bid for the total package. Usually they will submit a bid for each item, but these will be totaled up and one supplier usually wins the whole bid. (e.g. booking of shipping containers usually include port to port transport services, as well as truck container transport from port to local warehouse [3]). A cherry picked auction is slightly different. As the name suggests, suppliers have the opportunity to 'cherry pick 'certain lines from an auction and only bid for these. A buyer can then choose to award the contract to several different suppliers for different lots, or award to one supplier. (E.g. putting a more than one less than full truck loading in the same country could give you combined or separated logistic solutions [4]). A scorecard auction is slightly more complicated in that the buyer can assign an internal scorecard to each potential supplier. Each bid a supplier submits is then recalculated against the values assigned by the value on the scorecard to produce a weighted bid. The buyer may choose to share the scorecard information with the suppliers.This may improve their performance as they will know what they are up against and where they need to improve. (E.g. priority in some cases is speed of delivery over the lowest price [5]).

III. HOW IT STARTED?

There was a need for faster, easier, less time consuming booking of transport of goods, according to the budget or market price and level of service. I have used mix of free

on-line tools (Google spreadsheet , mailchimp.com) to create my own auction site [6] . The idea was to save time through transparent procedure, but after more than 100 auctions it show up that costs savings are significant i.e. 50-400EUR (up to 30% lower costs of transport) per booking compared to other offers or interCLEAN budget. Similar savings about 30% ware also found in other studies [7].
List of approved carriers is manually added according to previous contact or business experience with them. List of approved carriers is additionally segmented to contacts specialized for truck transportation, container transport and air cargo. Below (Picture 1.) is report from Mailchimp software about open rate of my quote request for each mode of transport (for air cargo/Avio, container/brodski and trucks/prevoznici) and statistics about the how many of contact carriers and freight forwarders (agents) have gone to my auction site.



Picture 1. MailChimp statistics

Offer from carriers are received through e-mail, skype, phone call and manually added to View only google documents, so carriers could all be aware of the best possible offer at that moment. From SME (shipper-owner of the goods) point of view additional decision parameters, beside price, are important e.g. loading date, unloading date, additional warehouse costs etc. In cooperation with other carriers and freight agents additional rules are added like ,,in latest 30 minute of time limited auction, reverse bidding ticker is -30EUR". The whole process could be work intensive and time consuming for carriers, so shipper transparency is highly recommended. Regarding, that interCLEAN has manually added all offers to auction site, the main problem was that time limited auctions ware the most intensive during the last minutes. The future solution for this problem will be automatization of the process using web application [8] (technology used: Laravel v.5.1 + JS - Jquery - Google Viz on Azure cloud platform).
Together with Mr. Perica Aleksov less than full truck loading calulator was developed in order to have idea about the referent market price [9] (google map API is used for distance km input). However, calculator development for full truck loadings seems even more challenging because other factors like trade deficit and surplus between two countries plays one of mayor rolls in determine market price as well as period of the year. Data sources like Timocom barometer and Statictical Office of Republic of Serbia will be used for fine tuning of truck transportation cost calculator.
Shipping freight rates for transporting containers from ports in Asia to Europe in US$ per 20-foot container (TEU) could be check weekly on Shanghai Shipping Exchange or through new ones [10] , but from point of SME (shipper) the main question is are you able to find even lower price pushing container agents to reverse auction bidding. The answer is yes, for both less than full containers loading and full container loadings [11].
Air freight (cargo) deliveries have been also tested successfully for reverse auction bidding [12].
Future testing will include rail transportation service.

IV. FUTURE

T As more of the spot market truckload freight transaction process moves online and gets conducted from a mobile device rather than a laptop, truckload freight matching gets closer to the Uber model. However, "Uberization of trucking" is maybe more buzzword than reality [13], but above analysis is from unique shipper point of view - the owner of goods who is paying at the end for the transportation services. However, there are research that are more focus on carriers' point of view [1]
Transparency is also a trend in container shipping industry with online platforms such as tryFLEET.com , Flexport.com , Freightos.com, Xeneta.com , Haveninc.com , 45hc.com , but this are still tools for innovators/early adopters in other regions.
There is a range of software and service providers offering e-auction capabilities Oracle, , SAP+Ariba, IBM+Emptoris , en.Vortal.biz , Bravosolution.us .
It is debatable whether it is better to enjoy the flexibility and functionality from providers who concentrate on eSourcing suites or if it is better to wait and take up functionality from existing ERP/eProcurement providers. In any event it would be unwise to ignore this issue altogether and let competitors take advantage of the potential savings which such functionality has to offer [14].
E-auction users should be aware of agreeing to pay a supplier a fixed percentage of savings (ie,5%) if outsourcing the e-auction process.
There are also advantages to buying the e-auction expertise and software to host e-auctions in-house. Although initially expensive and time consuming, the ability to build upon an e-auction programme year on year, introducing an expanding range of products and services provided, does have its advantages. Many large companies who have introduced an in-house e-auction programme started off outsourcing the process to test the process and prove the process internally.

V. CONCLUSION

During 18 months interCLEAN - import department succeeded from bringing idea to gain enough basic knowledge about reverse bidding auction approach. The first testing started at the beginning July 2014. and now after more than 100 auctions we could successfully confirmed proof of concept, about reverse auction bidding on spot transport market in region .

For companies that are just starting to conduct reverse auctions, the best practice would be to use an RFP (request for proposal) process followed by an auction [15].

Above interCLEAN solution is low cost - almost free, with small know-how barrier to test it as a pilot project for most of companies.

References

[1] Zhan Pang, Ming Hu, Oded Berman, B. Noble, and I. N. Sneddon, "Dynamic Bidding Strategies for Freight Transportation Services " , *The Manufacturing and Service Operations Management Society (MSOM)* , pp.2. http://msom.technion.ac.il/conf_program/papers/TB/5/196.pdf

[2] Helen Alder, Clerk Maxwell, *Introduction to e-auctions*, pp.4. , http://www.cips.org/Documents/Resources/Knowledge%20Insight/Introduction%20to%20e-auctions.pdf

[3] real world interCLEAN case samples of auction you could see using the following link https://goo.gl/7jro2h

[4] real world interCLEAN case samples of auction you could see using the following link https://goo.gl/wW3pAo

[5] real world interCLEAN case samples of auction you could see using the following link https://goo.gl/Rah78O

[6] real world interCLEAN case samples of auction you could see using the following link https://goo.gl/TpKyW3

[7] Semra Agralı, Barıs Tan, Fikri Karaesmen, "Modeling and analysis of an auction-based logistics market" European Journal of Operational Research (2007), doi:10.1016/j.ejor.2007.08.018 pp. 2. http://home.ku.edu.tr/~fkaraesmen/pdfpapers/ATK07v2.pdf

[8] previous http://netracuni.com/teret1.php and current development you could check on www.cargobid.me/home using following link as invitation http://goo.gl/CAsibs - registration is needed ; more commercial details in English www.lablogistix.com or in Serbian www.cargobid.co

[9] http://netracuni.com/quotes/test.html

[10] http://supplychainmit.com/2015/12/10/will-nyshex-lift-container-shipping-service-levels/

[11] real world interCLEAN case samples of auction you could see using the following links: https://goo.gl/no71FK , https://goo.gl/7jro2h

[12] real world interCLEAN case samples of auction you could see using the following link: https://goo.gl/zVBfGH

[13] Todd Dills, On-demand load matching: Where it's on the way , April 03, 2015 http://www.overdriveonline.com/on-demand-load-matching-where-its-on-the-way/ ; www.facebook.com/LabLogistix/

[14] e-Sourcing A BuyIT e-Procurement Guideline ; © 2004 IT World Limited on behalf of the BuyIT Best Practice Network Page 1 of 30 e P9 Issued: November 2004, https://www.cips.org/Documents/Knowledge/Procurement-Topics-and-Skills/12-E-commerce-Systems/E-Sourcing-E-Procurement-Systems/BuyIT_e-Sourcing.pdf

[15] Madhur Shah, , blog: Auctions in procurement Oct 11, 2010 , http://scn.sap.com/people/madhur.shah/blog/2010/10/11/auctions-in-procurement

Achieving interoperability of parking system based on *RFID* technology

Miloš Ivanović*

*University of Belgrade, Faculty of Organizational Sciences, Belgrade, Serbia

**Qfree-Elcom d.o.o, Belgrade, Serbia

smajhel.m@gmail.com

***Abstract* — The subject of this paper is based on communication between two heterogeneous systems. It is necessary to ensure the parking system or unit that will accept data from other systems and thus to ensure efficient data exchange to work together, without altering the exchanged information. The goal of this research is to present concrete example of parking system using RFID technology and its interoperability which will increase productivity and efficiency of the entire system.**

Figure 1. Integration of RFID technology with existing parking system

I. Introduction

The research in this paper is directed towards the ability of communication between two heterogeneous systems. On one side there are the concepts of RFID (Radio Frequency identification) technology, which should solve the problem of simple use of parking system when entering or exiting the vehicle via radio waves. Thereby it will reduce the waiting time and operating costs of parking and increase productivity and efficiency complete parking system as well. When thinking of a complete parking system, it means that the application of these technologies must be applied from the standpoint of interoperability so that these technologies can be integrated into existing parking systems and , using them information can be exchanged that actually allow operation of the entire system.

On the other side, there are existing parking systems that work using the device for issuing paper cards and devices to control barriers at the entrances and exits to the parking area. These parking systems contain certain protocols and rules that they operate on. It is necessary to carry out interoperability of the two systems by the concept that these two systems are heterogeneous and that it is necessary to ensure the parking system or unit that will accept data from the other systems and thus to ensure effective data exchange and joint operation of the system, without changing the structure of the exchanged information. RFID technology is a technology that uses radio frequency to exchange information between the portable devices.

This system consists of the tag (transponder), which is actually a data bearer, and the antenna that communicates with tags. This system has no utility value unless a unit is provided between the antenna and the existing parking system that will manage and monitor the communication between the antenna and the tag, in order to use resources from this transaction in other parts of the parking system more efficiently. This research problem is worth studying because of the quality, accuracy, and achieving productivity at the lowest costs of this system. This system must operate completely, without any possible exceptions because it can cause big problems in exploitation and work, because this is a system that works in real time. One of the problems that can occur due to inadequate communication of software in the controller with the antenna, is that the controller, which is a mediator between the two systems, incorrectly processed data from the tag. Further research will explain in detail in what way the software in the controller communicates with other components of the system and allows its functionality.

The rest of paper is organized as follows. In the second chapter will discuss about the concepts of RFID technology to be applied in the functioning of the parking system, and their advantages. In the third chapter will discuss about the ways of solving the problem of interoperability RFID technology in existing parking system. In the fourth and the fifth chapter will be presented a concrete example of interoperable parking system based on RFID technology.

II. RFID technology and a parking system

The application of RFID technology in the operation of parking systems, achieves a higher level of automation of all business processes and algorithms that make up the system. For example, in most cases, RFID in parking systems is used for automating the access control in closed or open parking spaces where the inputs and outputs are automatic barriers.

A vehicle, that has a transponder placed on the windshield in front of the driver, when approaching the entrance of the parking lot, has space where the RFID reader via radio waves reads data from the memory of the transponder and communicates with him dually. Then the software, that manages the reader, converts data from the transponder into useful system information that it stores, processes, shares with other systems and updates. After verification, system sends a signal to the automatic barrier to allow access to the vehicle in a parking space. Transponder has its own unique identification number that distinguishes it from all the other such devices, along with a valid data of the owner or of the vehicle.

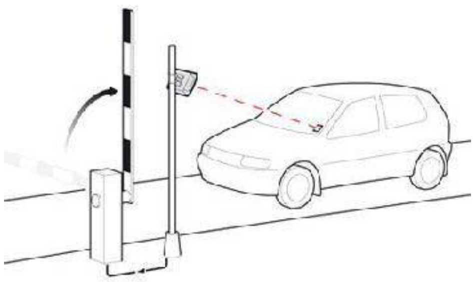


Figure 2. Identification of vehicle on the entrance/exit of parking lot

The advantages of using RFID technology in the parking system are:

1. Processing data from the transponder takes place without the influence of user, i.e. antenna reads data from the microchip in the transponder via radio waves.

2. There is a centralized software that manages the complete system and devices, and it collects, processes and exchanges data with other parts of the system, which makes this system interoperable.

3. Transponder is easy to use and by dimension it can be easily placed onto the inside of the vehicle.

4. Memory of transponders is large enough to be able to store all the necessary information about its holder.

5. Distance range in which the antenna detects the transponder is up to 10 meters, which is enough for a parking space.

6. The RFID system consists of components that can be easily installed in the field and do not require much space and cables to connect them to the network and power source.

7. Speed of reading data from the transponder is measured in milliseconds, which implies that communication is very quick and it saves user waiting time for the transaction.

8. The lifespan of a transponder is far greater than the barcode and can handle up to 100 000 transactions to its next replacement. That is, the user of a transponder can realize up to 100 000 entrance / exit of the parking space.

III. ACHIEVING INTEROPERABILITY OF RFID TECHNOLOGY WITH PARKING SYSTEM

The result of this research is a concrete example of parking system using RFID technology and how its interoperability increases the productivity and efficiency of the entire system by reducing the cost of investing in only the maintenance and operation of the entire system. The desired results of this research will be reached by applying the concept of interoperability between the two heterogeneous systems and the ways of exchanging data between them without altering the exchanged information and communication protocol.

European centre for interoperability stated that interoperability can be achieved using the following levels of interoperability [1]:

1. Technical interoperability

2. Syntax interoperability

3. Semantic interoperability

Using the above stated levels of interoperability , we will come to the unit system that will completely contain technical, syntactic and semantic interoperability and it will be operational for its safe and efficient exploitation. Interoperability is defined as a common set of processes, procedures and equipment adopted by more than one provider, to support and improve simple use for users and for data acquisition. In interoperable system, user can easily switch between heterogeneous parts of the system, and take data from them in order to integrate information suitable for achieving good results of the system. At the highest level of interoperability between heterogeneous system, boundaries are invisible to the user, in order that these background processes and procedures work together and provide the user with the transaction of system, data exchange, data warehousing and distributing data to other parts of the system.
RFID systems consist of transponders (tags), reader (antenna), programmed cell and various server software interfaces that enable interoperability with other systems. Over the management software interoperability of such a system is achieved, because this software is the main component in the management of RFID systems and exchanging information with other parts of the system. Many RFID systems have a server that collects information from the transponder in parking systems and maintains a complete database in one place.

Database as a first and important component of this parking system is the starting point in achieving interoperability. In fact, many parking systems are designed to store all transactions between the reader and transponder in a database, in order to monitor the history and create reports with the possibility to exchange such information with other systems.
To ensure interoperability, parking system needs to improve the capability and the technology used for storage, import and export data (in real time), to adopt standards and standardization for networking and data management, and to adopt the use of open source, so the user can enhance the functionality of the software.

Two possible ways in achieving interoperability in terms of data exchange between systems are [2]:

1. The exchange of data between systems via peer2peer technology is suitable for a small number of systems but manages to overcome the barriers between systems and enables them to communicate.

2. The database is the most up-to-date central component that allows you to interact with the systems, it is removable and standardized.

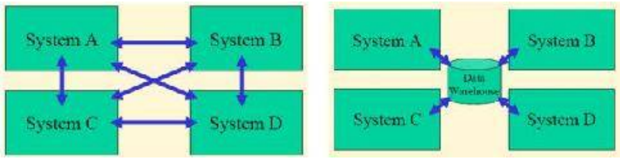


Figure 3. Achieving interoperability between heterogeneous systems

RFID interoperability refers to the ability to allow vehicles to identify the different needs and in different locations, using unique On-Board-Unit (transponder or tag) for each vehicle. In the transponder all the necessary information about its owner are found, in this case that is the vehicle. RFID interoperability obtains its role when the parking system expands to different locations, and when it is necessary that users can safely use any parking site that they have the privilege and that the data between the parking system can be viewed and processed.

Parking system that uses RFID technology as the holder of control and access to the vehicle parking area comprises three essential components that the user needs to have in order to participate in the system. Those are:

1. Bank account - to be able to pay the money for the use of parking spaces.

2. Transponder - a device used by a user identified at the entrance / exit of the parking lot.

3. Mobile phone - through which also more easily, via messages can provide funds in his account and he receives information timely on using transponders.

The above stated components of the system are the very interoperability of the system. Interoperability is the connection of these three components in a single unit which enables mutual cooperation and exchange of data and information, through appropriate software. It is a very significant technical, operational and administrative undertaking whose application is not to be taken lightly, but it is necessary to invest a lot of time and effort to implement and use in the best possible way.

Advantages of RFID interoperability achieved in the parking system are [2]:

1. Increase of loyalty and customer satisfaction.

2. Reduction of operating costs parking system, since they no longer use automatic billing for spent hours in the parking lot, where he spends far more, both the electricity and paper for printing bills.

3. Increase cooperation between the parking system.

4. Without interoperability, parking systems could function independently, without mutual cooperation that would lead users in an awkward position as it would have to communicate independently with any parking system.

5. Application of physical interoperability means that users can use their transponders for any parking system located in this community.

6. Back-office interoperability enables data exchange between the parking system on the status of transponders, the validity of transactions, payment status and account of the users.

IV. "LCC 550" PARKING SYSTEM

"LCC 550" parking system uses a universal controller, here in after referred to as the controller, which on one hand allows the integration of RFID technology in the entire system and on the other side, the same interoperability with other parts of the system. The controller in the parking system enables the exchange of data between heteregonous system, thus enabling the efficiency of the entire system, which leads to positive business results.

The controller is one of the currently available products on the market. Designed and developed to automatically connect to an existing parking systems, which makes it very efficient, because now there are largely on the market parking systems that operate successfully and the advantage of this controller is that it in a simple way of implementation allows achieving interoperability with RFID technology. The use of RFID technology, which was previously discussed, recently reached a high level of use in the systems of collection of parking and tolls.

This controller is called the universal as it can support the management of various applications, such as:

1. Parking systems,

2. Systems of access control,

3. Systems for automatic license plate recognition.

Figure 4. Universal parking controller "LCC 550", key for interoperability

The controller allows its interoperability with different producers parking systems currently on the market. So far, in practice it has been implemented with the following manufacturers of parking systems for the collection of use of parking spaces:

1. Skidata

2. Schiedt & Bachmann (S&B)

3. Parcare

Connection with other parking systems is realized via RS232 serial communication and via an ethernet Local Area Network. These modes of communication provide the basis for achieving interoperability with other parts of the parking system, because in this way they enable the exchange of data with the same through predefined communication protocols without changing the content and meaning of data exchanged.

The controller can be synchronized and implemented on all modern systems for the collection of parking services, including current and future systems that are installed in parking systems. It enables improvement of parking systems without replacing existing parts of the system, but to integrate with them and enhance the quality of results from the system. Also, the architecture of the system allows remote management of entire system from the so-called control center. The control center is mainly a large office with modern equipment to enable connections to other parking systems. It secures way of exchanging data for the purpose of processing, viewing and storage of data as well as the maintenance of the system to be functional and without exception in work.

The advantages of this system are:

1. The possibility of upgrading the system, depending on the additional requirements.

2. Management system from remote locations.

3. Processor, memory and communication system are stable and fast.

4. Controller is designed to work in real systems in different environments, temperatures and operating conditions.

5. Operating system of controller is based on the Linux platform.

6. Reduces operating costs of the parking system, as it automates all business processes for entry-exit of vehicles, payment of parking space usage and processing of data.

The objectives of the operation of this parking system can be multiple. First of all, users who decide to introduce this type of parking system, aware that it will speed up the processes of the system and create serving its customers, drivers, effectively. The advantage, that at the very beginning of the introduction of this system is achieved, is a reduction of operating costs and business automation.

The very architecture of the system allows data exchange not only with existing parts of the system, but also through the global Internet network, which creates a high level of efficiency of such a system. Complete system can be operated from a single control center, which is connected with the equipment via a wired or wireless Local Area Network, depending on the user's request.

This parking system removes vending machines for paying of spent hours in the parking lot, because in this system, all transactions between the driver and the parking system are carried out through practical transponders or tags. Transponder is located in the vehicle, mainly on the windshield, and it contains all the information about its owner such as name and surname, address, registration number, telephone number, account number, etc... The memory of the transponder allows the storage of sufficient data of the owner so that system can function normally, and most importantly that it can be improved in the future. The system is very dynamic and open for any additional customer requirements, making it flexible and interoperabile.

The main objective of this system justifies the subject of this work, and that it is interoperabile with other systems. Data exchange with other parts of the system is performed via serial communication and local network communication. This exchange of data is a key of interoperability, for this parking system has been developed to be easy and simple to integrate into existing communication protocols and comply with them. The goal of interoperability is data exchange between systems without disrupting and changing the content of the data to be exchanged in order to preserve the integrity of the system.

V. SYSTEM ARCHITECTURE

The components that make up this parking system and its operation efficient and flexible are the following:

1. Controller

2. Transporder (tag)

3. Antenna (RFID reader)

4. Control management center

5. Existing parking system

Figure 5., shows the architecture of "LCC 550" parking system, where we can see the role of the controller (in the middle), which is to maintain interoperability between system components. Namely, it enables the exchange of data between the RFID technology, existing parking systems and parking management center. In the current parking systems , there is a certain system that allows its operation, the aim of this controller is to integrate all parts of the system in one unit and that all have one common goal: to increase the efficiency and flexibility of the entire system while reducing operating costs and increasing profits.

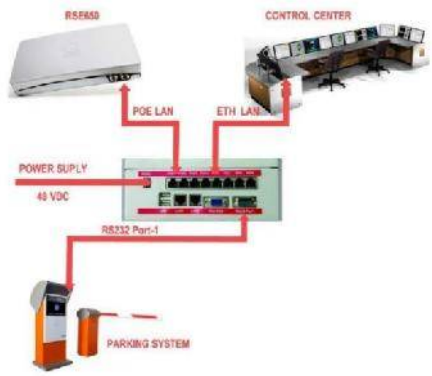


Figure 5. “LCC 550” system architecture based on RFID technology

The transponder or tag is the main component of this system because it is the only system component with which the user has direct contact and therefore it is essential to be very reliable, of appropriate design and dimensions. Specifically, the design and functionality of this component is determined by the quality and cost of the entire system because users use it to communicate with parking systems using RFID technology.

Tag is the holder of all required information about the user and it is located in the interior of the vehicle. RFID reader communicates by using radio waves with transponder and exchange data with it.

The advantages of the transponder in relation to the other on marke are as follows:

1. New chassis design, size 60 x 40 x 19 mm,

2. Easily attached with a single carrier,

3. New and faster processor that with the antenna exchanges data with a reader,

4. Weighing about 30 grams,

5. Lifetime = 10 years,

6. Active transponder, which means it has its own power supply using battery.

Figure 6. Transponder (Tag)

RFID antenna „RSE 650“, is the version that this parking system has in use currently, representing the components of this system whose main function is the identification and communication with the transponder. This antenna is placed at the entrance / exit of the parking space and by detection of transponders, it reads data from it and forwards them to the controller for further processing and exchange with other parts of the system.

The antenna has been optimized and programmed to operate parking systems, toll collection and access control systems. It is designed to be able to operate in all weather conditions, the so-called internal and external use. Works by standard IP65, which defines the requirement that the product is water-impermeable, as this allows the antenna to operate in outdoor conditions. Range up to which it can communicate with the transponder is 6 to 8 meters, which is quite sufficient condition for a parking system, where it is a lot less distance, up to 4 meters.

Power antenna is less than 4 watts, which is 70% less than competitors use. This is a very important characteristic of this component, because it reduces the cost of the power resources a lot, that are not convenient source today. In standby mode, the antenna draws less than 1 watt.The antenna is in this mode, until it is activated, and the controller activates it at the time when the vehicle encounters an inductive loop which represents the trigger for starting a transaction between the antenna and transponder. So, the controller, that is the key for achieving interoperability in this system, reduces cost efficiency of the power source, which is the most expensive for the operation in the system, and it is current. An additional feature of the antenna is that its power supply can be from batteries and solar panels.

Speed of activated antenna after switching from the defunct regime, is 200 milliseconds, and it is ready for the transaction. Speed of activated antenna after standby mode is 10 milliseconds. This feature is very important and it justifies the possibility of putting the antenna in standby mode when there is no vehicle on the inductive loop, because many users will not wait for the transaction, i.e. speed of service is very efficient.

Dimensions of antenna are 170 x 310 x 40 mm with housing, weighing 1.1 kilograms, and with angle-adjustable mounting bracket to the din rail. Design and antenna is a recent standard that allows this antenna to participate in market competition.

Figure 7. RFID Antenna

Universal parking controller is the central and main component of the system that allows interoperability of parking system with RFID technology. On one side it communicates and exchanges data with an RFID antenna that is reader, it retrieves data from the transponder, processes them and prepares for exchange with other parts of the system. Other parts of the system are the existing parking systems and management center. After receiving and processing data from the transponder, controller packs them in a defined protocol to communicate with the existing parking system. The existing parking system is a very important part of the system with which the controller must achieve two-way communication, because the system is managed by barriers (gates) at the entrances and exits of parking spaces and also communicates with inductive loops that give us information about when to activate the RFID antenna to start the transaction. The structure of the controller, which can be seen in Figure 8., consists of the following hardware and software components:

1. Industrial PC on which to install the appropriate software for the management controller. The configuration of this component is very important because the quality and speed of these components is dictated by how

effectively and quickly operates the software that performs the processing and exchange of information with other parts of the system, of course without failure.

2. Power Switch, which in addition to enabling connection to the network controller parking system, powers RFID antenna.

3. On storage drive is installed the operating system Linux and the software required for the management controller, a memory space of up to 120 gigabytes, which is enough for many years of operation of this controller.

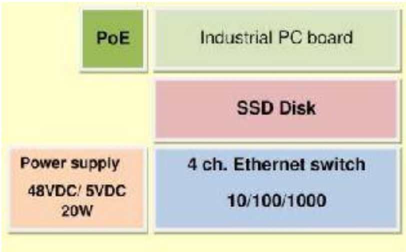


Figure 8. Structure of universal parking controller “LCC 550”

Controller communication with the RFID antenna is realized through Local Area Network (LAN), connecting the antenna in switcher controller which powers the antenna in addition to communication. The protocol used in communication is TCP client-server. The RFID antenna in software has running TCP client, through which it sends data from the transponder to the controller. The antenna is the client and in the controller there is a running server that listens to all clients on the network. This is done so that multiple antennas can be connected to one controller. The antenna after reading data from the transponder packes them into a single XML telegram and sends them to the controller. The controller takes XML telegram from antennas, processes and prepares a set of data that needs to be sent to the existing parking system so the transaction could be ended. Controller communication with the existing parking system is made via a serial RS232 communication, and through it forwards the telegram to the system as a series of bytes. Parking system takes the data and checks the identification number of transponder in the database and if the conditions for entry into the parking lot are completed, it sends a command to the barrier to open.

VI. CONCLUSION

Universal parking controller provides interoperability of parking system and RFID technology, through the levels of interoperability that fulfills completely. The levels of interoperability which the controller uses as an intermediary in the exchange of data between two heterogeneous systems are:

1. The technical interoperability refers to the level at which there are defined communications protocols for data exchange between systems, through which he established communication infrastructure for the exchange of bits and bytes. Controller sharing of bits and bytes is exercised in serial communication with the RFID antenna and the existing parking system.

2. Syntax interoperability refers to the ability of two or more systems to exchange data at its basis are defined data formats and communication protocols. Example for syntax interoperability is the XML standard. In the controller is a syntax interoperability contained in communication with the RFID antenna, where as the standard for data exchange uses defined data format XML.

3. Semantic interoperability ensures that the two systems are able to exchange information automatically, meaningfully and precisely, and also to mutually interpret the information exchanged in order to provide useful outputs or results that users expect. Controller automatically and without affecting the operator communicates with other parts of the system and over defined communication protocols realizes the transaction of the system and gives the effective output of the system which the user expects.

The benefit of this research is the implementation of solutions to the problem of interoperability in the parking system based on the use of RFID technology. The main components of RFID technology are RFID readers, RFID tags, computer units, barriers and software for managing the complete system. The software has been handled for the management, control and reporting of transactions for different parking spaces. Check-in/out of vehicles from the parking lot is exclusively controlled by RFID readers, RFID tags and barriers. This means that in the future the use of RFID technology can enable parking system that will fully operates unmanned, secure and automated. Check-vehicles will be accelerated without requiring that the vehicle stops to seize the card or pays service for using the parking lot and in order to reduce traffic congestion.

ACKNOWLEDGMENT

This project is supported and financed by the Norwegian company Q-free with headquarter in Trondheim. I would like, as the author of this system, to thank the company Q-free on the provided trust and support during the development of the system. This system is currently installed at dozens of parking lots in the cities of Spain and Chile.

REFERENCES

[1] Luras, M., Zelm, M., Archimede, B., Benaben, F. & Doumeingts, G., “Enterprise Interoperability”, London: ISTE Ltd., 2015.

[2] Lierz, L., “Unified parking systems”, unpublished.

[3] Cameron T. Howie, John C. Kunz, Kincho H. Law, “Software Interoperability”, Stanford University, 1997.

[4] Rahul, M., Phatale, P., Kulkarni P.H, “RFID technology for parking”, 2011.

[5] Pfleeger, S.L., i Atlee, J.M., “Softversko inženjerstvo”, 3.izdanje. Beograd: CET., 2006.

[6] Ivanovic, M., “Achieving interoperability of parking system based on RFID technology”, Belgrade: FON, Specialistic academic work, 2015.

Comparing Apache Solr and Elasticsearch search servers

Nikola Luburić, Dragan Ivanović
University of Novi Sad, Faculty of Technical Sciences, Novi Sad
{nikola.luburic, dragan.ivanovic}@uns.ac.rs

***Abstract* – The paper presents a comparative analysis of the leading two platforms for developing information retrieval systems, Apache Solr and Elasticsearch. We briefly examine other similar solutions, but focus on the previously mentioned solutions as they provide greater functionality with better performance. Our goal was to examine both systems, including what they offer and how they are used. We examine expert opinions on both systems, as well as concrete use cases. After that we make a comparative analysis focusing on many aspects, from usability to working at scale. Finally we conclude which system works better for which use case.**

INTRODUCTION

With the development of information communication technology more and more data is being written and stored digitally. While collecting data has become less of a problem, extracting useful information from massive volumes of digital documents has become a large issue. Relational databases, which were the previously leading solutions for data storage, aren't designed for such scale and big data searches. In order to solve this issue, a search engine needs to be built that efficiently stores, indexes and searches data, so that the end user can quickly access the information he\she needs.

Search engine indexing is a process in which data is collected, parsed and stored in order to support fast and accurate information retrieval [1]. Searching is a process which includes query processing and information retrieval based on the query. In order to be efficient, the process uses previously formed indexes, so as to avoid scanning every document in the corpus [2].

The previously mentioned functionality is essential for solving the problem of efficient information retrieval, but it is not sufficient in and of itself. The system needs to provide an easy to use, intuitive user interface, with which the user can utilize the search capabilities. The system needs to take into account not only the user's lack of knowledge about the underlying structure of the data set, but also his\her inability to form precise queries. Another issue, related to indexing and searching in general, is that the result set for a given query will be imprecise, regardless of the quality of the query or the implementation of the search system. The previously mentioned issues can be solved by introducing a ranking system, which will sort the results for a given query by relevance.

As information retrieval has become a serious problem, many solutions have arisen over the years trying to address it. In this large set of solutions it is hard to choose the right tool without previous experience.

This paper represents a comparative analysis of the leading systems which solve the problems mentioned above. The main motivation behind this paper is to provide developers and members of the scientific community an overview of the best solutions for developing information retrieval systems, as well as give insight for the best use cases of both solutions.

During our research we preformed manual inspection of various solutions, and in the end isolated the two systems Apache Solr [3] and Elasticsearch [4]. Solutions such as Sphinx [5] and Xapian [6], Whoosh [7], while still in active development, lack functionality compared to Solr and Elasticsearch, while solutions such as Swish-E [8] have stopped with development altogether.

In the following chapter we take a look at examples where Solr and Elasticsearch have been utilized. In chapter three we analyze both systems separately, while chapter four focuses on the actual comparison. This includes comparing the differences in design, offered functionalities, ease of use and resource consumption. Finally, we list the use cases for each system and make predictions about the future of these two systems.

RELATED WORK

We examined several solutions which used Apache Solr or Elasticsearch as their primary search engine. This includes solutions produced by the scientific community as well as several major organizations in the industry.

In [9] Atanassova and Bertin present an information retrieval system for scientific papers using Solr. Their approach provides a new way to access relevant information in scientific papers by utilizing semantic facets. Faceted search allows the user to visualize multiple categories and to filter the results according to these categories. In [10] Cuff and colleagues show a significant improvement in the search function of their CATH system when switching to Solr. CATH, which stands for class, architecture, topology and homology, is a hierarchical protein domain classification system.

Many organizations, such as Helprace, Jobreez, Apple, Inc., AT&T, AOL, reddit, etc. use Solr for search and faceted browsing [11]. Helprace uses Solr to power its search engine and search suggestions, while Jobreez uses Solr to search for jobs across 25 000 sources. AT&T uses Solr to run local searches on its yellowpages, while AOL utilizes Solr to power most of its channels.

In [12] Kononenko and colleagues describe the significant improvement in performance they got with their software analytics dashboard tool. This performance boost is solely due to switching from a traditional relational database to Elasticsearch. In [13] Thompson and

colleagues describe their use of Elasticsearch in querying graphical music documents. As far as organizations go, notable users include CERN [14], GitHub [15], Stack Exchange [16], Mozzila [17], etc. CERN uses Elasticsearch to efficiently manage and search through the various logs their devices produce. In 2013 GitHub started using an Elasticsearch cluster which indexes code as it gets pushed to the repository. This change marked an increase in search result relevancy and general search performance.

Lastly, there are organizations such as Foursquare which use both Solr and Elasticsearch in their infrastructure [18].

As both Solr and Elasticsearch are leading solutions for information retrieval this paper aims to compare the two systems. There are several articles written by experts in the industry which compare these systems [19-22], and this paper represents a synthesis of those articles, as well as our own experience.

ANALYZED SYSTEMS

Before examining Solr and Elasticsearch we take a look at Apache Lucene [23] which is the underlying information retrieval library for both systems. Lucene is a free, open source and independent library which has been widely recognized for its utility in the implementation of Internet search engines and local searching. The primary function of this library is indexing and searching and Lucene result ranking uses a combination of the Vector Space model and the Boolean model of information retrieval [24] to determine how relevant a given document is to a user's query.

Apache Solr is an open source enterprise search server originally written in Java. It runs as a standalone full text search server, using Lucene for indexing and search functionalities. The system exposes a REST-like API and most of the interaction between the user and Solr is done over HTTP. By sending HTTP PUT and POST requests it is possible to send documents for storage and indexing, while the HTTP GET request allows the user to retrieve results based on queries. The data sent and retrieved supports several formats, including XML, JSON and CSV. Not only does Solr store, index and search data, it also offers additional features, including analytics of the indexed data. Unlike Lucene which offers indexing and searching, but lacks the needed infrastructure to be a standalone application, Solr is a web application which can be deployed on any servlet container. This allows Solr to be used as a tool by people from various professions, as was shown in the related works section.

Similarly to Solr, Elasticsearch is also an open source enterprise search server written in Java. It provides a distributed full text search engine, with a REST-like web interface and uses JSON for the document format. It provides scalable search, has near real-time search and supports multitenancy. A significant feature of this system is massive distribution and high availability. Elasticsearch allows users to start small and scale horizontally as they grow. These clusters are resilient and will detect new or failed nodes and reorganize data automatically, to ensure that it stays safe and accessible. More so than Solr, this system offers real time advanced analytics of the indexed data. Elasticsearch can be used as a standalone system by people of various professions.

It should be noted that both systems evolved together and learned from another, reaching a point where they are very similar to one another.

COMPARATIVE ANALYSIS

Before diving into the comparative analysis it should be noted that both solutions in their current versions (Solr at 5.3.1 and Elasticsearch at 1.7.2 at the time the comparison took place) are similar systems, in the sense that they offer near-equal functionalities and have similar performance. It should also be noted that while some difference does exist, both systems can be suitable solutions for most common information retrieval needs. As both technologies are mature and stable and have a strong community behind them, most of the time the decision whether to use one solution over the other will come down to preference and the foreknowledge of the team.

The main difference between these solutions derives from their cores which significantly differ from one another. Looking at the distributions, Solr takes more space on the hard drive. This is primarily owed to the fact that the standard Solr distribution includes functionality, other than the base, which may or may not be useful to the end user, such as Map-Reduce, a testing framework, a web application which represents a GUI monitoring tool, etc. Unlike Solr, Elasticsearch's core consists of only the base code and documentation, which is why it needs one third of the space that Solr needs. Fig. 1 shows the composition of the web application archives of both systems.

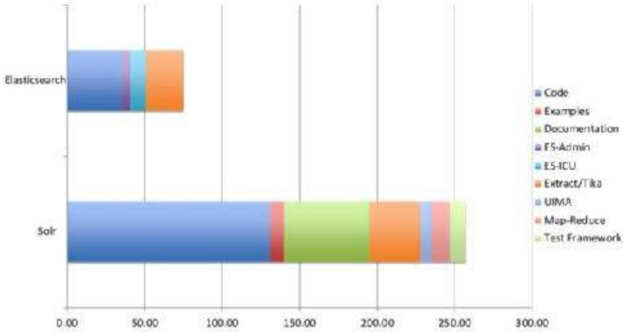


Figure 1. Solr and Elasticsearch .war composition[1]

Elasticsearch starts from the premise that the end user will always need the minimum functionality that Lucene offers and not much else. By configuring the system and using additional tools like Logstash, Kibana and Marvel, the user can expand the basic functionality to suit his needs. While Solr also offers a wide variety of plugins, its core includes modules which aren't always necessary. A problem that both systems have, but is more evident with Elasticsearch, is the lack of a centralized orchestration tool, for plugin and dependency management. If the user wants to create an Elasticsearch cluster he must manually install Elasticsearch and all the needed plugins on each node.

The second difference can be found in the cluster management subsystems of these two solutions. Solr relies on Apache ZooKeeper [25] which is a mature and tested technology, but more often than not offers little more than

[1] Source: http://www.slideshare.net/arafalov/solr-vs-elasticsearch-case-by-case

the much simpler built-in cluster management subsystem of Elasticsearch. ZooKeeper adds complexity to the system, as it is a separate component that needs to be managed. ZooKeeper also requires three nodes to form a cluster, while Elasticsearch can form a cluster with only one node.

Both Solr and Elasticsearch handle document preprocessing in a similar fashion. Both systems have configuration files in which analyzers, tokenizers and filters are declared. The declared components are used both during file and query preprocessing. Fig. 2 shows an example document containing the preprocessing configuration in Solr, while fig. 3 shows a similar configuration for Elasticsearch.

```
<fieldType name="nametext" class="solr.TextField">
  <analyzer type="index">
    <tokenizer class="solr.StandardTokenizerFactory"/>
    <filter class="solr.LowerCaseFilterFactory"/>
    <filter class="solr.KeepWordFilterFactory" words="keepwords.txt"/>
    <filter class="solr.SynonymFilterFactory" synonyms="syns.txt"/>
  </analyzer>
  <analyzer type="query">
    <tokenizer class="solr.StandardTokenizerFactory"/>
    <filter class="solr.LowerCaseFilterFactory"/>
  </analyzer>
</fieldType>
```

Figure 2. Solr preprocessor configuration file

```
index :
    analysis :
        analyzer :
            standard :
                type : standard
                stopwords : [stop1, stop2]
            myAnalyzer1 :
                type : standard
                stopwords : [stop1, stop2, stop3]
                max_token_length : 500
            # configure a custom analyzer which is
            # exactly like the default standard analyzer
            myAnalyzer2 :
                tokenizer : standard
                filter : [standard, lowercase, stop]
        tokenizer :
            myTokenizer1 :
                type : standard
                max_token_length : 900
            myTokenizer2 :
                type : keyword
                buffer_size : 512
        filter :
            myTokenFilter1 :
                type : stop
                stopwords : [stop1, stop2, stop3, stop4]
            myTokenFilter2 :
                type : length
                min : 0
                max : 2000
```

Figure 3. Elasticsearch preprocessor configuration file

Apart from the built-in preprocessing components, both systems allow for creation and use of custom components. It is even possible to move the preprocessing to an entirely separate system, and this is where the two systems differ. While Solr recommends a tight coupling between preprocessing and indexing and searching, Elasticsearch takes a more modular approach, and recommends a separate system (e.g. Logstash) for preprocessing. This approach increases the system complexity by introducing another moving part, but avoids bottlenecks by allowing each subsystem to scale separately.

In the context of document preprocessing it is also worth noting how both systems handle language recognition. Solr has this functionality built-in, while Elasticsearch requires a plugin. Several good solutions exist, and coupled with such a library, Elasticsearch handles this issue as well as Solr.

Both Solr and Elasticsearch can index digital documents such as PDF, MS Word document, etc. Once again, this is part of Solr, while Elasticsearch uses an external module. Both systems rely on Apache Tika for this functionality [23].

Highlighting is another feature both systems handle well, offering a high degree of flexibility in the configuration of summary creation and management. Solr handles this using the `hl` object [24]. By accessing the object's fields (e.g. `hl.formatter`, `hl.snippets`, etc.) the highlighter can be configured. Elasticsearch highlighter configuration is done by sending a `highlight` object [25] with the query request. This object contains most of the fields that Solar's `hl` object has. Fig. 4 shows an example of such an object. As the image shows, it is possible to form a query for the highlighter, making highlighting independent from searching.

```
"highlight" : {
    "order" : "score",
    "fields" : {
        "content" : {
            "fragment_size" : 150,
            "number_of_fragments" : 3,
            "highlight_query": {
                "bool": {
                    "must": {
                        "match": {
                            "content": {
                                "query": "foo bar"
                            }
                        }
                    },
                    "should": {
                        "match_phrase": {
                            "content": {
                                "query": "foo bar",
                                "phrase_slop": 1,
                                "boost": 10.0
                            }
                        }
                    },
                    "minimum_should_match": 0
                }
            }
        }
    }
```

Figure 4. Highligh object in an Elasticsearch query request

Elasticsearch has a few advantages when compared to Solr. The main advantage this system has comes from its simplicity. Elasticsearch is easier to install, setup and use. The REST-like services which work with JSON are not only simple to use, but are more aligned with the current trends in the industry that Web 2.0 has brought. This does simplify things for software developers, but it should be

noted that other industries use these systems as well. The JSON-based query language that Elasticsearch uses is also arguably simpler than the HTTP requests that need to be formed in order to query Solr.

Another difference comes from the general direction that both systems are moving towards. While Solr remains specialized for document indexing and searching, the Elasticsearch team puts significant efforts into improving and expanding their data analytics subsystem. This might very well be the most significant difference between these two systems.

When analyzing performance we found that both systems showed similar results on datasets of medium size. Elasticsearch did, however, surpass Solr when testing analytic queries, which was expected.

Finally, another key difference can be seen in the metrics that both systems offer. While Solr does provide key metrics, Elasticsearch (in its core, as well as by utilizing plugins) offers significantly more metrics.

CONCLUSION

The paper examines two leading solutions for information retrieval, Apache Solr and Elasticsearch, and presents a side by side comparison of these two systems. After manually inspecting both systems and researching the papers and articles on the subject we conclude that both systems are good choices when it comes to document indexing and searching.

Over the years both solutions learned from one another and significant improvements in both systems are partially due to the competition. While Solr still seems to be the more common solution for classic enterprise systems, Elasticsearch's simplicity, flexible design and modular architecture make this system a great choice for both prototyping and large, scalable information retrieval solutions. Elasticsearch offers far better data analytics, and when combined with Logstash and Kibana, the ELK stack surpasses Solr in many areas, including preprocessing, analytics and visualization.

Even though both teams continue to upgrade and develop new features, the fact of the matter is that Elasticsearch has a fresh, compact core, created after the various drawbacks of Solr were noticed. Solr hasn't stood still and while many improvements were made it can be concluded that Elasticsearch will in time surpass this system. This coupled with the fact that Elasticsearch relies on one man to approve or decline changes to the system, while Solr requires that every new features goes through a more rigid protocol of evaluation means that Elasticsearch, as it stands, will have quicker and more frequent updates.

The only real downside of Elasticsearch in its current version is that it lacks a centralized tool for managing the nodes of a cluster. This can easily lead to misconfiguration in clusters with many nodes, as a separate installation of the core Elasticsearch instance and all of its plugins is needed every time a node is added to the cluster. Without version control, installing updates for parts of the system present another problem.

While Elasticsearch does seem to be the go to solution for use cases where serious analytics are needed this does not mean Solr should be abandoned. The ELK stack might be a slightly more suitable solution, but reworking a system which already utilizes Solr will more often than not be pointless. Likewise, teams who have experience with Solr shouldn't switch to a new system without serious consideration, as both systems are near equal in most cases.

ACKNOWLEDGMENT

Results presented in this paper are part of the research conducted within the Grant No. III-47003, Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Z. G. Gonzalez, M. Kelly, T. E. Murphy Jr, and M. Nisenson, International Business Machine Corporation, 2012. *Search Engine Indexing*. U.S. Patent Application 13/713,765.

[2] S. T. Kirsch, W. I. Chang, and E. R. Miller, Infoseek Corporation, 1999. *Real-time document collection search engine with phrase indexing*. U.S. Patent 5,920,854.

[3] Apache Software Foundation, *Solr*, http://lucene.apache.org/solr/, retrieved: 20.10.2015.

[4] S. Banon, *Elasticsearch*, https://www.elastic.co/, retrieved: 20.10.2015.

[5] A. Aksyonoff, *Sphinx*, http://sphinxsearch.com/, retrieved: 20.10.2015.

[6] Xapian Team, *Xapian*, http://xapian.org, retrieved: 20.10.2015.

[7] M. Chaput, *Whoosh*, https://bitbucket.org/mchaput/whoosh/wiki, retrieved: 20.10.2016.

[8] K. Hughes, *Swish-E*, http://swish-e.org/, retrieved: 20.10.2015.

[9] I. Atanassova, and M. Bertin, Faceted Semantic Search for Scientific Papers. *PLoS Biology*, *2*, pp.426-522.

[10] A. L. Cuff, I. Sillitoe, T. Lewis, A. B. Clegg, R. Rentzsch, N. Furnham M. Pellegrini-Calace, D. Jones, J. Thornton, and C. A. Orengo, 2011. Extending CATH: increasing coverage of the protein structure universe and linking structure with function. *Nucleic acids research*, *39*(suppl 1), pp.D420-D426.

[11] Apache Wiki, *PublicServers*, https://wiki.apache.org/solr /PublicServers, retrieved: 20.10.2015.

[12] O. Kononenko, O. Baysal, R. Holmes, and M.W. Godfrey, 2014, May. Mining modern repositories with elasticsearch. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, pp.328-331.

[13] J. Thompson, A. Hankinson, and I. Fujinaga, 2011. Searching the Liber Usualis: Using COUCHDB and ELASTICSEARCH to query graphical music documents. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*.

[14] G. Horanyi, *Needle in a haystack*, https://medium.com/@ghoranyi /needle-in-a-haystack-873c97a99983#.autnzq6bt, retrieved: 20.10.2015.

[15] T. Pease, *A Whole New Code Search*, https://github.com/blog /1381-a-whole-new-code-search, retrieved: 20.10.2015.

[16] N. Carver, *What it takes to run Stack Overflow*, http://nickcraver.com/blog/2013/11/22/what-it-takes-to-run-stack-overflow/, retrieved: 20.10.2015.

[17] P. Alves, *Firefox 4, Twitter and NoSQL Elasticsearch*, http://pedroalves-bi.blogspot.rs/2011/03/firefox-4-twitter-and-nosql.html, retrieved: 20.10.2015.

[18] H. Karau, A. Alix, *foursquares now uses Elasticsearch*, http://engineering.foursquare.com/2012/08/09/foursquare-now-uses-elastic-search-and-on-a-related-note-slashem-also-works-with-elastic-search/, retrieved: 20.10.2015.

[19] K. Tan, *Apache Solr vs Elasticsearch - The Feature Smackdown*, http://solr-vs-elasticsearch.com/, retrieved: 20.10.2015.

[20] O. Gospodnetić, *Solr or Elasticsearch – that is the question*, http://www.datanami.com/2015/01/22/solr-elasticsearch-question/, retrieved: 20.10.2015.

[21] R. Sonnek, *Realtime Search: Solr vs Elasticsearch*, http://blog.socialcast.com/realtime-search-solr-vs-elasticsearch/, retrieved: 20.10.2015.

[22] A. Rafalovitch, *Solr vs. Elasticsearch – Case by Case*, http://www.slideshare.net/arafalov/solr-vs-elasticsearch-case-by-case, retrieved: 20.10.2015.

[23] Apache Software Foundation, *Lucene*, https://lucene.apache.org, retrieved: 20.10.2015.

[24] C. D. Manning, P. Raghavan, and H. Schütze, 2008. *Introduction to information retrieval*, Cambridge: Cambridge university press.

[25] Apache Software Foundation, *Apache ZooKeeper*, https://zookeeper.apache.org/, retrieved: 20.10.2015.

[26] Apache Software Foundation, *Apache Tika*, https://tika.apache.org/, retrieved: 20.10.2015.

[27] *Solr HighlightingParameters*, https://wiki.apache.org/solr /HighlightingParameters, retrieved: 20.10.2015.

[28] *Elasticsearch Highlighting*, https://www.elastic.co/guide/en/ elasticsearch/reference/current/search-request-highlighting.html, retrieved: 20.10.2015.

APPLYING SEO TECHNIQUES TO IMPROVE ACCESS TO A RESEARCH PROJECT WEBSITE

Baro Silvija*, Ivanović Dragan**
* Levi9 IT Services, Novi Sad, Serbia
** University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
silvija.baro@gmail.com, dragan.ivanovic@uns.ac.rs

***Abstract* — The paper describes the optimization of the research project DOSIRD UNS web site using Search Engine Optimization (SEO) techniques. The aim of this study was to better rank the website on the leading web search engines Google and Yahoo! using keywords that belong to the domain of the project. The website is developed using Drupal CMS version 7.35. On-Page, Off-Page SEO techniques and Drupal modules are used in order to achieve the better web site ranking.**

I. Introduction

SEO is a methodology of strategies, techniques and tactics used to increase the amount of visitors to a website in such a way that the website will get a high ranking on the result page of a search engine. SEO is a subfield of information retrieval which deals with techniques for the representation, storage, organization, access and retrieval of information [1; 11]. If the website is better ranked in the primary results there is a bigger chance that the user will visit that website. In this way, the discoverability of the information on the website is increased. The impact of search engine optimization on online advertising market is analyzed in the paper [12]. How scholarly literature repositories should be optimized to be better ranked on the Google Scholar search engine is the main topic of the paper [13]. Analysing Google rankings through search engine optimization data is presented in the paper [14].

The aim of presented study was to better rank the website DOSIRD UNS on the leading search engines Google and Yahoo! using keywords that belong to the domain of the project. In order to achieve the aim, On-page and Off-page SEO techniques are applied and Drupal modules are used. Google Webmaster [6], Google Analytics [7], Bing Webmaster Tools [8], HubSpot [9] and AuthorityLabs [10] are used for the purpose of tracking progress.

The structure of the paper is the following one. In the second chapter the website that has been optimized with SEO techniques is presented. The third chapter describes the importance of using SEO techniques, the necessary steps before effectively applying SEO techniques, Drupal modules that are used and SEO techniques used in order to achieve a better web site ranking. The methodology used for this study is presented in the fourth chapter, and results of this study are presented in the fifth chapter. The overview of the changes and optimization through the months April, May, June and July are presented, as are the position of the keywords on the search engines Google and Yahoo! during this period. In addition, significant data which was collected with the help of SEO tools are shown. The conclusion is given in the last chapter and it contains additional guidelines for further development.

II. DOSIRD UNS web site

The University of Novi Sad was founded on 28th of June 1960 and it represents an autonomous institution for education, science and arts. While software infrastructure for educational domain of the University of Novi Sad is well developed, there is lack of software infrastructure for science and arts. The main goal of DOSIRD UNS (Development Of Software Infrastructure for Research Domain of the University of Novi Sad) project is the development of software infrastructure for the research domain of the University of Novi Sad. The project has been started in the year 2009.

The mission of this project is the development of software infrastructure which will fulfill all local requirements prescribed by the University, government of the Autonomous province of Vojvodina and government of the Republic of Serbia. Besides that, the software infrastructure should be implemented in accordance with international well known standards and protocols belonging to the field of research domain. The goals of this project include: Development of a research information system (CRIS UNS), Development of a digital library of Ph.D. dissertations defended at the University of Novi Sad (PHD UNS), Development of an institutional repository of scientific research outputs of the University of Novi Sad, Export data to the following networks of digital libraries and institutional repositories: DARTEurope , OATD, OpenAIRE+ and Implementation of searching data through a web page and through a standardized protocol for Search/Retrieve via URL (SRU).

The web site *DOSIRD UNS* is developed using Drupal CMS (Content management System) [2], version 7.35. MySQL was used as DBMS [3] version 5.6.24. Apache [4] is used as the server, version 2.2.15 and PHP [5] version 5.3.3. The web site is available on the address http://dosird.uns.ac.rs and the homepage is shown in *Figure 1.*

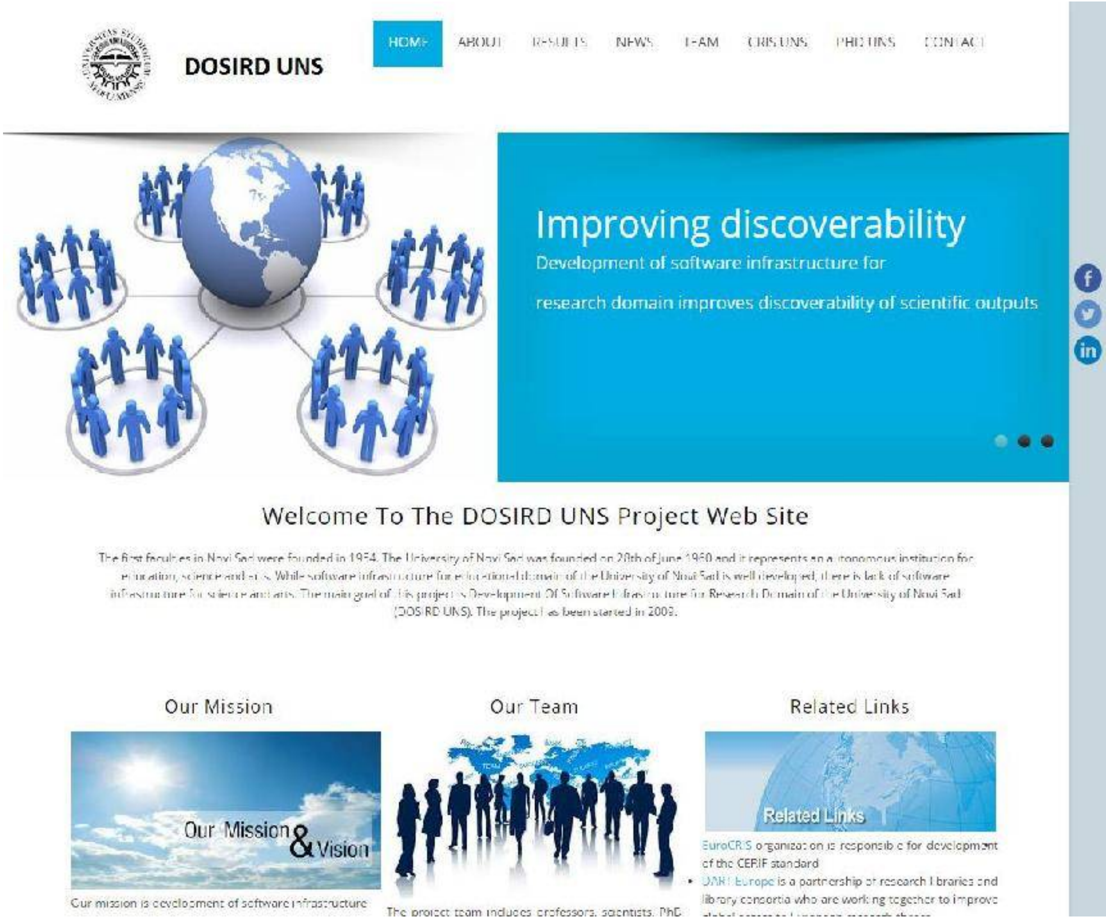


Figure 1 DOSIRD UNS web site

III. BACKGROUND

This section describes the importance of using SEO techniques, the necessary steps before effectively applying SEO techniques and the Drupal modules and SEO techniques used in order to achieve a better web site ranking by web search engines.

A. *The importance of using SEO techniques*

1. Most users look only at the first 5 results, therefore, it is crucial for the website to rank good.

2. SEO techniques adopt to search engines and users too; therefore usage of SEO techniques also improves user experience and the usability of the website.

3. Users trust search engines, so if a website is ranked among the first results this can increase the trust in the website and its content and services too.

4. SEO is beneficial for promoting a website. If a user finds a website via a search engine, and if he/she likes the content on the website, he/she will most likely share it via social networks.

5. SEO can help a website to be in a better position than their competitors. If two websites are selling the same product, the website that is more optimized is more likely to have a larger number of visitors and thus to increase sales of products.

6. With the help of various tools, such as Google Analytics, it is possible to conclude what type of users are visiting a website, which are the most popular pages and searched keywords and this information can be later used to improve the marketing strategy and to better create new content.

B. *Steps before applying SEO techniques*

1. **Identification of target audience**: Target audience is defined as a group of people who are interested in the content, product, or service that a website offers. Identification involves determining their sex, age, geographical location, financial status, education and most importantly- their needs: What do they want? What they don't have?

2. **Analysis of competing websites**: Finding out who the competitors are and which keywords they bid for. Collecting valuable information such as the competitor visibility on the search engine, the amount of their backlinks and traffic they get, the number of users on social networks and the content in their Meta tags.

3. **Research and discovery of keywords**: Finding out which words best describe the central theme of the website and its content and take into account the findings from steps 1 and 2. Create phrases and find synonyms. Use the keywords throughout SEO optimization.

C. *Drupal SEO modules*

Drupal offers a lot of possibilities to optimize a website for SEO. Modules used for this purpose are: Path Auto, Page Title, SEO Checklist module, Path and Global Redirect, Search 404, Metatag, XML Sitemap, Alchemy,

HubSpot and Google Analytics module, Clean URL, File Cache, Content Optimizer and SEO Compliance Checker.

D. On-page SEO optimization

On-Page SEO optimization refers to techniques applied directly on the website in order to improve its rankings on the search engine. On-Page SEO optimization can be grouped in three parts:

1. **Code optimization**: Adding and modifying the title of the website, Meta tags, alt tags, heading tags, XML sitemap and robots.txt file and optimizing the website speed.
2. **Content optimization**: Creating content which is rich with relevant keywords, density and allocation analysis of chosen keywords, avoiding content that cannot be indexed, content update frequency and content optimization for mobile devices.
3. **Link structure optimization**: URL structure and length, internal and external links, breadcrumb trail and 404 redirection.

E. Off-page SEO optimization

Unlike On-Page optimization, which is based on changing content, code and links of a website, Off-Page optimization is not visible on the website. Off-Page techniques are applied after On-Page optimization. Off-Page SEO optimization can be divided into three parts:

1. **Link building**: Outbound link quality, PageRank and Authority of the linking page, content relevancy of the linking page, anchor text.
2. **Creating and promoting content**: Social media, RSS feed, creating and updating a blog, submitting the website to directories and search engines.
3. **Website reputation**: Trust rank of the website, domain and page authority.

IV. Methodology

The experiment lasted for four months. The first two months, from April till May, were dedicated for creating the DOSIRD UNS website and applying SEO techniques. June and July were used to observe the data. The last measured progress was on July 31st.

A. Changes review

April - At the beginning of the month, with the help of Drupal CMS, the website was created. After that, the website was submitted to Google Webmaster, Google Analytics and Bing Webmaster Tools, and later HubSpot and AuthorityLabs tools were included. The XML sitemap was submitted to Google Webmaster and Bing Webmaster and the website was indexed. In the second half of the month SEO techniques were applied, first - On-Page Optimization. Keyword analysis was done. The most important keywords are: *CRIS*, *University of Novi Sad*, *digital library*, *PhD dissertation*, *scientific results*. Based on the keywords the Alt tags on the pictures, the Title tag, important Meta tags and the content were adapted and changed. Drupal modules, which were mentioned in chapter 3, were used in order to better optimize the website for this CMS. Social profiles were created and content was periodically published taking in account the content update frequency factor. Anchor texts were adjusted so that they carried semantics and contained keywords. Clean URL was used for better URL structure. The internal link structure was optimized so that the crawler can access every page. In the theme the Breadcrumb trail was activated allowing easier navigation and module Search 404 was used for eventual 404 situations, improving the user experience.

May – The speed of the website was tested with Google Page Speed Insights, and optimized using Leverage Browser Caching, Specify a Vary: Accept-Encoding header and FileCache module. Robots.txt file was adjusted by adding new rules. After On-Page optimization, Off-Page optimization was started. The website was referenced on the website of Chair of Informatics, Faculty of Technical Sciences, on the website of the University of Novi Sad, on the Drupal website and on the CRIS UNS website. Using Google Link Checker the links on the website were checked and adjusted to use less redirects and the broken links were fixed too. With the tool HTML checker, the code of the website was adjusted and the issues were fixed. RSS Feed was added which was linked on the News page.

June and July – During this period new content, in the form of dissertations from the University of Novi Sad, were periodically placed on the News page and that content was then promoted on social networks. The theme of the website was adjusted for mobile devices using the responsive theme approach.

V. Results

This section presents achieved results by the application of SEO techniques.

A. Keywords positions

To track the keyword positions a web browser was used with a cleared history and an anonymous profile (The user is not logged in and there are no cookies) so that the results can be as objective as possible. HubSpot and AuthorityLabs tools were used for tracking the keyword positions. Since there are a lot of variations only a few combinations of the keywords were selected that will be monitored. For Google search engine (*Table 1*) the measuring was performed three times per month (May, June, July), for each month on the first day, approximately in the middle and at the end of the month. For Yahoo! search engine (*Table 2*) the measuring was performed twice every month, once at the beginning of the month and once at the end. Both Google and Yahoo! show 10 results per page.

From the tables it can be concluded that the rankings were drastically improved during May and later on there were only minor fluctuations for one or two positions, which is normal. The keyword combination with the most competition was “PhD dissertation digital library”, especially on the Google search engine, where it achieved the worst ranking, taking into account the rest of the results. Unlike this, on the Yahoo! search engine these keywords ranked in second place. This can be explained by the fact that on Google there are more websites competing for these keywords then on Yahoo! and that these two search engines use different algorithms for ranking the results. Yahoo! Algorithm is more focused on analyzing the content of a website, while Googles algorithm is more complex, and it takes into account more factors like link popularity, social signals, content relevance and quality.

Table 1 The ranking results of the keywords on the Google search engine

Keyword/Date	1.5	21.5	1.6	19.6	30.6	1.7	31.7
university of novi sad phd dissertation	3rd page	2	1	1	1	2	1
university of novi sad digital library	2nd page	2	1	1	1	1	1
phd dissertation cris	2nd page	1	1	2	1	1	1
phd dissertation digital library	5th page	43	22	20	17	17	41
university of novi sad digital library cris	2	1	2	1	2	2	1
university of novi sad phd dissertation cris	3	2	1	1	1	2	1
university of novi sad digital library scientific results	1	1	1	1	1	1	1
phd dissertation digital library cris	2nd page	2	2	1	1	2	1

Table 2 The ranking results of the keywords on the Yahoo! search engine

Keyword/Date	1.5	21.5	1.6	30.6	1.7	31.7
university of novi sad phd dissertation	1	2	1	1	3	1
university of novi sad digital library	4	1	1	2	2	2
phd dissertation cris	4	2	2	1	2	1
phd dissertation digital library	10	2	3	2	2	2
university of novi sad digital library cris	1	1	1	1	1	1
university of novi sad phd dissertation cris	5	2	2	1	1	1
university of novi sad digital library scientific results	1	1	1	1	1	1
phd dissertation digital library cris	8	5	3	2	3	1

B. SEO tools data

Table 3 shows how the users discovered DOSIRD UNS website between April and July.

Table 3 Acquisition

Parameter/Month	April	May	June	July	Whole period
Direct	78	94	96	35	303
Organic	40	43	42	22	147
Social	28	33	20	6	75
Referral	17	21	1	4	55
Total sessions	163	191	159	67	580

The dominant way was through direct traffic (when a user types the URL directly in the address bar), then organic (when a user searched for a query on the search engine and clicked on the website link which was listed in the results), then social (when users come to the website via social networks), and last, referral (when users came to the website because they clicked on a backlink). *Table 4* shows the total number of users who visited the website and the total number of pages viewed within sessions per month. A session is a group of interactions that take place on a website within a given time frame.

Table 4 Number of users per month

Parameter/Month	April	May	June	July	Whole period
Users	69	98	105	50	322
No. of viewed pages	1.296	1.069	642	182	3.189

On the site http://tools.pingdom.com/fpt the speed of a website page can be checked. *Figure 2* shows the results of speed optimization for a page of DOSIRD UNS project.

The first score of 77/100 is without any optimization, measured on May 1st. A score of 87/100 je is achieved when Drupal caching and File Cache module were used, measured on May 7th. A score of 96/100 is achieved when using Leverage Browser Caching for images, CSS and JS and using GZip compression with the measure date on May 12. The other pages of the website have similar success and a significant increase in page speed loading.

Figure 2 The speed of DOSIRD UNS website

On the site http://www.seoreviewtools.com/website-authority-checker/ the authority of a website and a page can be checked. *Figure 3* shows the authority of the domain *dosird.uns.ac.rs* and the page authority of CRIS UNS page measured on May 5th. As it can be seen, the page authority is only 1 and the domain authority is 58. *Figure 4* shows the same test measured on July 31st. One can see an increase, with the page authority now 30 and the domain authority now 61. It should be kept in mind that the domain authority has a much slower increase.

Figure 3 Domain and page authority on May 5th

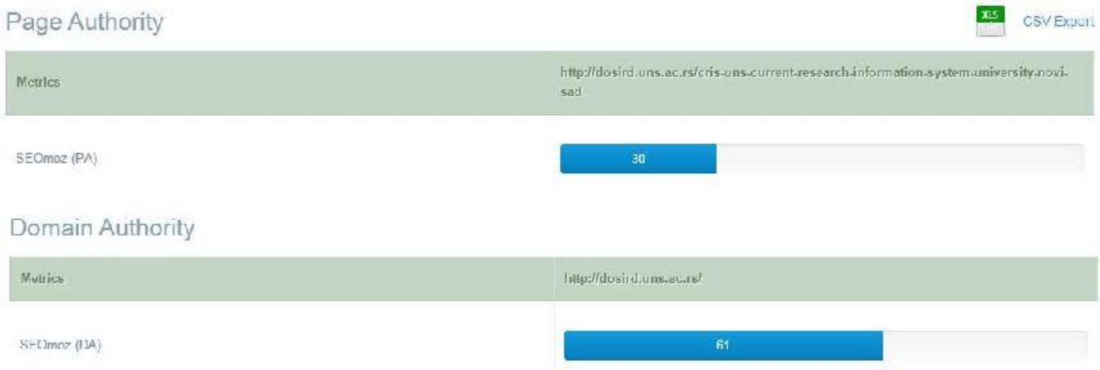


Figure 4 Domain and page authority on July 31st

VI. CONCLUSION

With many changes in the search engine algorithm, today the most important thing is creating interesting, informative and high quality content that users will want to read and share. Search engines want the most adequate and relevant content for the best ranking, and because of this, SEO has become more oriented towards optimizing content for the users and less for the search engines. How the algorithm of a search engine changes so does SEO change and adapt. SEO is a long term process. A website cannot be optimized and leaved as it is. This can be seen from the results, as the number of user sessions declined slowly if there wasn't created new, interesting content that will attract new users and motivate old ones to come back. SEO can make a huge difference, especially if the website is intended to sell products or services and there is high competition.

Further improvements should include the following:

1. Off-Page SEO optimization must be worked on, especially it should be put an emphasis on building links from relevant sources.
2. A separate blog should be created where new and interesting content will be created, rather than the creation of new content relying only on the News page.
3. When a quality blog is created, this would improve the aspects of social media too, which also needs more work.
4. It would be useful to add a forum to the DOSIRD UNS website where users could discuss the dissertations. This would allow a creation of a community of users, and most importantly, an exchange of knowledge and experience.
5. The theme should be changed to have a more modern design. Today One Page websites are popular, where all the key information like About Us, Team, News, statistics are shown on one big scrollable page.

REFERENCES

[1] Ivanović, Dragan & Milosavljević, Branko. *Upravljanje digitalnim dokumentima*. Univerzitet u Novom Sadu, 2015.

[2] Drupal CMS, https://www.drupal.org/

[3] MySQL, https://www.mysql.com/

[4] Apache server, http://httpd.apache.org/

[5] PHP, https://secure.php.net/

[6] Google Webmaster, https://www.google.com/webmasters/

[7] Google Analytics, http://www.google.com/analytics/

[8] Bing Webmaster tools, http://www.bing.com/toolbox/webmaster

[9] HubSpot, http://www.hubspot.com/

[10] AuthorityLabs, http://authoritylabs.com/

[11] Kent, Peter. *Search engine optimization for dummies*. John Wiley & Sons, 2012.

[12] Xing, Bo, and Zhangxi Lin. The impact of search engine optimization on online advertising market. In *Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet*, pp. 519-529. ACM, 2006.

[13] Beel, Jöran, Bela Gipp, and Erik Wilde. Academic Search Engine Optimization (aseo) Optimizing Scholarly Literature for Google Scholar & Co. *Journal of scholarly publishing* 41, no. 2 (2009): 176-190.

[14] Evans, Michael P. Analysing Google rankings through search engine optimization data. *Internet research* 17, no. 1 (2007): 21-37.

IMPLEMENTATION OF BOOKS DIGITAL REPOSITORY SEARCHING USING HIBERNATE SEARCH SOFTWARE LIBRARY

Šnajderov Jovan*, Ivanović Dragan**
* Levi9 IT Services, Novi Sad, Serbia
** University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
jovansna@gmail.com, dragan.ivanovic@uns.ac.rs

***Abstract*— A web application for searching through digital book repository is implemented. Server application indexes and persists book using Hibernate Search software library, exposing it's functionality through REST services. Client application developed using AngularJS, calls these services through web browser.**

I. INTRODUCTION

It was the month of March 1989, when British engineer Tim Berners-Lee finished his proposal project, for a system that would allow information flow between researchers at CERN physics department. Not long after that, Tim also developed first web browser, a tool for accessing web pages, written in format that would allow them to contain text, images, videos, and software components. This web browser made the communication with first internet server – possible. Date when the first internet web server at CERN became available over internet, 25th of December 1990, started a new era of global information.

Creating web, brought along new challenges. Amount of information that was enormous even before this global interconnecting, only kept growing with an unstoppable pace. Main challenge was extracting information from this new network. People searched for information on the web using links, mainly starting from hand created indexes from sites like Yahoo. These manually maintained and created lists covered the important topics efficiently, but subjectively. They were extremely expensive to create and maintain, and did not show any signs of improving or speeding up. Automatic search engines relied heavily on key words and most often presented too many bad matches. Until one day in 1996, two engineers from Stanford University presented a prototype of new search engine, with enormous scalability that relied heavily on text structure on web pages. This project started a revolution and changed the face of search, as we knew it.

Scale of this revolution is visible today, as well. Search is a key component of our digital lives (Google, Facebook, Amazon…). On every web page, in every application, we expect highly intelligent data search.

In this paper, we present a digital books repository implemented as modern web application, which in its core contains software tool that allows for advanced data search and low level data indexing using the Hibernate Search library.

Point of this paper is to present use of Hibernate Search Java library as well as to highlight advances in development of search tools, through a real world example. Quality of search and search results that we've got accustomed to, using modern search system, is now available for easier implementation in our own and specific applications.

II. BACKGROUND

Almost all applications require data persistence. If the information system would not persist data, all of the data would be lost on system shut down and with it, any practical use would be lost. When it comes to Java persistence, it usually relates to saving data into relational data bases using SQL. This process usually consists of large amount of repetitive boilerplate Java code. However, there is a solution that tackles this issue from Java code's perspective - object relational mapping.

A. Hibernate

Object relational mapping (ORM) is automatized and transparent persisting of objects into tables of relational databases, using metadata that describe mapping between these objects and the database [1; 2]. Hibernate is a complete ORM tool created as independent, non-commercial, open-source project from which the Java persistence specification came to be. This Java persistence API describes data access, data persistence, and data management between Java objects/classes and relational databases. Data extraction from relational databases should also allow end user to search for information. By that, we mean process of search of documents, search of information inside documents, and metadata about documents.

B. Lucene

Apache Lucene is very powerful and widespread tool for information search [3], but if we try to integrate it into existing software solutions, we quickly come across its flaws. Problems that Lucene faces are the following mismatches:

1. Structural mismatch – converting object domain into index in text form; dealing with connections between objects in index.
2. Synchronous mismatch – how to maintain database and index synchronized the entire time.
3. Retrieval mismatch – how to sustain smooth integration between domain models oriented methods for data retrieval and full text search.

In light of this mismatches, another project emerged, Hibernate Search.

C. Hibernate Search

Hibernate Search is Java library that represents integration of Hibernate ORM solution and Lucene [4; 5]. It is a project complementary to Hibernate ORM library, allowing full text search using queries directly over persistence domain model of the data. Technology that lies beneath Hibernate Search project relies completely on Apache Lucene. It hides complex usage of Lucene API, using simplified controls for advanced Lucene functionality, and allows us to index and retrieve persistence domain data from Hibernate, with minimal effort. The Hibernate Search software library has been previously used in various projects [6; 7].

III. System specification

Primary use cases necessary to implement in the digital library were:

1. Adding a book – adding book metadata and the book content itself.
2. Updating a book – updating existing data and metadata related for any and every book.
3. Book search – advanced book search using specter of different search parameters (metadata or the contents themselves).
4. Reviewing a book – inspecting data of existing books (look at the metadata and possibility of book download).
5. User subsystem – access to repository user records and tracking access permissions to different parts of the system.

IV. System architecture and implementation

System architecture is shown in Figure 1. System consists of server application and client application. Server application exposes services to handle books, implements business logic, enables user manipulation, persists data and it indexes books. Data is persisted in PostgreSQL database that we access from inside of the server application over the ORM tool (Hibernate). We index books (metadata and e-book contents) using Hibernate Search, and we keep its index on a server's file system. Main server application is available over application server that exposes interaction with available services over http protocol.

Client application uses these exposed server side REST services to allow users to add a new book, to search for a book, to update book metadata, to create user account, etc. Core of the project, simplicity of adding advanced search capabilities will be presented on the main data model, the EBook entity shown on Listing 1.

To enable full text capabilities over our entities, we first need to add couple of annotations. First, we annotate that the entity should be indexed using Hibernate Search @Indexed annotation. Second, we define entity identifier that Hibernate Search saves in its index for every entity. For this purpose we set @Id annotation on identification attribute of our class and our primary key in eBook database table.

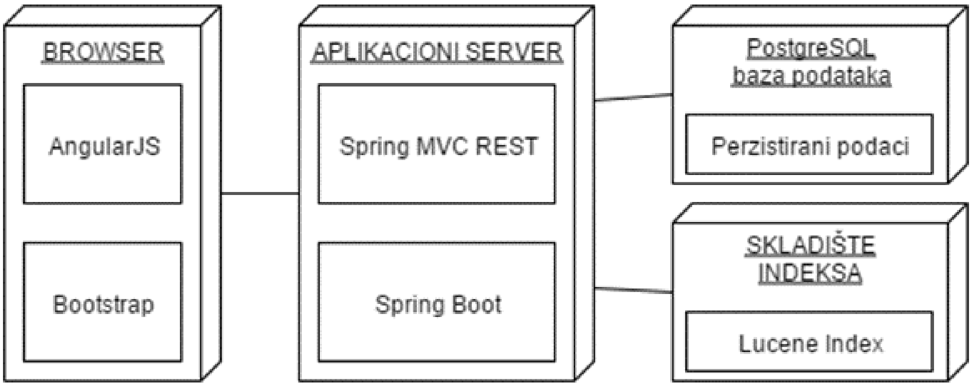


Figure 1 Deployment diagram

Next step is to determine which fields we want be able to search in our digital repository. Annotation that we use for this is @Field. Annotation parameters that we can set are as follows. Parameter index = Index.YES assures that the text will be indexed. Parameter analyze = Analyze.YES assures that the text will be analyzed using default Lucene analyzer. Third parameter store = Store.YES assures that the data themselves won't be saved in index. Saving attribute data inside index does not have anything to do with the ability to search them. Advantage that this parameter enables is ability to retrieve attributes value over projections. When we do not use projections, Hibernate Search uses Lucene queries to find entity identifiers in the database, so it could use them to retrieve managed objects from the database. Using projections, we could avoid this trip to the database and back, but not without cost: this way we would get array of objects and not managed object that we get with regular query. One option allows better performance, while the other enables ease in development.

```java
@Entity
@AnalyzerDef(name = "customanalyzer", tokenizer =
@TokenizerDef(factory = StandardTokenizerFactory.class),
filters = {
    @TokenFilterDef(factory = LowerCaseFilterFactory.class),
    @TokenFilterDef(factory =
SnowballPorterFilterFactory.class, params = {@Parameter(name
= "language", value = "English")})})
@Indexed
@Table(name = "EBOOK")
@Cache(usage =
CacheConcurrencyStrategy.NONSTRICT_READ_WRITE)
public class Ebook implements Serializable {

    @Id
    @GeneratedValue(strategy = GenerationType.AUTO)
    private Long id;

    @Field(index = Index.YES, analyze = Analyze.YES, store =
Store.NO)
    @Column(name = "title")
    @Analyzer(definition = "customanalyzer")
    private String title;

    @Field(index = Index.YES, analyze = Analyze.YES, store =
Store.NO)
    @NotNull
    @Column(name = "creator", nullable = false)
    private String creator;

    @Field(index = Index.YES, analyze = Analyze.YES, store =
Store.NO)
    @Column(name = "description")
    @Analyzer(definition = "customanalyzer")
    private String description;

    @Type(type =
"org.jadira.usertype.dateandtime.joda.PersistentLocalDate")
    @JsonSerialize(using = CustomLocalDateSerializer.class)
    @JsonDeserialize(using =
ISO8601LocalDateDeserializer.class)
    @Column(name = "date", nullable = false)
    private LocalDate date;

    @Field(analyze = Analyze.YES, store = Store.YES)
    @TikaBridge
    @Analyzer(definition = "customanalyzer")
    private String content;
```

Listing 1. The Ebook entity

Advanced search options over our entities are achieved using more advanced annotation parameters. Let us assume that we indexed the book entity with the title “Hibernate Searching” and that we want to have this book as a matched result even if we search for key words “search”, “searches”, “searched” or “searching”. To achieve this using Lucene; we need to set analyzer class just like in Listing 1. This analyzer sets word stemming while indexing and during search process. In our class, we use StandardTokenizerFactory that splits the words at punctuation marks and dashes, but maintains the format of the email addresses it comes across, as does with the internet domains. On this analyzer we chained two filters. First one converts all the letters to lower case (LowerCaseFilterFactory). Second filter adds specific grammar rules and syntax for the English language.

@TikaBridge annotation uses Apache Tika tool to allow for extracting text and metadata of the passed documents. This annotation uses attributes with type String, URI, byte[] or java.sql.Blob. Using String typed attributes, this tool recognizes values of attributes as file locations and tries to open them to parse the document saved at that location. During the eBook upload, we actually save the book on the server’s file system, and repository only grabs a hold of the file location address inside the String content attribute. After persisting the eBook object, Hibernate Search reads location address from the content attribute, finds the book on file system, parses document contents, indexes text contents and updates the index for that particular book. We only have to take care that the right book contents path is present in the content attribute. Figure 2 depicts the form for input book metadata and uploading a digital file which represents the certain book.

Figure 3 depicts the look of the digital repository search page. Using this page, we call the method:

```java
public List<Ebook> search(EbookSearchObject ebookSearchObject)
```

We pass the search object to this method. This method takes this object and parses its list of search criteria. Using these criteria, we get Lucene query suitable for searching Lucene index:

```java
luceneQuery = MultiFieldQueryParser.parse(paramValuesArray,
paramNamesArray, flagsArray, new StopAnalyzer());
```

Given query, we wrap inside Hibernate query:

```java
org.hibernate.search.jpa.FullTextQuery jpaQuery =
fullTextEntityManager.createFullTextQuery(luceneQuery,Ebook.
class);
```

Execution of previous jpaQuery retrieves the list of results.

We just showed how, using just a couple of methods, we raised the whole complexity of the entire advanced search on the higher level of abstraction. With this, we saved ourselves from writing great amount of boilerplate code, leaving that part of the job, to the specialized library – Hibernate Search.

V. Conclusion

Quality data search is necessary for managing almost endless ocean of information that is out there. To average user of the global network, it would be beyond imagination to retrieve any usable valued information if it is not available through simple search and retrieval system. This complete dependency on our search tool shows just how clear goal of further advances in this field is. Every meaningful advance in this field of computer science represents a gold mine of its own kind. Gold mine that could almost the entire world population is waiting to use regularly. This claim is backed by the value of the company behind the most influential Internet search engine - Google.

Search technology presented in this paper, using Hibernate Search library, represents simple all-round solution for quick and efficient enabling of advanced search capabilities to any system. Tools like this software library, open the door for new researches, and create new possibilities for engineers looking to push the envelope in the field of search.

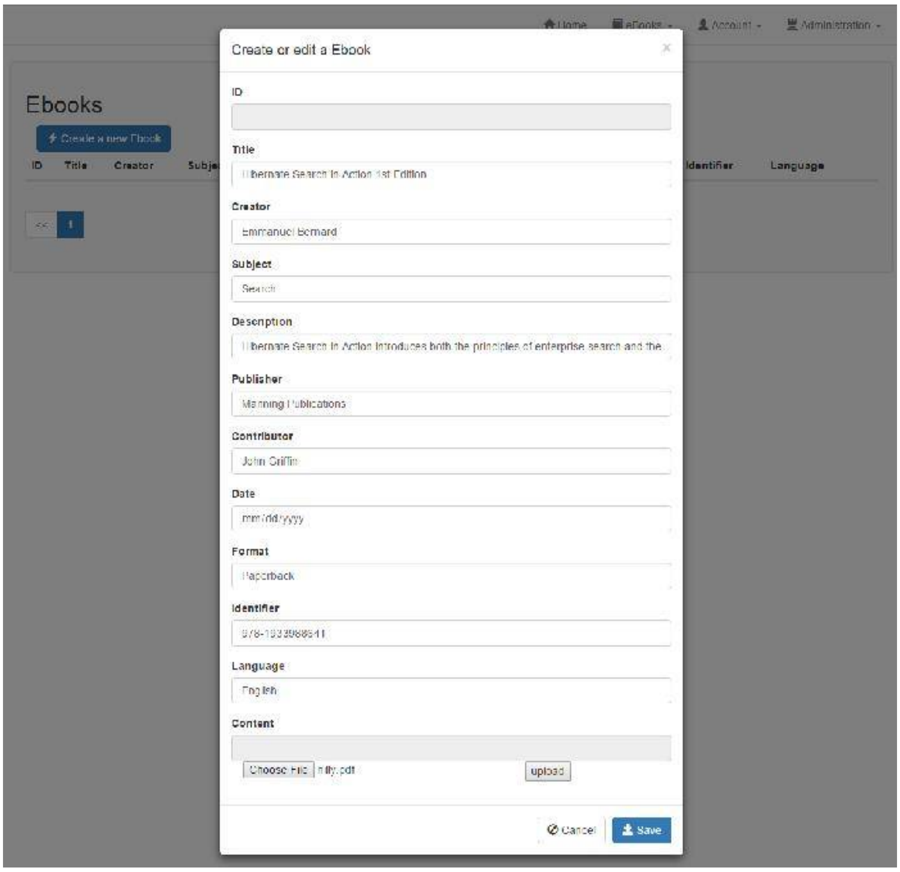


Figure 2 Dialog for input new book

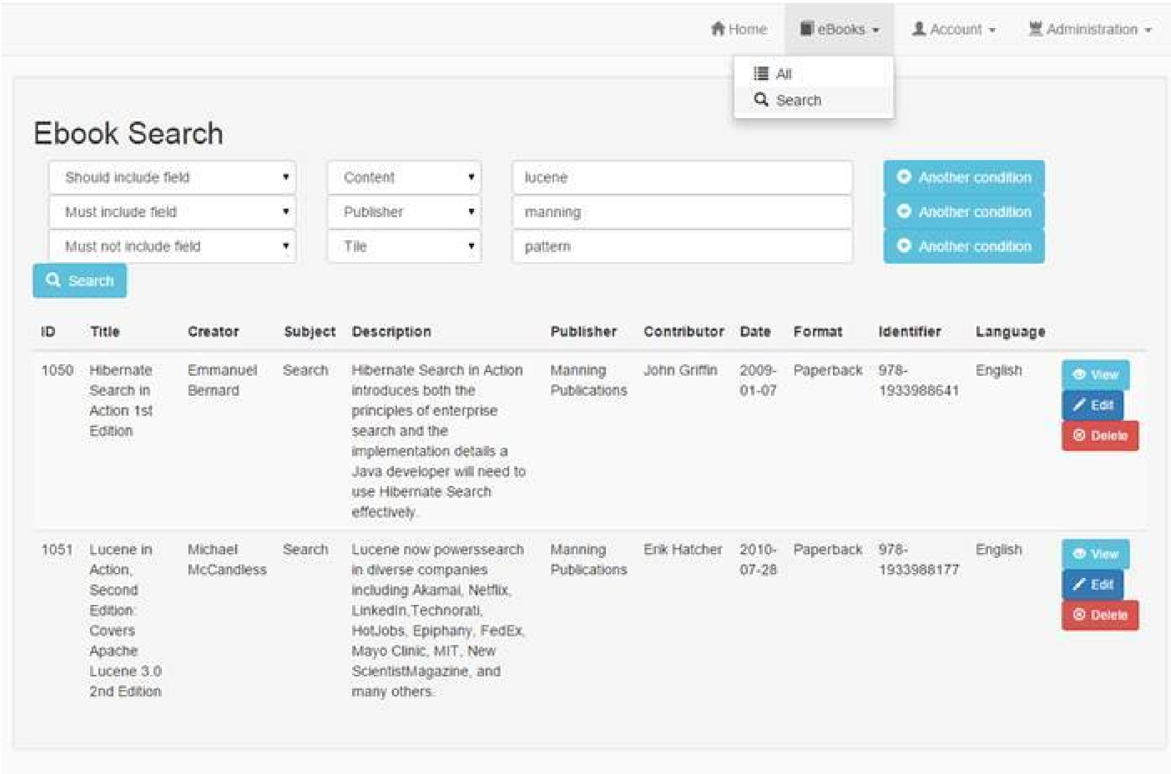


Figure 3 The repository search page

REFERENCES

[1] C. Bauer, G. King, G. Gregory, *Java Persistence, 2nd edition*. Manning Publication, 31. oktobar 2015.; ISBN 978-1617290459

[2] C. Bauer, G. King. *Hibernate in Action*. Manning Publication, 28. avgust 2004.; ISBN 978-1932394153

[3] E. Hatcher, O. Gospodnetić, M. McCandless, Lucene in Action, 2nd edition. Manning Publication, 28. jul 2010.; ISBN 978-1933988177

[4] E. Bernard, J. Griffin, *Hibernate Search in Action, 1st edition*, Manning Publication, 7. januar 2009.; ISBN 978-1933988641

[5] S. Perkins, *Hibernate Search by Example*, Packt *Publishing, 26. mart 2013.; ISBN 978-1849519205*

[6] K.M. Anderson, A., Schram, *Design and implementation of a data analytics infrastructure in support of crisis informatics research (NIER track)*. In *Proceedings of the 33rd International Conference on Software Engineering* (pp. 844-847). ACM. 2011.

[7] Y.A.N.G., Qi, L.I.N., Zhen-can, H.U.A.N.G., Fan, X.I., Jian-qing, *Database Full-Text Retrieval System Based on Hibernate Search*. *Computer Engineering*, *4*, p.029, 2010.

Science Network of Montenegro: Open government eService based on Open data and Open standards

mr Darko Petrušić, dr Zora Konjović, dr Milan Segedinac

Ministry of Science, Podgorica, Montenegro
Singidunum University, Belgrade, Serbia
University of Novi Sad, Novi Sad, Serbia

Abstract

Open movement is influenced by digital revolution and its technological foundation and socio-economic impact. Open data and Open standards are the integral components of Open government (OG) paradigm, together with Open architecture, playing an important role in the creation of OG e-services. One such OG e-service, that follows Open movement principles, is the *Science Network of Montenegro (SNM). SNM is an Open science type of application and also an Open government e-service, designed using Open architecture principles with an aim to publish Open data, following Open standards and using Open source software components. Its mission is to become a centralized virtual online platform for presenting and collecting information about researchers, research papers, research projects, equipment and scientific knowledge. In this paper we have presented the role that Open movements (with special focus on Open data and Open standards) play in Open government initiatives, and shown how Open data and Open standards combine in the case of Science Network of Montenegro.*

KEYWORDS: *Open data, Open standards, Open government, Open source, Open Science*

1 INTRODUCTION

Generally speaking, the term “Open” assumes freedom and independence from the potential influence and control of arbitrary power. Movements for “Open” strive for transparency, participation, empowerment of individuals and knowledge for all [1]. With Information revolution, rapid development of Internet and “Digital Economy”, social and technological foundation for enabling the vision of the “Open” movement was created. Open source, Open government and Open data are the most recognized concepts of “Open” movements alongside with Open access, Open science, Open innovation, Open education, Open knowledge, Open linked data and Open architecture. As the Internet became a fully open and global phenomenon paving the way to easily accessible digital technologies, governments now have the opportunity to transform their services by making them more efficient and more accessible to their users. Also, every government should be open and transparent with regard to its decisions. Furthermore, those decisions should be logical and easy to explain to citizens as well as evidence based i.e. data-driven. Many governments have recognized that information increases in value as it becomes shared. Therefore sharing and reusing data can reduce the effort in data production and it is introducing quality assurance process on published data by making them freely accessible. Publishing government’s data means opening opportunities to use data in new and innovative ways, thus creating economic value [2, 3] out of the published data. Since government is the major source of open data, open data provided by governments becomes Open government data (OGD). Open government data goes hand in hand with Open standards movement. The use of Open standards provides interoperability and enables open access to Open government data. Therefore the three main components of Open government are Open Data, Open Standards and Open Architecture [4]. The idea of Open Government is to establish cooperation between public administration, citizens and private sector by enabling transparency, participation and collaboration.
The aim of this research is to analyse the impact and present the role that Open movement (Open data, Open source, Open science and Open standards) plays in Open government initiatives and its services.

2 OPEN DATA

Open data (OD) is the concept by which data is made freely available to the public, and where ‘anyone is free to use, reuse, and redistribute it, without any legal, technological or social restriction’ [5, 6]. The term “Open data” refers to data beyond governmental institutions alone and includes those from other relevant stakeholder groups such as business, citizens, NGOs, science or

education. The goal of OD movement is to allow citizens and organizations outside government to use and re-use information extracted from data and/or combine it with other information in ways that provide added value to the public.
In 2007, thirty Open government advocates came together to develop a set of OGD principles that that would underscore the importance of OGD for the society (Fig. 1). Therefore, data is made public in a way

File Format	Recommendations
CSV	* * *
XLS	*
PDF	*
DOC	*
XML	* * * *
RDF	* * * * *
KML	* * * *
SHP	* * * *
ODS	* *
KMZ	* * * *
JSON	* * *
TXT	*
HTMl	*
TIFF	*
JPEG	*

Tab 1. Tim Berners-Lee's star scale for selected file formats [7]

that complies with the following principles [6]:

- Data must be complete,
- Data must be primary,
- Data must be timely,
- Data must be accessible,
- Data must be machine processable,
- Access must be non-discriminatory,
- Data formats must be non-proprietary, and
- Data must be license-free.

The first practical implementation of Open data initiative was launched in the USA in May 2009, and open data website (Data.gov) [8] aimed to improve access to data in machine-readable format. Following the successful launch of Data.gov, the United Kingdom's open data portal was launched in January 2010. Since then there have been many, internationally implemented, Open data initiatives. As a result of widespread recognition of the importance of Open data within context of Open government, a multilateral initiative Open Government Partnership was created in 2011 [9]. At the Gov 2.0 Expo in Washington DC in 2010, Sir Tim Berners-Lee presented his 5-Star Model [10] as a roadmap of how to move from Open (government) data to linked Open (government) data (LOD) (Tab. 1). LOD [11] facilitates knowledge creation from interlinked data. The idea of Open Data is developed as part of a social web, whereas the idea of Linked data is associated with the semantic web. Therefore, it can be noted that LOD brings these two major movements together. However our ability to fully utilize the value of government data depends on open standards [12] including data modeling standards, metadata

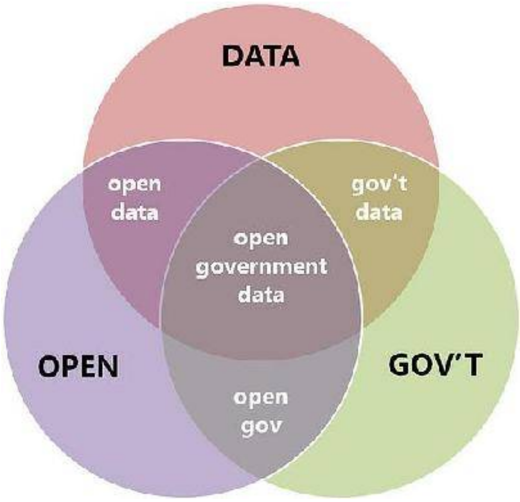


Fig 1. Open Government Data Venn diagram [17]

standards, standard for software and user interfaces etc.

3 OPEN STANDARDS

Even though the concept of "Open standard" does not have a universally accepted definition [13] we can accept the following one: "Open standards are transparent descriptions of data and processes that are enabling seamless exchange of information between different information systems".

In essence, Open standards provide the basis for interoperability, where interoperability means the ability of information and communication technology (ICT) systems and business processes they support to exchange data and enable information and knowledge sharing [14, 15]. Open standards can be implemented either by Open source or proprietary systems where there are no barriers in combining those systems, assuming that they use same Open standards [16]. Most of proprietary and Open source products have implemented Open standards to access Open data.

To reach interoperability in the context of pan-European e-Government services [18], the following are the minimal characteristics that an Open standard must have:

- The standard has to be adopted and maintained by a not-for-profit organisation, and its on-going development has to occur on the basis of an open decision-making procedure available to all interested parties (consensus or majority decision etc.).

- The standard has to be published and the related the standard specification document has to be available either for free or at a nominal charge. It must be permissible to all to copy, distribute and use it free of charge for no fee or at a nominal fee.

- Every intellectual property - i.e. patents possibly present within (parts of) the standard has to be made irrevocably available on a royalty-free basis.

- There must be no constraints to the re-use of the standard.

Open standards have provided support for the creation of many technological innovations. Even the Internet and the “www” are based on Open standards [13] such as TCP/IP, HTTP, HTML, CSS, XML, RDF, POP3 and MTP.

Open standards and Open source [18] are the key elements of an open concept which combines them and leverages openness while supporting interoperability, flexibility and choice.

The digital revolution has triggered a widespread use of technology providing citizens with continuous access to government services across all devices [19]. There is a growing trend in the use of Open source software for the creation of Open government service. Open source uses the same principles as Open government [20, 21] i.e. the code is transparent so that it can be used or tailored for a specific purpose. Open source software (OSS) is where the source code has been made available to licensed users, allowing them to tailor the software to their needs and make iterative improvements. The key benefits of Open source software are the promotion of innovation and cost saving but without compromising the usability or effectiveness.

4 Applying Open Data and Open Standards – The Case of Science Network of Montenegro

The main goal of the Open science movement [22] is to make scientific research data available and accessible to both scientific and non-scientific communities. Open science can therefore be observed as part of a wider Open government paradigm, considering the fact that most governments publicly support research programmes realizing that public sponsorships invested in research and development lead to economic and social development. The Open science concept [23] should adhere to a set of important principles such as: a) open access to research outputs; b) open access to the usage/impact statistics of research outputs; and c) open access to basic research assessment data accumulated for outputs, researchers and organizations.

In Montenegro it has been recognised that a practical approach to Open science and Open government should include the promotion of collaboration, participation, joint projects and joint use of the research infrastructure by developing a dedicated electronic service – SNM [3]. The second and enhanced release of the SNM project is a part of Higher Education and Research for Innovation and Competitiveness Project (HERIC) (www.heric.me) funded by Ministry of Science of Montenegro [24]. SNM is an interactive web portal that will enable scientists to exchange information, collaborate and find relevant information about research infrastructure in Montenegro. This information system was created under the strong influence of the “Openness” concept [25], therefore the concepts of the Open architecture, Open data

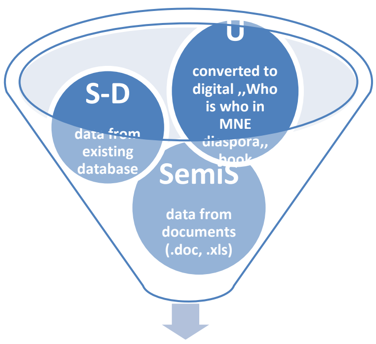


Fig 2. SNM data collection process

and Open standard were used where Open source software served as building block for the information system. SNM is a Current Research Information Systems (CRIS) (www.eurocris.org) and it represents an Open science type of application.

In addition to researchers and scientific research institutions, important users of the SNM portal at a national level will include: large, medium and small companies, entrepreneurs, agents from both public and private industry, local, regional and national governmental bodies and civil society. The Open data movement concerns datasets that are in a structured and tabular format. Therefore, we have stored our data in a structured way [26] having in mind the fact that we have dealt with three types of data: structured (from legacy database), unstructured (documents with data obtained from the book titled "Who is who from the Montenegrin diaspora") and semi-structured data from documents and spreadsheet files (Fig. 2). We have collected and converted all three kinds of data and then created SNM database which represents the core of the information system. It is important to mention that the database of the system was designed using entity-attribute–value (EAV) model [27].

SNM also uses CERIF (Common European Research Information Format) standard [28] as a model format for managing Research Information and it is used to connect OGD and data produced by scientific and research projects. The use of the CERIF format is an EU recommendation to its member states. The core CERIF entities are Person, Organisation Unit, Result Publication and Project.

Besides these four entities, in SNM we have included an additional entity representing the research infrastructure i.e. equipment. Part of the HERIC project includes a Study on the existing research equipment capacity and the creation of a joint research facility in Montenegro. Optimal use of research equipment is one of the greatest challenges faced by a majority of scientific groups.

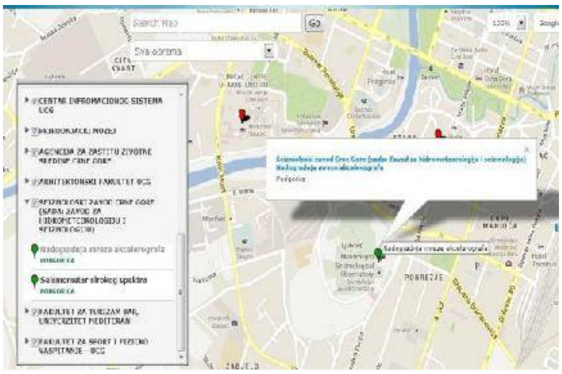


Fig 3. Pilot Google map presentation

Study on the existing research equipment capacity provides information about equipment in laboratories of scientific research institutions in Montenegro i.e. the documents and information about the relevant institution (address, e-mail, web) as well as detailed information about the equipment (purchase price, maintenance costs for the equipment, the condition of the equipment, how often it is used - weekly, monthly, daily, information about which areas of scientific equipment can use the primary function of that equipment. The collected data, as dataset, is presented in XML and CSV format in the first Montenegrin Open data web portal (http://www.open-data.me/) [7]. This data will be a part of our SNM database and it will be integrated with the rest of scientific data and visually presented using Google maps (Fig. 3). After its full implementation, SNM will become a trusted source of data related to equipment in laboratories of scientific research institutions and it will feed Montenegrin Open data web portal with up to date datasets. Once it is fully operational, SNM will help utilize and manage equipment effectively; research institutions will increase the quality and productivity, and reduce operational costs. Considering data formats presented so far and the fact that most of the information related to research papers and thesis will be presented in PDF format, we can have SNM star rated table (Tab. 2).

File format	Recommendation
PDF	*
Google map	*
CSV	**
XML	****

Tab 2. Open data formats in SNM

Furthermore, SNM uses a standardised Research and Development (R&D) classification, such as Frascati [29]. The Frascati Manual provides internationally accepted definitions of R&D and the classifications of its component activities. It is based on the experience gained from collecting R&D statistics in OECD member countries.

In addition SNM was designed according to standards set up by Open architecture with the aim of creating an Open government compliant application. It is important to mention that the software architecture of the system is based on DRUPAL-LAMP (www.drupal.org) [30] combination of Open source software components, or better known as DAMP. DRUPAL is an Open source Content Management System (CMS), whereas LAMP is the acronym of the names of its

original four Open source software components: the Linux operating system, the Apache HTTP Server, the MySQL relational database management system (RDBMS), and the PHP as the object-oriented scripting language. In this case we used the enhanced version of DAMP stack (Fig. 4) with a collection of the best known Open source tools to drive SNM web portal with focus on security, performance and functionality. Therefore to improve security features we added antivirus ClamAV, software firewall CSF and Apache ModSecurity modules.

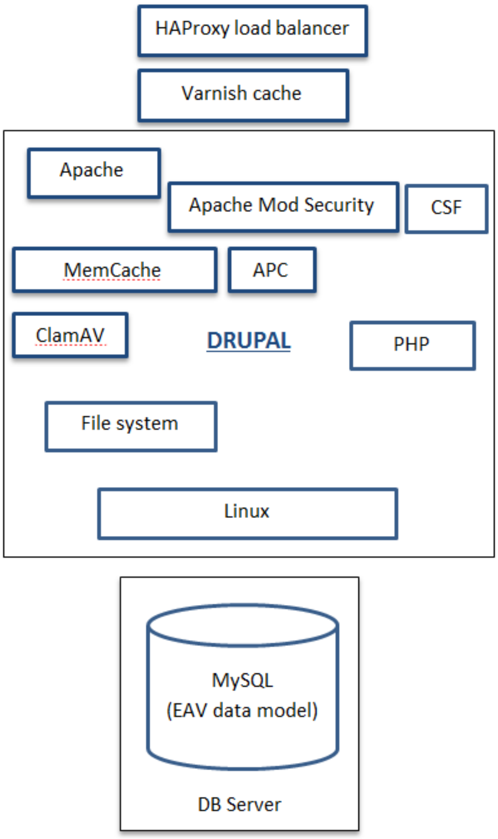


Fig 4. Enhanced DLAMP software architecture

To speed up the overall IS we used two complementary caching components APC and Memcache, whereas to improve the performance of SNM web portal we used Varnish which is an HTTP accelerator designed for content-heavy dynamic web sites. Furthermore to improve the reliability we used HAProxy, which stands for High Availability Proxy.

5 CONCLUSION AND FUTURE WORK

Open data and Open standards are the essential part of the Open government initiative. Combined with other Open movement initiatives, especially Open source initiative, they can lead to the development of powerful applications. Open source has proven to be a suitable choice for many Open government initiatives [31], due to the inherited use of Open standards, the rapid application development and ability to publish Open data. One such Open source platform, which is used for the development of SNM, is DAMP that consists of Drupal and LAMP. SNM is an Open science type of application and it is an Open government e-service designed using Open architecture principles with the aim of publishing Open data, following Open standards and using Open source software components. SNM is intended to become a centralized virtual online place to offer and collect information about researchers, research papers, research projects, equipment and scientific knowledge. In this paper we have presented how SNM, developed using the concepts of Open movements, is to be used at a national level. However, we anticipate that in future, the scope of SNP application can be expanded and internationalised through efforts invested in regional collaboration.

REFERENCES

[1] M. Peters, R. Britez , Open Education and Education for Openness

[2] G. Magalhaes, C. Roseira, L. Manley, Business models for open government data, ICEGOV, 2014

[3] Science Network of Montenegro – Naučna mreža, available online: http://www.naucnamreza.me, 2012

[4] D. Petrušić, Conceptual Model of Open Architecture for Open Government, ICIST, (pp148 - 155), Kopaonik, Serbia, 2013

[5] M. Fioretti, Open Data, Open Society, LEM, Scuola Superiore Sant'Anna, 2010

[6] The Annotated 8 Principles of Open Government Data, available online: http://opengovdata.org/

[7] F. Xavier Cardi, E. Garshina – Mapize, HOMER state of the art for PSI strategy, 2013

[8] The home of the U.S. Government's open data, available online: https://www.data.gov/, 2009

[9] Open government partnership official website, available online: http://www.opengovpartnership.org/

[10] Berners-Lee T. Design Issues: LinkedData, available online: http://www.w3.org/DesignIssues/LinkedData.html, 2006

[11] F. Bauer, M. Kaltenböck, Linked Open Data: Linked open data: The essentials: A quick start guide for decision makers, Vienna, 2012

[12] D. McKenzie, R. Exler, The Success of Open Data Depends on Open Standards, The Indiana Geographic Information Council (IGIC), 2014

[13] S. Abraham, Open standards, Centre for Internet and Society, Bangalore, 2008, available online: www.cis-india.org

[14] Open Standards principles, updated 2015, available online: www.gov.uk

[15] Open government standards, available online: http://www.opengovstandards.org/

[16] J. Dalziel, Open Standards versus Open Source in E-Learning, EDUCASE QUATERLY, Number 4, 2003

[17] J. Park, Workshop on "Open Government: Open Data, Open Source and Open Standards", 2011

[18] European interoperability framework or pan-european eGovernment services, available at IDABC website (http://europa.eu.int/idabc), 2004

[19] M. Thompson, Open Source, Open Standards: Reforming IT procurement in Government, Judge Business School, Cambridge University, 2007

[20] C. Smith, Open source translates to Open government, GTEC, 2016

[21] P. Maassen, Open Source and Open Government , Open World Forum, Paris, 2014

[22] Directorate-general for Research and Innovation (RTD), New policy initiative: The establishment of an Open Science Policy Platform, E.C., 2015

[23] S. Parinov, Open repository of semantic linkages at Socionet CRIS, euroCRIS, 2011

[24] Higher Education for Research and Innovation project – HERIC official website, Ministry of Science, Montenegro, available online: www.heric.me, 2012

[25] L. Woodall, M. Marius, Free and Open Source Software, Open Data, and Open Standards in the Caribbean: Situation Review and Recommendations, UNESCO, August 2013

[26] R. Sint, S. Schaffert, S. Stroka, R. Ferstl, Combining Unstructured, Fully Structured and Semi-Structured Information in Semantic Wikis, 4th Semantic Wiki Workshop (SemWiki 2009), Heraklion, Greece, 2009.

[27] J. Anhøj, Generic Design of Web-Based Clinical Databases, Journal of Medical Internet Research, 2003

[28] B. Jörg, K. Jeffery, Asserson et al., CERIF 1.3 Full Data Model (FDM), 2012

[29] OECD, Frascati Manual: Proposed Standard Practice for Surveys on Research and Experimental Development, 6th edition, 2002

[30] Open source technology, available online: www.kankaanpaa.fi/a&o

[31] P, Waugh, Open government, what is it really?, available online: https://opensource.com/government/open-government-what-it-really

BPMN Serialization - Step Toward Busssiness & IT alignment

Gordana Prlinčević*, Sonja Oklobdzija*, Danilo Oklobdzija*
* Business school of professional studies - Blace, Serbia
gordanaprlincevic@yahoo.com, sonja.oklobdzija@gmail.com, danilo.oklobdzija@vpskp.edu.rs

***Abstract*—This paper points out the importance of using standardized tools for notation and exchange or execution of business processes. Presented service oriented platform enables business IT alignment and make companies open for collaboration in global business environment. Special attention is given to methods for serialization or converting notation of business processes to forms suitable for computer processing. This business-IT alignment enables the efficient use of all resources in order to achieve business goals in real time.**

I. INTRODUCTION

The need for alignment of business and IT domain is not a new need. Due to the growth of information systems in companies during the seventies and the eighties of the last century, the alignment of the possibilities of IT solutions with business requirements is becoming a key challenge. As a solution to this, many technologies of IT planning and development of information systems have been developed, such as Business System Planning [1], Information Engineering [2] and other technologies and studies. Since the main objective of these methodologies was to standardize the development of large information systems, the attention was mostly paid to the analysis of the data and their structure of the organization. A.J. Gilbert Silvius [3] found through the analysis of the main characteristics of the methodologies of that period that their rigid approach did not enable the harmonization of the business requirements and IT solutions in practice. Namely, these methodologies were primarily "Procedures of IT professionals by IT professionals" [4]. That approach resulted in the development of applications which primarily generate schedules, summaries and reports, but significantly limit the role of the user in defining the objectives of improving or further development.

Less formal approach to IT and business domains alignment originated during the mid-90s of the last century. This primarily involves focusing on business requirements and objectives and their conversion into innovative IT solutions. The reference model of the strategic harmonization was given by Henderson and Venktaman [5], Fig. 1.

The strategic model of alignment shown identifies two types of integration between business and IT domains. The first type, which is called strategic integration, represents the link between business and IT strategy and reflects the external components. More precisely, it represents the possibilities of the IT domain to shape and support the business strategy. Another type of integration, called functional integration, refers to the internal domains and represents the link between organizational infrastructure and processes on one side, and IT infrastructure and IT processes on the other.

The modern approach to alignment of business and IT domain differs from the traditional one in the following aspects [6]:

- Focusing on strategy. Unlike the traditional methods which were focused on analysis of the data and their structure, modern technologies of IT planning are focused on the business strategy of the organization. This allows the "translation" of business strategy into measurable performance indicators (KPI – Key Performance Indicator) which show the relationship between business goals and IT solutions.
- Pragmatic approach. Planning of IT solutions is gradually becoming less formal in methodology and it is primarily oriented towards finding the applicable solutions.
- Modernization of existing solutions. Given that most companies have significant capacities of inherited hardware systems and software solutions, modern methodologies of development of IT infrastructure must be based on improvement of existing resources. This certainly includes the development of completely new IT solutions in line with business objectives.
- Continuous innovation. Due to the significant impact of IT domains on the overall business, it is of great importance to create such IT solutions which would support business innovation, and not just give solutions for production improvement. This implies the involvement of business professionals in the IT domain, as well as understanding of business objectives by IT professionals.

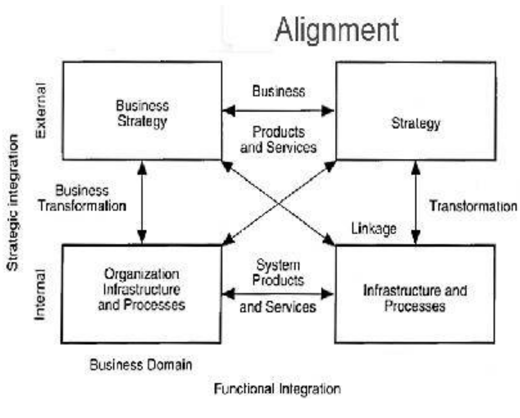


Figure 1. Functional integration

II. Bussines Proces and Service Oriented Architecture

The term Business Process Modeling (BPM), or as it is also called Business Process Management, refers to the design, management and execution of business processes. According to OMG (Object Management Group) a process is a "flow of work and information in business." A business process is defined as a set of logically related tasks that are executed in order to attain a predefined business outcome. A process is a structured set of activities designed with the aim to produce a certain result for a customer or a market.

A business process can be very complex in its structure. It may have multiple participants, such as people, organizations and systems which perform multiple tasks. In order to accomplish the given task, the participants must perform defined tasks in coordinated manner, usually grouped into sub-processes. In some situations, the sub-processes are executed in parallel, in others they are sequential. Some processes require re-execution of sub-process. Most processes include decision points, which lead to the branching of the flow depending on the previously set, satisfied or not satisfied conditions. Within some processes, the participants must share certain information. The transfer of information can have the role of triggering or starting a new task. Some processes are ad-hoc processes, which means that their sub-processes do not have defined triggers. The participants do not have to complete all the defined tasks before they, or another participant, start the execution of another dependent sub-process. The business process consists of atomic steps, which are interconnected with business rules. A step in business process is a basic part which represents a unit of work and cannot be broken down into smaller steps. It is also called the activity or transaction, and since it is atomic, it uniquely identifies all the events in the process. The human participants are an integral part of any process which is not fully automated, so that non-automated steps have a user interface by which the process is monitored and controlled by the participants. The user interface allows the participant to input and modify attributes, as well as to initiate, suspend, establish, stop and end the processes. In addition, each process has access and the possibility to change the condition of any business entity which is linked to that process.

For business analysts, BPM means understanding the organization as a set of processes that can be defined, managed and optimized. Instead of the traditional orientation where the operations are divided by organizational units, BPM is oriented towards operations, regardless of the organizational unit they are executed in. For technical staff, BPM is a group of technologies focused on defining, executing and overseeing the process logic. Regardless of the different perspectives, both business analysts and technical staff have the objective to improve business processes.

Business process management is an evolutionary advance in alignment of business and IT domains. Full alignment of business and IT domain is possible only in a flexible and collaborative working environment. Technological basis of this paradigm are SOA (Service-oriented architecture) and semantic Web 2.0. Service-oriented architecture is a business-oriented approach to architecture of IT solutions, which enables the integration of businesses in the form of related business services. At the same time, SOA is an evolutionary approach to building an integrated information system, which is focused on solving business problems [7].

In theory, BPM is a continuous process of information exchange and application of harmonized technologies through the stages of design, development and monitoring of the life cycle of a business process. However, in practice, the transition between phases is usually not continuous, primarily due to different methodologies, standards and languages used, which can cause semantic problems. These semantic differences primarily arise during the transition from the business domain to the IT domain, and vice versa, due to the different nature of BPMN (Business Process Model and Notation) and BPEL (Business Process Execution Language) standards. Business analysts tend to present business processes as business diagrams (workflow diagrams). On the other hand, in the IT sector, the existence of strictly defined procedures and formalized languages for their presentation is essential. This "conceptual mismatch" can cause problems during the transition from one phase of the life cycle of a business process to the other. A detailed identification of the origin of the conceptual inconsistency (conceptual mismatch) of the languages used for modeling business processes, as well as a generic solution to the problem was given by Recker and Mendling [8].

III. BPMN

BPMN is an international standard for modeling and notation of business processes developed and controlled by the Object Management Group. BPMN is widespread both in practice and in academic circles, as evidenced by the large number of tools for BPMN modeling, as well as a number of academic articles on various aspects of the BPMN [9].

The first version of BPMN 1.0 by Stephen White of the IBM Company in 2004 primarily represents the standardization of a set of graphical forms for documentation and visualization of business processes. BPMN version 2.0 [10] was developed in order to provide notation for defining business processes, easily understandable by all business users, business analysts who need to outline the processes, IT and other technical staff who are responsible for implementing the technology for executing processes, and business people who will manage and monitor business processes. BPMN 2.0 enables simple and easily understandable way of creating business process models, without limiting the complexity of the process being modeled. These two contradictory demands were reconciled by defining only five basic categories of elements, which can be easily recognized in diagrams and their meaning can be intuitively understood. Within the basic category of elements (Flow Objects, Data, Connecting Objects, Swimlanes and Artifacts) there are additional variations. In addition, additional information can be defined in order to support complex business processes, without dramatic changes to the basic layout of the process model.

Business process modeling covers a very wide range of participants, information and ways of exchanging that information between the participants of the business process. BPMN 2.0 is designed to respond to these different requirements in modeling, and considering the

different objectives, the following BPMN models can be differentiated:

• Processes. This model represents the orchestration of activities and the data flow between the entities involved in the process. This model includes:

o Internal business processes which are not executed;

o Internal business processes which are executed;

o Public business processes.

• Collaboration. Collaboration describes the process of interaction between two or more business entities. The interaction is shown as a sequence of activities, that is, a set of messages exchanged between the participants. The activities for the participants in the collaboration can be seen as the point of separation or the point of merger. Therefore, this type of diagram shows that public aspect of the process, visible to each of the participants in collaboration.

• Choreography. Although the choreography is consisted of a network of basic BPMN elements, the same as the private business processes are, the main focus is on presenting the messages that the business entities interchange.

BPMN 2.0 offers dramatic improvements in (unfortunately non-formal) defining of the semantics for the automation of execution of business processes. Automation of execution of business processes can be implemented using any of the well-known programming languages (Java, C #, etc.). However, the use of programming languages for services orchestration, which are parts of the business processes, often has the effect of creating inflexible solutions. The main cause of the inflexibility of this approach is due to the unclear boundaries between the process flow and the business logic, which must not be firmly connected. The general need for creating solutions for the automation of business processes demands standards, as well as specialized language for the composition of services into business processes. BPEL is the kind of language which provides the possibility of executing business processes in a standardized way using the generally accepted language.

IV. BPEL

BPEL (short for BPEL4WS - Business Process Execution Language for Web Services) is a specification created as a synthesis of the positive sides of the past specifications, Microsoft's XLANG and IBM WSFL. BPEL4WS is in fact a convergence of the structural orientation of displaying business processes of XLANG and the graphical orientation of WSFL-in. With the acceptance of BPEL specification by the OASIS standards organization and the support of all the major software vendors (Oracle, Microsoft, IBM, SAP, Sun, BEA, etc.), BPEL specification became the standard model for linking individual Web services in order to create a reliable business solution [11]. Using the principles of the process oriented applications when connecting Web services and the business processes, it provides the portability and interoperability of the implemented business processes. Process oriented application structures have two strictly separated levels: the higher level of the business process, which represents the logic of the business process by using BPEL language, and the lower level of Web services, which represents the functionality of business processes by using Web services. The basic concept of BPEL can be used in two ways:

• BPEL process can define business protocols. The business protocol specifies the potential sequence of messages exchanged between partners (Web services) in order to achieve a business objective. It also defines the order in which a particular partner sends and expects messages from others, in accordance with the specific business context. Since the business protocols are not executable, they are often called abstract business processes;

• BPEL specification can also define executable business processes. With the executable processes, the activities (messages that are exchanged) between the partners (Web services that participate in the implementation of the business process) are defined by logic and the current state of the business process. The access to a particular Web service is enabled by using portType element of the WSDL service.

In order to use the BPEL specification and execute business processes defined in BPEL specification, appropriate graphical tools for defining business process are necessary, as well as an efficient executive environment, that is, a BPEL server. BPEL servers provide the execution environment for performing BPEL business processes. BPEL is firmly connected with the XML Web services and modern software platforms which support the development of XML Web services, especially with Java Enterprise Edition and Microsoft .NET platforms. This connection allows BPEL servers to use the services, which are provided by the aforementioned platforms, such as security, transaction management and connection with databases [12].

V. BPMN SERIALIZAION

BPEL is the de facto standard for the implementation of business processes using Web services technology. There are many solutions which support the execution of BPEL processes, and many of these solutions include the possibility of graphic editing. However, all these tools do not support the higher level of abstraction necessary for the design process of business analysts. On the other hand, BPMN as a language for defining business processes provides certain interaction between business analysts and IT analysts. However, none of the ubiquitous tools that support BPMN allow the execution directly, but they support the translation from BPMN to BPEL. The main role of XPDL (XML Process Definition Language) and BPEL languages is to enable the transition of the business process from the business domain to the IT domain, and vice versa, Fig.2.

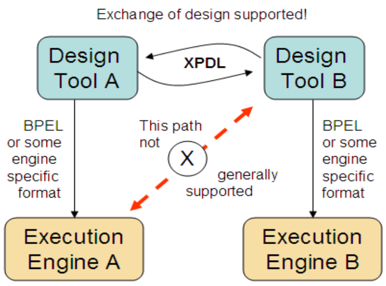


Figure 2. The roles of BPEL and XPDL languages.

The problem of translation from BPMN to BPEL or XPDL is the translation from graph-oriented language to block-oriented language. Graph-oriented languages define the control flow by using the links which are the logical connection between nodes of different types. On the other hand, block-oriented languages define the control flow by using the commands and their structures for representing competitiveness, sequences, loops, etc. [13].

As shown in Fig.2. [14], on the top (development) level, there are various tools for designing business processes, while on the bottom there are the executable versions of the respective processes. The role of XPDL is to facilitate the exchange of design of business processes between different design tools, while the primary role of the BPEL is business process execution.

With the analysis of existing tools, Chum Quyang [15] found that none of the tools meets the following requirements:

- Completeness, that is, the possibility to be applied on any BPMN model;
- Automation, that is, the possibility of production of the final code without the human intervention in identifying and adapting parts of the source code;
- Clarity, that is, the possibility of producing the code which is understandable to humans as well. This requirement stems from the fact that BPEL definitions obtained by translation also require the definition of additional information.

Table 1 shows the characteristics and the purpose of certain languages in the design of systems for managing business processes [16].

VI. CONCLUSIONS

The complexity of the modern business environment is characterized by an extremely strong causal connection of all the business activities, from the most general processes of long-term planning to the technological procedures. Timeliness and continuity of these activities are the basic conditions of successful overall business. In order to meet these conditions, their integration is necessary, and to the extent to which it is possible to achieve unobstructed communication between all relevant business functions. It is undisputed that the management of business processes is one of the most important aspects of integration of the organization. In addition, business processes are becoming a place where business and IT world meet. Therefore, the definition of the true notations and languages which are equally understandable to business managers and to IT personnel is of very great importance. The possibility of converting graphical representation of business processes to the BPEL language, which is standardized in the IT domain, is the first step in the integration of these two domains [17].

TABLE I. LANGUAGES CHARACTERISTICS

	BPMN	BPEL	XPDL
Characteristics	**Notation for drawing**	**Format to save processes**	**Format to save images**
Primary Purpose	**Visualization**	**Control over other systems**	**XMLization of BPMN diagrams**
Diagram Information	**Suitable**	**Impossible**	**Possible**
Data Format	**Image**	**XML**	**XML**

REFERENCES

[1] John Zachman, “Business Systems Planning and Business Information Control Study: A comparision”, *IBM Systems Journal*, vol 21, no 3, 1982. p. 31-53.

[2] James Martin, *Information Engineering, Book I: Introduction*, Prentice Hall, New Jersey, 1 edition (July 7, 1989), ISBN-10: 0538825332.

[3] Drs. A.J. Gilbert Silvius, “Business & IT Alignment in theory and practice”, *Proceedings of the 40th Hawaii International Conference on System Sciences* – 2007, [Available: https://www.computer.org/csdl/proceedings/hicss/2007/2755/00/27550211b.pdf, [Accessed December , 2015].

[4] POLS, R. VAN DER ‘Nieuwe informatievoorziening; informatieplanning en ICT in de 21e eeuw’, Academic Services, The Hague, 2003.

[5] J.C. Henderson, N. Venkataman, “Strategic alignment: Leveraging information technology for transforming ogranization”, *IBM System Journal*, Vol 38, Nos 2&3, 1999, pp 472-484, 1999, Available: [http://gunston.gmu.edu/ecommerce/ITGovernance/doc/henderson.pdf], [Accessed December , 2015].

[6] A.J.G. Silvius, “Business & IT Alignment in theory and practice”, *1st International Conference on Information Management & Business (IMB) Taipei*, Taiwan, 2005.

[7] S. Carter, *The New Language of Business SOA & Web2.0*, Crawfordsville, Indiana, USA, Pearson Education, Inc., 2007, ISBN 13: 987-0-13-195654-4.

[8] [9] Jan Recker, Jan Mendling, “On the Translaion betwen BPMN and BPEL: Conceptual Mismatch betwen Proces Modeling Languages”, *18th International Conference on Advanced Information Systems Engineering. Proceedings of Workshops and Doctoral Consortiums*, 2006, pp. 521-532, [Available: http://ceur-ws.org/Vol-364/paper15.pdf], [Accessed December , 2015].

[9] Matthias Geiger, Guido Wirtz “BPMN 2.0 Serialization – Standard Compiliance Issues and Evolution of Modeling Tools”, *5th International Workshop on Enterprise Modelling and Information Systems Architectures*, St. Gallen, Switzerland, September 2013, pp. 177–190, [Available: https://www.uni-bamberg.de/fileadmin/uni/fakultaeten/%20 wiai_lehrstuehle/praktische_informatik/Dateien/Publikationen/emisa13-geiger-camReady.pdf], [Accessed December , 2015].

[10] OMG Document Number: formal/2011-01-03, “Business Process Model and Notation (BPMN) Version 2.0”, January 2011 [Available: http://www.omg.org/spec/BPMN/2.0/PDF/], [Accessed December , 2015].

[11] OASIS, „Web Services Business Process Execution Language Version 2.0 - OASIS Standard“, 11 April 2007, [Available: http://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.pdf], [Accessed: January 2016].

[12] Snežana Popović, „Modelovanje poslovnih procesa korišćenjem BPMN i BPEL jezika“, *14. konferencija YuInfo, Kopaonik*, 2008, ISBN: 978-86-85525-03-2.

[13] Aleksandar Nedeljković, “Jedan novi pristup u prevođenju iz BPMN-a u BPEL”, *Zbornk radova 18. konferencije YuInfo 2012*, pp 606-611

[14] Thinking Matters, “The BPMN-XPDL-BPEL value chain”, [Available: http://social-biz.org/2006/05/26/bpmn-xpdl-and-bpel/], [Accessed: January 2016].

[15] Chun Ouyang, Wil M.P. van der Aalst, Marlon Dumas, Arthur H.M. ter Hofstede, “Translating BPMN to BPEL”, BPM Center Report BPM-06-02, BPMcenter. org, 2006 [Available: http://www.processmining.workflowpatterns.com/documentation/documents/BPM-06-02.pdf], [Accessed: January 2016].

[16] Business Process Library Q Business Process Management, Q-BPM.org, “BPEL”, [Available: http://en.q-bpm.org/mediawiki/index.php?title=BPEL], [Accessed Mart, 2015].

[17] Kapil Pant, Matjaz B. Juric, *Business Process Driven SOA using BPMN and BPEL*, Packt Publishing Limited, Birmingham, UK, August 2008, ISBN 139781847191465

Smart watch access control application based on Raspberry Pi platform

Marina Antić*, Miloš Kosanović*, Slavimir Stošović*
*College of Applied Technical Sciences / Computer sciences, Niš, Serbia
marinamantic94@gmail.com, milos.kosanovic@vtsnis.edu.rs, slavimir.stosovic@vtsnis.edu.rs

***Abstract*—In this paper we described a system that allows access to the laboratory by unlocking the electromechanical lock with the smartwatch device, collecting the data about the working conditions (temperature and humidity) in the laboratory, keeping records of the present members and the preparation of the reports of the working time. The main hardware component of this system is the Raspberry Pi, which serves as the local server for the database of the registered users. Creating the user accounts and the registration of smartwatch devices for the control is performed in the PHP Web application. The software modules are realized by HTML, CSS, JavaScript, jQuery, Python, and PHP with the integration of the MySQL database.**

I.INTRODUCTION

The modern era has introduced the computers in all the areas of life which leads to complete automatization. The concept of connecting embedded devices within the existing Internet infrastructure is called the Internet of Things *(IoT)* [1]. It allows objects to be sensed and controlled remotely across the existing network infrastructure, thus creating opportunities that surpass the current communication of the two machines and cover various protocols, domains, and applications. Thereby, it refers to a wide range of devices such as air conditioner, washing machine, implants for cardiac monitoring, bio-chips installed in the domestic animals, vehicles that help in saving or acquisition of data in inaccessible terrain, such as VTS Explorer [2]–[3].

In order to enhance the learning and the understanding of computing, in the recent years many concepts of the cheap hardware have been developed that allow users to glimpse “under the hood“ without fear of making a damage to the expensive device. The two most popular concepts are known as Arduino [4] and Raspberry Pi [5].

Arduino is a physical computing platform based on open source panel with the simple input/output pins. Through digital input/output and analog input connectors, an Arduino microcontroller can receive signals from the environment or control the other electromechanical components, but still has more modest performances than the hardware called Raspberry Pi.

Raspberry Pi (RPi) is a computer of the small dimensions, with the aim of promoting computer science in schools [6] [7]. This credit card sized computer can perform many tasks just as a desktop computer, such as creating the spreadsheets, the word processing, playing the video games, etc. It also has the ability to play HD videos. Raspberry Pi can run several versions of Linux and it is used around the world in the various fields, from teaching programming to children, to building robots, to the home entertainment system.

Mobile phones are increasingly being used to control and manage the aforementioned systems. Nowadays, as the popularity of the wearable devices is also in constant increase, the smartwatch has become one of the commonly used devices in such systems. As for the smartwatch devices, nowadays, many companies have stood up for realization of such devices (LG, Samsung, Sony, Asus, Apple, etc.). In this project, we exclusively used the Samsung's wearable device, smartwatch model known as Samsung Gear S [8]. Applications for this device are being made for Tizen operating system, using the following technologies: HTML, CSS, JavaScript, and jQuery.

Within the system described in this paper, for the control of working time and the access to laboratory, Raspberry Pi is connected with the electromechanical assembly and is controlled by the application created for smartwatch and smartphones devices. PHP Web [9] application is also created and used to control the database of user accounts. Data recorded in MySQL [9] database can be used to create appropriate reports on working time.

In the second chapter, the basic hardware features of Raspberry Pi microcomputer used in this project are described. After the third chapter and a description of the system, in the fourth chapter there is a detailed scheme of the implemented electromechanical assembly. Chapter five gives a detailed description of the functionality of PHP Web application, while chapter six describes the specifics of the smartwatch application.

II.RASPBERRY PI

Ability to install several different operating systems (Debian GNU / Linux, Fedora, Arch Linux, Risc OS, and other), as well as programming in many languages (Scratch, C, C++, JAVA, Perl, Ruby, Python) with low price and small size, have made the RPi computer very popular.

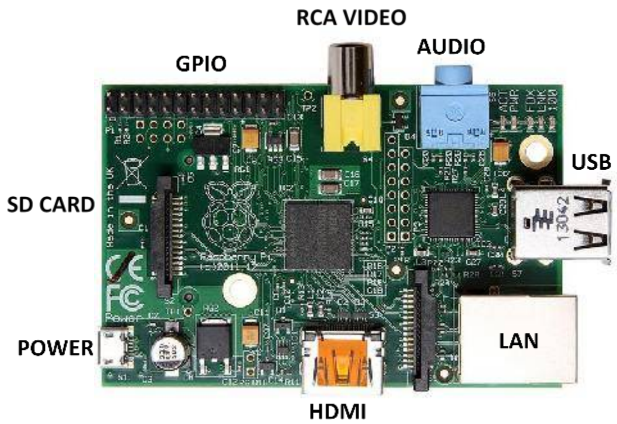


Figure 1. Raspberry Pi model B

The main component that allows small size of this computer, while keeping the powerfulness at the same time, is Broadcom BCM2835 system-on-chip containing RM1176JZFS processor core with floating point and clock speed of 700MHz (which is much better compared to Arduino's 16MHz), and VideoCore 4 graphics processing unit (GPU). Graphics processing unit makes it possible to use Open GL ES 2.0 and enables the decoding of 1080p30 H.264 signals and calculating general purpose at speeds of 1Gpixel/s, 1.5Gtexel/s or 24 GFLOPSs. This means that the Raspberry Pi can be connected to an HDTV and Blu-Ray high-quality video can be watched using the H.264 codec to 40Mbits/s.

Raspberry Pi Model B, used in this project (Figure 1), has 10/100 Ethernet port to allow surfing the Internet or using it as a web server, which is utilized in this project. Most of the Linux systems for RPi can be easily stored on SD card of 2GB, but larger cards are also supported. It also has a standard connector for Raspberry Pi camera. RPi model B has two integrated USB ports for connecting mouse and keyboard, but in order to connect multiple devices, USB HUB is used. It is recommended to use a hub with power supply in order not to overload the voltage regulator on the motherboard. Power supplying the Raspberry Pi is very simple; it is only needed to turn on any USB power in the micro USB port. The power button doesn't exist, so the RPi launches as soon as it is connected to the power supply, and it shuts down by simply removing the power supply.

What makes this computer exceptional is GPIO port of general purpose, with eight I/O pins through which additional sensors, relays or complete devices can be controlled. In this project, these pins are used for signal transmission in order to control electrical circuits, in this case an electric lock. The necessary power for Raspberry Pi (5V) is brought by this connector.

III. ELECTROMECHANICAL SYSTEM DESCRIPTION

The scheme of electromechanical assembly is shown in Figure 2. Communication of passive and active components with Raspberry Pi is achieved through a simple interface created on grid plate (Figure 3).

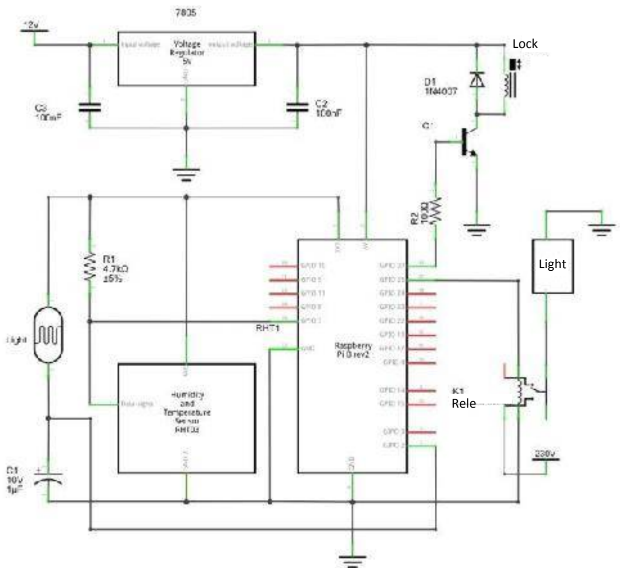


Figure 2. Scheme of electromechanical assembly

The entire circuit is powered with a 5V voltage regulator that provides L7805. At the entrance of this circuit we can bring maximum 40V, and then at the output we get stable 5V voltage. It is recommended to use a block capacitor at the input and output of the L7805 circuit, which ensures stable function and prevents regulators to self-oscillate.

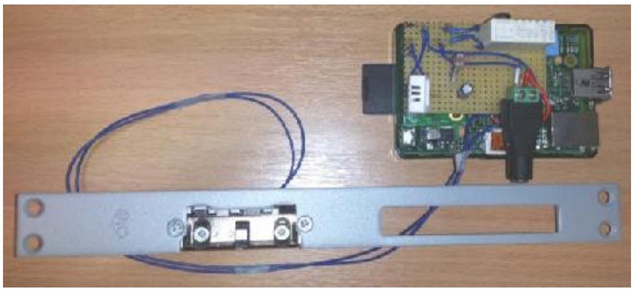


Figure 3. Electromechanical assembly Raspberry Pi with peripherals

Raspberry Pi and transistor switch are supplied by voltage of 5V, and their role is to control the power for activating the electromagnet in the lock. The switch is created using NPN transistor BDP947, marked on the chart as Q1, and its benefits are very small size and large collector current which is essential for activating electromagnet. 1N4007 diode is connected on electromagnet of the lock and its role is to protect the transistor from the reverse electromotive force that is present while switch is in the OFF mode. The impulse for "opening" the lock is brought with pin 13 of Raspberry Pi through the resistor R2 (100 Ω) to the base of a switching transistor. At the moment, when impulse is on the base, current flows through the transistor and electromagnet of the lock, which leads to unlocking.

Sensors allow data acquisition in the laboratory, while Raspberry Pi camera's role is to control video surveillance. Sensor that is used in this paper is temperature and humidity sensor DHT-22, and light-dependent resistor (LDR), which is also realized on RPi, thus giving the value of brightness in specific area.

IV. SYSTEM FUNCTIONALITY DESCRIPTION

Functionality of the entire system is illustrated in Figure 4. All system users must be registered in the database of laboratory members. Since the database is unique for working time and access control system and VTS Apps Team's Portal PHP application, this base is located on a web server and also on the Raspberry Pi.

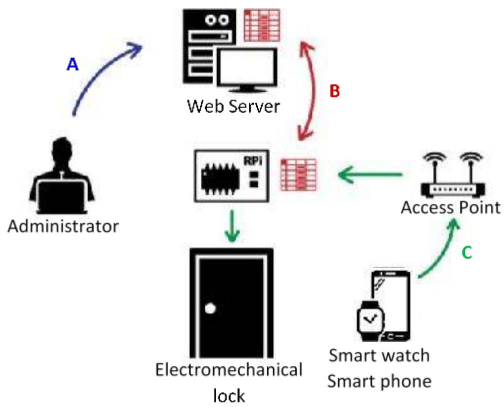


Figure 4. Scheme of realized system

So, the first step for administrator is to create user account (Figure 4 step A) by filling in Web form in PHP application that is shown in Figure 5. In this way, the user gets a universal username and password to access all the above mentioned applications of the Team.

While creating account, PHP application checks whether the entered username for a new user is already in use. At the same time, it checks form validation, in terms of checking the required fields and whether the data is in the correct format. If the created account meets all the conditions and safety checks, PHP stores account with detailed information of user in MySQL database and synchronizes the base with the one located on Raspberry Pi server (Figure 4 step B). As a consequence, the system can be used even without active internet connection.

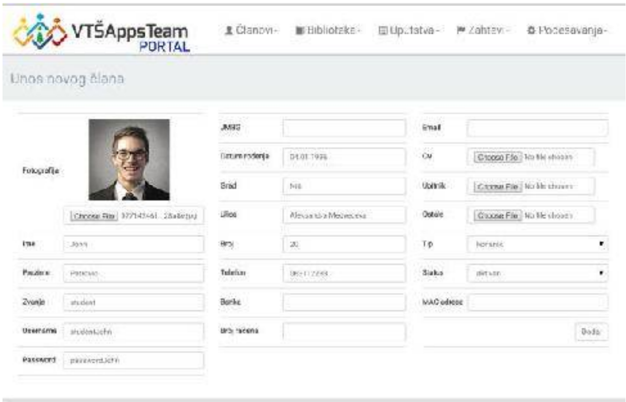


Figure 5. PHP Form for creating user account

After that, a user installs VTSAccessControl application on smartwatch device. The user can register only one device at the moment. Changing a registered smartwatch device for system control includes checking out of the previously registered device. When smartwatch application starts, Splash Screen appears first (Figure 6).

Figure 6. Splash Screen

Figure 7. Login Form

Appearance of login form that will appear after the Splash Screen depends on the results of the following checks:

1. Whether it is the first time application is started,
2. Whether the user was registered through smartwatch application,
3. Whether the user is approved by the administrator.

These three checks are performed by reading the file located in the internal memory of the device, which is encrypted by the decryption key, SHA-256 hash.

If the application has started for the first time, which means that user account has been created but the device has not been registered yet, the login form appears (Figure 7). User needs to fill in two required fields for username and password. By clicking on button “Sign up“, HTTP POST request is sent from smartwatch application to PHP application with the following parameters:

- token (the security key) encrypted by sha1 hash
- MAC address of the device
- username and
- password

PHP application gets HTTP request by smartwatch application and then processes it by examining all parameters that have been sent and searches for users in MYSQL database according to certain criteria (username and password), and then corresponds with specific *response* code.

Based on the received code, smartwatch application makes a decision which procedure should be realized. There are several codes that smartwatch application can receive from PHP application and these are:

- user with specific username and password doesn't exist in the database,
- token is not valid (security key),
- attempt of signing in with information of existing user but without the same registered device (attempt will be saved in the table of unauthorized users),
- user sent request for device registration,
- user is successfully registered.

If smartwatch application gets a response that user has successfully sent a request for device registration, new form will appear, called Wait Screen, with a message that the request for registration is successfully sent and that account will be activated as soon as administrator confirms it. At this point, PHP application sends a notification email with request for device activation. Each time the application is started, but after the device registration step, Splash Screen immediately redirects the user to the Wait Screen and checks whether the account is activated through HTTP POST request, after which the user is redirected to the Home Screen (Figure 8).

On the Home Screen the user has the information of the current time and access to three buttons for unlocking the door, check-in, and check-out. While Home Screen form is displayed, background checks whether there is an active Internet connection (WiFi, GPRS) and whether the user is connected to a particular VTSAppsTeam Access Point network device with predefined MAC address (Figure 4 step C). If there is an interruption of Internet connection, a dialog box appears with notification and shortcut for reactivating the Internet.

Only under the condition that the user's smartwatch device is connected to the proper Access Point, all the options on the Home Screen will be enabled. In that way safety is provided so the door cannot be unlocked unless the user is physically nearby the laboratory. In order to unlock the door, check-in or check-out, the smartwatch device in local network communicates with RPi through a

socket whose server is written in the Python programming language. The procedure is as follows:

1. JSON object is created on smartwatch application with following parameters: username, password, MAC address, action (action that needs to be done), token (the security key),
2. JSON object is forwarded via socket,
3. RPi decodes the received parameters (decode them also in JSON) and checks the token and action that needs to be done,
4. When all checks are done, RPi sends an HTTP request to a remote server with PHP application and MySQL database,
5. PHP processes the request and inscribes data such as username, time of check-in/check-out, and responds with particular code message,
6. If the request does not pass the security check, the device is disconnected from socket.

Figure 8. Home Screen

V. SMARTWATCH APPLICATION DESCRIPTION

Lately, watches are not only devices for indicating time, but are also significant in providing brief information and have ability to serve as a kind of remote control that is always available at hand. Due to increasing popularity of application development for smartwatch, in this paper, a special emphasis is on this type of control.

Tizen is open and flexible operating system designed to enable support for different devices (mobile phones, watches, bracelets, cars, gaming consoles, televisions, etc.). Tizen is developed by a community of developers under an Open Source license and is open to all members who wish to participate. Tizen operating system comes in multiple profiles to meet all the requirements of industry. The current profiles are Tizen IVI, Tizen Mobile, Tizen TV, and Tizen for Wearable. It also offers complete flexibility with HTML5 support.

For Tizen application development, in this case smartwatch application, Tizen SDK is used because of the wearable development environment. Application in this project is developed using HTML5, CSS3, JavaScript, and JQuery technology. For example, to create a form for entering a username or password, it is necessary to include the following line of code:

```
<form>
    Sign up:<br>
    <input type="text" name="userN" value="Username"><br>
    <input type="password" name="psw" value="Password"><br>
</form>
```

Once the structure is generated, the layout and design of HTML elements are readjusted with help of CSS (Cascading Style Sheet). All the functionality is created using JavaScript language and its library, JQuery. JQuery simplifies and shortens the writing code.

Due to the small size of smartwatch screen (360x480) in pixels, a user interface creation represents a particular challenge. As mentioned earlier, there are three buttons for unlocking the door, check-in, and check-out that perform particular action depends on which button is pressed. These buttons are realized as *div* elements in HTML and CSS adds style to them through classes and IDs. HTML structure is done with following code:

```
<div id="unlock" class="container"></div>
<div id="checkIn" class="container-small"></div>
<div id="checkOut" class="container-small"><div>
```

Sliding finger from the top edge of the screen to the bottom presents leaving the application (exit).

VI. CONCLUSION

The system described in this paper was created out of the need to allow more users using the laboratory (one room) with a certain level of control. Smartwatch and smartphones devices or more precisely smart applications, are used as a “key“ to unlock the electromagnetic lock. Raspberry Pi microcomputer is the core of this system and has the task of keeping the data in user database. PHP Web application is also developed and enables managing user accounts, sending particular email messages and notifications. MySQL database is in charge of keeping records of arrival and departure times.

All described functionalities are confirmed in numerous experiments in real time, and another advantage of this system is its modularity, i.e. ability to add new functionalities.

ACKNOWLEDGMENT

The application was realized by members of the VTS Apps Team within the Samsung Apps Laboratory at College of Applied Technical Sciences in Nis. The authors are thankful to Samsung Electronics Adriatic, Belgrade Branch for their support in this work.

REFERENCES

[1] O. Vermesen, P. Fries, *Internet of Things - Converging Technologies for Smart Environments and Integrated Ecosystems*, River Publishers, 2013.

[2] VTŠ Explorer 1 promo film, http://www.youtube.com/watch?v=yKZJHq831Vg, january 2016.

[3] VTŠ Explorer presentation film, https://www.youtube.com/watch?v=Ph6pR9QKd88, january 2016.

[4] John Baichtal, Arduino for Beginners: Essential Skills Every Maker Needs, Amazon, 2013.

[5] http://www.raspberrypi.org/, january 2016.

[6] http://www.bbc.co.uk/blogs/legacy/thereporters/rorycellanjones/2011/05/a_15_computer_to_inspire_young.html, january 2016.

[7] Russel Barnes, *The Official Raspberry Pi Projects Book*, Raspberry Pi Trading Ltd, 1st edition, 2015.

[8] http://www.samsung.com/global/microsite/gears/gears_featurfe.html, january 2016.

[9] Hugh E. Williams and David Lane, Web Database Applications with PHP & MySQL, O'Reilly Media, may 2004.

Detection and analysis of aperiodic ionospheric D-layer disturbances

Dušan Raičević *, Jovan Bajčetić **, Aleksandra Nina***,

* Department of Telecommunications and Information Science, University of Defence, Military Academy, Belgrade, Serbia

** Department of Telecommunications and Information Science, University of Defence, Military Academy, Belgrade, Serbia

*** Institute of Physics, University of Belgrade, Belgrade, Serbia

ricek9@gmail.com
bajce05@gmail.com
sandrast@ipb.ac.rs

***Abstract*— This paper provides overview of a procedure developed for detection and analysis of non-periodic ionospheric D-layer disturbances induced by solar X-ray flares, intensive γ-ray bursts, cyclones, earthquakes, etc. This procedure is universally applicable for primary VLF and LF signal processing, which we use for spatial radio probing of the lower layers of ionosphere.**

I. Introduction

From the aspect of the telecommunication Earth atmosphere represents complex medium and introduces number of the adverse factors impacting on telecommunications signals. Within Earth Atmosphere, there is a number of the ongoing processes as well as intermittent events which as a consequence, among other things, affect the propagation of electromagnetic waves. Earth Atmosphere is generally decomposed in five layers: troposphere (lowest atmospheric layer, averaging 12 km from the Earth's surface, thinner at the poles, and thicker at the equator), stratosphere (from around 12 km to up to 50 km from the Earth's surface, ozone layer is located here and temperature averages 0 Celsius)**,** mesosphere (which is located above the stratosphere, from 50 km to 80 km above Earth surface inside this layer atmosphere is tin and cold, around -90 Celsius), thermosphere (starts above stratosphere, from 80km to 700 km above Earth surface, this layer is characterized with high temperature, reaching 1000's degrees Celsius, and energetic movement of gases), exosphere (most outer layer, starts above the thermosphere, from 700 km up to 10 000 km). The air and gases within Earth's Atmosphere are ionized by the Sun's radiation. This ionisation creates ionosphere. Ionosphere overlaps the multiple atmospheric layers. Ionosphere is continuously changing, for example there are difference between ionosphere during day time (spreading through mesosphere, thermosphere, and parts of the exosphere), and night time (mainly affecting thermosphere and lower exosphere). Ionosphere is also affected by variations in solar activity and Earth's climate. In general Ionosphere is divided in to three layers: the D, E (Heaviside-Ennelly), and F (Appleton). These ionized layers affect the propagation of telecommunication waves. Ionosphere as a part of the Earth's atmosphere, in particular affects the propagation of ULF (300 Hz – 3kHz) ,VLF range (3-30kHz) and LF range (30-300kHz) [1]. By observing the emission of real and known electromagnetic signal in ULF, VLF or LF range on the receiving end we can analyze processes which take place in the lowest layer of the ionosphere by varying the parameters of the known signal. In order to detect and later explore the short lasting phenomena which affect the change of electron density it is necessary to determine a procedure which will effectively identify the phenomenon according to its duration as a desired input parameter. Known mathematical and programming tools are used in the parameter analysis and they will be explained in detail later on.

II. Definition and Structure of the Ionosphere

The Earth's Atmosphere is under influence of many natural processes that cause changes of the chemical and physical characteristics of each layer. In addition Atmosphere is under constant influence of solar and cosmic radiation which represent major factors of frequent changes in atmosphere layers. Such changes and mutual interdependency of the Earth Atmosphere layers characterize Earth's Atmosphere as frequently changing and unstable environment.

From telecommunications perspective, atmosphere is used as a medium and it is frequently a subject of many scientific and research papers [2]. In order to better understand the specific characteristics of the Earth's atmosphere and its layered structure and how the changes in Earths atmospheres influence radio signal propagation, different parameters have been devised such as the following: dielectric constant, electron density specific for each layer individually and those that originate from them due to ionization, refractive index - which depends on electron density and geomagnetic field.

Based on those parameters, it is possible to create a model which spatially and time-wise describes the propagation of electromagnetic waves [3].

Ionosphere is a layer whose charged particle density significantly influences the propagation of the electromagnetic waves. As mentioned ionosphere consists of 3 layers: D- layer, E- layers i F- layer. D-layer is the lowest one, and its characteristics above 70 km are influenced by solar radiation the most, which explains why it exists only during the day. During the night, at these heights cosmic radiation is the only source of ionizing radiation which influences the D-layer which is not enough for its existence [1]. Concerning the propagation of the electromagnetic waves, this layer reflects very low frequency waves (VLF/LF), while high frequency waves pass through this layer attenuated.

Like the previous one, E-layer exists only during the day. It is formed when solar X-ray flux is high enough to ionize the proper neutral particles. For the purposes of telecommunications the emergence of sporadic E-layer (E_s) is significant. E_s lasts for a short period of time and is considered to occur as a consequence of solar activity and lightning [4].

The F-layer is characterized by nonhomogenic chemical composition during the day which is why in that period it is divided into two sub layers: F1 and F2 sub layer. Each of these sub layers is influenced by different types of radiation, which can lead to the occurrence of multiple deflections from each of the sub layers individually, especially in such a way so that, during the propagation of the electromagnetic waves on certain distances, the radio signal is not even detectable on Earth due to its retention in a temporary waveguide (a phenomenon known as Ducting) [5, 6, 7, 8]. During the night, F1and F2 sub layers are combined and make up the F-layer. High frequency electromagnetic waves are reflected off this layer [9].

All radiation waves which get into the atmosphere are spatially and time-wise variable. Besides solar radiation, the influence of natural phenomena, such as lightning, represent a considerable factor in analyzing changes in ion density in D-layer of the Ionosphere. During the last couple of decades the changes in the ionosphere are influenced through human activity (rapid development of technology and industry, the human influence on the ionosphere is increasing). Recent scientific research has provided results which indicate increasing influence of nuclear and strong chemical explosions on atmospheric particle density [10, 11].

Concerning telecommunications, the most significant parameter of the ionosphere is electron density (N_e) which is used to calculate electron plasma frequency (w_p):

$$w_p = \frac{4\pi N_e q_e^{2}}{\varepsilon_0 m_e} \qquad (1)$$

Where q_e is the elementary charge, ε_0 the medium permittivity and m_e the mass of an electron. Electron density increases with the increase of altitude. Ionizing radiation, solar flares and γ-ray bursts can all affect electron density. Lightning exerts the most significant influence of the lower layers of the atmosphere and on D-layer parameters. The most significant consequences of this natural phenomenon are ionization and warming. Lightning also causes the creation of quasi-electrostatic field, whose presence can trigger an array of other disturbances [12]. The causal link of natural phenomena proves the complex nature of the Earth's atmosphere as well as the complexity of its research.

The analysis of the electron density in the D-layer of the ionosphere is based on the theory of the propagation of the VLF/LF electromagnetic waves. During the propagation of the VLF/LF radio signals, the amplitude of their surface component is equalized with the noise level and becomes undetectable [2]. Unlike the surface component, the spatial component of VLF/LF radio waves reaches the D-layer of the ionosphere where it is reflected back to the Earth's surface where it is once again reflected and returned to the atmosphere. The space limited by the D-layer's upper border layer on one end, and the Earth on the other is treated as a waveguide in the analysis of the propagation of the radio wave of this frequency. The electromagnetic field which comprises of the individual spatial components (mods) is registered by an adjusted receiver. It is exactly this method of measurement that enables us to conduct the analysis of nonperiodic and short lasting D-layer changes from the Earth's surface.

III. EXPERIMENTAL SETUP AND THE ANALYSIS OF ELECTRON DENSITY IN THE D-LAYER

Layered analysis of the D-layer is possible because of the widespread VLF transmitters and receivers. There are several of international networks, some of the best known being: AWESOME (Atmospheric Weather Electromagnetic System for Observation Modeling and Education), AARDDVARKK (Antartic-Artic Radiation-belt (Dynamic) Deposition – VLF Atmospheric Research Konsortium) and SAVNET (South America VLF NETwork).

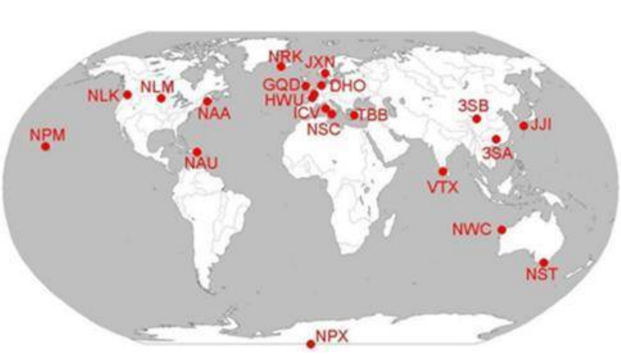


Figure 1. AWESOME transmitter map [13]

The receiver used for the purpose of this paper is a part of Stanford/AWESOME (Figure 1 and Figure 2) network which is located in the Institute of Physics in Belgrade.

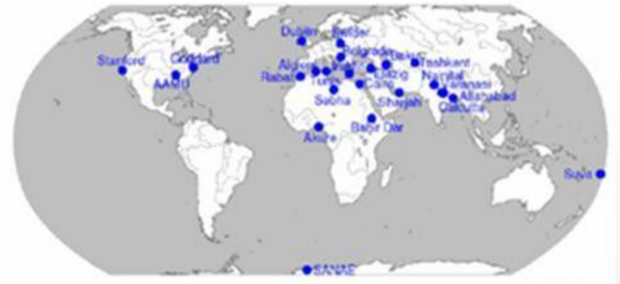


Figure 2. AWESOME receiver map [13]

Intense radiation which causes significant changes of the amplitude and phase of the VLF signal and originate from the Sun and the space are easily detectable and noticeable during the analysis. In the case of lower intensity radiation and changes considered to cause insignificant variations in electron density various statistical methods and transformations have been utilized.

The post-processed receiver data has been recorded as day-long measurements in individual files with the .mat extension for the purposes of further processing and analysis using Matlab software (as one of the programs compatible for further analysis).

The first problem to be solved using Matlab was related to loading the data from .mat files, (eleven representative days of measurement (Figure 3)).

```
d=dir('*.mat');
x=[];
for i=1:length(d)
x=[x; load(d(i).name)];
y(i).data=x(i).data;
end
```

Figure 3. Code section used for loading .mat files

The data contained inside mentioned .mat file is comprised of many data records from which the received signal level was to be filtered out. Figure 3 shows that the values of the received signal amplitude are assigned to matrix y. For the purposes of further processing it was necessary to define the time interval depending on the phenomena of interest. For example, if we want to observe the influence of short lasting phenomena, such as lightning, we will use the shortest time interval possible, which, in a limiting case can be reduced to 20ms, which is the recording resolution of the mentioned parameters.

It was also necessary to define the signal level which can be considered an error, i.e. made under the influence of human factor. On two occasions there was no reception which caused gaps in data records and further complicated processing of the data and detailed analysis (Figure 4). The cause of these gaps is the fact that the transmitter stops working for an hour in the morning, while the receiver stops working for an hour in the afternoon.

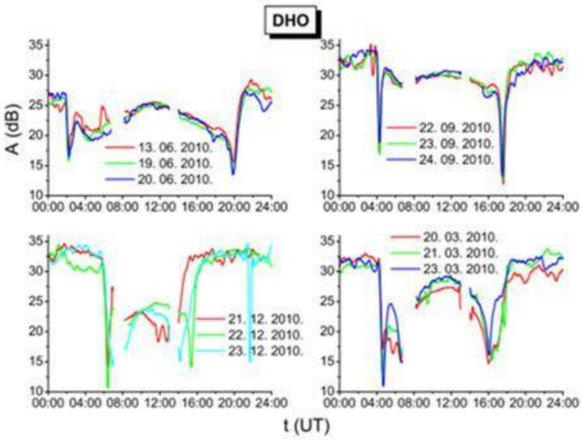


Figure 4. Received signal levels and “gaps” created by the discontinuation of the receiver’s and transmitter’s operation [2].

A more detailed analysis of the phenomena of interest, whose duration interval covers a larger portion of the day, required the gaps to be filled with adequate data, so that the measured signal can be represented with a fitted function, which authentically displays the daily change trend, as accurately as possible. In the beginning all the received signal’s amplitude values are loaded. Each of these values is compared to the defined minimum and if the level of the received signal value is lower, its value is increased by one tenth of the previous one. In this manner it is possible to eliminate the problem in the program loops’ operation, so that the „NaN“ values are assigned real values which could later be used in processing. In this manner, regarding the short lasting phenomena which affect the change in the electron density of the D-layer, the time period in which there was no measurement was defined as „calm state“, but in later analysis retained detectability with the linear manner of its amplitude’s increase. When the values of the amplitudes are examined, a matrix in which all gaps are populated is obtained. In order to define the phenomenon of interest, for the purposes of detection it is necessary to enter the time frame in which the expected phenomenon is registered as an input parameter. For detection purposes, as well as increased detection speed, averaging in the interval defined by the time frame is performed in order to compensate for the high measurement resolution which causes large quantities of measurement data. In order to obtain the average value, the „mean“ function is used (Figure 5).

```
for i=1:length(y)
    for j=1:length(y(i).data)
    if y(i).data(j)<minimum
       y(i).data(j)=y(i).data(j-1)+0.1;
    else
    end
        usrednjeno(i,:)=mean(y(i).data,prozor);
    end
end
```

Figure 5. Code section which compensates for the lost signal values and calculates the average value of the given time frame

The next step in processing was fitting – approximation of the measured data with a function which most accurately depicts the timewise change trend of the measured data. Fitting is done so that, using the obtained results, after averaging on the time interval defined by the time frame which defines the expected time interval of the subject phenomenon, a function, whose later analysis of the extreme value and inflection points can be used to automatically identify the phenomena in the large amount of data acquired by measurement, can be provided. In this manner further analysis becomes simpler, while the time needed for the subject phenomenon detection is significantly reduced. Matlab provides a large number of functions used for fitting. For the analysis described in this paper the choice narrows down to two: fitting using a polynomial of a certain degree and fitting using a Fourier series with a specific number of coefficients. The former is implemented by function „polyfit“. The function requires three arguments the first of which is a vector which contains data based on which the fitted function is created. The second parameter defines the data interval in which fitting is conducted. Depending on the phenomenon whose influence on the D-layer is observed, the interval is defined as a number of points of measurement in accordance with the already defined time frame. The last parameter provides the degree of the polynomial which is used to represent the requested function.

```
for i=1:length(y)
    for j=1:length(y(i).data)
        t(i)=[1:m(i)];
        z(i)=fit(t(i)',usrednjeno(1,j)','fourier8');
        y=polyfit(usrednjeno,t,5);
    end;
end;
yprim=polyder(y);
plot(t,usrednjeno(i))
hold on;
plot(z(i),'r');
```

Figure 6. Code section performing the averaging of the function in the given time frame

Using this type of approximation the values, which would be satisfactory for short lasting changes are not obtained. Their similarity to the measured values is incresed in accordance with the increase of the function degree. Because of that, the procedure of forming a Fourier series proved to be the adequate method of detecting phenomena whose duration is approximately one minute. After obtaining the results using this method, it was observed that they were considerably more accurate than those obtained by the previouse one and that this method required a function of a considerably lower degree. The shortcoming of averaging using a Fourier series is reflected in the fact that the obtained results are considerably more difficult to process compared to those obtained by fitting using a polynomial.

IV. CONCLUSION

This paper depicted an automated procedure of identifying a phenomenon which causes a disturbance in the D-layer electron density. Considering the large quantity of data obtained by recording via a receiver of a high time resolution, it was necessary to identify and separate the phenomenon in question in the large quantity of data for the purposes of later analysis. The described procedure is universally applicable to an array of measured data in which inconsistency is present (so called measurement gaps), where data can originate from any source used for detecting nonperiodic changes of the measured quantity.

The general shortcoming of this procedure is processing time. The analysis of data contained in eleven files each containig two million recordings in average, using a workstation with an Intel Core i7 4710HQ processor and with 8GB RAM took approximately 48 hours. An attempt has been made to compensate for this by code optimization, using fewer loops, but this exercise did not result in significant improvement in processing time.

V. ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Education, Science and Technological Development of the Republic of Serbia for the support of this work within the projects III-44002, 176002, 176004 and TR-32030.

REFERENCES

[1] Cummer, S.A, Modeling electromagnetic propagation in the Earth-ionosphere waveguide, Antennas and Propagation, IEEE Transactions on Vol. 48, Issue: 9J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Budden, K. G. (1988), The Propagation of Radio Waves, Cambridge University Press, Cambridge.K. Elissa, "Title of paper if known," unpublished.

[3] Aleksandra M. Nina, Diagnostic of plasma of ionospheric D region by electromagnetic VLF waves, Faculty of physics, University of Belgrade, 2014.

[4] Christos Haldoupis, Dora Pancheva and N. J. Mitchell, „A study of tidal and planetary wave periodicities present in midlatitude sporadic E layers“, Journal of Geophysical Research,Vol. 109, A02302, 2004

[5] A. V. Moshkov, „Influence of the curvature of geomagnetic field lines on the trapping criterion for Low-Frequency electromagnetic waves entrapped into wave ducts with increased electron density in the Earth's magnetosphere“, Journal of Communications Technology and Electronics, 2008, 53, 5, 529

[6] P. A. Bernhardt, C. G. Park, „Protonospheric- ionospheric modeling of VLF ducts“, Journal of Geophysical Research, Vol. 82, Pages 5222-5230, 1 November 1977

[7] Shigeru Fujita, „Duct propagation of hydromagnetic waves in the upper ionosphere, 2, Dispersion characteristics and loss mechanism“ , Journal of Geophysical Research: Space Physics, Vol. 93, Issue A12, Pages 14674-14682, 1 December 1988

[8] G. M. Milikh, K. Papadopoulos, H. Shroff, C. L. Chang, T. Wallace, E. V. Mishin, M. Parrot, J. J. Berthelier, „Formation of artificial ionospheric ducts“ , Geophysical Research Letters, Vol.25, Issue 17, September 2008

[9] Donald L. Nielson, Michel Crochet , " Ionospheric propagation of HF and VHF radio waves across the geomagnetic equator" , Reviews of Geophysics, Vol.12 , Issue 4, pages 688-702, November 1974

[10] V. Drobzheva and Y. M. Krasnov, „The acoustic field in the atmosphere and ionosphere caused by a point explosion on the ground“ , Journal of Atmospheric and Solar- Terrestrial Physics, Volume 65, Issue 3, pages 369-377, February 2003

[11] V. Drobzheva and Y. M. Krasnov, „The acoustic field in the ionosphere caused by an underground nuclear explosion“ , Journal

of Atmospheric and Solar- Terrestrial Physics, Volume 67, Issue 10, pages 913-920, July 2005

[12] K. I. Kucherov, A. P. Nikolaenko, „Heating of electrons in the lower ionosphere by horizontal lightning discharges“ , Radiophysics and Quantum Electronics, Vol. 22, Issue 7, pages 621-623, July 1979

[13] http://nova.stanford.edu/~vlf/IHY_Test/TechDocs/

Creating a Decision Making Model Using Association Rules

Višnja Istrat
Faculty of organizational sciences, University of Belgrade, Serbia
Visnja.istrat@gmail.com

Abstract*: Being a highly significant and complex function of management, decision making requires methods and techniques which simplify the process of selecting one choice among all available options. Decision making is therefore selection of that particular choice over any of several alternatives. Due to the process complexity, a continuous research and improvement of the methods and techniques modern decision making involves is required. One of many modern business challenges is to discover any possible improvement in the decision making process managers shall use in making the right decision. Any decision made by managers directly impacts the realized profit, business and company's position on the market. The contribution of the paper is showing the importance of association rules in modern decision-making.*

Key words: business intelligence, association rules, management, decision making.

1. Introduction

Three dimensions determining complete development of this discipline need to be highlighted: qualitative, quantitative and information-communication aspect. These three aspects of decision making completely satisfy all the concepts of modern decision making development, both theoretically and practically.

The fact is that mankind faces the decision making problem in each phase of its social development, which has resulted in increased need for learning more about it. In this work both the significance and application of association rules will be analyzed on an example of car sales business. The research was conducted on a sample of 1728 transactions in order to recognize and establish the association rules and then determine their impact on the sales and profit. For the purpose of this research, a large car sales database was used as a source of information, which is also described in this work. Once these association rules were established, they were then used to create a better and more complete market supply.

Quantitative approach in modern decision making defines the basic formalism of general decision making problem [1]. According to [2], the decision making problem is a five item problem $(A, X, F, \Theta, \succ)$ where:

> A: represents a definite set of available alternatives (actions) ranked by a session participant in order to select the most acceptable one;

X : represents the set of possible outcomes as a consequence of selecting an alternative;
Θ: represents a set of world states, and depends on the unknown state $\theta \in \Theta$ because the consequences of selecting alternative $a \in A$ may differ;
F: $A \times \Theta \rightarrow X$, for each world state ϖ and for each alternative a, determines the resulting consequence $x = F(a, \varpi)$;
$\succ$: weak order relation on X, i.e. a binary relation that satisfies the two following criteria:

- Completeness: either $x \succ y$ or $y \succ x$, $\forall\, x, y \in X$;

- Transitivity: if $x \succ y$ and $y \succ z$ then $x \succ z$, $\forall$ x, y, z $\in$ X.

Relation $\succ$ features the decision maker and is called a preference relation. Two other important relations can be derived from the preference relation. The first is the strict preference relation where $x \succ y$ if and only if both $x \succ y$ and not $y \succ z$. The second is the indifference relation where $x \sim y$ if and only if $x \succ y$ and $y \succ z$. The most often way of solving decision making problems is the transformation of weak order $\succ$ on X into normal order $\geq$ over the field of real numbers by the means of utility functions.

As the perfect solution rarely exists, a decision making process will be deemed successful if it produces the most acceptable decision for given problem. As stated in [1], the moment of taking such a decision is unquestionably both the most creative and most critical moment in the complete process of decision making.

II THEORETICAL FRAMEWORK

According to [1], Agrawal was first to establish association rules back in 1993 in order to perform the analysis of market basket. Cornerstone literature for fundamental basis of association rules use is [12]. Association rules are data mining technique where the goal is to find the data structure, regardless of the variables value. The task of the association rules is to determine which items go together. Typical example is grouping the items that can be bought together in shopping in supermarket – market basket analysis.

Supermarket chains use association rules to plan the timetable of the items on the shelves or in the catalogue so that the items that can be bought together are most often together on the shelves. They are used to identify chances for cross-selling and designing of attractive packaging or grouping the products or services. It is simply approach on how to create the rules from large databases. If two items, i.e. computer and web camera are often bought together, we can produce two association rules:

- Buyers that buy computer can also buy web camera with probability of P1
- Buyers that buy web camera can also buy computer with probability of P2

The market basket method has predominantly been used in sales analyses since, although its significance has been proved in other fields such as the analysis of credit card sales, identification of insurance companies committing a fraud or analysis of telecommunication services. However, this method cannot be explicitely used in case of simultaneous events, but only in case of successive events, which may be very useful in marketing for instance.

Numerous examples show how wide the application of the association rules methods in business can be. The researchers have drawn some significant conclusions based on which the right business decision can be taken. The methods proved to be most effective when applied to the market basket analysis, where association rules were used to identify patterns in purchasing products related to the car sales business.

III METHODOLOGY

Data mining is a process of analysing large data sets in order to discover significant patterns and rules. As modern companies are constantly seeking for higher goals, which especially refers to their productivity, it becomes absolutely necessary to improve the functioning of their organization through better understanding of their customers' needs. Association rules determine which items have been purchased together. According to [12], the term *association rules* was first introduced by Agrawal. The task is to identify a set of rules which co-exist in some data set. According to [13], knowledge is a theoretical or practical understanding of facts and information, while wisdom is the synthesis of knowledge and experience that deepens our understanding of connections between different entities and uncovers some hidden message in their

existence. If knowlege is considered tools, then wisdom can be considered a set of skills using knowledge as its tools. There is a case of association rules in the field of market basket [10], where these rules classify into groups the items purchased together in supermarkets.

IV RESULTS AND DISCUSSION

The data used for the purpose of this research are all car sales related. A total of 1728 transactions were processed. Attributes were assigned to shopping (very high, high, medium and low rate of transactions), car maintenance (very high, high, medium, low level), number of car doors (two, three, four or five), number of seats (two, three, four), trunk size (small, medium, large) and safety (low, medium, high). All the data were classified as: accurate, inaccurate, good and very good.

In this paper one application of business intelligence techniques on a real business problem has been described. Some modern software architecture has been used for this purpose, namely *Orange* data mining software, which is a very popular data mining tool among managers, who use it massively to support themselves in decision making. *Orange* is extremely handy and user-friendly software which contains numerous data mining options for data analysis, such as model integration, testing, data visualization, solution application, etc. One of the most popular and well-proven technologies for data mining is CRISP-DM methodology. In order to discover the association rules in a large database of car sales transactions, the data have been processed using *Orange* software. The association rules can further help customers select a car to purchase by providing them with relevant sales data based on the information gathered from previous sales.

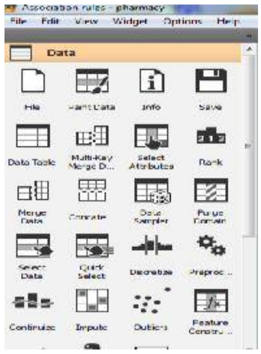


Figure 1: Desktop appearance of Orange software

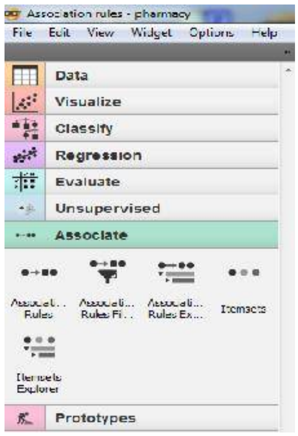


Figure 2: Data entry options of Orange

Figure 1 shows the desktop appearance of Orange software with its data search options. Underneath the menu bar there is the tools menu with the main project functions. Data mining functions such as uploading data, visualization, classification, evaluation or association rules are also available [14]. This is where the model of association rules applied to this research data was created. For the purpose of identifying the association rules, a predefined data set has been used.

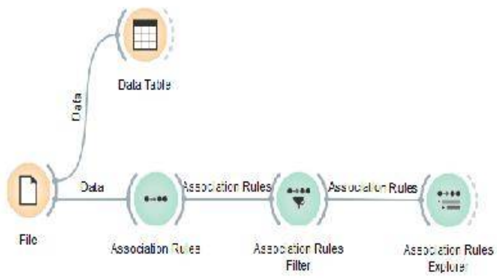


Figure 3: Model of association rules created in Oranges’canvas interface

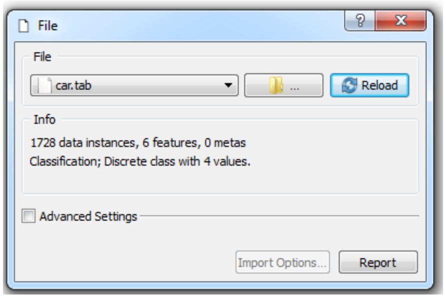


Figure 4: Uploading the car sales data file

The first step was uploading the data file into the program. The Support and Confidence parameters were defined next. As the research progressed, these parameters were altered in order to determine their impact on the final result of the association rules searching process.

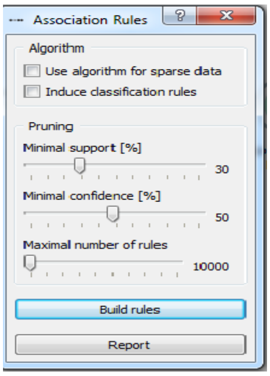


Figure 5: Defining Support and Confidence parameters

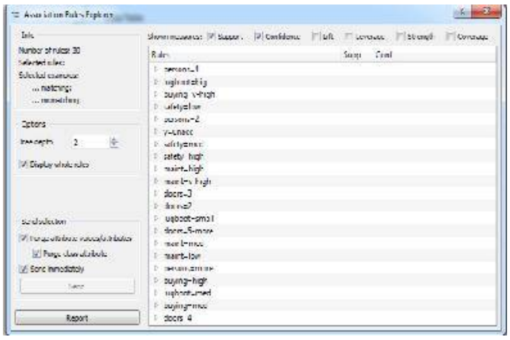


Figure 6: Association rules

Figure 6 shows the preview of obtained association rules. Considering the fact that the minimal Support had been set at relatively high 30%, only two association rules were obtained. The first association rule, with this parameter set at such a high value, was indicating that the two-seater car was classified as ***inaccurate***. Moreover, 30% of all the car sales transactions for the two-seater car fell into this class. With the Support changed at 50%, there was 50% probability that a two-seater car would be purchased. The second association rule referred to the safety issue and established that the low safety cars were most often classified ***inaccurate***.

Orange has built-in some excellent options for visual previewing of obtained results. A report will show both the number of obtained association rules and their description.

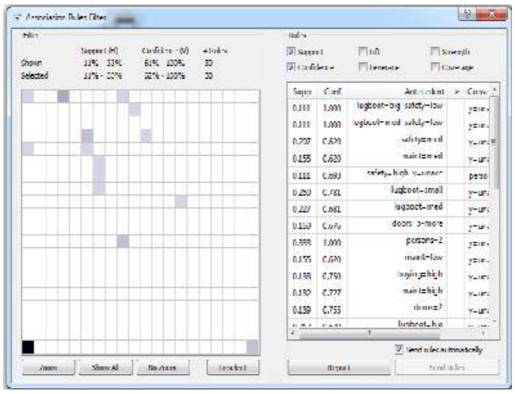


Figure 7: Association rules

Different results were obtained when the minimal Support was reduced from 30% to 10 %, and the minimal Confidence was increased from 50% to 60%. The number of discovered association rules consequently increased. There were thirty new association rules with the Support ranging from 11% to 33% and Confidence ranging from 62% to 100%. Further analysis of the obtained results revealed the purchasing patterns to be used in forming the optimal set of products in order to increase the sales rate and satisfy the customers to the furthest extent possible. Highlighted is the association rule with the highest Support and Confidence, of 0.333 and 1 respectively. The analysis of each association rule showed precision in characteristics and patterns in the customers' behaviour.

Car attributes (characteristics) used for the purpose of this research were: price, maintenance costs, number of doors, number of seats, trunk size and safety. The obtained association rules suggested that those customers interested in purchasing a low safety two-seated car would most likely purchase a car classified

as ***inaccurate***. Marketing experts and experts from the sales sector are recommended to work to attract potential customers by evaluating their needs and offering them a suitable car with the highest probability of purchase. Furthermore, the obtained association rules also show that when a car is purchased, the car radio is commonly purchased as well being part of the car's additional equipment. Therefore any marketing strategy should be developed with respect to this information on order to be successful.

V Conclusion

By analyzing the current research work of various domestic and foreign experts in the field of business intelligence, and despite the fact this field is relatively new in some segments of application, a great popularity and potential it carries have been proved. The obtained results have been analyzed using modern scientific methods and several recommendations for the future research course have been made.

For the purpose of this research, a large car sales database has been used for extracting data and discovering association rules. The obtained association rules have been used to create better and more comprehensive market offer. The process of improving the car sales model has been presented having in mind that managers learn about their customers' purchase habits through association rules. The significance of multidisciplinary approach has been explained – the dependence between decision making, business intelligence, human resource management, knowledge management and modern ICT has been proved to be effective in the case of business system management.

VI References

[1] Suknović M., Delibašić B., Business intelligence and systems for support of decision making, *FOS*, Belgrade, 2010.

[2] Suknović M., Development of methodology of support to group decision making, PhD dissertation, *FOS*, Belgrade, 2001.

[3] Mihailović, D., Methodology of scientific research, *FOS*, Belgrade, 2004.

[4] Čupić M, Suknović M., Decision making, *FOS*, Belgrade, 2008.

[5] Drucker P., „My view on management“, *Adiges*, Novi Sad, 2003.

[6] Grabmeier, J., & Lambe, L. Decision trees for binary classification variables grow equally with the Gini impurity measure and Pearson's chi-square test. *Int. Journal of Business Intelligence and Data Mining, Volume 2*, 2007, pp.213-226.

[7] Hornick, M., Marcade, E., & Venkayala, SJava Data Mining: Strategy, Standard, and Practice. *Elsevier*, 2007.

[8] Delibašić B., Suknović M., Jovanović M., Algorhitms of machine learning for data mining, *FOS*, Belgrade, 2009.

[9] Saaty, T., "An Eigenvalue Allocation Model for Priorization and Planning", *University of Pennsylvania*, 1972.

[10] Chauhan, A., Cerpa. N., "A comparison for procurement models for B2B Electronic Commerce", Proceedings of the OPTIMA Conference, October 10-12, 2001, Curico, Chile.

[11] Astudillo, C., Bardeen, M., Cerpa, N., "Editorial: Data Mining in Electronic Commerce – Support vs. Confidence", Journal of Theoretical and Applied E-Commerce Research, ISSN 0718-1876, Vol 9, Issue 1, Jan 2014, pp. I-VII

[12] Agrawal, R., et al, "Mining association rules between sets of items in large databases. In Proc. 1993 ACM-SIGMOD, *Int. Conf. Management of Data*, pp.207-216, Washington, D.C., May 1993.

[13] Davenport, T.H., Prusak, L., “Working knowledge: How organizations manage what they know”, *Harvard Business Press*, 2000.

[14] http://orange.biolab.si/

Sensor Signal Processing for Biofeedback Applications in Sport

Anton Umek, Anton Kos, and Sašo Tomažič
Faculty of Electrical Engineering, University of Ljubljana, Slovenia
anton.umek@fe.uni-lj.si, anton.kos@fe.uni-lj.si, saso.tomazic@fe.uni-lj.si

***Abstract*—This article explains technological challenges of real-time biofeedback in sport. Motion tracking systems, in connection to the biomechanical biofeedback, help in accelerating motor learning. Requirements about various parameters important in real-time biofeedback applications are discussed. Studies are done on specific biofeedback problems in various sports. Problems addressed are sensor accuracy, movement dynamics, system data rate, and processing demands. Inertial sensor tracking system accuracy is tested in comparison with a high performance optical tracking system. Sensor signal acquisitions and real-time processing challenges, in connection to biomechanical biofeedback, are presented.**

I. INTRODUCTION

Technology and science are being increasingly valued in modern sports. They offer new knowledge, expertise, and tools for achieving a competitive advantage. One such example is the application of biomechanical *biofeedback systems*. In this paper, the word biofeedback denotes a body activity in the sense of physical movement and it is classified as a biomechanical movement biofeedback [1].

One of the most common uses of biomechanical biofeedback is motor learning in sports, recreation, and rehabilitation [2]-[6]. A combination of wearable devices and ubiquitous computing can provide the means for the mobile implementation in motor learning tasks. The process of learning new movements is based on repetition [1]. Numerous correct executions are required to adequately learn a certain movement. Biofeedback is successful if the user is able to either correct a movement or abandon its execution given the appropriate biofeedback information.

The concurrent biofeedback can reduce the frequency of improper movement executions and speed up the process of learning the proper movement pattern. Such movement learning methods are suitable for recreational, professional, and amateur users in the initial stages of the learning process [2].

The general configuration of the biomechanical biofeedback system is illustrated in Figure 1. It includes sensors, a processing device, a biofeedback device, and communication channels. Together with a user they form a biofeedback loop.

Sensors represent the capture side of the system and are usually attached to the user's and/or integrated in sport equipment (sport shoes, tennis rockets, golf clubs, boots, skis, ski-boards). They are the source of different type of signals and data used by the processing device.

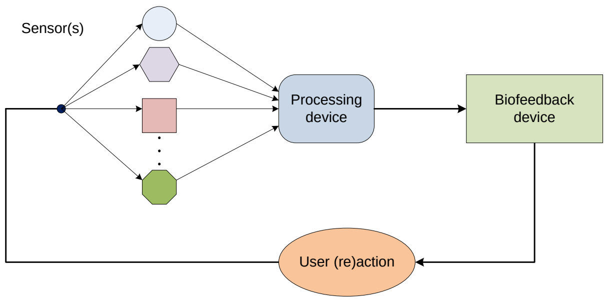


Figure 1. Architecture and operation of a biomechanical biofeedback system. Multiple sensors feed their signals to the processing device for real-time signal analysis. Analysis results (biofeedback signals) drive the biofeedback device activity. User's (re)action alters sensor signals, thus closing the biofeedback loop.

Motion capture systems employ various sensor technologies for motion acquisition. High precision motion tracking systems are camera based systems that use passive or active markers for determining their position in space and time. Inertial sensor based motion tracking systems are generally mobile and have no limitation in covering space. Modern inertial sensors are miniature low-power devices integrated into wearable sensor devices. Sport equipment sensory can be supplemented with flex sensor, force sensors, pressure sensors, etc.

The *processing device* is the core of the system. The processing device analyses sensor signals, generates, and sends feedback signals to the biofeedback devices. The employed processing devices should have sufficient computational power. While this is generally not critical with terminal biofeedback that uses post-processing, it is of outmost importance with concurrent biofeedback that uses real-time processing. When sampling frequencies are high, this demand can be quite restricting, especially for local processing devices attached to the user.

The *biofeedback device* uses human senses to communicate feedback information to the user. The most commonly used senses are hearing, sight, and touch. It is desirable to use the sense with the least cognitive load induced by other activities. For skiing learning support application illustrated in Figure 2, headphones can be used as a simple feedback device. In more complex biofeedback applications audio feedback can be supplemented with visual information by using a head-up display helmet.

Communication channels enable communication between biofeedback system devices. Although wireless communication technologies are most commonly used, wired technologies can also be used if processing device is installed locally.

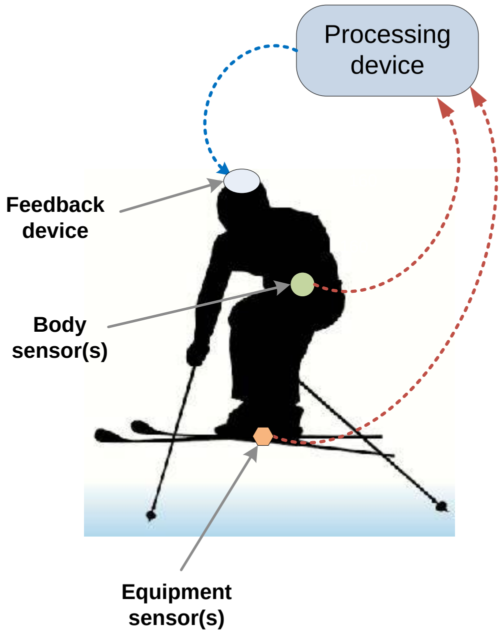


Figure 2. Skiing assistant biofeedback concept. User movement and equipment performance are captured by sensors and their signals are sent to the processing device for analysis. Feedback device can use one or more human modalities: audio (headphones) and visual (helmet with integrated head-up display).

II. Challenges in real-time biofeedback

An ideal real-time biomechanical biofeedback system is an autonomous, wearable, lightweight system with large enough number of sensors to capture all the important motion parameters. Sensor signals exhibit high enough sampling frequency and accuracy. Processing is done instantly and the feedback modality is chosen in a way that it is not interfering with the principal modality of the motion. The main challenges in this effort are various and often contradictory. For example, under the constraints of technology, the ideals of being wearable and lightweight contradict the ideals of autonomy and processing power because of the battery time.

The first challenge is to achieve the desired accuracy of motion capture. Inaccuracies and errors present in various capture systems limit the usability in certain cases. For example, the direct use of MEMS accelerometers for position tracking is problematic because even a small inaccuracy in sensor readings will induce a rapid, square-time positional error.

Another challenge is the sampling frequency and with it related issues. While achieving high enough sampling frequency is generally not a problem, it leads to large amounts of sensor data that needs to be first transferred to the processing device and then analyzed. Problems that may occur are available bandwidth of the communication channels and the computational power of the processing device. The latter is especially a problem in real-time biofeedback systems.

Communication channel bandwidth, range and delays are yet another set of potential problems. Low power wearable devices usually have low channel bandwidth, with very limited communication range.

III. Capturing of human motion

An important area of research connected to biofeedback is various motion capture systems. The majority of motion capture systems are based upon various optical systems and inertial sensors, such as accelerometers and gyroscopes. Motion is captured through measurement of various physical quantities such as acceleration, velocity, position, angular velocity, rotation angle.

Experimentally we have evaluated two different motion capture systems: (a) passive marker based optical system, and (b) MEMS gyroscope based system.

We used a professional optical motion capture system Qualisys™**.** This is a high-accuracy tracking system [8] with eight Oqus 3+ high-speed cameras that offers real-time tracking of multiple marker points as well as tracking of pre-defined rigid bodies. Sampling frequency of the system is up to 1000 Hz. As stated in [10] the measurement noise for a static marker is given by its standard deviation for each individual coordinate: stdx=0.018 mm, stdy=0.016 mm and stdz=0.029 mm. In view of the given results, we can regard the measurement inaccuracy of the optical tracking system as negligibly small. Inertial sensor accuracy is limited by the precision of self-adhesive reflective marker positioning.

Despite the fact that Qualisys has video frame rates of up to 1000 Hz, the comparison with inertial sensors could be done only up to sampling frequencies of 60 Hz. We identified the reason in processing load of real-time calculation of the 6DoF orientation that could not be met by laptop processing power. It should be mentioned here that Qualisys is by itself already a HPC system. It has 8 cameras with integrated Linux system doing parallel processing of captured video. The results of marker positions are communicated to the central processing device (laptop) for synchronization and further processing.

Figure 3. Experimental setup for golf swing motion. Four infrared reflecting markers are attached directly to smartphone to form the defined QTM rigid body orthogonal vector basis.

For inertial sensor based motion tracking a smartphone iPhone 4 is used. Is has the embedded the following MEMS inertial sensors: ST Microelectronics LIS331DLH accelerometer, and STMicroelectronics L3G4200D gyroscope [11].

For the golf swing movement the smartphone was attached directly onto the forearm of the player, see Figure 3. Four infrared reflecting markers are attached directly to the smartphone in a way to form the orthogonal vector basis of the x-y plane of the local coordinate system of the rigid body. Gyroscope was first calibrated to reduce the errors imposed by biases, scaling factors and body axes misalignment [12].

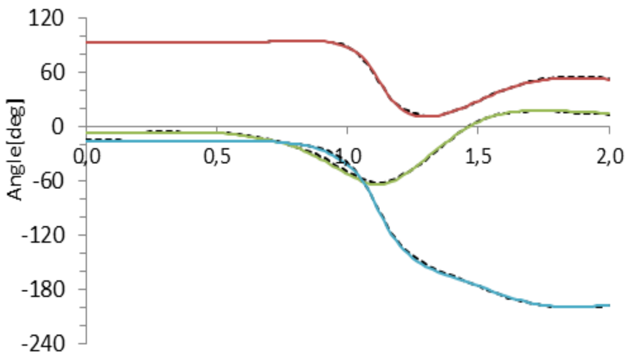


Figure 4. Comparison of smartphone embedded gyroscope (doted black plots) and QTM body rotation angles (solid colored plots: red=roll, green=pitch, blue=yaw). Figure shows the first part of the golf swing movement in 2 seconds time interval from address to top of the backswing.

A testing rotation pattern was generated in golf swing movement, recorded from the beginning (address) phase to top of the backswing. Figure 4 shows the comparison of the body Euler rotation angles, measured with both tracking systems. The calculated root-mean square deviation is 1.15 degrees. Such accuracy is good enough for golf biofeedback application.

A. Motion dynamics and sampling frequency

Due to real-time communication speed limitations of Qualisys and inertial sensor device, the above experiments are performed at sampling frequencies of 60Hz [8], [9]. While such sampling frequency is sufficient for evaluation of capture system accuracies, it is too low for capturing high dynamics movements in sport. To estimate the required sampling frequencies for capturing human motion in sport, we performed a series of measurements with wearable Shimmer3™ inertial sensor device. Shimmer3 allows accelerometer and gyroscope sampling frequencies of up to 2048 Hz.

A set of time and frequency domain signals for a handball free-throw movement is shown in figure 5. The sensor device was attached at the dorsal side of the hand. Measured acceleration and rotation speed values shown in Figure 5(a) are close to the limit of the sensors dynamic range. High sampling rate enables the measurements of actual spectrum bandwidth for both physical quantities. Most of the energy of finite time signals on is within upper limited frequency range, as shown in Figure 5(b). The bandwidth containing 99% of signal energy is a useful measure of signal bandwidth as shown in Figure 5(c). The signal spectrum bandwidths differ in each dimension and are higher than for absolute 3D values. The highest measured values on Figure 5(c) are 59Hz for acceleration and 40Hz for rotation speed. For some other, more dynamic, explosive movements we have measured the energy spectrum bandwidth f (99%) that exceeds 200 Hz, and thus requiring sampling frequency of 500 Hz or even more. All the experiments were performed by the amateurs and it is expected that professional athlete’s movements are even more dynamic.

IV. Real-time Processing of Human Motion

To assure real-time operation of the system, all operations on received data frame must be done within one sampling time, before the next frame arrives. The threshold of real-time operation of the processing device depends on many factors: computational power of the processing device, sampling time, amount of data in one streamed frame, number of algorithms to be performed on the data frame, complexity of algorithms, etc. It is therefore difficult to set an exact threshold or values of each parameter of the processing device.

Delay is the primary parameter defining the concurrency of a biofeedback system, as viewed from the user's perspective. The feedback delay that is the sum of all delays of the technical part of the biofeedback system (sensors, processing device, actuator, communication channels), should not exceed a small portion of the user's reaction delay. To present an exemplary calculation, let us set the sampling frequency at 1000 Hz and maximal feedback delay at 20% of user's reaction delay. Considering that reaction delay is around 150 ms [7], the maximal feedback delay is at most 30 ms.

While many simple examples of biofeedback applications, that do not require huge amounts of processing, exist, one can easily find enough examples of use that do need high performance computing. One such example is a high performance real-time biofeedback system for a football match. Parameters in the capture side of the system are: 22 active players, 3 judges, 10 to 20 inertial sensors per player, 1000 Hz sampling rate. The data includes 3D accelerometer, 3D gyroscope, 3D magnetometer, GPS coordinates, and time stamp. The first three sensors mainly produce 16 bit values in each of the axes, GPS coordinates are 64 bits each, and timestamp is 32 or 64 bit long. Taking the lower values of parameters (10 sensors, 32 bits for time stamp) the data rate produced is 92 Mbit/s. The presented example clearly implies some form of high performance computing and some form of high speed communication, especially when complex algorithms and processes are used on them.

Algorithms that are regularly performed on a streamed sensor signals are: statistical analysis, temporal signal parameters extraction, correlation, convolution, spectrum analysis, orientation calculation, matrix multiplication, etc. Processes include: motion tracking, time-frequency analysis, identification, classification, clustering, etc. Algorithms and processes can be applied in parallel or consecutively.

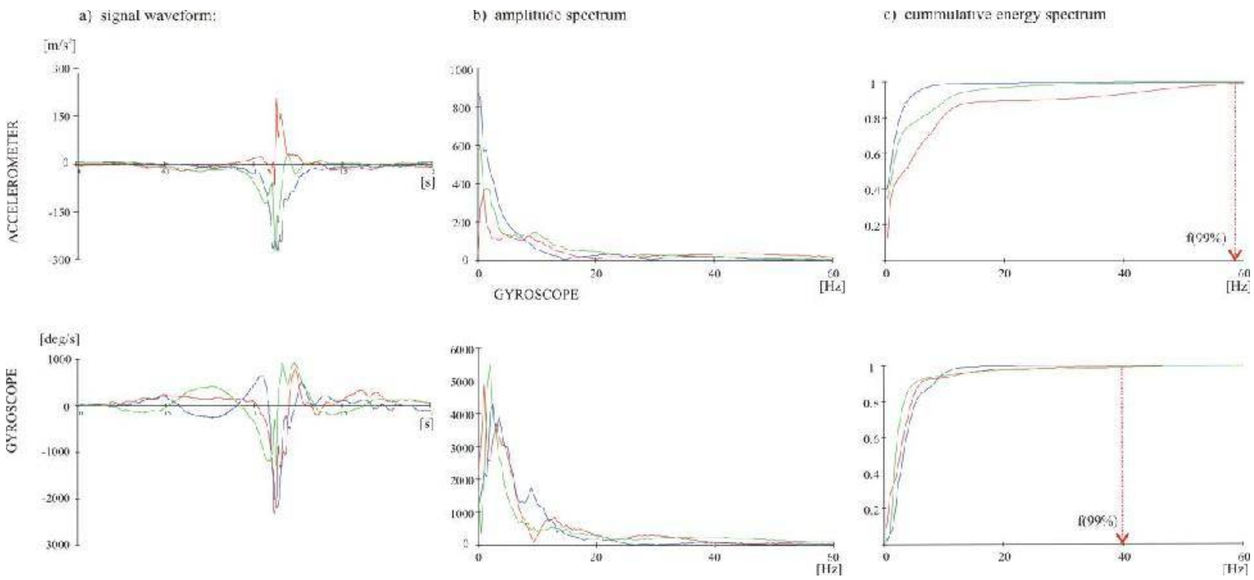


Figure 5. An example of a high dynamic movement: a handball free-throw hand movement measured by a 6DoF sensing device: (a) Accelerometer and gyroscope signals are sampled with 1024 Hz. (b) Signal spectrum (DFT) is calculated on the sequence of 2048 data points inside the 2 s time frame. c) Signal bandwidth is measured and calculated by the relative cumulative energy criterion: f(99%).

V. CONCLUSION

In sport sensory systems signal processing is a crucial process. Many specific biofeedback problems in various sports exist. We addressed several problems, using a different sport example for each one. Sensor accuracy is tested on a golf swing, where we found that the rotation accuracy requirement can be met by smartphone gyroscopes. Movement dynamics on handball free-throw is measured; we found that the sensor dynamic range of professional body attached sensor device hardly meets the experiment requirements. A multi-user signal processing in football match is recognized as an example for high performance application that needs high speed communication and high performance remote computing. With growing number of biofeedback applications in sport and other areas, their complexity and computational demands will grow as well.

REFERENCES

[1] Oonagh M Giggins, Ulrik McCarthy Persson, Brian Caulfield, Biofeedback in rehabilitation. Journal of NeuroEngineering and Rehabilitation, 2013

[2] Sigrist, R., Rauter, G., Riener, R., & Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review.Psychonomic bulletin & review, 20(1), 21-53

[3] Baca, A., Dabnichki, P., Heller, M., & Kornfeind, P. (2009). Ubiquitous computing in sports: A review and analysis. *Journal of Sports Sciences*, 27(12), 1335-1346.

[4] Lauber, B., & Keller, M. (2014). Improving motor performance: Selected aspects of augmented feedback in exercise and health. European journal of sport science, 14(1), 36-43.

[5] Wong, Charence, Zhi-Qiang Zhang, Benny Lo, and Guang-Zhong Yang. "Wearable Sensing for Solid Biomechanics: A Review." *Sensors Journal, IEEE* 15, no. 5 (2015): 2747-2760.

[6] Umek, Anton, Sašo Tomažič, and Anton Kos. "Wearable training system with real-time biofeedback and gesture user interface." Personal and Ubiquitous Computing, 2015, 10.1007/s00779-015-0886-4.

[7] Human Reaction Time, The Great Soviet Encyclopedia, 3rd Edition 1970-1979.

[8] Qualisys Motion Capture System. Available online: http://www.qualisys.com (accessed on 10 September 2015).

[9] Lars Nilsson, QTM Real-time Server Protocol Documentation Version 1.9, Qualisys, 2011, http://qualisys.github.io/rt-protocol/ (accessed on 10 September 2015).

[10] Guna, Jože, Grega Jakus, Matevž Pogačnik, Sašo Tomažič, and Jaka Sodnik. "An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking." *Sensors* 14, no. 2 (2014): 3702-3720.

[11] ST Microelectronics. MEMS motion sensor: ultra-stable three-axis digital output gyroscope, L3G4200D Specifications. ST Microelectronics, December 2010

[12] Stančin, Sara, and Sašo Tomažič. "Angle estimation of simultaneous orthogonal rotations from 3D gyroscope measurements." *Sensors* 11, no. 9 (2011): 8536-8549.

Mapping scheme from Greenstone to CERIF format

Valentin Penca*, Siniša Nikolić *, Dragan Ivanović *

* University of Novi Sad/Faculty of Technical Sciences/Department of Computing and Automatics, Novi Sad, Serbia
{valentin_penca,sinisa_nikolic, chenejac}@uns.ac.rs

Abstract— **This paper describes the basics of the Greenstone institutional repository and CRIS systems and their data models. The result of this research is mapping scheme of the data from Greenstone to the CERIF standard.**

I. INTRODUCTION

Quick development of science and technologies resulted with vast amount of various data. One of the most important tasks is how to preserve and make that data accessible. Institutional Repository (IR) can solve the mentioned issue. In [1], an IR is addressed as an electronic system that captures, preserves and provides access to the digital work products of a community. The three main objectives for having an institutional repository are:

1. creating global visibility and open access for an institution's research output and scholarly materials.
2. collecting and archiving content in a "logically centralized" manner even for physically distributed repositories.
3. storing and preserving other institutional digital assets, including unpublished or otherwise easily lost ("grey") literature.

The availability of open-source technologies affect on the rapid development of IRs worldwide, particularly among academic and research institutions. Thus, it is not surprising the existence of several open-source software platforms available for developing IRs like *Greenstone* (GS) [2], *EPrints* [3], *DSpace* [4], *Fedora* [5] and *Invenio* [6]. GS large popularity in large number of countries is addressed in [7]. Although IRs had been used for a long time, they still don’t have a mutually agreed and standardized representation of their data. This can cause difficulties in data exchange between diverse IR systems

So, to overcome those difficulties in the data exchange, one of the possible solutions is to rely on some predefined standard outside IR domain. Common European Research Information Format (CERIF) standard [8], which is the basis of Current Research Information Systems (CRISs), is used for data exchange from scientific-research domain and can be utilised in IR domain.

In this paper the scheme for mapping data from Greenstone IR to CERIF format is proposed. That scheme can be used as a guideline, supporting the exchange between Greenstone repositories and CRIS systems. Motivation for this work was also to extend and improve research from [9] [10].

II. GREENSTONE IR

The Greenstone digital library software is a system for construction and presentation of information for digital resources. Digital resources can range from newspaper articles to technical documents, from educational journals to oral history, from visual art to folksongs, etc. So, Greenstone is an IR that can contain data which can vary in types and formats. In GS, each single digital resource is described with appropriate metadata and has a link to a stored physical document of that digital resource.

The software provides a way of organizing information and publishing it on the Web in the form of a fully-searchable, metadata-driven digital resource. Greenstone was one the first IR software packages to appear and has been available for more than 15 years. First widespread version was Greenstone 2.0 that was released under the GNU General Public License in September 1999. Greenstone has been developed and distributed in cooperation with UNESCO and the Human Info NGO in Belgium. It is an open-source multilingual Institutional repository software. One of Greenstone's unique strengths is its multilingual nature where their interface is available in over 60 languages [11].

The Greenstone software operates under most variants of Unix (including Linux, FreeBSD and MacOS X) and all versions of Microsoft Windows. Current Greenstone version is "Greenstone3" that is a complete redesign and reimplementation of the original digital library software developed (Greenstone2) back in 2000 and the last stable release came in 9th September, 2015. So, it can be concluded that Greenstone is continually updated and reliable software that has a long successful history of development.

Popularity of software is confirmed thought its usage on worldwide scale. Greenstone software has spread in over 90 counties (e.g. Canada, Germany, New Zealand, Romania, Russia, United Kingdom, United States, etc.). According to the official data presented in [12] there are roughly 3800 active software instances. Popularization of Greenstone has been have been increased by realizing workshops all over the World [13]. Comparison and advantages of Greenstone to other IR repositories is described in [14].

In Greenstone all digital resources (**documents**) are organized into **collections** [Figure 1]. Those collections are additionally organized into a library (called a **site**), and

GS can have multiple sites in a single Greenstone3 installation. A site can be seen as a collection of collections. Site actually provides one upper level of data organization. For e.g. one site can house all of image collections, another site all of pdf collections, and yet another can be used for all of multimedia collections. Or, perhaps sites can be used to separate collections based on topics. Technically GS could have a separate site for each collection (this may not be particularly useful, but is possible).

The collection database dictates how data about a collection is stored. Greenstone comes with two database options: GDBM and JDBM. They are essentially the same. GDBM (or GNU Database Manager) is a simple flat-file database engine which Greenstone uses as the default database for new collections. JDBM is simply a Java implementation of GDBM. It is also possible to use a relational database system, like Microsoft SQL (MS-SQL) or MySql to create a collection's database. The collection database dictates how data about collection is stored.

The physical documents for digital resources are stored as such (PDF, DOC, HTML, XML…). In GS there are different 'Plug-ins' [15] that are used for different types of physical documents in order to extract textual content from those documents. Extraction of words from physical document is essential in digital resource creation. Extracted words represent the basis for constructing metadata (e.g. title, author, creation date, etc.) which is used for describing and indexing of those digital resources. In Greenstone supplying metadata to digital resource is an extremely important part of building digital collections. It provides important contextual and provenance information about that recourse. Essentially, it helps users to navigate collections and find the information/documents they need.

In Greenstone, all metadata fields belong to a metadata set, which is simply a pre-defined collection of metadata fields. Because sets often have metadata fields with the same name (for instance, most sets will have a 'Title' field), namespaces are used to distinguish between metadata from different sets. For instance, all metadata fields in Dublin Core are preceded by word dc followed by dot symbol (e.g. dc.Title, dc.Creator, etc.). In the process of creating a new collection in Greenstone, the default metadata sets that are available are Dublin Core (dc), the Greenstone Metadata Set (gs), and the Extracted Greenstone Metadata Set (ex). The Extracted set is specific because it contains metadata automatically generated during the collection building process and cannot be edited by user. List of all available sets that come with Greenstone installation is found on [16].

For every stored document within Greenstone a XML-format file called metadata.xml is used. That file contains metadata of the electronic resource and applies to all files that are in the same directory, or in one of its child directories.

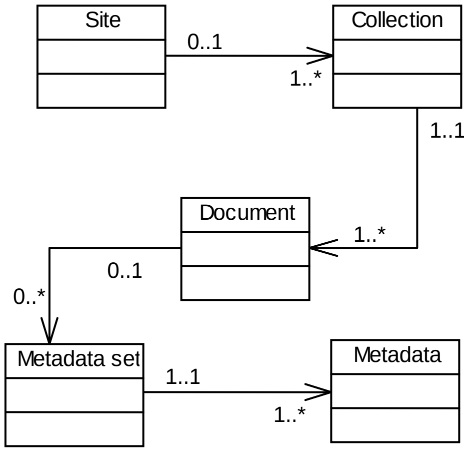


Figure 1 - Greenstone data model

III. CERIF MODEL

CERIF is a standard that describes data model which can be used as a basis for an exchange of data from scientific-research domain. CERIF Standard describes the physical data model [17] and the exchange of XML messages between the CRIS systems [18]. The best feature of CERIF is that it can be expanded and adapted to different needs. In practice, CERIF is often mapped to other standards that also represent the data of scientific-research domain, for example CERIF/MARC21 mapping described in [19]. Authors of [20] recommend an extension of CERIF that incorporates a set of metadata required for storing theses and dissertations. Another example is [21] where authors argue how CERIF can be used as a basis for storage of bibliometric indicators.

Hereinafter we will present main entities of the CERIF data model version 1.5

- Base Entities - represent the core (basic) model entities. There are only three basic entities *cfPerson*, *cfOrganizationUnit* and *cfProject*.
- Result entities - A group of entities which includes results from scientific research like publications, products and patents. Representatives of this group are: *cfResultPublication*, *cfResultProduct* and *cfResultPatent*.
- Infrastructure Entities - represent a set of infrastructure entities that are relevant for scientific research. The entities which belong in this group are: *cfFacility*, *cfEquipment* and *cfService*.
- 2nd Level Entities - Entities that further describe the Base Entities and Result Entities. E.g. *cfMedium* can be physical representation of some Result Entity.

Link Entities - are used to link entities from different groups. Typical entities of this group are: *cfOrganizationUnit_OrganizationUnit*, *cfOrganizationUnit_ResultPublication* and *cfResultPublication_DublinCore*. Link Entities allow for a generic classification mechanism to define their meaning, indicating the role for each entity instance in a relationship. Every Link

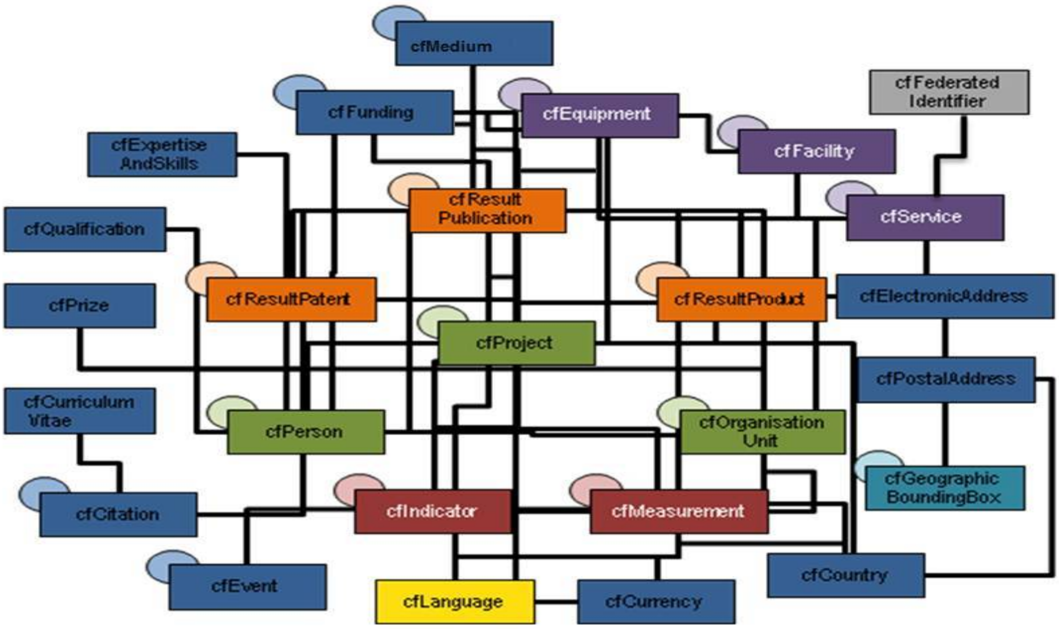


Figure 2 – CERIF model

entity is described with a role (*cfClass*, *cfClassScheme*), timeframe of relation (*cfStartDate*, *cfEndDate*), value (*cfFraction*) and identifiers of elements creating relation (e.g. *cfOrgUnit*, *cfResPublId*). The 'role' in link entities is not stored directly as attribute value, but as reference to Semantic layer.

- Multiple Language Entities - These entities provide multilingualism in CERIF for some entities.
- Semantic Layer Entities - Provide different kinds of semantics in CERIF model. The entities in this group are *cfClassificationScheme* and *cfClassification*. Those entities are used to describe classes and classification schemes for link and other entities. CERIF prescribes a controlled vocabulary to describe some of the classifications.
- Additional Entities - Currently in this group are classified entities that represent DC record.

Figure 2 [Figure 2] shows some of Base, Result, Link and Multiple Language Entities which are relevant for the mapping proposed in this paper.

IV. MAPPING SCHEME FOR GREENSTONE TO CERIF

The motivation of authors for mapping Greenstone data to CERIF data model is found in the fact that they are part of the development team of CRIS UNS system [22], which currently does not have ability to obtain data from the Greenstone system. In the paper [23], a CERIF compatible research management system CRIS UNS is presented, which can be accessed at [22]. Currently, the system stores over 14,500 records of scientific publications (papers published in journals, conference proceedings, monographs, technical solutions and patents etc.). CRIS UNS system is under development since 2008 at the University of Novi Sad in the Republic of Serbia. Former development of that system covered implementation of the system for entering metadata about scientific research results [24]. Later phases in the development of CRIS UNS system included integration of various extensions that relay on CERIF model.

Proposed mapping scheme did not include any customization of Greenstone distribution since the most users use just the default installation.

Greenstone can store various physical documents that came from scientific research domain, , such as journal articles, conference papers, whitepapers, monographs, different forms of reports, etc. Unlike CERIF that has different entities (e.g. *cfResPubl*, *cfResProd*, *cfPers*, etc.) for keeping information about different record types (e.g. publications, products, persons, etc.), GS only has digital recourses in which the record type is not explicitly defined. Therefore, it is essential to derive the data mapping in a manner to comply with the structure and semantics of both models. Mapping of the data from GS model [Figure 1] will begin with key GS entity (Document) which carries the relevant data about the document.

Table 1 shows part of mapping of the GS metadata for digital recourses to the appropriate entities in CERIF. In proposed mapping each digital resource from GS is represented as an instance of CERIF *cfMedium* entity, in accordance to EuroCRIS suggestion [25]. “GS Metadata Sets” column indicates which metadata set form GS is used for a particular metadata that is displayed in the column “GS Metadata Element”. Columns “multiple” give information whether metadata may appear more than once in the GS and/or model CERIF.

One of the simplest examples of metadata mapping can be shown on document title which is represented in CERIF with the concrete value of title that is stored in *cfTitle* attribute of entity *cfMediumTitle* which is certainly linked with *cfMedium*. Representation of the organization that is resource owner requires creation of CERIF entity *cfOrgUnit* and *cfOrgUnit* Name which will store a name

TABLE I.
GREENSTONE DOCUMENTS METADATA FIELDS

GS Metadata Sets	GS Metadata Element	multiple	CERIF core result 2nd level entites	CERIF Link Enities	Used CERIF Classification	multiple
DLS	Title	X	cfMedium (cfMediumId) cfMediumTitle (cfMediumId, cfTitle, cfTrans="l")			X
DLS	Organization	X	cfMedium (cfMediumId) cfOrgUnit (cfOrgUnitId) cfOrgUnitName (cfOrgUnitId, cfName)	cfOrgUnit_Medium (cfOrgUnitID, cfMediumId)	Scheme:CERIF Entites Class: Organization	X
DLS	Language	X	cfDC(cfDCId, cfDCSchema="DLS") cfDCLanguage(cfDCId, cfDCSchema="DLS", cfDCValue)	cfMedium_DC (cfMediumId, cfDCID)	Scheme:Medium Metadata Class:has metadata	X
DCE	Relation.Has Part	X		cfMedium_Medium (cfMediumId1, cfMediumId2)	Scheme:Inter-Medium Realations Class:has part	

for organization/institution. A connection between the resource (*cfMedium*) and organization/institution (*cfOrgUnit*) in CERIF is implemented with link entity *cfOrgUnit_Medium* (column CERIF Link Enities). CERIF model relies on Link and Semantic Layer Entities to provide additional semantic between entities and for some particular entities. So, it is to be assumed that a large portion of metadata fields from GS documents will be stored as instances of those CERIF entities. In our case it is necessary to classify the link entity *cfOrgUnit_Medium* with appropriate classification scheme and class. (column Used CERIF Classification). Some of CERIF link entities and classification can be connected with the identifier (e.g. *cfMediumId*, *cfOrgUnitId*).

A certain metadata (e.g. *Language*) from GS cannot be directly mapped to *cfMedium* and/or its link entities from CERIF model. However, in CERIF there is a group of "Additional Entities" that is used for the purpose of mapping metadata that is coming from other external systems. These CERIF entities are built on model of a widely known DC metadata set [26]. So, language value will be stored in attribute *cfDCValue* of *cfDC_Language* entity. In order to adequately map mentioned metadata (e.g. *Language*) it was needed to add only one linking entity element (*cfMedium_DC*) between the *cfDC* and a *cfMedium*. Also, it was essential to use the appropriate classification (column used CERIF Classification) which provides the semantic for that relation.

A specific case is the mapping metadata that indicates a logical connection between resources. So, metadata *Relation.HasPart* from the extended DC metadata set [27] is actually defining the link between two physical resources (*cfMedium*), where the relationship is represented with entity *cfMedium_Medium* that is classified with the appropriate classification scheme (Inter-Medium Realations) and class (has part).

As previously mentioned in the description of the GS model, several electronic resources can form a single collection [Figure 1]. Metadata that describes the collection is stored in file "collectionConfig.xml". Table 2 shows a part of metadata mapping for GS collection. Keeping in mind that collection in GS represent only a hierarchical level of organization for electronic resources, collections in CERIF are represented with CERIF semantic layer. For that purpose a new classification scheme (*cfClassScheme*) *GS_Collection* is defined, and each collection is stored as a new classification (*cfClass*).Each GS collection has its own name that can be multilingual. The collection name is preserved in *cfTerm* attribute of *cfClassTerm* entity Language in which the name is stored and preserved via attribute *cfLangCode*. Metadata field "description" is mapped to an entity *cfClassDescription*, in similar manner as metadata element "name" to *cfClassTerm*. Metadata creator of the collection can be equally mapped to either *cfPers* or *cfOrg_Unit*, depending on the contents of the field. Collection is assigned to person or organisation with entities *cfPers_Class* or *cfOrgUnit_Class*. The public status of GS collection is defined in metadata field

TABLE II.
GREENSTONE COLLECTION METADATA FIELDS

GS Collection XML Element	multiple	CERIF	Used CERIF Classification	multiple
collection		cfClassScheme, cfClass	Scheme:GS_Collection Class: NEW CLASS INSTANCE	
"name" "lang"	X	cfClass, cfClassTerm (cfTerm, cfLangCode)		X
"description" "lang"	X	cfPers, cfPersName (cfFirstName, cfLastName, cfOtherName) cfPers_Class cfOrgUnit cfOrgUnitClass		X
"creator" "lang"	X	cfPers cfPersName (cfFirstName, cfLastName, cfOtherName) cfPers_Class cfOrgUnit cfOrgUnitClass	Scheme:Inter-Medium Realations e Class:has part	
public	X	cfClassScheme cfClass cfClass_Class (cfClassId,cfClassShemeid, cfClassId1, cfClassShemeid1, cfClassShemeid2, cfClassId2)	Scheme:GS_Collection_TYPE Class: Public Access or Private Access Scheme: GS_Collection_Type Class: Public Access or Private Access e Scheme: Inter_Collection_Realation Class has relation	

TABLE III.
GREENSTONE SITE METADATA FIELDS

GS Collection XML Element	multiple	CERIF	Used CERIF Classification	multiple
site		cfClassScheme cfClass cfClass_Class (cfClassId, cfClassShemeid, cfClassId1, cfClassShemeid1, cfClassShemeid2, cfClassId2)	Scheme:GS_Collection Class: EXISTING INSTANCE Scheme: GS_Site Class: NEW INSTANCE Scheme: Inter_Site_Realation Class has relation	
"name" "lang"	X	cfClass cfClassTerm (cfTerm, cfLangCode)		X
"description" "lang"	X	cfClass cfClassDescr (cfTerm, cfLangCode)		X
"creator" "lang"	X	cfPers cfPersName (cfFirstName, cfLastName, cfOtherName) cfPers_Class cfOrgUnit cfOrgUnitClass	Scheme:GS_Collection Class: EXISTING CLASS INSTANCE	

"public" in GS. In CERIF model the mentioned GS metadata information requires a creation of a new classification scheme (*GS_Collection_TYPE*) and class (Access Public or Private Access). Also, it was necessary to link that new classification to one that that represent the collection. The interconnection between two classifications in CERIF is achieved with link entity *cfClass_Class*. Assigning a digital recourse to collection is achieved by classifying *cfMedium* instances with a newly created CERIF classes that are GS collections.

As mentioned before multiple collections may be additionally organized within entities called "sites". (Figure 1). Each site is described with particular metadata that is stored in file siteConfig.xml. Similar to collections, the mapping of site (Table 3) entity can be done by relying on CERIF semantic layer. The main role of the site entity is to describe a group of collection. Like a collection site has the multilingual metadata name, description, and creator that can mapped on the model CERIF in a same manner as GS collection. All Greenstone mapping are presented on [28].

The main purpose of the presented scheme is mapping data from GS to a concrete CRIS UNS. The mere fact is that GS can export/imported data to/in certain format. According to that, this can open up a couple of potential opportunities for transferring data from one system to another. The first potential opportunity is the fact that it has the ability to export data by MARCXMLPlugin. CRIS UNS has MARC21 compatible data model [19] that could enable mapping the exported data from the GS. Exported data to DSpace model is also acceptable for CRIS UNS considering that the authors have already suggested mapping of DSpace data to CERIF [9].

There is another potential direction of interoperability between these two systems, from CRIS UNS to GS, where Greenstone is able to obtain data from CRIS UNS system. As a matter of fact, the Greenstone has SRU/W client so it can possible connect to CRIS UNS, considering that the system provides the data in SRW XML format in accordance to the specific profile [29]. Import of such data can be easily accomplished with GS SRU client and adequate plug-in in GS [30]. Some type of data (doctoral dissertation) from CRIS UNS that is available in OAIPMH format [31], can be useful for GS that have OAIPMH import plug-ins. Last but not the least, Greenstone can import a data (records) directly from the database (MySQL) using DatabasePlugin (DBPlug). The exported data from the CRIS UNS MySQL database could serve as a starting point for DBPlug. Greenstone has a large number of plugins [15] to import different kinds of data, so it has the potential to take over data from other systems that are not only CERIF like.

V. CONCLUSION

The importance of institutional repositories and CRIS systems for scientific research data is enormous. Making data accessible between these systems is unavoidable. Therefore, this paper presents a mapping scheme for Greenstone data to CERIF model where all GS data is available to CRIF like systems (CRISs, IRs that support CERIF, etc.).

The main contributions of this research are:

• Proposal for mapping data from Greenstone repository to the current 1.5 CERIF model

• Potential possibility for creation Import/Export plug-in for making full interoperability between this systems.

Future work will be directed towards mapping the data from other IRs like Fedora, and Invenio to CERIF format.

ACKNOWLEDGMENT

Results presented in this paper are part of the research conducted within the Grant No. III-47003, Ministry of Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] N. F. Foster and S. Gibbons, "Understanding faculty to improve content recruitment for institutional repositories," -Lib Mag., vol. 11, no. 1, pp. 1–12, 2005.

[2] "Welcome :: Greenstone Digital Library Software." [Online]. Available: http://www.greenstone.org/. [Accessed: 20-Dec-2014].

[3] "EPrints - Digital Repository Software." [Online]. Available: http://www.eprints.org/. [Accessed: 20-Dec-2014].

[4] "DSpace | DSpace is a turnkey institutional repository application." [Online]. Available: http://www.dspace.org/. [Accessed: 20-Dec-2014].

[5] "Fedora Repository | Fedora is a general-purpose, open-source digital object repository system." [Online].

Available: http://fedora-commons.org/. [Accessed: 20-Dec-2014].

[6] “Invenio.” [Online]. Available: http://invenio-software.org/. [Accessed: 20-Dec-2014].

[7] I. H. Witten and D. Bainbridge, “Creating digital library collections with Greenstone,” Libr. Hi Tech, vol. 23, no. 4, pp. 541–560, Dec. 2005.

[8] “Common European Research Information Format | CERIF,” 2000. [Online]. Available: http://www.eurocris.org/. [Accessed: 18-Jan-2014].

[9] V. Penca and S. Nikolić, “Scheme for mapping Published Research Results from Dspace to Cerif Format,” in 2. International Conference on Information Society Technology and Management, 2012, pp. 170–175.

[10] Valentin Penca, Siniša Nikolić, and Dragan Ivanović, “Scheme for mapping scientific research data from EPrints to CERIF format,” 2015.

[11] “Greenstone Language Support.” [Online]. Available: http://wiki.greenstone.org/doku.php?id=en:language_support.

[12] “Greenstone Factsheet.” [Online]. Available: http://www.greenstone.org/factsheet.

[13] “Greenstone Workshop Map.” [Online]. Available: http://www.greenstone.org/map.

[14] Y.-L. Theng, S. Foo, D. Goh, and J.-C. Na, Eds., Handbook of Research on Digital Libraries: Design, Development, and Impact. IGI Global, 2009.

[15] “Greensone plugins.” [Online]. Available: http://wiki.greenstone.org/doku.php?id=en:plugin:index.

[16] “Greensone Metadata sets.” [Online]. Available: http://wiki.greenstone.org/doku.php?id=en:user:metadata_sets.

[17] B. Jörg, K. Jeffery, J. Dvorak, N. Houssos, A. Asserson, G. van Grootel, R. Gartner, M. Cox, H. Rasmussen, T. Vestdam, L. Strijbosch, V. Brasse, D. Zendulkova, T. Höllrigl, L. Valkovic, A. Engfer, M. Jägerhorn, M. Mahey, N. Brennan, M. A. Sicilia, I. Ruiz-Rube, D. Baker, K. Evans, A. Price, and M. Zielinski, CERIF 1.3 Full Data Model (FDM) Introduction and Specification. 2012.

[18] J. Dvořák and B. Jörg, “CERIF 1.5 XML - Data Exchange Format Specification,” 2013, p. 16.

[19] D. Ivanović, D. Surla, and Z. Konjović, “CERIF compatible data model based on MARC 21 format,” Electron. Libr., vol. 29, pp. 52–70, 2011.

[20] L. Ivanovic, D. Ivanovic, and D. Surla, “A data model of theses and dissertations compatible with CERIF, Dublin Core and EDT-MS,” Online Inf. Rev., vol. 36, no. 4, pp. 548–567, 2012.

[21] S. Nikolić, V. Penca, and D. Ivanović, “STORING OF BIBLIOMETRIC INDICATORS IN CERIF DATA MODEL,” Kopaonik mountain resort, Republic of Serbia, 2013.

[22] “Current Research Information System of University of Novi Sad.” [Online]. Available: http://www.cris.uns.ac.rs/. [Accessed: 18-Jan-2014].

[23] D. Surla, D. Ivanovic, and Z. Konjovic, “Development of the software system CRIS UNS,” in Proceedings of the 11th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, 2013, pp. 111–116.

[24] D. Ivanović, G. Milosavljević, B. Milosavljević, and D. Surla, “A CERIF-compatible research management system based on the MARC 21 format,” Program Electron. Libr. Inf. Syst., vol. 44, no. 3, pp. 229–251, 2010.

[25] Houssos, Nikos, Jörg, Brigitte, and Matthews, Brian, “A multi-level metadata approach for a Public Sector Information data infrastructure,” presented at the CRIS2012 Conference, 2012.

[26] “DCMI Home: Dublin Core® Metadata Initiative (DCMI).” [Online]. Available: http://dublincore.org/. [Accessed: 11-Sep-2014].

[27] “Dublin Core Extended.” [Online]. Available: http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/.

[28] Valentin Penca, Siniša Nikolić, and Dragan Ivanović, “GreenStone Mapping Full.” [Online]. Available: https://drive.google.com/open?id=19Cni775c2nWubbFTlOtgUGm4PkoI3vrAS3NHl4OjD0Y.

[29] V. Penca, S. Nikolić, D. Ivanović, Z. Konjović, and D. Surla, “SRU/W Based CRIS Systems Search Profile,” Program Electron. Libr. Inf. Syst., 2014.

[30] “Greenstone SRU plugin.” [Online]. Available: http://wiki.greenstone.org/doku.php?id=en:user_advanced:z3950#download_through_sru.

[31] L. Ivanović, D. Ivanović, and D. Surla, “Integration of a Research Management System and an OAI-PMH Compatible ETDs Repository at the University of Novi Sad,” Libr. Resources Tech. Serv., vol. 56, no. 2, pp. 104–112, 2012.

THE ROLE OF PIVOT METAMODEL IN AUTOMATIC GENERATION OF METAMODEL EXTENSIONS

Igor Zečević, Branko Perišić, Petar Bjeljac, Danijel Venus
Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
{igor.zecevic, perisic, pbjeljac, danijelvenus}@uns.ac.rs

Abstract – **Until recently, the approaches of creating domain specific languages "from scratch" and extending existing languages represented adverse solutions that cannot be integrated in a unique model driven process. The main reason was the lack of a suitable formal theoretical framework that, based on the domain specific languages metamodel, enables the automatic generation of a final number of syntactically valid metamodel extensions. In this paper, the use of the Pivot metamodel concept is described. This metamodel enables the transformation of elements from one metamodel into extensions of another, regardless the compatibility level of their meta-metamodel. The rules and restrictions for transforming random models into the Pivot metamodel are also presented.**

I. Introduction

One of the main goals in the Model Driven Engineering (MDE) approach to software development [1] is the specification of Domain Specific Languages (DSLs) [2] which enable a semantically rich description of a designed software product at high abstraction level. Among many approaches to DSL development, two are dominant:

- adapting (extending) existing languages, that are commonly used in concrete problem domains [3] or
- creating completely new modeling languages, based on the characteristics of the application domain and the base rules for formal languages specification [4].

Until recently, in practice these two approaches were considered adverse and impossible to synchronize in a unique MDE process [5]. This standpoint was justified by the lack of a methodology that could provide a good enough level of interoperability for different approaches, methods, techniques and tools. It is possible to establish the synergic use of these two approaches through a formal theoretical framework that, based on the DSL metamodel, supports the automatic generation of a final number of syntactically valid metamodel extensions. The implementation of this formal theoretical framework, in the form of a workspace, enables:

- Interoperability of different meta-modeling tools based on the transformations established between their meta-models;
- Automatic model transformations, for models representing the instances of meta-models for which transformations have been defined;
- The integration of artifacts originating from models, which represent instances of meta-models for which transformations have been defined.

The specification of a formal theoretical framework requires the definition of rules that support the transformation of the DSL metamodel elements into extensions of existing modeling languages meta-models. Based on analysis of published research [6, 7, 8, 9, 10, 11, 12, 13], it is impossible to completely automate the transformation process between meta-models that are derived from different meta-meta-models. In general, there exist a large number of different combinations between meta-models derived from different meta-meta-models. This makes the individual transformations an exhausting activity that is hardly economically justifiable.

In this paper, an approach that raises the level of transformation automation based on a mediator (Pivot metamodel) is described. Using this approach, it is only necessary to establish a two-way transformation between a metamodel and its Pivot metamodel. In order to fulfill their role, in the formal theoretical framework context, the Pivot meta-models need to have the capacity to define any arbitrary metamodel that is the subject of transformation.

In this paper, the Pivot metamodel is described in the context of a formal theoretical framework for generating metamodel extensions. Furthermore, the rules and restrictions for transforming meta-models into the Pivot metamodel are presented.

The paper is organized as follows: in Section 2, the definition of models in a three-level model architecture and the concept of Technological spaces are presented. In Section 3, the theoretical framework for metamodel extension generation is described. The definition of the Pivot metamodel and the rules that guide the transformation of arbitrary metamodel into the Pivot metamodel are presented in Section 4. In section 5, based on the mappings established on meta-metamodel level, the conditions for generating the Pivot metamodel are described. In Section 6, an example of transforming a simple metamodel into the Pivot metamodel is presented. Section 7 concludes the paper and describes the directions of future work.

II. Technological spaces and metelevel based architectures

From an organization point of view a model can be defined as follows [2]:

Definition 2.1*:* A directed multigraph $\mathrm{G} = (\mathrm{N_G}, \mathrm{E_G}, \mathrm{F_G})$ consists of a finite set of nodes $\mathrm{N_G}$, final set of edges $\mathrm{E_G}$ and function $\mathrm{F_G} : \mathrm{E_G} \rightarrow \mathrm{N_G} \times \mathrm{N_G}$ which maps edges to their source and target nodes.

Definition 2.2: A model $\mathrm{M} = (\mathrm{G}, \mathrm{M_A}, \mu)$ is an ordered triplet, where $\mathrm{G} = (\mathrm{N_G}, \mathrm{E_G}, \mathrm{F_G})$ is a directed multigraph, $\mathrm{M_A}$ is a model (called the referential model), which is defined using a directed multigraph $\mathrm{G_A} = (\mathrm{N_A}, \mathrm{E_A}, \mathrm{F_A})$, $\mu : \mathrm{N_G} \cup \mathrm{E_G} \rightarrow \mathrm{N_A}$ is a function, which transforms elements (nodes and edges) of G multigraph of the model M into nodes of the $\mathrm{G_A}$ multigraph of the model $\mathrm{M_A}$.

The relation between model M and referential model M_A is called conformance (conformsTo). The elements of metamodel M_A (nodes N_A and edges E_A) are called metaelements.

This definition of a model allows an unlimited number of referential model levels. By restricting the number of levels to three, meta-metamodel and metamodel may be defined as follows:

Definition 2.3: A meta-metamodel is a model that represents its own reference model (it conforms to itself).

Definition 2.4: A metamodel is a model such that its reference model is a metametamodel (it conforms to the meta-metamodel).

In a three-level model architecture, levels are usually marked as M1, M2 and M3. M1 level contains models which confirm to concepts of M2 level and are not metamodels. M2 level contains models which are not meta-metamodels and confirm to M3 concepts, while the M3 level is self-defined and contains meta-metamodels.

Technological spaces (TS) is a concept defined by Kurtev et al. in a discussion on problems during the integration of different technologies [14]. The technological spaces concept was initially defined as “a working context with a set of associated concepts, body of knowledge, tools, required skills, and possibilities”. Certain technological spaces can be intuitively recognized. It is possible to isolate technological spaces such as executable programming languages TS (JAVA, C#), database management system TS or modeling framework TS (UML, MDA/OMG). In Fig. 2 the three-level model organization for two different TS is presented.

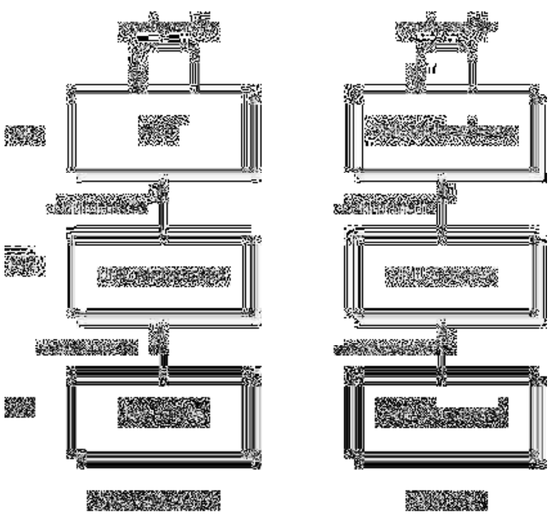


Figure 2. MDA and XML technological spaces [14]

III. FORMAL THEORETICAL FRAMEWORK FOR GENERATING METAMODEL EXTENSIONS

The aim of creating a formal theoretical framework for generating metamodel extensions is to enable a final number of syntactically valid metamodel extensions based on another metamodel. The metamodel from which these extensions are derived represents the source metamodel, and the metamodel subject to the extension represents the destination metamodel. It is assumed that the source and destination metamodel are both part of technological spaces that have a three-level meta-architecture (Fig. 1).

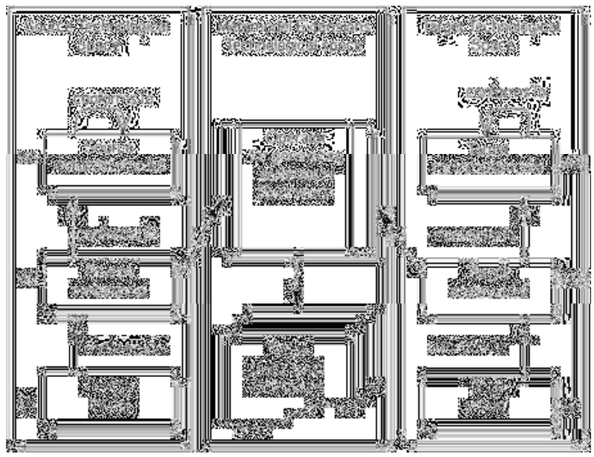


Figure 1. Formal theoretical framework for generating metamodel extensions

The formal theoretical framework should support the use of different TS-s, allow different TS-s for the source and destination metamodel, as well as to allow the source and destination metamodel to conform to different meta-metamodels.

A concrete implementation of the theoretical framework, containing these characteristics, supports interoperability between different MDE implementation strategies [4, 15, 16] (in the case of different source and destination TS), as well as generating metamodel extensions based on other metamodels (in the case of different source and destination meta-metamodels).

In this formal theoretical framework, the source and destination metamodels are presented using the KM3 metamodel definition language [17]. The purpose of KM3 is to give a relatively simple solution to define the Domain Definition MetaModel of a DSL. It is based on the MOF meta-metamodel [18] and the ECORE meta-metamodel [19]. The source metamodel in the KM3 format represents the source Pivot metamodel, whereas the destination metamodel in the KM3 format represents the destination Pivot metamodel. The Pivot metamodel can be any source or destination pivot metamodel. Each Pivot metamodel conforms to the KM3 meta-metamodel (Fig. 3).

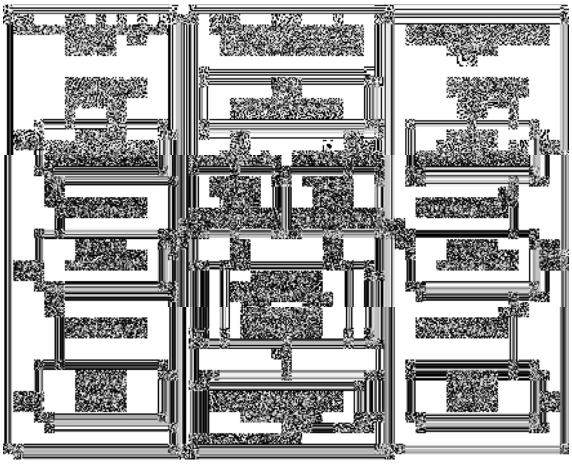


Figure 3. Pivot metamodels in the formal theoretical framework for generating metamodel extensions

IV. PIVOT METAMODELS

According to the definition, every KM3 metamodel is a metamodel. However, not every KM3 model is a Pivot metamodel. Only metamodels from the M2 level can be Pivot metamodels, and KM3 metamodels need to be additionally restricted in order to represent valid pivot metamodels.

Definition 4.1: Pivot metamodel is a KM3 model which represents a metamodel from the three-level metalevel hierarchy.

Based on the definitions of models in different meta levels (Def. 2.2), a metamodel in the three level metalevel hierarchy can be represented as an ordered triplet:

$$M_{MM} = (G_{MM}, M_{MMM}, \mu_{MM}) \tag{1}$$

where $G_{MM} = (N_{MM}, E_{MM}, F_{MM})$ is a directional multigraph containing elements from the M_{MM} metamodel (nodes N_{MM} and edges E_{MM}), $M_{MMM} = (G_{MMM}, M_{MMM}, \mu_{MMM})$ is the referential model for the M_{MM} metamodel (meta-metamodel), and μ_{MM} is a function which transforms elements (nodes and edges) from $\mathrm{G_{MM}}$ multigraph into nodes of $\mathrm{G_{MMM}}$ multigraph ("conformsTo" relation).

Pivot metamodel:

$$piv(M_{MM}) = (G_{PIV}, M_{KM3}, \mu_{KM3}, \omega_{PIV}) \tag{2}$$

is an ordered quadruplet, where $G_{PIV} = (N_{PIV}, E_{PIV}, F_{PIV})$ is a directional multigraph, which contains elements (nodes N_{PIV} and edges E_{PIV}), M_{KM3} is a KM3 metamodel, $\mu_{\mathrm{KM3}}: N_{PIV} \cup E_{PIV} \rightarrow N_{KM3}$ is a function which transforms elements of the $\mathrm{G_{PIV}}$ multigraph into nodes of $\mathrm{G_{KM3}}$ multigraph, and $\omega_{PIV} : G_{MM} \rightarrow G_{PIV}$ is a bijective function, which transforms elements from G_{MM} multigraph into elements of G_{PIV}, multigraph (Fig. 4).

In general, in order for a model to be considered a Pivot metamodel, two conditions need to be met. First, the model needs to conform to the KM3 metamodel. Second, in a three tier metalevel architecture, a metamodel needs to exist such that all elements from this metamodel have exactly one matching element in the Pivot metamodel and vice versa. The KM3 model presented in Fig. 7 represents a metamodel for the faculty organizational structure.

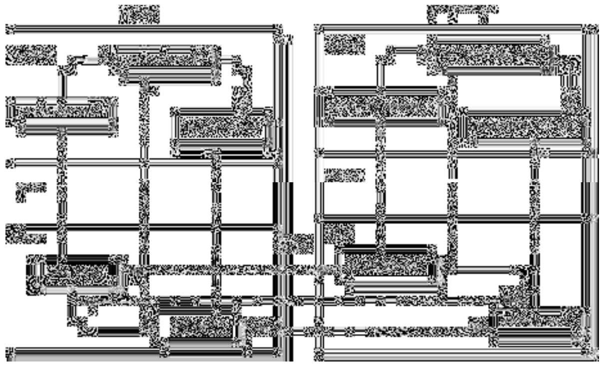


Figure 4. Pivot metamodel

For M_{MM} metamodel, with a total of $|\mathrm{G_{MM}}|$ elements, there are at most

$$|piv(M_{MM})| = |\omega_{PIV}| * |\mu_{\mathrm{KM3}}| \tag{3}$$

pivot metamodels. Function ω_{PIV} transforms each element from the M_{MM} metamodel into exactly one element of the pivot metamodel, and there is a total of $|\omega_{PIV}| = |\mathrm{G_{MM}}|!$ different bijective functions, whereas each of the pivot metamodels contains a total of $|G_{PIV}| = |G_{MM}|$ elements. On the other hand, function μ_{KM3} defines an element of the KM3 meta-metamodel to which each element of the pivot metamodel conforms.

For a pivot metamodel with a total of $|G_{PIV}|$ elements and $|N_{KM3}|$ nodes in the KM3 metamodel, there are at most

$$|\mu_{\mathrm{KM3}}| = |N_{KM3}|^{|G_{PIV}|} \tag{4}$$

different functions μ_{KM3}. Finally for M_{MM} metamodel with $|\mathrm{G_{MM}}|$ elements there are at most:

$$|piv(M_{MM})| = |\mathrm{G_{MM}}|! * |N_{\mathrm{KM3}}|^{|G_{MM}|} \tag{5}$$

different pivot metamodels. Depending on the abstract and concrete M_{MM}, metamodel syntax, the total number of possible pivot metamodels can be smaller. Based on (5), the maximum number of different pivot metamodels for a certain metamodel has an exponential rise in relation to the number of metamodel elements $|G_{MM}|$.

Definition 4.2: The source pivot metamodel is the pivot metamodel of the source metamodel.

Definition 4.3: The destination pivot metamodel is the pivot metamodel of the destination metamodel.

In order to be able to transform a metamodel into a KM3 model, the model needs to have mappings from each element of the metamodel to an element in the KM3 metamodel, which conforms to the KM3 meta-metamodel. Based on these mappings, it is possible to define a transformation of a metamodel into a KM3 model. This means that a new transformation needs to be defined each time a new source or destination metamodel,

which needs to be represented in the form of a pivot metamodel, is introduced.

V. USING M3B APPROACH IN GENERATING PIVOT METAMODELS

Using the M3B approach [20,21] it is possible to reach a certain level of automation in transforming metamodels into KM3 models. Using this approach, it is possible to automatically generate model transformations on lower metalevels, based on established mappings between elements of models on the highest metalevel.

By establishing the mapping between meta-metamodels to which the source or destination metamodel conform, and the KM3 meta-metamodel, each metamodel which conforms to this metamodel can be transformed into KM3 metamodel (Fig. 6). In order to make mappings between the meta-metamodel and the KM3 meta-metamodel, it is required that each element in the meta-metamodel, has a matching element in the KM3 meta-metamodel. This means that when a new source or destination meta-metamodel is introduced, a new transformation needs to be defined.

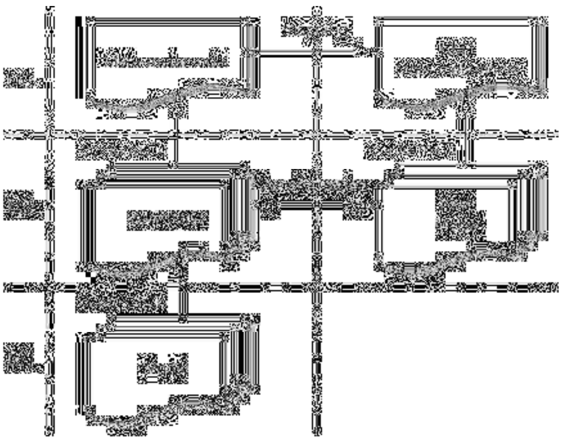


Figure 6. M3B approach in the design of Pivot metamodel

The MB3 approach significantly reduces the number of transformations that need to be implemented in order to represent the source and destination metamodel in KM3 format. Instead of defining new transformations for each new metamodel, only one transformation on meta-metamodel level needs to be defined. If there are two technological spaces with different meta-metamodels, one only needs to define two transformations enabling the automatic transformation of compliant metamodels into KM3 format. As a consequence, meta-metamodels consist of a smaller number of elements than metamodels, thereby simplifying the transformation process.

By establishing correspondence between a meta-metamodel to which a metamodel conforms and the KM3 meta-metamodel it is possible to reduce the number of pivot metamodels. This correspondence can be presented as a function $\omega_{KM3} : G_{MMM} \rightarrow G_{KM3}$, which transforms each element from the G_{MMM} multigraph into exactly one element in the G_{KM3} multigraph.

For a single function ω_{KM3}, there is exactly one pivot metamodel $piv(M_{MM})$ for each metamodel M_{MM} which conforms to the M_{MMM} meta-metamodel. Namely, if there is exactly one matching element in the KM3 meta-metamodel for each element in the meta-metamodel, then each element from a metamodel which conforms to the element in the meta-metamodel has a matching element in the pivot model which conforms to the same element in the KM3 meta-metamodel.

For M_{MMM} metamodel with a total of $|G_{MMM}|$ elements, there are at most

$$|\omega_{KM3}| = |G_{KM3}|^{|G_{MMM}|} \tag{6}$$

pivot metamodels.

Based on this formula, we can conclude that the maximum number of pivot metamodels for a metamodel depends on the number of elements in the meta-metamodel to which this metamodel conforms.

VI. AN EXAMPLE OF PIVOT METAMODEL CREATION

In Fig. 5 a metamodel for a simple faculty organization structure is presented. The metamodel conforms to ECORE meta-metamodel.

The fact that KM3 meta-metamodel is based on ECORE and MOF concepts simplifies transformations of metamodels, which conform to these meta-metamodels,

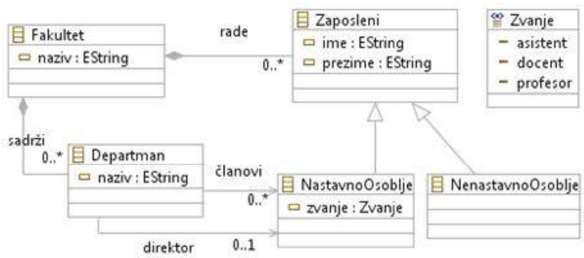


Figure 5. Metamodel for a simple faculty organization structure

into KM3 models. Each class in the ECORE metamodel (Faculty, Department, Employee, Teaching staff, Technical staff) is transformed into an element which conforms to the Class element of the KM3 meta-metamodel. The ECORE metamodel primitive type (EString) is transformed into an element which conforms to the DataType metaelement in the KM3 meta-metamodel. Each ECORE attribute is transformed into an element which conforms to the metaelement Attribute in the KM3 meta-metamodel, and belongs to the Class element, which is an equivalent to the ECORE class, to which the attribute belongs.

The type of the KM3 attribute is a matching DataType element (if it is a primitive type), or a matching KM3 class. The same transformation rules can be applied to ECORE references. Class inheritance from the ECORE metamodel is represented by an attribute from a matching KM3 class. Finally, ECORE enumeration (Title) is transformed into an element of the KM3 metamodel, which conforms to the Enumeration metaelement of the KM3 meta-metamodel. The Pivot metamodel representing a complete metamodel of the faculty organizational structure is presented in Fig 7.

VII. CONCLUSION

In this paper, the Pivot metamodel concept in the context of a formal theoretical framework for generating

```
package OrganizacionaStruktura{
	datatype String;
	enumeration Zvanje{
		literal asistent;
		literal docent;
		literal profesor;
	}
	class Fakultet{
		attribute naziv [1-1] : String;
		reference sadrzi [*] container: Departman;
		reference rade [*] container: Zaposleni;
	}
	class Departman{
		attribute naziv [1-1] : String;
		reference clanovi [*] : NastavnoOsoblje;
		reference direktor [0-1] : NastavnoOsoblje;
	}
	abstract class Zaposleni{
		attribute ime [1-1]: String;
		attribute prezime [1-1]: String;
	}
	class NastavnoOsoblje extends Zaposleni{
		attribute zvanje [1-1] : Zvanje;

	}
	class NenastavnoOsoblje extends Zaposleni{
	}
}
```

Figure 7. Pivot metamodel for the faculty organizational structure

metamodel extensions is presented. This formal theoretical framework allows the generation of a finite number of syntactically valid extensions of a metamodel (destination metamodel), based on another metamodel (source metamodel).

Pivot metamodels have a mediator role in this formal theoretical framework. These metamodels enable transformation of elements from one metamodel into extensions of another, regardless the meta-metamodels to which they conform. By using strictly defined rules and limitations, Pivot metamodels allow the description of an arbitrary metamodel that needs to be transformed in the formal theoretical framework context. Pivot metamodels are represented in the form of KM3, a textual language for DSL metamodels specification.

Apart from the definition of the Pivot metamodel, rules and limitations for transforming any metamodel into a Pivot metamodel are formulated. Furthermore, a possibility of reaching a certain level of automation in the process of transforming metamodels into KM3 models, based on the M3B approach, is described. Pivot metamodels can be formed directly (by transforming metamodels into KM3 format), or indirectly (using transformations based on correspondences between appropriate meta-metamodels).

In the concrete implementation of the formal theoretical framework, metamodels are formed indirectly, based on mappings between the source meta-metamodel and the KM3 meta-metamodel. The main reason is a potentially unlimited number of different source metamodels which conform to one source meta-metamodel. In practice, only a relatively small number of destination metamodels are subject to extending. Due to this fact, the concrete implementation of the formal theoretical framework contains a finite set of prepared destination pivot metamodels.

The direction of future work is focused the definition of transformations between a certain number of meta-models and KM3 meta-metamodel. This would enable automatic transformation of arbitrary metamodels which conform to selected meta-metamodels into a set of Pivot metamodels.

REFERENCES

[1] Atkinson, C. and Kühne, T., Model-driven development: a metamodeling foundation. Software, IEEE, 20(5), pp.36-41, 2003.

[2] Kurtev, I., Bézivin, J., Jouault, F. and Valduriez, P., 2006, October. Model-based DSL frameworks. In Companion to the 21st ACM SIGPLAN symposium on Object-oriented programming systems, languages, and applications (pp. 602-616). ACM.

[3] Fuentes-Fernández, L., Vallecillo, A.: An Introduction to UML Profiles. The European Journal for the Informatics Professional (UPGRADE), vol. 5, pp. 5–13, 2004.

[4] Luoma, J., Kelly, S., Tolvanen, J.-P.: Defining Domain-Specific Modeling Languages: Collected Experiences. 4th OOPSLA Workshop on Domain-Specific Modeling (DSM'04), 2004.

[5] Watson, A.: UML vs. DSLs: A false dichotomy. Whitepaper, 2008.

[6] Abouzahra, A., Bézivin, J., Del Fabro, M.D. and Jouault, F.: October. A practical approach to bridging domain specific languages with UML profiles. In Proceedings of the Best Practices for Model Driven Software Development at OOPSLA (Vol. 5), 2005.

[7] Del Fabro, M.D., Valduriez, P.: Towards the efficient development of model transformations using model weaving and matching transformations. Software & Systems Modeling, 8(3), pp.305-324, 2009

[8] Wimmer, M.:. A semi-automatic approach for bridging DSMLs with UML. International Journal of Web Information Systems, 5(3), pp.372-404, 2009.

[9] Giachetti, G., Marín, B. and Pastor, O.: Integration of domain-specific modelling languages and UML through UML profile extension mechanism. *IJCSA*, *6*(5), pp.145-174, 2009.

[10] Selic, B.: A systematic approach to domain-specific language design using UML. In Object and Component-Oriented Real-Time Distributed Computing, 2007. ISORC'07. 10th IEEE International Symposium on (pp. 2-9). IEEE, 2007.

[11] Bézivin, J., Bruneliere, H., Jouault, F. and Kurtev, I.: Model engineering support for tool interoperability. In Proceedings of the 4th Workshop in Software Model Engineering (WiSME 2005), Montego Bay, Jamaica (Vol. 2), 2005.

[12] Robert, S., Gérard, S., Terrier, F. and Lagarde, F.: August. A lightweight approach for domain-specific modeling languages design. In Software Engineering and Advanced Applications, 2009. SEAA'09. 35th Euromicro Conference on (pp. 155-161). IEEE, 2009.

[13] Lagarde, F., Espinoza, H., Terrier, F., André, C. and Gérard, S.: Leveraging patterns on domain models to improve UML profile definition. In Fundamental Approaches to Software Engineering (pp. 116-130). Springer Berlin Heidelberg, 2008.

[14] Kurtev, I., Bézivin, J. and Akşit, M.: Technological spaces: An initial appraisal, 2002.

[15] Pastor, O., Molina, J.C.: Model-Driven Architecture in Practice: A Software Production Environment Based on Conceptual Modeling. Springer, New York, 2007.

[16] Selic, B.: The Pragmatics of Model-Driven Development. IEEE Software, vol. 20, pp. 19–25, 2003.

[17] Jouault, F. and Bézivin, J.: KM3: a DSL for Metamodel Specification. In Formal Methods for Open Object-Based Distributed Systems (pp. 171-185). Springer Berlin Heidelberg, 2006.

[18] MOF, O., 2002. OMG Meta Object Facility (MOF) Specification v1. 4.

[19] Eclipse Foundation, "Eclipse Modeling Framework.", http://www.eclipse.org/modeling/emf/ Accessed Januar, 2016.

[20] Kern, H. and Kühne, S.: Integration of microsoft visio and eclipse modeling framework using m3-level-based bridges. In Proceedings of Second Workshop on Model-Driven Tool and Process Integration (MDTPI) at ECMFA, CTIT Workshop Proceedings (pp. 13-24), 2009.

[21] Kern, H., Stefan, F., Dimitrieski, V. and Čeliković, M.: Mapping-Based Exchange of Models Between Meta-Modeling Tools. In Proceedings of the 14th Workshop on Domain-Specific Modeling (pp. 29-34). ACM, 2014

LCR of Wireless Communication System in the Presence of CCI in Dissimilar Fading Environments

Caslav Stefanovic
Faculty of Natural Sciences and Mathematics, University of Pristina, Department of Informatics, Kosovar Mitrovica, Serbia

Abstract—**In this paper, the ratio of the Rician random variable and the product of two Rayleigh random variables is considered. Moreover, the average level crossing rate of the ratio using Laplace approximation formula for two-fold integrals is calculated. Obtained results can be applied in performance analysis of wireless mobile communication system operating over LOS (line-of-sight) multipath fading channel in the presence of CCI (co-channel interference) which originate from wireless relay communication system with two sections subjected to NLOS (non line-of-sight) multipath fading. The influence of different system parameters on the average level crossing rate of the ratio of Rician random variable and product of two Rayleigh random variables is graphically presented and discussed.**

***Keywords*:** Average Level Crossing Rate, Rayleigh Random Variable, Rician Random Variable, Wireless Relay System

I. Introduction

Probability density function and cumulative distribution function of the product, but also the ratio of two random variables are often considered in open technical literature. The first and second order performance measures of two or more random variables have application in performance analysis of wireless communication systems. In paper [1], average level crossing rate of product of *N* Rayleigh random variables is evaluated. Moreover, this expression is used for calculation of average fade duration of wireless relay communication system with two, three and more section operating over multipath fading channel. Average level crossing rate and average fade duration of product of two Nakagami-m random variables are calculated in paper [2], and obtained results are applied in performance analysis of wireless relay system with two sections. In paper [3], the first order statistical characteristics of the ratio of the product of two random variables and random variable using Maijer functions are calculated.

The average level crossing rate (LCR) is the second order performance measure of wireless communication system defined as the number of crossings of random process on determined level in positive or negative direction. LCR can be calculated as average value of the first derivative of random process. The average fade duration is also important second order performance measure of wireless communication system defined as meantime that the signal envelope stays below the determined threshold and can be evaluated as ratio of the outage probability and average level crossing rate. The outage probability is the first order performance measure defined as probability that the signal envelope average value falls below the defined threshold and can be calculated using cumulative distribution function.

The Rician distribution can be used to describe small scale signal envelope variation in multipath, LOS fading channel with one propagation cluster and constant power. The Rician factor is the parameter of Rician distribution and can be calculated as the ratio of dominant component power and scattering component power. In the case, when Rician factor goes to zero, Rician fading channel becomes Rayleigh fading channel and when Rician factor goes to infinity, Rician fading channel becomes no fading channel [4-5]. Rayleigh distribution can be used to describe small scale signal envelope variation in multipath, NLOS fading channel [4-5].

In this paper the ratio of the Rician random variable and the product of two Rayleigh random variables is considered. The average level crossing rate of the considered ratio is evaluated where two-folded integrals are solved using Laplace approximation formula. Numerical results can be used in performance analysis of wireless communication system operating over Rician multipath fading channel in the presence of CCI originating from wireless relay communication system with two sections subjected to Rayleigh multipath fading.

II. Ratio of the Rician Random Variable and the Product of Two Rayleigh Random Variables

The ratio of the Rician random process and the product of two Raleigh random process is:

$$z = \frac{x_1}{x_2 x_3},\ x_1 = x_2 x_3, \tag{1}$$

where x_1 is Rician random variable and x_2 and x_3 are Rayleigh random variables. The x_1 random variable follows Rician distribution [4-6]:

$$p_{x_1}(x_1) = \frac{2x_1}{\Omega_1} e^{-\frac{{x_1}^2 + A^2}{\Omega_1}} I_0\left(\frac{2x_1 A}{\Omega_1}\right)$$

$$= \frac{2}{\Omega_1} e^{-\frac{A^2}{\Omega_1}} \sum_{i=0}^{\infty} \left(\frac{A}{\Omega_1}\right)^{2i} \frac{1}{(i!)^2} {x_1}^{2i+1} e^{-\frac{{x_1}^2}{\Omega_1}},\ x_1 \geq 0. \tag{2}$$

Random variables x_2 and x_3 are Raleigh distributed [4-5]:

$$p_{x_j}(x_j) = \frac{2x_j}{\Omega_j} e^{-\frac{x_j^2}{\Omega_j}}, x_j \geq 0, \ j = 2,3. \quad (3)$$

The first derivative of *x* is:

$$\dot{x} = \frac{\dot{x_1}}{x_2 x_3} - \frac{x_1}{x_2^2 x_3}\dot{x}_2 - \frac{x_1}{x_2 x_3^2}\dot{x}_3. \quad (4)$$

The first derivative of Rician random variable and the first derivative of Rayleigh random variable is Gaussian random variable. The linear transformation of Gaussian random variables has Gaussian distributions. Therefore, the first derivative of the ratio of Rician random variable and the product of two Rayleigh random variables has conditional Gaussian distribution. The variance of the first derivative of x is:

$$\sigma_{\dot{x}}^2 = \frac{1}{x_2^2 x_3^2}\sigma_{\dot{x}_1}^2 + \frac{x_1^2}{x_2^4 x_3^2}\sigma_{\dot{x}_2}^2 + \frac{x_1^2}{x_2^2 x_3^4}\sigma_{\dot{x}_3}^2, \quad (5)$$

where:

$$\sigma_{\dot{x}_1}^2 = \pi^2 f_m^{\ 2}\Omega_1, \sigma_{\dot{x}_2}^2 = \pi^2 f_m^{\ 2}\Omega_2,$$

$$\sigma_{\dot{x}_3}^2 = \pi^2 f_m^{\ 2}\Omega_3. \quad (6)$$

After substituting (6) in (5), the expression for variance becomes:

$$\sigma_{\dot{x}}^2 = \pi^2 f_m^{\ 2}\left(\frac{\Omega_1}{x_2^2 x_3^2} + \frac{x_1^2\Omega_2}{x_2^4 x_3^2} + \frac{x_1^2\Omega_3}{x_2^2 x_3^4}\right)$$

$$\frac{\pi^2 f_m^{\ 2}\Omega_1}{x_2^2 x_3^2}\left(1 + x^2 x_3^2\frac{\Omega_2}{\Omega_1} + x^2 x_2^2\frac{\Omega_3}{\Omega_1}\right). \quad (7)$$

The joint probability density function of *x*, derivative of *x*, x_2 and x_3 is:

$$p_{x,\dot{x},x_2,x_3}(x, \dot{x}, x_2, x_3)$$

$$= p_{\dot{x}|xx_2x_3}(\dot{x}|xx_2x_3)p_{xx_2x_2}(xx_2x_3)$$

$$= p_{\dot{x}|xx_2x_3}(\dot{x}|xx_2x_3)p_{x_2}(x_2)p_{x_3}(x_3)p_x(x|x_2x_3), \quad (8)$$

where:

$$p_x(x|x_2x_3) = \left|\frac{dx_1}{dx}\right| p_{x_1}(xx_2x_3), \quad (9)$$

and:

$$\frac{dx_1}{dx} = x_2 x_3. \quad (10)$$

After substituting (10) in (9), the expression for $p_{x,\dot{x},x_2,x_3}(x, \dot{x}, x_2, x_3)$ becomes:

$$p_{x,\dot{x},x_2,x_3}(x, \dot{x}, x_2, x_3) = x_2 x_3 p_{\dot{x}|xx_2x_3}(\dot{x}|xx_2x_3)p_{x_2}(x_2)$$

$$\times p_{x_3}(x_3)p_{x_1}(xx_2x_3). \quad (11)$$

The joint probability density function of *x* and derivative of x is:

$$p_{x,\dot{x}}(x, \dot{x}) = \int_0^\infty dx_2 \int_0^\infty p_{x,\dot{x},x_2,x_3}(x, \dot{x}, x_2, x_3)dx_3$$

$$= \int_0^\infty dx_2 \int_0^\infty x_2 x_3 p_{\dot{x}|xx_2x_3}(\dot{x}|xx_2x_3)$$

$$\times p_{x_2}(x_2)p_{x_3}(x_3)p_{x_1}(xx_2x_3)dx_3. \quad (12)$$

The average level crossing rate of the ratio of Rician random variable and the product of two Rayleigh random variables is [7]:

$$N_x = \int_0^\infty |\dot{x}|\, p_{\dot{x},x}(\dot{x}, x)d\dot{x}$$

$$= \int_0^\infty dx_2 \int_0^\infty x_2 x_3 p_{\dot{x}|xx_2x_3}(\dot{x}|xx_2x_3)$$

$$\times p_{x_2}(x_2)p_{x_3}(x_3)dx_3 \int_0^\infty \dot{x}\, p_{\dot{x}|xx_2x_3}(\dot{x}|xx_2x_3)d\dot{x}$$

$$= \int_0^\infty dx_2 \int_0^\infty x_2 x_3 p_{\dot{x}|xx_2x_3}(\dot{x}|xx_2x_3)$$

$$\times p_{x_2}(x_2)p_{x_3}(x_3)dx_3 \frac{1}{\sqrt{2\pi}}\sigma_{\dot{x}}, \quad (13)$$

since [8]:

$$\int_0^\infty \dot{x}\frac{1}{\sqrt{2\pi}\sigma_{\dot{x}}} e^{-\frac{\dot{x}^2}{2\sigma_{\dot{x}}^2}}d\dot{x} = \frac{1}{\sqrt{2\pi}}\sigma_{\dot{x}}. \quad (14)$$

After substituting (2), (3) and (7) in (13), the expression for N_x becomes:

$$N_x = \frac{8}{\Omega_1\Omega_2\Omega_3} e^{-\frac{A^2}{\Omega_1}}\pi f_m \frac{\sqrt{\Omega_1}}{\sqrt{2\pi}}\sum_{i=0}^{\infty}\left(\frac{A}{\Omega_1}\right)^{2i}\frac{1}{(i!)^2}x^{2i+1}$$

$$\times \int_0^\infty dx_2 \int_0^\infty \sqrt{1 + x^2 x_3^2\frac{\Omega_2}{\Omega_1} + x^2 x_2^2\frac{\Omega_3}{\Omega_1}}$$

$$\times e^{-\frac{(xx_2x_3)^2}{\Omega_1} - \frac{x_2^2}{\Omega_2} - \frac{x_3^2}{\Omega_3} + \ln(x_2x_3)^{2(i+1)}}dx_3. \quad (15)$$

The integral:

$$J = \int_0^\infty dx_2 \int_0^\infty \sqrt{1 + x^2 {x_3}^2 \frac{\Omega_2}{\Omega_1} + x^2 {x_2}^2 \frac{\Omega_3}{\Omega_1}}$$

$$\times\, e^{-\frac{(xx_2x_3)^2}{\Omega_1} - \frac{{x_2}^2}{\Omega_2} - \frac{{x_3}^2}{\Omega_3} + \ln(x_2x_3)^{2(i+1)}} dx_3, \quad (16)$$

can be solved using Laplace interpolation formula [1]:

$$\int_0^\infty dx_2 \int_0^\infty g(x_2, x_3)\, e^{-\lambda f(x_2,x_3)} dx_3$$

$$\approx \frac{2\pi}{\lambda} \frac{g(x_{2m},x_{3m})}{\sqrt{detB}} e^{-\lambda f(x_{2m},x_{3m})}, \quad (17)$$

where:

$$B = \begin{vmatrix} \frac{\partial^2 f(x_{2m},x_{3m})}{\partial {x_{2m}}^2} & \frac{\partial^2 f(x_{2m},x_{3m})}{\partial x_{2m} \partial x_{3m}} \\ \frac{\partial^2 f(x_{2m},x_{3m})}{\partial x_{3m} \partial x_{2m}} & \frac{\partial^2 f(x_{2m},x_{3m})}{\partial {x_{3m}}^2} \end{vmatrix}, \quad (18)$$

and variables x_{2m} and x_{3m} can be solved from system equation:

$$\frac{\partial f(x_{2m},x_{3m})}{\partial x_{2m}} = 0, \quad (19)$$

$$\frac{\partial f(x_{2m},x_{3m})}{\partial x_{3m}} = 0. \quad (20)$$

For the example taken into consideration, functions $g(x_2,x_3)$ and $f(x_2,x_3)$ are:

$$g(x_2, x_3) = \sqrt{1 + x^2 {x_3}^2 \frac{\Omega_2}{\Omega_1} + x^2 {x_2}^2 \frac{\Omega_3}{\Omega_1}} \quad (21)$$

$$f(x_2, x_3) = \frac{(xx_2x_3)^2}{\Omega_1} + \frac{{x_2}^2}{\Omega_2} + \frac{{x_3}^2}{\Omega_3}$$

$$-2(\mathrm{i}+1) ln y_1 y_2. \quad (22)$$

III. Numerical Results

The normalized average level crossing rate versus the ratio of the Rician random variable and the product of two Rayleigh random variables (SIR-Signal to Interference Ratio) for different values of Rician envelope average power, interference envelope average powers and constant dominant component envelope is shown in Fig. 1.

Average level crossing rate increases when SIR increases, reaches its maximum,while average level crossing rate decreases for higher values of SIR, For lower values of SIR, when Rician envelope average power increases, average level crossing rate decreases, while for higher values of SIR, when Rician envelope average power increases, average level crossing rate increases. Furthermore, average level crossing rate increases for lower values of SIR when Rayleigh envelope average power increase, while average level crossing rate decreases for higher values of SIR when Rayleigh envelope average power decrease.

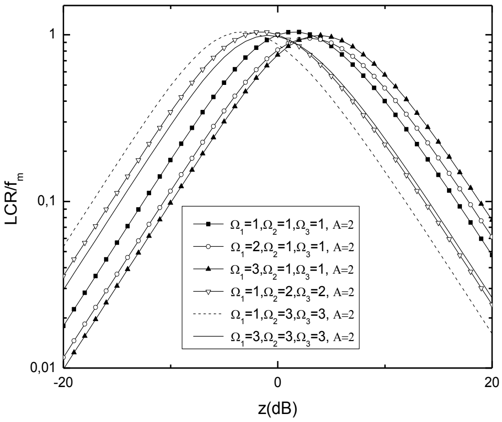


Figure 1. Normalized LCR for different values of Ω_1, Ω_2, $\Omega_3$and constant value of A.

Average level crossing rate increases when SIR increases, reaches its maximum,while average level crossing rate decreases for higher values of SIR, For lower values of SIR, when Rician envelope average power increases, average level crossing rate decreases, while for higher values of SIR, when Rician envelope average power increases, average level crossing rate increases. Furthermore, average level crossing rate increases for lower values of SIR when Rayleigh envelope average power increase, while average level crossing rate decreases for higher values of SIR when Rayleigh envelope average power decrease.

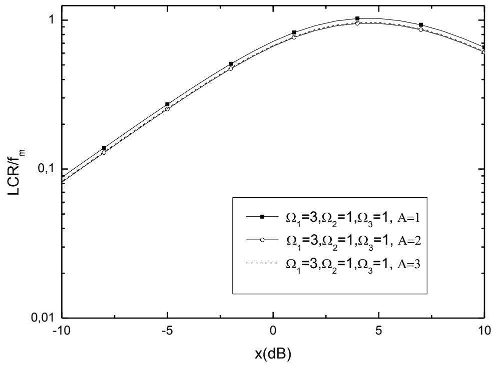


Figure 2. Normalized LCR for the same values of Ω_1, Ω_2, $\Omega_3$and different value of A.

The Fig. 2. shows the normalized average level crossing rate versus SIR for the constant values of Rician average envelope power, constant values of Rayleigh envelope average powers and different values of dominant component power. Average level crossing rate is approximately similar for the lower and the higher values of parameter A. Precisely, when dominant component amplitude increases, average level crossing rate slightly decreases for higher as well for lower values of SIR.

IV. Conclusion

The average level crossing rate of the ratio of the Rician random process and the product of two Rayleigh random processes by applying Laplace approximation formula for evaluation of two-fold integrals is evaluated. Obtained results can be used in performance analysis of wireless communication system operating over Rician multipath fading channel in the presence of co-channel interference originating from wireless relay communication system with two section subjected to Rayleigh multipath fading. In this system, ratio of the Rician random variable and two Rayleigh random variables is important parameter of a signal-to-interference-ratio (SIR) of wireless system. Average level crossing rate of the ratio of the Rician random variable and the two Rayleigh random variables can be used for evaluation of average level crossing rate of ratio Rayleigh random variable and two Rayleigh random variables as special case. Numerical results are graphically presented to show the influence of Rician envelope average power, Rayleigh envelope average powers and dominant component of Rician envelope on average level crossing rate. Normalized average level crossing rate is shown as signal to interference ratio of the Rician random variable and two Rayleigh random variables.

References

[1] Z. Hadzi-Velkov, N. Zlatanov, and George K. Karagiannidis, "On the second order statistics of the multihop rayleigh fading channel," *IEEE Transactions on Communications*, vol. 57, no. 6, pp 1815 – 1823, June 2009.

[2] N. Zlatanov, Z. Hadzi-Velkov and G. K. Karagiannidis, "Level Crossing Rate and Average Fade Duration of the Double Nakagami-m Random Process and Application in MIMO Keyhole Fading Channels, " *IEEE Communications Letters*, vol. 12, no. 11, pp. 822 – 824, Nov. 2008.

[3] A. Matovic, E. Mekic, N. Sekulovic, M. Stefanovi, M. Matovic and C. Stefanovic, "The distribution of the ratio of the products of two independent α-μ variates and its application in the performance analysis of relaying comunication system," *Matehmatical Problems in Enginering*, ISSN: 1563-5147, Hindawi Publishing, vol. 2013, Article ID 147106, 2013.

[4] M. K. Simon and M. S. Alouini, *Digital communication over fading channels*, John Wiley & Sons, 2005.

[5] G. L. Stüber, *Principles of mobile communication*. Springer Science & Business Media, 2011.

[6] I.S Gradshteyn and I.M. Ryzhik, *Table of integrals, series, and products*, 2000.

[7] S. Panic, M. Stefanovic, J. Anastasov and P. Spalevic, *Fading and interference mitigation in wireless communications*, CRC Press, 2013.

[8] D. Middleton, *An introduction to statistical communication theory*, New York: McGraw-Hill, 1960.

E-government based on GIS platform for the support of state aid transparency

Mirjana Kranjac*, Uroš Sikimić**, Ivana Simić***, Srđan Tomić****

* University of Novi Sad, Faculty of technical sciences Novi Sad, Department for transport, Novi Sad, Serbia
** Singapore Management University, Institute of Innovation & Entrepreneurship, Singapore
***Provincial secretary for economy, employment and gender equality, Novi Sad, Serbia
**** University Nikola Tesla, Faculty for engineering management, Belgrade, Serbia
mirjana.kranjac@uns.ac.rs, uros_sikimic@hotmail.com, ivana.simi@vojvodina.gov.rs, srdjan.tomic@fim.rs

Abstract **– Authors of the paper present solution for the transparency of various types of state aid to the citizens. They are using ICT concept which consists of the geographic information systems (GIS) application incorporated within e-government platform for the geo-referenced presentation of the different types of financial supports. The use of GIS in this way is presented through the case study of the subsidies for self employment of unemployed persons given by the Provincial government of Vojvodina.**

I. INTRODUCTION

Development of e government is a long lasting process enabled by the fast changes in the ICT sector. E government is focused on the services aimed to facilitate communication between government and citizens. A part of the e government should be review of the state aid delivered to the users. The significant part of the state budgets is spent to the various state aids. State aid serves to support realization of the strategic goals on different society levels. States have their national development programs in line with general strategic goals of the society. To reach them, they direct activities of many social groups like: public and private companies, entrepreneurs and civil society. Many specific social actors are under umbrella of the state aid. Some of them, like associations, NGOs are very dependent of the aid which they become. Some, like old handicrafts, remain working and saving old professions only due to this aid. State aid transparency is important and beneficial for the whole society. It provides an opportunity for citizens to be better informed about public policies. It promotes accountability as an essential factor within society. It also helps create more effective dialogue between citizens and government. It offers better policy decisions. Civil society and governments in all countries have made great advances in these effortsand increased transparency at local, regional and national levels, what happened in last few years. Spending of public resources is a very sensitive issue and it needs responsibility and participation in the terms how resources are allocated. Indeed, citizens have the right to know how their money is being spent [1]. A primary objective of the state aid modernization is to make the granting of a state more transparent. Transparency helps to improve the quality of public policy and strengthen the effectiveness of State aid control [2]. Technological progress is a base of the visible improvements of transparency.

This paper tackles the issue of the new ICT tools which make the state aid data available to the costumers.

The research object is the presentation of the state aid within e government platform.

Hypothesis: The GIS tool implemented within e-government platform contributes to the transparency of the state aid data.

Research tasks is to set up a new ICT tool which will improved presentation of geo referenced data within e-government platform and by this visibility and transparency of the governmental activities..

Authors used the following research methodology: Data collecting and analysis, geo referenced visualization of data.

The main goal of the paper is to ensure transparency of public data concerning to the state aid to the citizens. This will expand functionalities of the e government from the present functions related to " the utilization of information technology (IT), information and communication technologies (ICT), and other web-based telecommunication technologies to improve and/or enhance on the efficiency and effectiveness of service delivery in the public sector" [3] to the new attitude, to enhance visibility of the results of the state aid.

The specific goal is to present model which incorporates geographic information systems - GIS into the e government platform to enable clear presentation of the government activities, visually understandable to each consumer, especially those related to the state aid given to the different consumers groups, like: NGOs, SMEs, entrepreneurs, farmers…

Task of the paper is to present visualized data about state aid which financially supported self employment in the region of Vojvodina through subsidies to the users, for the year 2014. This is a good example which proves expanding possibilities of e-government by using GIS applications.

II. LITERATURE OVERVIEW

The geographic information systems are software which enables visualization of spatial data. Man is more sensitive to the visualized presentation of the data. He is easier solving problems which could be visually presented [4]. Geographic location is the element that distinguishes geographic information from all other types of information. Location as a data brings many benefits

of GIS like geographic presentation, planning, comparison, following time changes, overlapping of the spatial planes, cross cutting different data [5, 6].

GIS has a meaningful position within e-government service.

There are many discussions within literature about exact explanation of the e-government concept. Some definitions characterized e-government as simply the use of information and communications technology, like the Internet, to advance the processes within government activities. It aims to make government primarily responsive, than more efficient. It should enable transparency of public administration results. E-government offers a variety of new business opportunities.

E-government is having more functions. It is condition for reforming of the work processes among governmental institutions. It is improving services to enable collaboration not only with citizens, but with the business, and nonprofit organizations, too.

Very popular public service delivered to citizens is one-stop shop which can be used through the World Wide Web portals. These portals are aimed to provide a place which enables all the steps of a complex administrative process. It gives approach to the many government offices at one site. Users are not more running visiting offices, they have immediate, fast contact with the service [7].

Fast involvement of e-governance is closely related to the emerging development of IC technologies which basic added value is Internet platform. New level of the e-government processes is the involvement of the geographic information systems which facilitates interactions with the society in a user friendly way giving spatial information a visible reproduction. New Web sites with GIS services provide the following applications:

- Government-to-business: to present economic and land development, licensing.
- Government-to-citizen: to enable online service information.
- Government-to-government: to improve communication inside and among the government departments.
- To enable governments and citizens wisely exploitation of resources, be involved in planning actions and effectively respond to emerging situations [8].

Kumar (2012) presents his vision of the typical GIS architecture for the state implementation (Figure 1). It shows the strict links between government, private sector and citizens on a GIS platform design to support e government functioning. He stresses the big importance of the automatization of the government service by implementing of e-government framework with GIS inside.

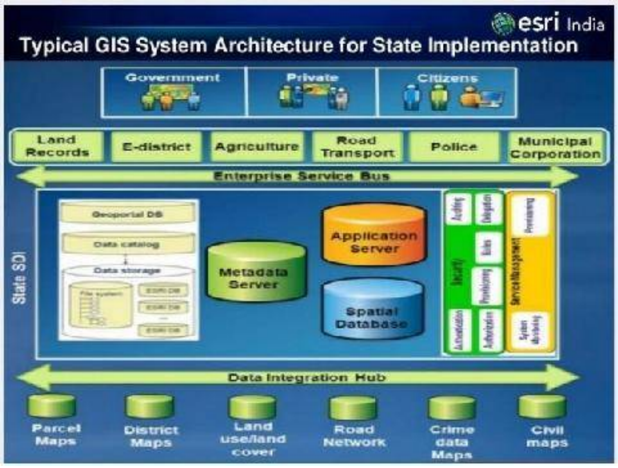


Figure 1. GIS architecture for the State implementation [10]

Participation of local community at different data inputs in GIS enable to produce a holistic and efficient output. GIS platform should be expanded with special software tools for group and individual decision support, geographic analysis and presentation. The specific tasks which can be realized with GIS are related to ownership problems, location problems, technical design, planning. With experts knowledge it helps in designing development plans, monitoring long-term changes and implementing corrections [10].

Why using GIS within e-government concept? Some authors explain this:

- "e-government involves GIS because the relationship between government and citizen is frequently geographically based
- e-government places a greater reliance on the robustness of information – geographic information has considerable scope for error
- e-government requires an e-geography i.e. a National Spatial Data Infrastructure [11]."

Geographical information systems has one additional role. This is to develop policy design and e-democracy. Some authors describe case studies which should have analyzed how use of GIS could be used to encourage participation of citizens in urban planning and redevelopment. They proved that potential of e-government is dependent on the citizens culture and knowledge, but that e-government can help bridge the gap between government and citizen and even reduce the scope of conflict between the two [12, 13].

Some authors [14] present the elementary requirements and the best technology for e-government information management. They propose a useful framework of unified e-government management platform based on the Web-GIS which could be easily implemented.

Web based economic development with GIS will qualify communities to compete with each other. Investment of private sector will be dependent on the spatial data which are presented on e–government platform. Investigations, analysis, monitoring, evaluation will be possible without visiting the place [15].

III. Possible scenarios of GIS implementation as tool for e government transparency

Secretariat for Economy, Employment and Gender Equality, on the basis of the decision of the Provincial Government of Vojvodina set aside funds to support self-employment, employment, professional practice, and public works on territory of Vojvodina. The funds has been allocated since 2009. The goal of such financial support is to reduce unemployment through fostering entrepreneurship and to increase number of jobs within existing companies. The target groups are two groups of unemployed people:

1. Those who have entrepreneurial potential to set up their own business
2. Those who already have consensus of entrepreneurs to be employed.

This paper analysis susidies for self employment. The authors present the possibility to enable that data about donations would be available to everybody who is interesting to have insight into results of public open competitions. They create visual presentations by using GIS which enable every person who has interest to have clear picture about activities of the provincial government.

Also, such spatial mapping of the results enables analysis of different data by comparing them from year to year, reached outcomes of the given subsidies, appeared problems, visible political influences etc. And, what is the most important, it gives bases for the improvement in many ways.

The following text contains maps of spatial data which were made to show some possibilities of GIS.

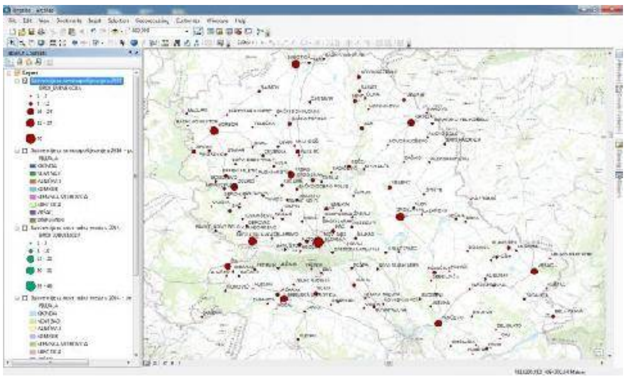


Figure 2. Subsidies for self employment in 2014 (source: authors)

Figure 2 represents results of the competition which refers to the subsidies for self-employment in 2014. In the table of contents located at the left side of the Figure is marked appropriate layer. Graduated circles were used to symbolize outcomes due to the locations. All circles are red and represent self employment. The size of the circle reflects the number of subsidies. The green color is used for the number of subsidies for new employment within existing companies.

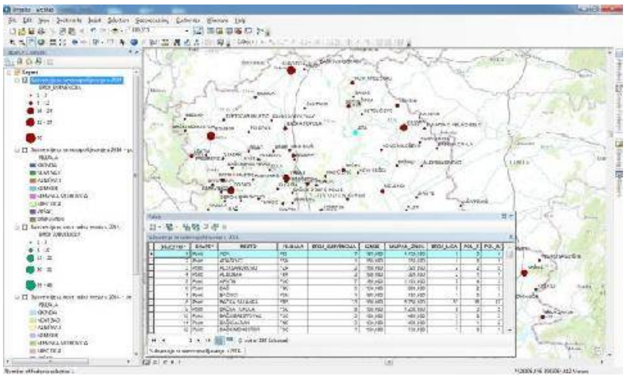


Figure 3. Subsidies for self employment in 2014 with attribute table (source: authors)

In the Figure 3 could be seen the attribute table which appears when one town is selected, in this case, Ada. Its circle changes color (into yellow) and the program opens the table with data. The table contains data with:

- The name of the local office
- Number of subsidies
- Sum of each
- Total sum for this local office
- Number of persons who got it
- Number of female
- Number of man.

When an object in the table is selected, this object is automatically selected on the map and visible.

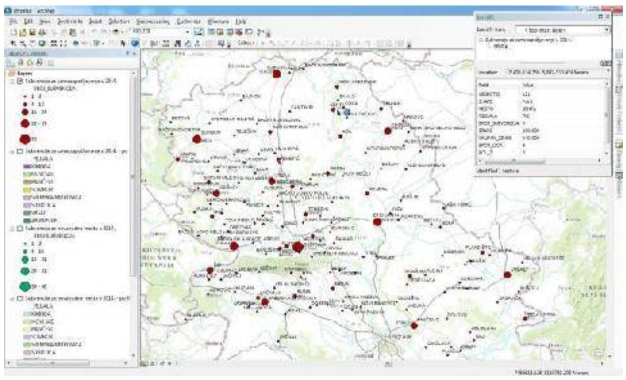


Figure 4. Subsidies for self employment in 2014 with identification of one town (source: authors)

Figure 4 illustrates the use of tools, in this case: Identify. It the user clicks on the desired place he gets all information like in the previous table, but this time only for the selected town and without opening the table.

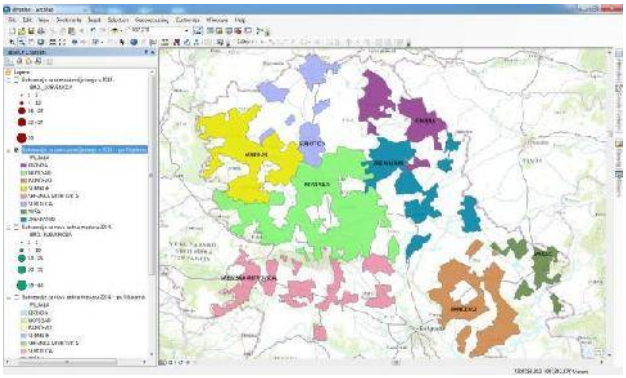


Figure 5. Subsidies for self employment in 2014 within branches of National Employment Agency (source: authors)

Figure 5 shows the subsidies for employment by branches. In table of content on the left side of the Figure is marked appropriate layer. Different colors were used to mark 8 branches of the National Employment Agency (NEA) in Vojvodina region.

All possibilities with presentation data from the table and data for each selected branch of the NEA are available also for the branches.

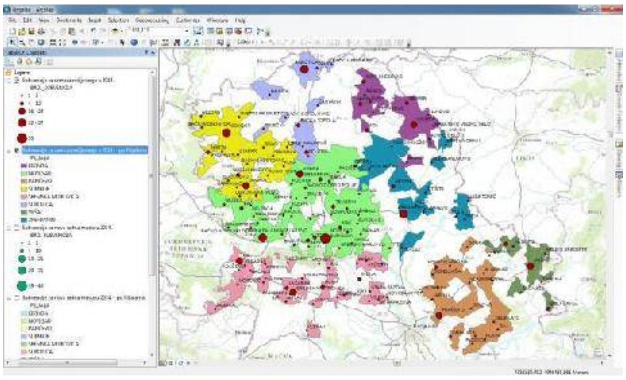


Figure 6. Subsidies for self employment in 2014 with overlapping of NEA branches with locations of villages (source: authors)

Figure 6 shows simultaneously two layers of the same public call - competition. By using overlapping tool option of the GIS superposition can be seen which shows which villages in each branch received grant and the size of the each grant in each village corresponding to the size of the red circle.

Figure 7 displays the zoomed branch of the town Sombor with belonging settlements that have received subsidies. It was also used a tool called: Identify for the selected town Sombor. The same procedure can be used for each town or settlement. The table in the right corner contains all details about total subsidies for the whole branch Sombor.

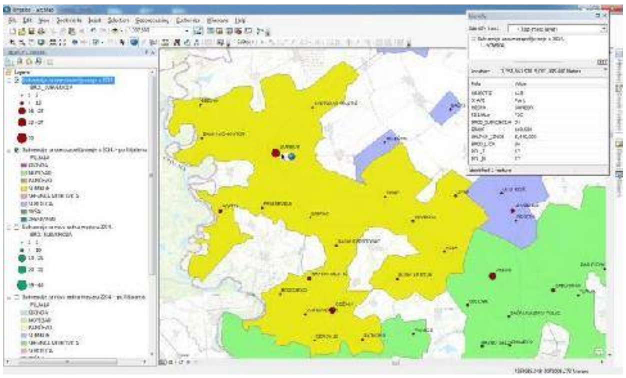


Figure 7. Subsidies for self employment in 2014 within zoomed branch of town Sombor with settlements (source: authors)

V. Conclusion

The authors of the paper showed one possible scenario which presents geographic information systems as a useful tool for the development of such e- government which will foster e-democracy. Such approach which enables citizens complete insight into the state decisions which must be independent and made according to previously set up criteria should be vision of the e-government development. The visualization in figures 2 to 7 prove the hypothesis that GIS tool implemented within e-government platform contributes to the transparency of the state aid data.

Such a platform should be further developed to an interactive tool which users could use to investigate, analyze, overlap layers with various data and request new state actions based on the conclusions made by GIS tools implemented. The authors already presented in articles few possible uses of GIS in different sectors of human activities, like at telecom market [16], postal service network [17] and new urban transport systems [18]. This paper is adding new value and an innovative approach dedicated to such an important social area, such is democratic decisions and human rights.

References

[1] „State aid transparency for taxpayers", EC Competition policy brief, http://ec.europa.eu/competition/publications/cpb/2014/004_en.pdf

[2] Nicolaides, P, "Transparency is also needed at the European Commission level", *State aid hub EU*, 2014. http://stateaidhub.eu/blogs/stateaiduncovered/post/22

[3] Hai, J, "*Fundamental of development administration*", Selangor: Scholar press. ISBN 978-967-5-04508-0, 2008, pp. 125-130

[4] Goodchild, M, "Twenty years of progress: GIS science in 2010". *Journal of Spatial Information Science*, 2010, pp.12-13.

[5] Heywood, I., Cornelius, S., Carver, S, *„An Introduction to Geographical Information Systems",* (3rd ed.). Essex, England: Prentice Hall, 2006, pp. 230-232.

[6] Perry, M., Hakimpour, F., Sheth, A, „Analyzing theme, space and time: An ontology-based approach", Proceeding: *ACM International symposium on Geographic information systems* (PDF), 2006, pp. 147–154.

[7] Gordon , T, "E-government- introduction", *ERCIM News* No.48, Fraunhofer, FOKUS, 2012, pp. 211-213.

[8] http://www.esri.in/industries/government/egovernance

[9] Nikhil, K, “Citizen centric e-government”, *State consultation workshops in partnership with Nasscom,* India, 2012.

[10] Kumar, A, “Participatory GIS for e-governance and local planning: A case study in different environments on its applicability”, *Centre for Study of Regional Development*, New Delhi, India, 2013.

[11] Barr, R, “*E-government and GIS”*, The British Experience

Manchester Regional Research Laboratory, School of Geography, University of Manchester, UK, 2013, pp.15-20.

[12] Moody, R, „Assessing the role of GIS in E-government: A tale of E-participation in two cities“, https://www.scss.tcd.ie/disciplines/information_systems/egpa/docs/2007/Moody%2007.pdf , 2005, (accessed 12. 10.2015)

[13] Fu, P., Sun, J, “*Web GIS: Principles and applications*”, ISBN:158948245X 9781589482456, Esri Press, New York, 2010, pp. 25-28.

[14] Lu, X, “A GIS-Based integrated platform for E-government application, Information Science and Engineering (ICISE), *1st International Conference on* GIS, IEEE, ISBN 978-1-4244-4909-5, 2009, pp. 200-218.

[15] Reddick, Ch, “*Handbook of research on strategies for local e-government adaptation and implementation: Comaparative studies”*, University of Texas, Information science reference, USA, 2009, pp. 632-635.

[16] Kranjac, M., Sikimić, U., Dupljanin, Dj., Tomić, S, “Use of Geographic information systems in analysis of telecommunication market”, *Proceeding ICIST 2015*, Kopaonik, 2015, pp. 303-308.

[17] Kranjac, M., Sikimić, Vujaković, M, “Model for calculation of post service efficiency by using GIS”, *Proceeding Mipro 2015*, Opatija, ISBN 978-953-233-083-0, pp. 353-359.

[18] Kranjac, M., Sikimić, U., Vučinić, M., Trubint, N, „The use of GIS in the model of implementation of the light train transport in Novi Sad”, *Conference Towards a human city,* 2015, Novi Sad, FTN, ISBN 978-86-7892-541-2, 2015, pp. 124-130.

Process modeling method for higher education institutions based on BPR

Evandro Henrique Cavalheri*, Marcelo Rudek*, Osíris Canciglieri Junior*, Ana Maria Magalhães Correia**, Marcia Cristina Alves dos Anjos Almeida***
* Pontifical Catholic University of Paraná – PUCPR / PPGEPS, Curitiba, Brazil
** Universidade Federal Rural do Semi-Árido – UFERSA, Mossoró, Brazil
*** Pontifical Catholic University of Paraná – PUCPR / PPGAD, Curitiba, Brazil
marcelo.rudek@pucpr.br, evandro.cavalheri@pucpr.br, osisris.canciglieri@pucpr.br, anamagalhaes@ufersa.edu.br, anjos.cristina@pucpr.br

***Abstract* — Reengineering means radically rethinking the processes. In this sense, the objective of this article is to propose a conceptual model of a method of modeling processes adapted to Higher Education Institutions - HEI, based on the BPR - Business Process Reengineering. HEI is a type of company that has activity in order to provide services, but their operations have specific characteristics and a high level of complexity. The method can be applied in different processes in an HEI, because it covers specifics of the activity for possible significant improvements. To develop the method, concepts such Operations Strategy added to concepts of Business Process Management, based on the reengineering of processes were applied. The key advantage is that the application of proposed method, can ensure that the process is aligned with the strategy. For the application of the method 6 stages are defined. Therefore, the method can add the strategic result.**

I. Introduction

The process modeling is a tool used to represent the current situation and describe the future vision of the business process [1] [2] [3]. For business processes, it is understood that these are real world activities, consisting of a set of logically related tasks and, when executed in a suitable sequence and in accordance with the business guidelines, they generate a settled result. [4] Among the main objectives of process modeling as listed: i) improvement of processes, aiming to eliminate obsolete and inefficient processes and unnecessary rules and management; ii) standardization documents; iii) ease of documentation; iv) reading skill; v) homogeneity of knowledge for all team members and vi) full complement in the documentation of processes. For an example, at an HEI, you can set its strategy for a certain period; have a certain number of students enrolled in a particular discipline for a certain language. Therefore, the main activity of the HEI is teaching, since the student must attend a number of disciplines, the institution will offer these disciplines, meeting their demand. However, to move it forward on its strategy, the HEI must not only offer this discipline, but offer it in a definite language of the strategy plan as well.

The business process modeling has been used in various organizations in the production of goods (manufacturing) or services (operations) in various branches of activity. According to [5] traditionally, operations management was seen as being operational, but this view has changed from the '60s, when the necessity emerged the definition of strategic objectives for the activities of operations because of lack of resources and the need to produce in large scale to minimize costs. The objectives are now ensuring that the value of production and delivery processes to the customer are aligned with the strategic intentions and the markets that the company intends to take over.

The modeling of business processes has improved the interactions between the interfaces, providing important information on the implementation of its operations. These improvements allow better understanding of the characteristics of the processes making clear the responsibilities of those involved in the operations that comprises the same. [6] [7]. In the interaction, the ability to implement processes and redesign costs and optimizing critical timing process are the focus point of this work.

II. Background

Among the existing modeling techniques are the BPM (Business Process Management) BPR - Business Process Reengineering and BPMN (Business Process Modeling and Notation). The Process Management, or BPM [8] [9] is defined as a set consisting of methodologies and technologies that aims at enabling business processes to integrate logic and chronological way, customers, suppliers, partners, influencers, employees, and each and every element with which they can, or want to have to interact, giving the complete picture organization and essentially integrated internal and external environment of operations of each participant involved in the business processes.

BPR is defined as "the fundamental rethinking and radical restructuring of business processes to achieve dramatic improvements in critical, contemporary indicators of performance such as cost, quality, service and speed". [10] And BPMN is used as an important tool in the implementation and restructuring processes. [11]. A modeled process will run their demands using fewer resources, be it financial, material or human. Once the operation is not dependent on the skill of the user that is running. This method uses the Operations Strategy concepts added to the concepts of Service Management Process based on reengineering processes expressed in BPMN notation. The notation was developed to provide

the use of the management of business processes through the establishment of standards [12].

The integration of reengineering and process modeling is applied as a method to improve the implementation of the operations involved in the processes. Thus makes it possible to manage the processes ensuring that its implementation will be aligned with the strategy defined by the organization. As if the strategy was the top of a mountain, or where the organization wants to reach, and processes are the paths that must be covered. Line up strategy to the process is to ensure that every step will shorten the distance to the top [13].

For an application where there is a requirement for modeling a process, many techniques may be engaged to better illustrate the process. A HEI, as well as other service providers are vulnerable to market demand, changes in the economy, seasonality and the inviolability [14]. And still, due to its type of business, higher education is a susceptible to specific issues such as the demand for courses, programs, performance indicators, standardized tests, scholarships, earnings, philanthropy, community service and a number of variables, such as: students, teachers, subjects, classrooms, research, training activities, readiness for market, internationalization, the incoming and outgoing exchange students, its infrastructure, the academic management systems, and the different levels of teaching undergraduate and graduate programs [15] [16]. However, the methods available in the literature, does not include the specific characteristics of a HEI. Thus, complexity levels, and specificities existing in a service process in a HEI [16] show the hidden requirement for a separate modeling method that is built according to its individualities.

III. Proposed Method

The proposed method, different from generic modeling methods has the feature of encompassing the most diverse variables present in a HEI, and to propose a modeled process and eliminates rework, miscommunication, lack of formalization of activities, it will ensure the process and strategy are walking together. Through theoretical framework and literature of intersection shown in Fig. 1: Operations Strategy, Process Management and Business Process Reengineering.

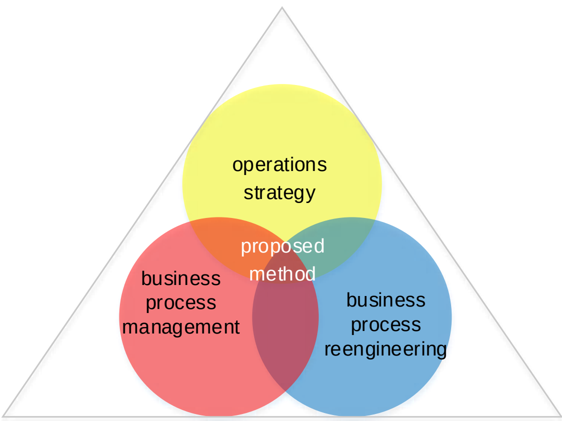


Figure 1. Intersection of literature that support the method

For its development, shown on Fig. 2, the proposed method comprises some steps that was withdrew from literature and organized systematically to facilitate its application, in which activities are carried out aimed at providing elements that will be needed during its execution.

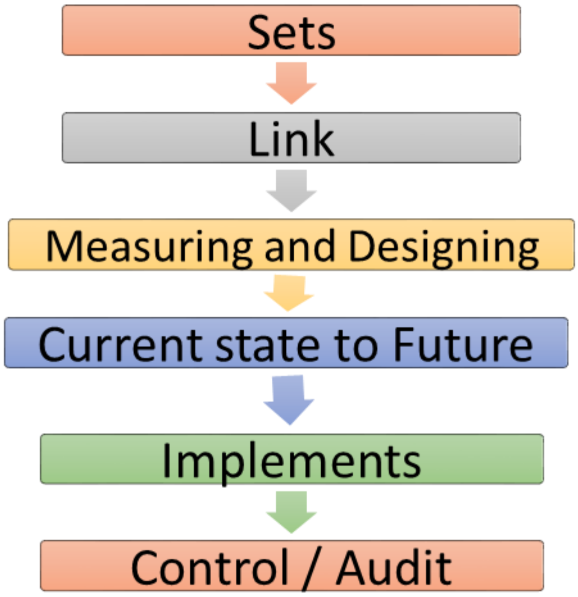


Figure 2. Steps execution of method

The established sequence to perform the steps of the method shown above, it was designed so that the elements of the processes for a HEI would be easily identified, as to define the process to be modeled for the activity “scholarship”. It is possible to understand what is the connection with other variables in the second step, point it out how students grants will contribute to the organization's strategy, as an example, the HEI defines its strategy to have a number of students and out of that number, 40% of grant students, the next step will quantify this process, their time and pauses, as the interfaces connect and how this connection’s arrangement. For a better example: when a student applies for a scholarship, he does it at the designated area, the department collects the request, forwards it to another area that runs a financial analysis, along with another area that does a credit check for the economic background of the student. For the next step, the process is depicted to show how process happens, or how it is "mapped out" and based on its results a new application is proposed to change the "modeling", this change can include an algorithm that automates the analysis involved in, streamlining activity, and makes it less susceptible to human mistakes. It is therefore validated the possibility of the modeled process, highlighting their contributions. Finally, it is necessary to control the changes, check if failures are occurring, due to problems of culture or resistance from those involved. The established sequence is shown on Fig. 2, and described their following steps:

A. 1st Process to be modeled

[17] and [18] suggests that the mechanism of the strategy of operations may be understood as follows: the organization defines the long-term investments, how and where they want to operate and in which businesses

operate, which is the corporate strategy. Therefore, at this stage, the goals to be achieved will be defined with the improvement, its scope, the formation of the change management structure to support the improvement, allocation of roles and responsibilities among those involved. For this phase, the activities will be as to:

- define who will be the sponsor/ owner/ leader of the process and what can it ensure its commitment;
- train and integrate the team with planning tools for the implementation of improvement;
- define the scope of the improvement to be held;
- create a work schedule.

B. 2nd Connection with Organizational Strategy

For the process to be modeled in order to not only meet their demands, but rather fulfill a demand in order to be aligned with the organizational strategy. [13] Thus contributing to achieving the strategy defined by the organization, the goals become the guarantee that the value of processes and delivery to the customer are aligned with the strategic intentions and the markets that the company intends take over as detailed by [19].

- creating flowcharts illustrating the connection between process and organizational strategy using BPMN;
- analyze the influence of the strategy in the process;
- understand the process of contribution in strategy.

C. 3rd Time measuring and flow designing

The business process management is a consolidated structured approach that meets standards and some of the possible best practices to identify, design, execute, to document, measure, monitor, control and improve business processes, automated or otherwise, to achieve the results consistent and aligned with the strategic goals of an [20] [21] organization. From the affirmation, at this stage, the time required for performing the process are measured, the breaks are described in the process and the bottlenecks showing the dependence on connections with other interfaces and the gained results. The purpose of this step is that starting from the collected information about timing and results, everyone involved must be acquainted with the current stage of the process. For such results, some activities were carried out:

- create metrics to measure performance;
- define which services are delivered to the customer, resulting from the process;
- collect the information to create the Workflow Map;
- Identifying bottlenecks;
- design workflow using BPMN notation.

D. 4thModel the current status (AS IS) and propose future status based on Reengineering (TO BE)

From the implementation of the Business Process Modeling Notation BPMN, consolidated as the most important standard open graphical notation for drawing and modeling business processes [22], it makes it possible to design business processes as business process diagram and so enabling capture and document the current status (AS-IS) in clear diagrams, and design and describe the ideal status (tO-BE) [8] [23] [24]. Highlighting the contrast between the "mapped out" process (AS-IS) and the application for the "modeling" process (TO-BE), according to the example shown on Fig. 3 as follow:

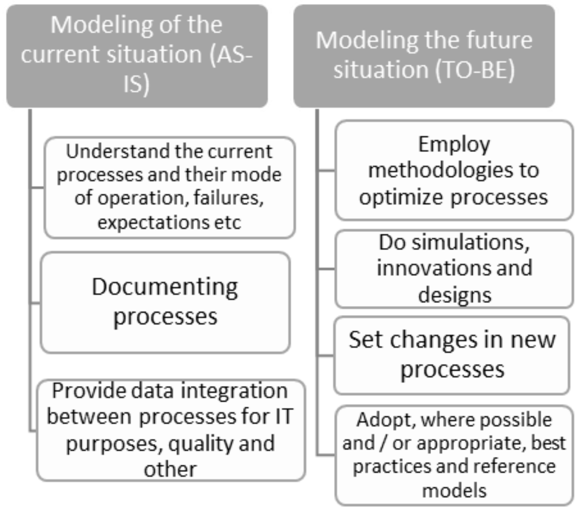


Figure 3. Current modeling for future modeling

For the development of this phase the information collected in the previous phase are used for the plan through technological tools which means that will be used to reach the improvement. The activities in this phase are:

- to clarify the source of bottlenecks;
- to analyze the flow map current process;
- to plan the future process flow map;
- to project the future layout;
- and to plan the flow movement from operations and people.

E. 5th Implement the generated model

[10] Claim that to guide the work for processes, rather than departments, abolishes the functional areas, and people now belong to activities of the work flow, not a department that once welcomed people with similar specialties. The goals, which were previously departmental and different for each point of the work flow, are now combined by process, leading team activity. The very nature of the work changes. People who are specialized in only one type of activity, they are no longer desirable by the company. By establishing teams, the responsibility for implementing the process falls on everyone involved, and even if the person always try to keep the team together, one always seek to accomplish, within the team, tasks where someone is more skilled, there are no more limits to what each individual performance of responsibility. Everyone is responsible for the progress of the process, and everyone should know how it is done, and assist in the implementation. Guided by these concepts this stage aims to plan the execution and explicitly improves planned previous phases. Such activities consist of:

- set the improvement schedule to be implemented;
- plan the implementation necessaries;
- create the deployment teams;
- train the deployment teams;
- train the operating of the targeted area of employees.

F. 6th Control / Audit

In modern companies, business processes, range from broad to a median location are typically multifunctional, that is the activity is fully carried out, and it must runs by professionals with different skills and therefore is usually processed by different departments and organizational areas. [18] That means that the process is managed by different areas, which conducts are determined by performance indicators that allow no connection to the process, and are used to measure the efficiency of the area's resources, manage the process at different stages of the area. The lack of metrics to monitor the process performance and the lack of business processes management, responsible for analyzing the entire flow means that there is not effectively managing business processes [25]. On this last phase, right after the implementation the established improved process after its modeling, through the pairing of the data is carried out as detailed survey of the obtained results. Also after the implementation phase, an awareness campaign should be conducted with those involved, so that everyone understands the contribution and resulting improvement in the implementation of the proposed modeling. Thus becoming part of the organizational culture the vision processes based on the concepts of Business Process Reengineering.

The activities in this phase are:

- to create an advertising campaign that expresses the results;
- to highlight the importance of the view of culture by processes;
- to transfer the responsibility to a process flow manager;
- conduct the audits;
- create change patterns;
- create a manual for future modeling.

Therefore, the last phase stabilizes the changes made in the previous phase and prepares the organization for another improvement cycle.

IV. APLICATIONS

The method has the characteristic of making it possible to model any process of an HEI, because it was designed to include its peculiarities, this is a variable that involves student billing, scholarships or exchange students. However, as the main academic contribution, their application enables modeled after the process, that had come to contribute to organizational strategy, consolidating its connection.

For the reception process of the exchange students the HEI will be able to, starting from the application process the method proposed to operationalize this step so that the assistance for this students should be standardized because different from regular student who is enrolled at the institution and will not require a special treatment, and exchange student will not be familiar with the school environment, and if each student in this situation requires a different attendance, there will be a major bottleneck. Another example is the use/implementation of management systems, the process is already exhibited, all bottom up will be available, along with the workflow process, so that the HEI can evaluate how your process will have to adapt to its new system, or highlight the necessity to build the system oriented to the fit the requirements of the process.

Anyhow, the operation of the process must be intrinsically in harmony with the strategy. However, to move it forward on the way of its strategy, the HEI must not only offer this discipline, but also offer it in certain language in the strategy.

Consequently, the main advantage is that the method will ensure that the process is aligned with the strategy defined by HEI, as the result of a process will add the result to the strategy. The method includes the most diverse variables present in a HEI, and to propose a modeled process and eliminates rework, miscommunication, lack of formalization of activities, it can ensure that the process and the strategy are side by side.

The restraint of the article is that the method has not been applied yet in a case study. The contribution of the paper is to facilitate the integration between processes and strategic alignment by the method of process modeling for HEI. Future research may apply to the method in order to better verify the results, advantages and possible adjustments that may arise.

ACKNOWLEDGMENT

The authors would like to thank to Pontifícia Universidade Católica do Parana - PUCPR by financial and technical support.

REFERENCES

[1] J. Jeston, J. Nelis. "Business Process Management, practical guidelines to successful implementations". 2ª. ed. Oxford: Elsevier Ltda, 2008

[2] B. Andersen. Business process improvement toolbox, Milwaukee,Wisc., ASQ, 233p. 1999

[3] M. Havey. Essential Business Process Modeling. O'Reilly Media 2005

[4] G. Lomow, E. Newcomer. Service-oriented architecture with web services. 1. ed. São Paulo: Pearson, 2004.

[5] R. H. Hayes, G. P. Pisano, D. M. Upton, S. C. Wheelwright. Operations, Strategy and Technology: Pursuing the Competitive Edge. John Wiley&Sons, 2004.

[6] A. Dutra. Fundamentos da gestão de processos. out 2008. Disponível em: <http://www.portalbpm.com.br/servlet/leartigo?qual=/WEBINF/artigos/digital/>. Acesso em 01/11/2011

[7] A. MELO. Afinal o que a minha empresa ganha com gerenciamento de processos? Mar, 2008.

[8] R. L. Baldam, et al. Gerenciamento de processos de negócios: BPM - business process management. São Paulo: Érica, 2007.

[9] J. F. Chang. Business process management systems: strategy and implementation. Boca Raton, FL: Auerbach Pub, 2006.

[10] M. Hammer, J. Champy. Reengineering the Corporation - A Manifesto For Business Revolution. Harper Business, NewYork, USA, 35 -49, 2001.

[11] M. Muehlen, J. Recker. How Much Language is Enough? Theoretical and Practical Use of the Business Process Modeling Notation. In Proceedings of 20th International Conference on Advanced Information Systems Engineering, pp. 1-15. Montpellier, France, 2008.

[12] OMG. Available Specification. Business Process Modeling Notation, v 2.1., fevereiro 2008.

[13] R. Hayes. Strategic Planning ? forward in reverse? Harvard Business Review, p.111-119, nov.-dec., 1985.

[14] V. A. Zeithaml, M. J. Bitner. Marketing de serviços: a empresa com foco no cliente. 2. ed. Porto Alegre: Bookman, 536p., 2003

[15] W. Leal Filho. Sustainability and university life. International Journal of Sustainability in Higher Education, v. 1, n. 2, p. 168-181, 2000.

[16] R. W. Scholz et al. Higher education as a change agent for sustainability in different cultures and context. International Journal of Sustainability in Higher Education, v. 9, n. 3, p. 317-338, 2008.

[17] R. H. Hayes, G. Pisano. Beyond world-class: the new manufacturing strategy. Harvard Business Review, p.77-86, jan./feb., 1994.

[18] R. Kaplan, D. Norton. Having problem whit your strategy? Then map it. Harvard Business Review, p.167-176, sep./oct., 2000.

[19] W. Skinner. Manufacturing: missing link in corporate strategy. Harvard Business Review, p.136-145, may./jun., 1969.

[20] J. O'Connell, J. Pyke, R. Whitehead. Mastering Your Organization's Processes: A Plain Guide to Business Process Management. Cambridge University Press. 270pp, 2006

[21] ABPM. Association of Business Process Management Professionals. Guide to the BPM common body of knowledge, 2009.

[22] OMG, Business Process Modeling Notation (BPMN), versão 1.2., 2009.

[23] B. Silver. BPMN Method and Style. 2nd edition, Cody-cassidy Press, 2011.

[24] B. G. Dale. et al. Fad , fashion and fit : An examination of quality circles , business process re-engineering and statistical process control. "International Journal of Production Economics", v. 73, n. 2, p. 137-152., 2001

[25] T. H. Davenport. Process innovation: reengineering work through information technology. Boston: Harvard Business School Press, 337 p., 1993.

Network drawing application for use in modern web browsers

Marko Letić, Branislav Atlagić, Dragan Ivetić

Abstract — **The problem of client-side drawing of electrical network elements on a city map in a web browser is analyzed in this paper. Element drawing is programmed in JavaScript language with the support of Cascading Style Sheets and Hyper Text Markup Language 5 elements. Algorithms for object clustering and best fit view are implemented in this solution. Application performance is analyzed in different web browsers and on different platforms.**

Keywords — **Client-side drawing, Hyper Text Markup Language 5, JavaScript**

I. Introduction

ONE of the most important information that an interactive web geographic application should provide about some object is its location. Also, it should show its connections with the other ones, and provide various kind of data associated with it. Whether it's an application that gives information about the location of ATMs, all the restaurants near user's current position or more complex systems such as electrical grid systems, the user must have a clear overview of where the relevant objects are located. The application has to make the view responsive, informative and accurate, independently on the type of the target device [1].

Main subject of this research is the drawing of parts of an electrical grid that are geographically positioned on a city map. For the base city view, on which the grid is drawn, background layers implemented in *OpenStreetMap* [2] are used. These layers are provided by *MapBox* [3] service provider.

Drawing of electrical elements and their integration with the background layers is implemented using *JavaScript* language with the support of *Leaflet.js* library [4]. *Leaflet.js* library presents an open-source *JavaScript* library that implements drawing of objects on HTML5 canvas and their integration with map layers. Special attention is dedicated to the performance optimization of drawing and displaying elements on the canvas while maximizing the usage of resources provided by the client device. Picture 1 shows the application process flow diagram.

Picture 1. *Application process flow diagram*

II. Input data preparation

As an input data XML (*Extensible Markup Language*) document with the description of the city grid is used. Each of the elements in this document, besides the attributes that are specific to that element type, contains a unique geographic location of that element with which the mapping of that element on the city map will be implemented. Coordinates provided are in UTM (*Universal Transverse Mercator*) format. As the *Leaflet* library uses latitude and longitude input format a conversion from UTM had to be done.

III. HTML5 Canvas

Leaflet library supports two types of map and element drawing: using HTML5 canvas element or using SVG (*Scalable Vector Graphics*) format. An explanation of their differences is required to explain why the canvas was used instead of SVG.

SVG image format is an XML format, so the SVG object are directly inserted into the DOM (*Document Object Model*) three. It is possible to manipulate SVG objects using CSS and *JavaScript*. SVG represents a data model that persists in memory. As HTML creates an object model from elements, attributes and styles so does the SVG. When the SVG element is found inside the HTML document it behaves as the part of the DOM three.

HTML5 creates a bitmap inside the canvas of the browser which is easily manipulated. On the canvas it's possible to draw vector graphics but it is rendered using a raster format. One of the main differences compared to the SVG format is that the only way to manipulate the canvas is via *JavaScript*, but it adds only one line in the DOM three. Since canvas represents an HTML5 object, it's supported only in the newer web browsers. Also it is important to add that the SVG is based more on the

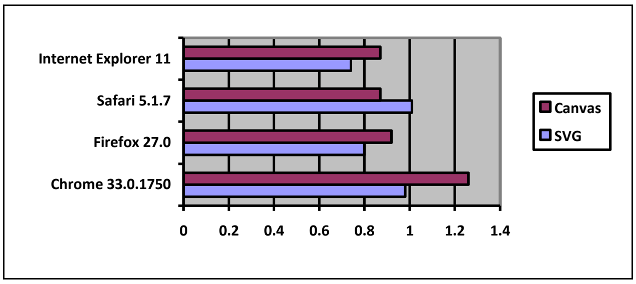
Marko Letić, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia (e-mail: marko.letic@uns.ac.rs).

Branislav Atlagić, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia (e-mail: branislav.atlagic@gmail.com).

Dragan Ivetić, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia (e-mail: ivetic@uns.ac.rs).

description of the shapes itself and the canvas is more pixel oriented.

For the need of performance comparison a simple test scenario is created using *jsPerf* web application. In the test two completely equal HTML container elements are created. Inside the first a map is drawn using the canvas while inside the other one a map is created using SVG. On the both maps the same number of elements is added. Response time of the application while performing a panning action is measured. Picture 2 shows the test results obtained by the use of different web browsers.



Picture 2. *Test results*

It is noticeable that the canvas performs better then SVG in most web browsers. The only browser that shows different results is *Safari* but as it takes only 4% of the market. Surely, much more relevant information is the performance measured in *Chrome* browser which takes 56% [5]. When the number of drawn elements is increased, the difference in performance is also higher. Information relevant for the electrical grid that is displayed on the map needs to be updated in certain time intervals. This type of update would be an extremely expensive operation if, for every change in the grid, the entire DOM three should be re-created, as the SVG demands. Also it's important to state that web browsers such as *Mozilla Firefox* don't perform well with large amount of SVG elements.

HTML5 canvas is able to draw large number of elements and updates the relevant information about those elements in short amount of time and it's supported in almost all web browsers used today.

IV. Drawing Implementation

Before introducing the functionalities available in *Leaflet*, the structure of the page must be defined. All the necessary libraries, basic HTML structure, styles and *JavaScript* functions needed for the map drawing must be included.

After the main page structure is created, a layer that will represent the base on which all the elements will be drawn must be defined. As this application is based on *OpenStreetMap* format, a provider must be found that will on request return an image tile containing map base layer. Requested layer is made of image tiles with the resolution of 256x256 pixels that are merged inside the browser to create a map. This application used *MapBox* service provider, but the choice of the provider is entirely left up to the user.

Layer types that the application will use should be chosen based on the user's needs and preferences. Besides the basic layers that display the vector data for the buildings and streets on the map, it is also possible to choose the layer for night time visibility, some geographic information such as rivers, mountains etc.

There are two types of elements chosen to be displayed onto the map: substation elements that represent medium voltage transformer stations and electrical lines.

Before the actual drawing of the element itself, all the relevant information must be loaded from the XML file, most important of which is the location. Application reads the information from the XML file, creates arrays of all the elements contained inside and then passes these arrays to the functions that are responsible for the drawing.

Function responsible for the substation element drawing as the only parameter takes an array of all the elements of substation type and draws them onto the map. For each element a conversion from UTM to latitude/longitude coordinates takes place. A blue circle marker is chosen as the shape that will represent a single substation. After the marker has been created it is added onto the map.

For a small number of elements which have large enough distance from each other, this way of drawing is quite good. But with the increase of the number of elements that are displayed at once onto the map, the performances start to decline. The solution is to clusterize (group) the elements of the same type, in this case the substations. Criteria for the grouping is the location of the elements on the map [6]. If the elements are close enough and zoom level is high enough, they will be clustered. As the zoom level gets lower, the elements will de-cluster. After positioning the mouse cursor over one of the clusters, the application will mark the surface it takes on the map. Picture 3 shows substation elements added onto the map and their clusterization.

Picture 3. *Substation elements drawn on the map*

Electrical line elements contain more information than the substations. One line is consisted of more connected segments. Also the difference between overhead and cable lines must be displayed. This difference is shown in different line coloring.

Algorithm used for the drawing of electrical lines is different from the one used for substations. Every point of a single line segment must be added to an array that creates

a polyline object that represents line segments. Before the drawing, a check must be made to see if the type of the line matches applied coloring. A legend is added to the bottom right side of the application to connect the line type with the used color. After this object is created, it needs to be added to an array that contains all the polylines that will eventually be drawn on the map. These actions are done for all the segments in iterations.

V. Element Manipulation

As frequent change of values describing the system are expected, these changes must be visible on the map as soon as the information of the change is received.

This means that the application should only update the data on the map and not redraw the entire view with every new change. So with periodic request for change updates, only those element that were affected by the change will be updated. This type of information can be sent in any format readable for the web browser such as XML or JSON format. With the decrease of sent information, so does the size of the data sent to the client decreases while in the same time updates are being applied faster. Global performance of the application is increased as less data is needed for the update.

Server processes are simulated on the client side with a random updates of the values contained in the substation elements. Each of the elements has a unique identifier. With the form available for the element update, the user can change values of a substation with the appropriate identifier. Element changes the color to better illustrate the speed required for the update operation.

Leaflet library doesn't contain the possibility for the map to run in full screen mode. That's why a plugin that enables this is added to the application. Simple click on the control opens the map in full screen.

Surface that is covered by the electrical grid is usually equal if not larger than the surface of the city itself. On the lower zoom levels user is not aware of the grid size and this is why a control is created to set the zoom level on the optimum level.

With the increase of element types displayed on the map, a user must be able to choose which elements should be visible and which should not. The map layer itself carries information such as street and building names, different colorings of those objects etc. When the information about the electrical grid is added on top of the map, the view becomes overpopulated and unreadable.

This is why the user is allowed to show or hide the available element types that are also grouped into layers. But also an option of alternative map base layer should be available. A functionality that enables a simple layer switch is implemented. Picture 4 shows the map view with implemented plugins and controls.

Picture 4. *Map with available layers and added controls*

VI. Result Comparison

For the testing purposes a plugin that measures frame rate of the rendered view is included in the application. Measured values are recorded for two scenarios. In Scenario 1 there is no value updates, while in Scenario 2, 5 random elements are updated on each 5 seconds. Values shown in the table represent average frame rate per second (fps) in the browser in which the application was tested. Measurement results are shown in the Table 1.

Web browser	*Scenario 1*	*Scenario 2*
Google Chrome 33	*54 fps*	*43 fps*
Mozilla Firefox 27	*32 fps*	*21 fps*
Safari 5.1.7	*36 fps*	*22 fps*

Table 1. *Measurement results*

VII. Conclusion

Although better performances can be acheved using hardware accelerated drawing, this type of client side drawing can take the load off the server and use the all the potential of modern web browsers to acheve more than satisfying rendering speeds at low cost. Also with this kind of implementation the user can choose any device that has a web browser and not worry about the compatibility of the software with the used device.

Besides the shown use in displaying the electrical grid of the city, this application can be used to display telecomunication, water or gas infrastructure status in real time in the web browser.

References

[1] W. Zenghong, "An emerging trend of GIS interaction development: Multi-touch GIS", Information Science and Control Engineering 2012 (ICISCE 2012)

[2] OpenStreetMap's community, "OpenStreetMap". Available: www.openstreetmap.org/about

[3] MapBox company, "MapBox maps". Available: www.mapbox.com

[4] Vladimir Agafonkin, "Leaflet - a JavaScript library for mobile-friendly maps". Available: www.leafletjs.com

[5] W3Schools, "Browser Statistics", Available: www.w3schools.com/browsers/browsers_stats.asp

[6] H. Yan and R. Weibel, An algorithm for point based cluster generalization based on the Voronoi diagram, Computers and Geosciences 34, 2008

WEB APPLICATIONS DEVELOPMENT USING A COMBINATION OF JAVA AND GROOVY PROGRAMMING LANGUAGES

Katarina Mitrović*, Danijela Milošević*, Nenad Stefanović*
*Faculty of Technical Sciences Čačak, University of Kragujevac, Serbia
mitrovic.kaca@yahoo.com
danijela.milosevic@ftn.kg.ac.rs
nenads@kg.ac.rs

Abstract *- The area of this research is Web application development in Grails framework using Java and Groovy programming languages. This paper briefly describes the basic features of Grails framework, as well as the Groovy programming language, and the development of Web application using these technologies. Also, this paper presentes comparative analisys of developing different functionalities in Java and Groovy programming languages. The aim of this work is to study similarities and differences of Java and Groovy programming languages, as well as to analyse the advantages and disadvantages of applying a combination of these languages in the development of Web applications.*

1. INTRODUCTION

Ever since the programming languages were designed, the focus in the application development was on a single language. However, programmers who used only one programming language were not able to maintain high quality in each part of application, and very often they could not implement certain features, partially or completely. Over time, a lot of programming languages specialized for different purposes have been developed. It has come to the conclusion that the most effective software development could be achieved only by combining several programming languages. This trend was called polyglot programming.

This work is devoted to the analysis of a polyglot programming, in which are used Java and Groovy as a key programming languages. An overview of existing results and the literature pertaining to the subject of this paper is provided in the next chapter. The Groovy programming language and Grails framework are briefly described in the third chapter, while the development of Web applications using the above mentioned technologies is analysed in the fourth chapter.

2. BACKGROUND RESEARCH

Nowadays, the development of Web applications without combining at least two different programming languages is being difficult to imagine. Learning multiple programming languages is a demanding process, but when the necessary programming languages are mastered, developers are able not only to meet the requirements of customers, but also to speed up the process of creating applications, and at the same time to make code more concise and understandable. *Having basic knowledge of more than one language can never be a con, but being expert on just one will limit you to find work for you.*[1]

In multi-language programming paying attention to the level of interoperability of systems is necessary. In some cases, programming langages can be integrated smoothly and easily, but there are cases where the integration can be very difficult to achieve and where is such a combination of programming languages being unprofitable to apply in a single application. This is being determined by the programming languages themselves, necessary application functionalities, framework usage etc.

The example of polyglot programming, in which interoperability between systems takes place in a simple and effective way, is development of applications by using a combination of Java and Groovy programming languages, to whom this work is dedicated.

Development of **Groovy programming language** was started in 2003, with the goal of integrating dynamic functionality into Java programming language. The idea was initialized by programmer James Strachan, and the first version was realized in 2004. *Groovy is an agile dynamic language for Java platform, with many features that are inspired by languages like Python, Ruby and Smalltalk, making them available to Java developers using Java syntax.*[2]

Most of the functionalities offered by the Java programming language are included in Groovy, but Groovy is a lot more concise, because it takes less code for implementing them. XML parsing and SQL data manipulating is facilitated with Groovy, and a solution for scripting is provided by Groovy. Also, features that are not enabled in Java are provided by Groovy, such as writing code without semicolon, writing code without brackets, *def* data type for dynamic type assigning, maps, operator overloading, Star operator, Elvis operator, Spaceship operator, support for regular expressions, polymorphic iterator, incorporation of expressions in String data types, safe navigation operator, modularity, type checking, markup languages support, etc.

In a very short time Groovy programming language became worldwide, and some of the largest companies embedded Groovy in their business, such as Netflix, Cisco, IBM, LinkedIn, Master Card, Oracle, Sony, UBS, etc.[3]

Grails framework represents an open source Web framework based on Groovy and Java programming languages. Grails is a powerful Web framework for the Java platform, whose goal is to duplicate the productivity of programmers. This is achieved with the "*convention over configuration*" approach, smart quality standards and APIs.

The whole Grails concept lies on Java virtual machine. Several software frameworks and technologies are included in Grails, such as Java EE, Spring, Hibernate and Sitemesh, and their functionalities are integrated in Grails, beside the Groovy programming language. The speed of scripting languages and the stability and security of the Java programming language are combined in Grails framework. A number of important technologies are provided to Grails by Java EE, such as agile software development, Web services and testing units. Grails is based on the MVC architecture, which is built through the abstraction of Spring's MVC, while ORM which is used in Grails is an abstraction built over Hibernate.[4]

The existing results and the literature pertaining to the subject of this paper are listed below:

- Benjamin J. Evans, Martijn Verburg, *The Well-Grounded Java Developer: Vital techniques of Java 7 and polyglot programming*, Manning Publications Co. Greenwich, CT, USA, 2012
- Adam L. Davis, *Modern Java: Java 7 and Polyglot Programming on the JVM*, CreateSpace Independent Publishing Platform, USA, 2014
- Dean Vampler, Tony Clark, *Guest Editors' Introduction: Programming Multiparadigm*, IEEE Computer Society, 2010
- Juhana Harmanen, *Polyglot Programming in Web Development*, Faculty of Computing and Electrical Engineering, 2013
- Hans-Christian Fjeldberg, *Polyglot Programming: A business perspective*, NTNU, 2008

The development of applications using combination of Groovy and Java programming languages in Grails framework is described in the following section, through examples and comparative analysis. Also, the advantages and disadvantages of this type of programming are listed in the next chapter.

3. THE EXAMPLE OF WEB APPLICATION DEVELOPMENT USING COMBINATION OF GROOVY AND JAVA PROGRAMMING LANGUAGES

For the purposes of this paper, explorative application, which implements Sales module, was developed. The application consists of activities regarding order processing. The preview of the explorative application is shown in Figure 1. Functionality is order entry functionality is demonstrated.

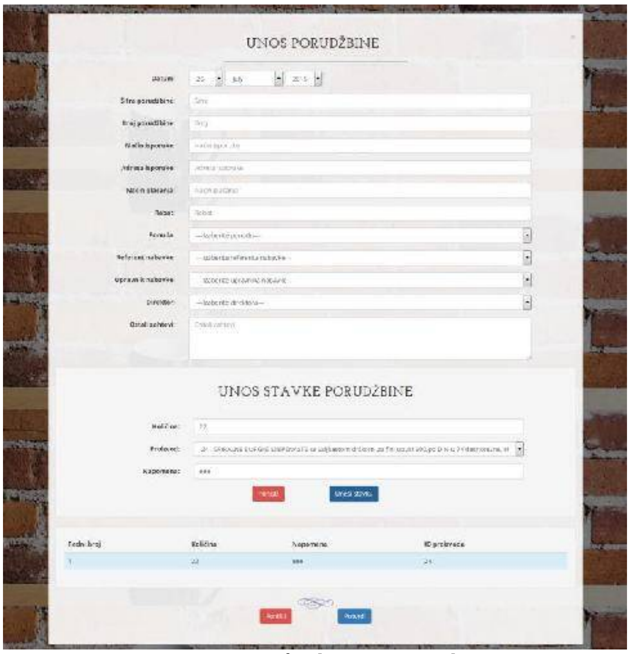


Figure 1 - Preview of order insert in application

The class diagram for the purchase order, with its entities, relations and other elements, is shown in the Figure 2.

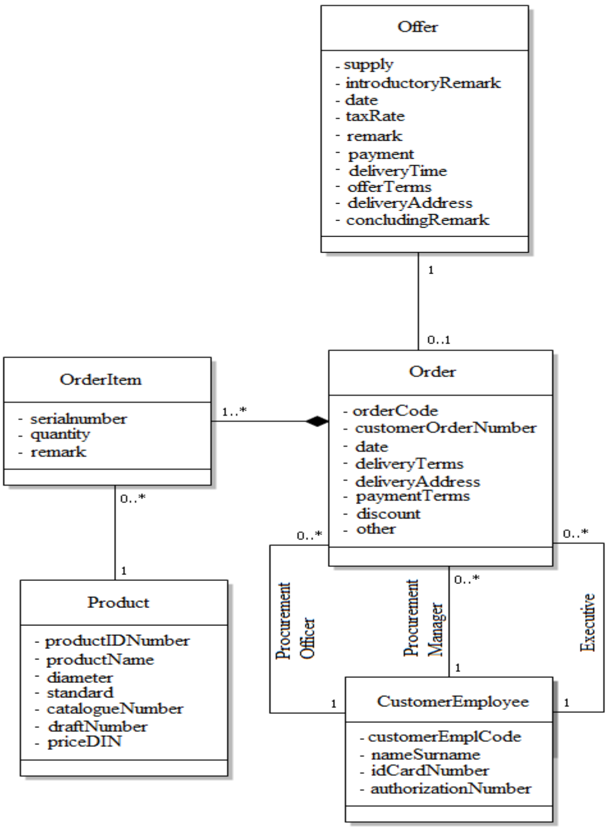


Figure 2 – Class diagram for Order

In explorative application development is used following:

- NetBeans 8.0.2. development environment based on JDK 1.7 Java platform (https://netbeans.org/downloads/, http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html).

• Grails 2.5.0. framework (https://grails.org/download.html).
• Groovy 2.4.3. programming language (http://www.groovy-lang.org/download.html).

Several directories with a particular purpose are constituting Grails project, such as Configuration, Domain Classes, Controllers, Services, Views and Layout, Libraries etc. Groovy and Java programming languages can be combined in Controllers and Domain Classes directories, so the maximum benefit from both languages can be obtained.

When writing code, the errors are not invoked if the code was written either in Java or Groovy programming languages. Each line of code can be written in any of these two programming languages, the semicolon at the end of the code line is optional, one method may consist in both Java and Groovy commands etc. There has been a great flexibility when writing code and different combinations of functionalities of these two programming languages can be used.

The main question that arises after the conclusion above is **how is the program compiling the class in which are mixed two different programming languages without any problems**?

The answer to this question lies in following: first of all, Java objects and classes are percieved by Groovy programming language as its own; second, the Groovy code is converted into Java byte code after compiling of Groovy classes, which is different from any other language that supports the JVM.[5] Since the Java programming language is compiled into Java byte code, it can be easily concluded that integration between Groovy and Java programming languages is achieved on the highest level.

4. RESULT ANALYSIS

4.1. Implementation of Groovy programming languages

One of the functionalities of Groovy programming language which is very useful and which was used multiple times in demonstration application is **map**. The tag "`[:]`" is used for defining map and any number of members can be contained within a map. Each map member consists of a key and the assigned value. The default key type is String.

In the following example from the demonstration application, map is used for the presentation of the items of the document, whose members are attributes of item:

```
def stavka = [:]
stavka << [redniBroj:sp.getRedniBroj()]
stavka << [kolicina:sp.getKolicina()]
stavka << [napomena:sp.getNapomena()]
stavka << [idProizvoda:sp.getIdProizvoda()]
```

In the following example from the demonstration application, map is used for transfering message about operation execution successfulness:

```
def message = [:]
message = [ispis:"Uspešno ste obrisali
proizvod!",warning:"alert alert-dismissible
alert-success"]
```

For database **queries,** simplified Groovy commands are used. A query for displaying codes and names of all the employees of customers from the database is shown in following example:

```
def db = new Sql (dataSource)
def result = db.rows("SELECT
SifraRadnikaKupca, ImePrezime FROM
RadnikKupca ORDER BY SifraRadnikaKupca")
```

For view displaying and passing values into them, Groovy keyword `render` is used. In the following example, **render** command is used for displaying view "unos" and forwarding message values and list "listaJM" in that view:

```
render(view: "unos", model:
[message:message, listaJM:listaJM])
```

Passing through the members of the collection is simplified and improved in Groovy programming language by using `each` command. The following code shows how **each** member of the collection listaJMSQL is placed in the list listaJM, ie. how the list of units of measure is converted to a list of text values:

```
listaJMSQL.each { row ->
JedinicaMere jm = new
JedinicaMere(row.SifraJM, row.NazivJM)
listaJM << jm.toString()
}
```

A comparative analysis of Groovy and Java functionalities can be made after researching the features and performances of Groovy programming language on the examples.

4.2. Comparative analysis of Groovy and Java programming languages

More flexible, shorter and readable code can be achieved by using following Groovy features:[5]

- Groovy is a *dynamic language*, which means that data types do not have to be assigned while writing the code; they can be assigned automatically during compilation. Groovy does not perform the standard call of methods. The call is sent as a message to an object, which can respond to it or not. In dynamic code writing, the data type is not important, but only a response message from the object.[6]
- The default access modifier is *public*.
- Instead of System.out.println, it is sufficient to write just *println*.

- The *semicolon* at the end of code line is not mandatory.
- Due to the dynamic nature of this language, the methods do not need to have a precisely defined return value. It is sufficient to use the universal data type *def*.
- The last line of code in the method is considered as return value of the method, so the usage of *return* command is optional.
- The upgraded version of the String data type – GString – is another useful innovation which Groovy brings. In this data type, concatenation is not performed with double quotation marks, but with literal "`$ {text}`".
- Properties are used instead of getters, setters and private access modifier of the attribute.

The following paragraphs will be devoted to a comparative presentation of certain functionalities in Groovy and Java programming languages with examples from demonstration applications.

Working with lists

Groovy

```
listaRKSQL.each { row ->
...
}
```

Java

```
for(RadnikKupca rk : listaRKSQL){
...
}
```

Groovy programming language is using the keywords `each` and `row`, while Java performs the same functionality with `for` loop.

Definition of method

Groovy

```
def proveriString(def s) {
...
}
```

Java

```
public String proveriString(String s) {
...
}
```

Due to the dynamic writing of Groovy programming language and default public access modifiers, Groovy method definition is shorter than in the Java language. Declaring of return value data type of methods and data types of parameters is not necessary.

The return value of method

Groovy

```
def getSifraJM() {
sifraJM;
}
```

Java

```
public int getSifraJM() {
return sifraJM;
}
```

The last code line in the method in Groovy programming language is automatically regarded as the methods return value, while in Java language is necessary to declare the data type of the return value and to specify the `return` keyword.

Working with database

Groovy

```
def vratiListuRadnikaKupca() {
def db = new Sql (dataSource)
def result = db.rows("SELECT
SifraRadnikaKupca, ImePrezime FROM
RadnikKupca ORDER BY SifraRadnikaKupca")
result
}
```

Java

```
public ResultSet vratiListuRadnikaKupaca()
throws ClassNotFoundException, SQLException
{
Class.forName("com.microsoft.sqlserver.jdbc
.SQLServerDriver");
String url =
"jdbc:sqlserver://localhost:1433;databaseNa
me=SOA_Baza;user=Dell-
PC\\User;password=;integratedSecurity=true;
";
Connection konekcija =
DriverManager.getConnection(url);
String upit = "SELECT SifraRadnikaKupca,
ImePrezime FROM RadnikKupca ORDER BY
SifraRadnikaKupca";
Statement ssql =
konekcija.createStatement();
ResultSet rs = ssql.executeQuery(upit);
return rs;
}
```

In Groovy programming language, connection to the database is defined only once in the DataSource configuration file, so connection with database is established with a single line of code. The execution of queries in Groovy programming language is also automated by using *rows* keyword, so the obtaining of query results is performed with one line of code. When comparing the amount of code written in Groovy and Java language, it can be concluded that the innovations implemented in Groovy programming language influenced to significantly reduce the amount of code needed for connecting to database and executing SQL queries.

Concatenation of Strings

Groovy

```
def toString (){
"${sifraJM } - ${ nazivJM }"
}
```

Java

```
String toString (){
return sifraJM + " - " + nazivJM;
}
```

In Groovy programming language the attributes can be directly called within the string (ie. Gstring) by combinig the characters - “`${}`”, while writing attributes outside the quotes is requierd in Java programming language.

Get and set methods (properties)

Groovy

```
def nazivProizvoda
```

Java

```
private String nazivProizvoda;
public String getNazivProizvoda() {
    return nazivProizvoda;
}
public void setNazivProizvoda(String nazivProizvoda) {
    this.nazivProizvoda = nazivProizvoda;
}
```

In Groovy programming language, instead of the standard Java getters and setters, properties are used. They represent something between methods and attributes and they are implicitly called through their get and set methods.

The advantages and disadvantages of applying a combination of Groovy and Java programming languages in developing applications is listed in the next chapter.

4.3. Advantages and disadvantages of combining Groovy and Java programming languages

The basic features and benefits of Groovy programming language are following [3][7]:

- Flat learning curve (Groovy is easy to learn for Java-programmers);
- Support for domain specific languages (flexible syntax, advanced integration and customization mechanisms for integrating readable business rules in application);
- Compact syntax (concise, readable and expressive syntax);
- Support for dynamically writing;
- Powerful processing of primitives (basic interface, or the part of code that is upgraded in order to create more sophisticated interface or code);
- Facilitating development of Web applications;
- Support for a unit testing;
- Perfect integration with Java (Groovy is transparently and smoothly connected with Java programming language and any library);
- Dynamical and rich ecosystem (development of Web applications, concurent, asynchronous, parallel execution, test frameworks, additional tools, code analysis, support for development of user interface – these are just a few of features that Groovy covers);
- Powerful functionality (meta-programming, functional programming, static compiling and many others);
- Support for scripting.

Some of the most important advantages, which are achieved using a combination of Groovy and Java programming languages in Grails frameworks for Web applications development, are:

- A concise, simple and clean code.
- Increased productivity of programmers.
- The ability to adapt quickly to change.
- High performance and scalability.
- Cost reduction, time savings and improved software quality.
- Modularity and reusability.
- Very active and up-to-date communities.
- Maximum use of the benefits of the frameworks over which the Grails framework is made.
- The ease of setting up and starting the work.
- Focus on domain classes.
- Minimized need for restarting the application upon changes.
- Built-in testing.
- No configuration.
- Built-in support for REST.
- Easy integration.
- Large number of available plug-ins for Grails.

Some of the ***negative*** effects which are common when using a combination of Groovy and Java programming languages in Grails framework for Web applications development are:

- Learning multiple programming languages.
- If the domain class object is placed in the session, there might be a separation of the object from the session.
- Some parts of the implementation "leak" from time to time.
- Declaring all variables with def is unsustainable and illegible.
- Since the number of lines of code is reduced to a minimum, it can not be effectively applied to a large number of patterns.
- GORM is problematic in dealing with multi-threaded applications.
- It does not support other ORM technologies.

5. CONCLUSION

This research has analysed one type of polyglot programming, i.e. programming using a combination of Java and Groovy programming languages in Grails framework. It has been shown that the polyglot programming technique results in a very flexible, high quality, fast and concise software. Polyglot programming is a programming by using more than one programming language, which means that each part of the application is developed with programming language that provides the needed functionality in the best possible way. Grails framework offers the "foundation" for the development of applications, and Java and Groovy programming

languages can be used and combined, depending on the functionality that needs to be implemented.

Once the programmer learns the way how Java and Groovy programming languages work together in Grails framework, and get to know the structure and functioning of the Grails frameworks, he is able to create a Web application in such a way to gain a maximum benefit, and bypass deficiencies of single-language programming.

With single-language programming, Web applications can be created, but with using Java and Groovy programming languages application creation is easier, both in temporal and functional sense. Web application development time is shorter, implementation of functionality is simplier, each programming request can be resolved in at least one way, code is concise, clear and easy to maintain when these programming languages are being used, which leads to development of high quality applications and satisfaction of both programmers and users.

REFERENCES

[1] *What are the pros and cons of mastering a single programming language??*, https://www.quora.com/What-are-the-pros-and-cons-of-mastering-a-single-programming-language, [online] january 2016.
[2] *Groovy Language History*, http://gr8labs.org/getting-groovy/, [online] january 2016.
[3] *About Groovy*, http://www.groovy-lang.org, [online] january 2016.
[4] Rajeshkumar I.P, *Grails Web framework*, Sapient, 2013.
[5] Andres Almiray, *Introduction to Groovy*, http://www.javalobby.org/articles/groovy-intro1/, [online] january 2016.
[6] *How does Groovy work in a dynamic way?*, http://stackoverflow.com/questions/13538571/how-does-groovy-work-in-a-dynamic-way, [online] january 2016.
[7] Margaret Rouse, *What is Groovy?*, http://whatis.techtarget.com/definition/Groovy, [online] january 2016.
[8] *Introduction to the Grails Web Framework*, https://netbeans.org/kb/docs/web/grails-quickstart.html, [online] january 2016.
[9] Scott Davis, Jason Rudolph, *Getting Started with Grails*, Second Edition, InfoQueue, 2010.
[10] *Groovy and Grails application development*, http://people10.com/blog/groovy-and-grails-application-development-2/, [online] january 2016.
[11] *What is Grails and Groovy?*, http://uws-software-service.com/en/services/java-competence/software-development-with-the-grails-framework.html, [online] january 2016.
[12] Venkat Subramaniam, *Programming Groovy*, The Pragmatic Bookshelf, 2008.

Generate User Interface Using Xtext Framework

Dragana Aleksić[1], Dušan Savić[1], Siniša Vlajić[1], Alberto Rodrigues da Silva[2] Vojislav Stanojević[1], Ilija Antović[1], Miloš Milić[1],
Faculty of Organizational Sciences, University of Belgrade [1]
Department of Computer Science and Engineering,IST / University of Lisbon[2]

Abstract- A very important aspect of the software development process is the cost of completing the software project. The development of the user interface is a significant part of such a process. Typically, the graphical user interface of an interactive system represents about 48% of the source code, requires about 45% of the development time and 50% of the implementation time. Therefore, if this part of the software development process can be automated in some way, it can reduce the cost of a software project. In this paper we propose a small domain-specific language for specification of a graphical user interface and present a generator for Java desktop applications. The generator should facilitate the development of a software system by automatic creation of immutable code and creation of variable code based on parameters of the input specification of the generator. For this purpose we have used an Xtext open-source framework. A small but representative example is shown to describe the whole process.

Keywords: generator, prototype, user interface, domain specific language

I. INTRODUCTION

While creating software one comes across various problems. Many of them are related to collecting user requirements and overall time needed for designing and creating graphical user interface (GUI). User requirements are often represented by verbal description provided by end user or/and domain expert together with programmer. Discovered in later phase of lifecycle development of software system, incomplete and incorrect user requirements can increase cost and overall time needed for development of entire system.

In accordance with this a very important aspect in software development is the cost of completion of a software project. Development of the user interface is a significant part of such a process [1].Development of user interfaces (UIs), ranging from early requirements to software obsolescence, has become a time-consuming and costly process. Typically, the graphical user interface of an interactive system represents about 48% of the source code, requires about 45% of the development time and 50% of the implementation time, and covers 37% of the maintenance time [2].

A step forward in software industry towards solution of above problems is creation of various code generators which generate parts or even entire software systems of particular domain. Code generator is a component that generates appropriate software system from input specification model usually expressed in some modeling language. Modeling language as a set of all possible models that are conformant with the modeling language's abstract syntax, represented by one or more concrete syntaxes and that satisfy a given semantics [3]. The code that is generated by generator is divided into: 1) Generic codes that represents a code which is independent of concrete applications and can be used in other applications of the same type without changes, and 2) Specific codes that represents the code that is specific to a particular application or can be seen as some templates, upon which program code is created. In this paper we introduce a UIL domain specific language (DSL) for user interface specification implemented using Xtext framework, and appropriate generator. Xtext is based on openArchitectureWare generator framework, the Eclipse Modeling Framework and Another Tool for Language Recognition (ANTLR) parser generator. This paper is organized as follows. Section II describes the background of this work while Section III present related work. Section IV presents proposed DSL and appropriate code generator. Section V concludes the paper and outlines future work.

II. BACKGROUND

According to [4], domain specific language is a programming language of limited expressiveness focused on particular domain. The term “limited expressiveness” refers to the fact that DSLs only have minimal features needed to support its domain as opposed to general purpose languages (GPLs) which provide a lot of capabilities that can be applied on various domains.

Language workbenches were created after some time as supporting tool for creation of external DSLs. They are specialized environments which facilitate development of DSL by providing a good support for creation of both concrete and abstract syntaxes, automatic creation of underlying parser and complete editing environments for using DSLs with all the necessary support needed for creating a program like syntax highlighting, code

completion, customizable outline and real-time constraint checking.

Model Driven Development (MDD) looks at the software development lifecycle through modeling and transformation of one model into another. MDD's goal is to replace traditional methods of software development with automated methods with domain-specific languages which efficiently express domain concepts and directly represent domain problem. Such development directs traditional programming into creating domain-specific models and code generators which further leads into higher software productivity, therefore, increased satisfaction of the end user.

Code generators make software development process easier by shortening time necessary for writing programming code as well as time for testing the generated code, considering that generated parts of system was tested already. Generators can be used for creating parts of the system or for creating whole system as a prototype which can be rejected, upgraded or kept as a final solution.

III. RELATED WORK

Today there are so many modeling languages. Modeling languages might be classified as general purpose or domain-specific modeling language [5], [6] The one part of them are User-Interface Modeling Languages such as UMLi, UsiXML, WebML (with IFML), XIS, DiaMODL, XIS-Mobile and etc. Morais and Silva [7] introduce ARENA framework and evaluate the quality and effectiveness of modeling languages [8]. Ribeiro and Silva in their paper [9] present XIS-Mobile language which is defined as UML profile and supported by the Sparx Systems Enterprise Architect. XIS-Mobile is part of the broader XIS project that considers three major groups of views: Entities, Use-Cases and User-Interfaces. [10] [11]. XIS focuses on the design of interactive software systems at a Platform-Independent Level according to MDA. XIS-Mobile language reuses some concepts proposed on the XIS language and introduces new ones in order to be more appropriate to mobile applications design. It has a multi-view organization and supports two design approaches: the dummy approach and the smart approach. Dejanovic at al. [12] present an extensible domain specific language (DOMMLite) for static structure definition of database-oriented applications. Also, they developed appropriate textual Eclipse editor for DSL from which they generate complete source code for GUI forms with CRUDS operations. [13]. Popovic at al. [14] proposed IIS*CFuncLang DSL to enable a complete specification of application-specific functionalities at the PIM level and developed algorithms for transformation of IIS*CFuncLang specifications into PL/SQL program code. Furthermore, they presented tree-based and textual editors that are embedded into IIS* Case development tool. Tran at al. REF [15], [16] proposed a process that combines the task, domain, and user model in order to design user interface and generate source code. They propose framework as well as software prototype tools named as DB-USE. Kennard has emphasized the need for UI generation within mainstream software development and explains five characteristics, which he believes, are key to a practical UI generator. These characteristics are discovered through interviews, adoption studies and close collaboration with industry practitioners [17]. Cruz an Faria proposed an approach to create platform independent UI models for data intensive applications, by automatically generating an initial UI model from domain and use case models. [18], [19] defined extensions for domain model and use case model to support user interface generation. They describe how both of these models could be mapped in user interface features. Pastor at al. proposed full Model Driven Development approach from requirements engineering to automatic code generation by integrating two methods: Communication Analysis and OO-Method [20]. Smiałek and Straszak presented a tool that automates transition from precise use case and domain models to code [21]. Smiałek at el. defined RSL as a semiformal natural language that employs use case for specifying requirements [22]. Each scenario in a use case contains special controlled natural language SVO (O) sentence. RSL has been developed as a part of ReDSeeDS project [23]. ReDSeeDS approach covers a complete chain of model-driven development – from requirements to code [24]. Requirements is described in form of use cases using constraint requirements specification language which can be used to generate complete MVC/MVP code structure.

Inside our Laboratory we developed integrated SilabMMD approach [25] that use SilabReq language [26], [27] for specifying user requirement which is implemented inside JetBrains Meta Programming System and can be used as plug-in for IntelliJ IDEA or for the MPS tools. In [28] we proposed use cases specification at different levels of abstraction to promote better integration, communication and understanding among involved stakeholders. Using this approach different software development artifacts such as domain model, system operations, user interface design can be automatically generated or validated. In [29] we identified the correlations between the use case model,

data model and the desired user interface, and purpose different ways of user interaction with the system and recommended the set of most common user interface templates.

IV. DEVELOPMENT OF DSL AND USER INTEREFACE GENERATOR

Developed generator creates GUI based on model description which is user defined with UIL DSL. Depending on model description, developed generator can produce limited scope of different user interfaces.

Desired DSL should be able to describe domain model like any other modeling language (for instance UML) in addition with information how domain object and their attributes are represented in GUI. Based on detail model description, developed generator automatically generates executable code, providing user with functional GUI in a few seconds. If not satisfied with the look, user can easily change DSL script and quickly see how changes he or she made are reflected in the final design.

Creation of user interface generator can be divided into several phases. First phase represents development of domain specific language for describing input specification of generator. The second phase refers to designing the architecture of software which is the expected output from the generator. Third phase covers the analysis of reference code implementation during which are noticed general and specific parts of the system. And final, fourth phase represents defining code generator.

A. DEVELOPMENT OF DOMAIN SPECIFIC LANGUAGE

As [4] points out, development of external DSL is very similar to development of GPL in a way that both have key parts that makes programming language: abstract syntax, one or more concrete syntaxes and whole process of parsing input including lexical and syntactic analysis.

For each domain object, user chooses whether the object will have representation or not. Attribute type supported are `int`, `double`, `boolean`, `String`, `Date` and reference type. Based on type of attribute, UIL DSL provides several graphical implementation choices, from which user, while describing attribute of certain object, can choose. Graphical representations supported are `TextField`, `TextArea`, `NumberPicker`, `RadioButton`, `CheckBox` and `ComboBox` which is default representation for reference. For establishing relation between object UIL DSL provides two graphical components: table and list which are visually placed in separate window and populated with collection of appropriate objects Also, user is able to customize certain configuration of graphical component, like minimal or maximal value in `NumberPicker` component or add some validation like regular expressions to text based components. An extract of our UIL DSL grammar is represented below.

```
Model: 'Domain objects:'
          objects+=Object*;
Object:
'{'  (createForm?='create form' (',''number of columns:'
columns = INT ',''number of rows:' rows = INT)? ',')?
'name:'name = ID ',''attributes:'  attributes+=Attribute*
'}';
Attribute: (PrimitiveType | ReferenceObject);
ReferenceObject:
          '{''name:'name=ID ':' type= [Object] '['minLimit=
Limit '..' maxLimit = Limit ']'
           (',''representation:' representation =
('ComboBox') (',''label name:' label= STRING)?)?
           (',''position:''['column = INT ';' row = INT']')?
           (',''selection:' selection = SelectionType)?
'}';
PrimitiveType: '{''name:'name=ID ':' type=  Type
(',''position:''[' column = INT ';' row = INT ']')?
(',' required?='required')?
'}';
SelectionType: name = ('List' | 'Table');
Type: (TypeString | TypeInt | TypeBoolean | TypeDouble |
TypeDate) ;
TypeString:name ='String' (',''representation:'
representation = (TextArea |TextField))?;
TypeInt: name ='int' (',''representation:' representation =
(TextField | NumberPicker))?;
TypeDate: name ='Date' (',''representation:' representation
= (TextField | NumberPicker))?;
TypeBoolean: name ='boolean' (',''representation:'
representation = (RadioButton | CheckBox))?;
RepresentationType: TextField | TextArea | NumberPicker |
RadioButton |CheckBox;
TextField: name ='TextField' (',''label name:' label=
STRING)? (',''regex:' regex= STRING)?;
NumberPicker: name = 'NumberPicker' (',''label name:' label=
STRING)? (',''initial value: ' initial = STRING)? (',''min
value: ' min = (STRING /*|'null' */ ))?
(',''max value: ' max = (STRING /*| 'null' */))? (',''step:
' step = STRING)? (',''format: ' format = STRING)?;
RadioButton: name = 'RadioButton' (',''label name:' label=
STRING)? (',''option true:' optionTrue = STRING)?
(',''option false:' optionFalse = STRING)?;
```

Let's define domain object called Pet with attribute race as type String, for which we don't want graphical representation. Next, we define Person with attributes first name and last name as type String, attribute date of birth as type Date and relation towards Pet with cardinality zero – to – many since in our model person can have none or many pets. While defining attributes of object, we have to specify required graphical representation that is `textfield` for first and last name, `numberpicker` for date of birth and `combobox` for relation. Since Person has relation with Pet we have to specify desired selection, in this case we choose table. For all graphical representation we can specify accompanying text as decryption of attribute that will be seen on GUI as label. Further, for `numberpicker` we can specify initial, minimal and maximal value as the format in which date will be represented. Also, we have

an option to specify positions where concrete attribute will be placed inside of GUI by providing desired number of columns and rows and exact index of pair column/row for each attribute. We want all attributes of Person to be arranged in one column.

```
Domain objects:{
         name: Pet,
         attributes:
         {name: race: String} }
{create form,
name: Person, number of columns:1, number of rows:4,
attributes:
{name: firstName: String, label name:"First name",
representation:TextField,position: [1; 1],required }
{name: lastname: String, label name:"Last name",
representation:TextField,position: [1; 2]}
{name: dateOfBirth: Date,
representation:NumberPicker,
         label name:"Date of bith",
         initial value: "05.02.1900",
         step: 'Calendar.YEAR',
         format: "dd.MM.yyyy",
         position: [1; 3] }
{name: pets: Pet [0..*],
         representation:ComboBox,
         position: [1; 4],
         selection:Table}
}
```

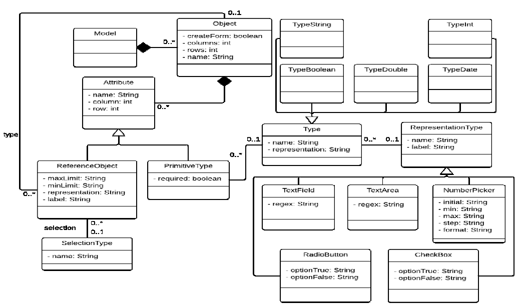


Figure 1 – Meta – model of UIL DSL

Abstract syntax describes concepts of programming language and their relationships. Fig.1 presents a meta – model of our UIL DSL. Concepts that are used in developed domain specific language are concepts that are tightly connected to the concepts of problem domain for which generator is developed. Given that the main concepts of UIL DSL are objects of domain model, their attributes, types and graphical representation of attributes. Concepts of abstract syntax can be represented with various concrete syntaxes. In Xtext, abstract syntax is represented in a form of ecore model and it is automatically created by Xtext based on defined concrete syntax.

Semantics defines the meaning for syntactically valid program by providing a set of rules which unambiguously specifies the meaning of a program. According to [30], semantics of language is translation of inputs into some target language which already has some behavior definition for its elements.

One way that DSLs usually differ from GPLs is that DSLs introduce one more intermediate layer called *semantic model* which is created after the parsing process is done and populated with data from abstract syntax tree. [4] points out that the semantic model defines semantic of DSL.. We implemented various validation rules, which will prevent end user from making semantically incorrect statements. Below is given one validation rule for cardinality between objects which prevents upper bound to be smaller than lower bound and also prevent upper bound to be zero.

```
@Check
publicvoid checkMinAndMaxLimit(ReferenceObject object) {
         if (object.getMaxLimit().equals("0")) {
error("Max limit cannot be 0",
FormDslPackage.Literals.REFERENCE_OBJECT__MAX_LIMIT);}
         if(object.getMinLimit().equals("*") &&
!object.getMaxLimit().equals("*")){
error("Min limit cannot be greater than max
limit",FormDslPackage.Literals.REFERENCE_OBJECT__MIN_LIMIT);
}}
```

After all changes in DSL script have been saved, generator begins with parsing process. Behind the scene, Xtext is integrated with ANTRL which generates parser for defined concrete syntax and population of abstract syntax tree with input data. Semantic model is automatically built from abstract syntax tree.

B. *DESIGNING OF ARCHITECTURE OF SOFTWARE*

After development of DSL, next phase is designing desired architecture of software system which will be generated. Software system needs to be well structured so it can be easily extended and maintained. We accomplished good structure of the system by implementing design patterns and SOLID principles.

C. *ANALYSIS OF REFERENCE CODE IMPLEMENTATION*

Third phase deals with creation and analysis of reference code implementation. Reference code implementation is done manually and serves as reference to what is the expected output from the generator. Reference software system should completely implement architecture designed in previous phase and should be extensive enough to cover all different use cases of generator and their implementations. After creation of reference software system, analysis is carried out during which specific and generic parts of system are diagnosed. Generic parts are ones that stay unchanged for different domains while specific parts are ones that change for particular domains.

D. *DEFINING CODE GENERATION*

Creation of code generator covers defining a transformation of concepts of DLS's meta – model into appropriate concepts of target platform – Java. Defining a

transformation means defining a set of rules of translation of inputs into outputs. For example we have to define translation between `NumberPicker` concept of UIL DSL and `JSpinner` component in Java. Given transformations, i.e. code generation is considered as translation semantics of programming language. Actual input in code generator is semantic model, created and populated with data (description of model which user defined) by XText after the parsing process is done. In other words, code generator is tightly coupled to semantic model since it reads all input data from it

For generic and specific parts diagnosed in the third phase, we look up for pattern they represent not related to any particular domain and upon that pattern we write transformations which will produce desired output. Code generation can be done in two ways, via transformer generation and template generation. In transformer generation we write programming code – transformation which represents logic of transforming input data into instructions of desired language. In template transformation we write templates which consist of static, generic parts that are same for all inputs and dynamic, specific parts that vary based on input. Specific parts, often called markers, retrieve data from semantic model and during compilation are replaced with real data. In our generator we use both approaches since they both have their advantages. Below is given a part of one template which generates table model for table component of UIL DSL. Table models consist of parts that vary based on different domain object, but there are still parts that are same for all, together those parts make a pattern – template upon which all table models will be generated.

```
public class
TableModel«UtilMethods::toFirstUpperCase(obj.name)» extends
TableModel { String columnNameFK[] = { "yes/no",
«UtilDomainAttribute::getAttributesName(obj)» };
public
TableModel«UtilMethods::toFirstUpperCase(obj.name)»(List<Wra
pperObject> lista) {
                this.columnName = columnNameFK;
                this.lista = lista;}
@Override
public Object getValueAt(int rowIndex, int columnIndex) {
        WrapperObject object = lista.get(rowIndex);
                switch (columnIndex) {
        «UtilDomainAttribute::getCasesForGetValueAt(obj)»
        }
        return "error";
}}
```

Based on output type of generator, transformation belongs to group M2C (model - to - code) since we transform input semantic model into Java source code. Generated source code is ready for execution without any modification, although user needs to populate a list of domain objects which will be used for selection if object has one – to – much relationship and wants to show that relation in GUI. Mixing generated and hand written code is not a good practice simply because all hand written changes will be overwritten next time generator runs. Given problem is solved by implementing Generation Gap pattern which with concept of inheritance decouples generated from hand written code. Fig.2 shows the results of code generation that are two forms, main form for inserting data about person and selection form for choosing objects of pets. User needs to populate decoupled list of pets with desired objects.

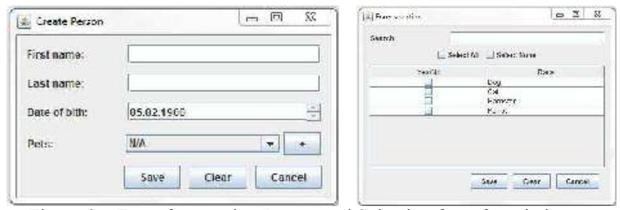


Figure 2 – Form for creating Person and Selection form for relations

IV. CONCLUSION

Fully developed generator eases the process of creating UI by freeing the user from direct implementation of programming code. Authors have acknowledged strengths of domain-specific languages, huge advantages of developing a code generator, in this case, GUI generator and great potential of MDD.

For creating a case study found in this paper, we have spent far less amount of time than for the development of similar projects using a traditional approach (general-purpose language). It can be concluded that the development of domain-specific language and code generator is definitely one of the ways to solve the problem of total amount of time necessary for full system software development and significantly reduce number of problems related to collecting of user’s requests.

REFERENCES

[1] Kennard, R., Leaney, J.: ‘Towards a general purpose architecture for UI generation’, J. Syst. Softw., 2010, 83, (10), pp. 1896–1905

[2] Myers, B.; Rosson, M. B.: Survey on User Interface Programming. Proc. of the 10th Annual CHI Conference on Human Factors in Computing Systems, pp. 195-202, 2000.

[3] Alberto Rodrigues da Silva, Model-Driven Engineering: A Survey Supported by a Unified Conceptual Model, in Computer Languages, Systems and Structures, Elsevier, 43, 2015

[4] Martin Fowler, Domain-Specific Languages. The Addison-Wesley signature series, Addison-Wesley 2011, ISBN 978-0-321-71294-3, pp. I-XXVIII, 1-597

[5] Mernik, M., Heering, J., and Sloane, A. M. (2005). When and how to develop domain-specific languages. ACM Computing Surveys, 37:316–344.

[6] Luoma, J., Kelly, S., and Tolvanen, J.-P. (2004). Defining domain-specific modeling languages: Collected experiences. In OOPSLA 4th Workshop on Domain-Specific Modeling.ACM.

[7] Francisco Morais and Alberto Rodrigues da Silva, Assessing the Quality of User-Interface Modeling Languages

[8] André Ribeiro, Alberto Rodrigues da Silva, Evaluation of XIS-Mobile, a Domain Specific Language for Mobile Application Development

[9] André Ribeiro & Alberto Silva, XIS-Mobile A DSL for Mobile Applications

[10] Silva, A.R.: The XIS Approach and Principles, Proc. of the 29th Conf. on EUROMICRO (EUROMICRO '03), IEEEComputer Society (2003)

[11] Silva, A.R. et al.: XIS-UML Profile for eXtreme Modeling Interactive Systems, Proc. of the 4th International Workshop on Model-Based Methodologies for Pervasive and Embedded Software (MOMPES'07), IEEE Computer Society (2007)

[12] A Domain-Specific Language for Defining Static Structure of Database Applications

[13] Milosavljević, G., Perišić, B.: A method and a tool for rapid prototyping of large-scale business information systems, Computer Science and Information Systems ,vol. 1, pp. 57-82 (2004)

[14] A DSL for modeling application-specific functionalities of business applications

[15] Generating User Interface from Task, User and Domain Models

[16] Generating User Interface for Information Applications from Task, User and Domain Models with DB-USE

[17] Kennard, R., Leaney, J. Is there convergence in the field of UI generation, Journal of Systems and Software, 2011

[18] Cruz, A.M.R., Faria, J.P..Automatic generation of user interface models and prototypes from domain and use case models. In Proceedings of the 4th ICSOFT (ICSoft 2009) , vol. 1, pp 169-176, Sofia, Bulgaria, INSTICC Press, July 2009.]

[19] Cruz, A.M.R. Automatic generation of user interfaces from rigorous domain and use case models. PhD thesis F.E.U.P., University of Porto, Portugal, 2010.

[20] Óscar Pastor, Marcela Ruiz, Sergio España (2011) From requirements to code: a full model-driven development perspective; Keynote talk at ICSOFT 2011.

[21] Michał Smiałek, Tomasz Straszak, Facilitating transition from requirements to code with the ReDSeeDS tool

[22] M. Smiałek, J.Bojarski, W.Nowakowski and T. Straszak, "Scenario construction tool based on extended UML metamodel". Lecture Notes in Computer Science, 3713:414–429, 2005.

[23] M.Smialek and T.Straszak, "Facilitating transition from requirements to code with the ReDSeeDS tool". RE 2012: 321-322

[24] M.Smialek, W.Nowakowski, N. Jarzebowski, A Ambroziewicz," From use cases and their relationships to code" MoDRE 2012: 9-18

[25] Dušan Savić, Siniša Vlajić, Saša Lazarević, Vojislav Stanojević, Ilija Antović, Miloš Milić, Alberto Rodrigues da Silva, SilabMDD - A Use Case Model Driven Approach

[26] D.Savic, I. Antovic, S. Vlajic, V. Stanojevic and M. Milic, "Language for Use Case Specification ", Conference Publications of 34th Annual IEEE Software Engineering Workshop, IEEE Computer Society, 2011, Pages: 19-26, ISSN : 1550-6215, ISBN: 978-1-4673-0245-6, Limerick, Ireland, 20-21 June 2011, DOI: 10.1109/SEW.2011.9

[27] D.Savić, A.Rodrigues da Silva, S.Vlajić, S.Lazarević, I.Antović, V. Stanojević, M.Milić, Preliminary experience using JetBrains MPS to implement a requirements specification language, in Proceedings of QUATIC'2014 Conference, 2014 , IEEE Computer Society

[28] Savić, A.R. Silva, S.Vlajić, S.D.Lazarević, V.Stanojevic, M.Milić, M.Milic, "Use Case Specification at Different Abstraction Level", 8th International Conference on the Quality of Information and Communications Technology, Lisbon, Portugal, 03-06.09. 2012

[29] I.Antović, S.Vlajić, M.Milić,D.Savić and V.Stanojević, "Model and software tool for automatic generation of user interface based on use case and data model," Software, IET , vol.6, no.6, pp.559-573, Dec. 2012 doi: 10.1049/iet-sen.2011.0060D

[30] A Taxonomy of Model Transformation. Mens, Tom / Gorp, Pieter Van. Amsterdam: Elsevier Science Publishers , 2006, Electronic Notes in Theoretical Computer Science, T. 152. 1571-0661

An approach to the semantic representation of the planning strategies

Sasa Arsovski[*], Zora Konjović[**],Branko Markovski[***], Miodrag Ivković[****], Dejan Madić[*****]
[*] Guarantee fund of APV, Novi Sad, Serbia
[**] Singidunum University, Belgrade, Serbia
[***] University of Novi Sad/Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
[****] University of Novi Sad/Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
[*****] University of Novi Sad/Faculty of Sport and Physical Education, Novi Sad, Serbia
sasa.arsovski@gmail.com

Abstract - This paper aims to describe a methodological approach to creating ontology that should assist in creation of state development strategies and corresponding action plans (ORS). The ultimate goal of the ontology is to provide semantic description of development strategies, as well as action plans for provincial administrative bodies, i.e., basis for formal, machine-readable representation of development priorities, specific goals as well as answers to questions such as: what development priorities are, how they will be implemented, who is responsible for their implementation, and why they are being implemented in the first place. The proposed approach enables a common dictionary of terms describing the goals and priorities in the process of strategic planning to be formed, which could in turn be used by all participants involved in strategic planning.

I. INTRODUCTION

The research described in this paper encompasses the development of semantic descriptions of state development strategies as a part of strategic planning processes of local and provincial state bodies.

A strategic plan is a document which contains a certain number of strategic goals. It gives guidelines for domain development activities in a given time period (from 3 to 5 years) and determines the direction, priorities, actions and responsibilities of the implementation.

Currently, common development goals and priorities of state and local administrative bodies are [14]: *smart growth:* the development of an economy based on knowledge and innovation. *Sustainable growth:* promotion of an economy which is more efficient in using its resources. *Inclusive growth:* promoting the growth of an economy with a high employment rate and which accomplishes social and territorial cohesion.

The strategic planning process (SP) is defined as the process by which managers analyze internal and external environment in order to formulate the strategy and the allocation of resources to develop a competitive advantage in the industry that enables the successful achievement of the objectives of the organization [11].

From the above definitions and the new needs of the organization, two key issues of formalization should be addressed and resolved in the context of the strategic planning [10]:

1) Formally define conceptual framework that is used to represent information / knowledge extracted from internal and external environment of the organization and each participant in the process as to create a common vocabulary, which will enable all managers to express/share a unique vision and use this vocabulary to communicate and collaborate.
2) To formalize the strategic planning process itself to determine the steps that comprise this process, type of information / knowledge to be used and who is involved in each case. A schematic description of the course of the strategic planning process is given by Figure 1 [15].

Figure 1. Schematic description of strategic planning workflow [15]

In order to allow for a certain degree of automation and formalization of the strategic planning process, the different tools and software have been utilized like Competitive Intelligence tools (CI) [5] and Business Intelligence (BI) [6]. These tools have not been, however, integrated into all phases involved in strategic planning.

The author of [7] classified ontologies by their level of generalization and has proposed four different types: Upper, domain, task and application ontologies. The development of ontologies in the field of planning is still in an early phase.

The majority of created ontologies is designed for specific purposes: budget planning [1], public health care projects [2], public transport organization, action planning [3]. Methodologies for the development of ontologies used specifically for planning have not been sufficiently researched yet.

The development and transformation of knowledge are the main reason for the relatively small number of different methodologies for creating planning-oriented ontologies. The main advantage of using ontologies to describe knowledge of a continuously changing domain is their ability to develop and transform knowledge [4].

This paper will describe the application ontology and develop a terminology for describing development strategies as a part of the SP. The use of the described ontology in an ontology driven informational system creates possibilities for automation and unification of creating development strategies and action plans in the process of strategic planning in the development area of the analyzed region.

II. Related work

Authors of [8] define the conditions which need to be satisfied by the developed strategy:

1) It needs to be founded on a clear understanding of the current situation and need for development;
2) It needs to be relevant and clearly defined;
3) Involves both people and institutions;
4) It is logically consistent;
5) It is realistic and applicable;
6) It needs to have a real influence on development processes;
7) It enables monitoring and creates the foundations of responsibility and transparency whilst being susceptible to change;
8) It is broad enough to encompass all main development problems;
9) It is authentic, i.e., reflects the state and development ideas of the specific system that is a planning subject;
10) It is understandable for the wider public.

According to [8], the process of strategic planning has the following phases of development:

Preparation, which encompasses forming of the management structure that will direct the process. Concerning the approach proposed in this paper, this should be done in a way to ensure both public and political support, but also to ensure adequate executive infrastructure that will support planning process.

Collection and analysis of information. Information should be collected and analyzed in a way to provide for understanding of the current state and recognition of risks/critical success factors. The approach proposed in this paper suggests SWOT analysis (Strengths, Weaknesses, Opportunities, Threats) as means to ensure that the analysis encompasses all questions emerging from data.

Development of a strategy (visions and goals of the action plan). Strategies need to represent a sequence of goals which grow progressively more detailed, and finally develop into a detailed action plan. There are three levels of goals: strategic goals, priorities and specific goals. The model of strategic goals, priorities and specific goals and, finally, the action plan needs to be logical and hierarchical. Higher goals need to logically lead to lower goals and lower goals need to contribute to the accomplishment of higher goals. Priorities and specific goals in the framework of every strategic goal need to support and supplement priorities within these strategic goals.

Finalization (conclusion and implementation of the strategy). Finalization of the Strategy document and its adoption by the competent bodies/individuals.

Again [8] defines the requirements every development strategy needs to fulfill:

1) *Consistency* - All strategic goals, priorities and specific goals must be in mutual accordance.
2) *Relevance* - Goals must be relevant to the given situation. *Specifiability* - Priorities and goals need to be specific enough in order to ascertain whether or not they have been accomplished.

Authors in [9] define the principles of designing planning-oriented ontologies:

1) Ontologies need to have a well-defined goal and support a defined group of usage cases;
2) Ontologies need to have a minimal amount of different concepts and traits;
3) Ontologies need to be higher than simple domain concept taxonomies;
4) They need to enable import/export of concepts and traits from other ontologies.
5) The goal of COA ontology described in [9] is to describe plans which could be used by different tools and applications.

Authors in [10] define the strategic planning (SP) process as one which analyses internal and external factors with the goal of forming strategies and allocating resources. Figure 2 shows a graphical representation of the model of SP as proposed in [11].

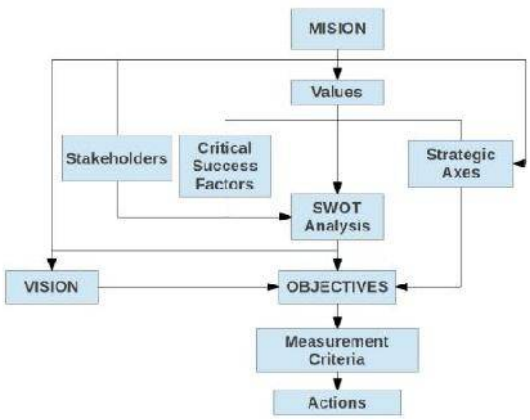


Figure 2. SP model proposed in [11]

Same source presents the ontology formalized by OWL which meets the standards approved by the World Wide Web Consortium (W3C) and is used for the formalization of the SP process, the knowledge that is created, and flows between the participants in the process.

Figure 3 shows the concepts and relationships contained by ontology structure, as well as entries of lexicon, which constitute the common vocabulary with which to refer to them.

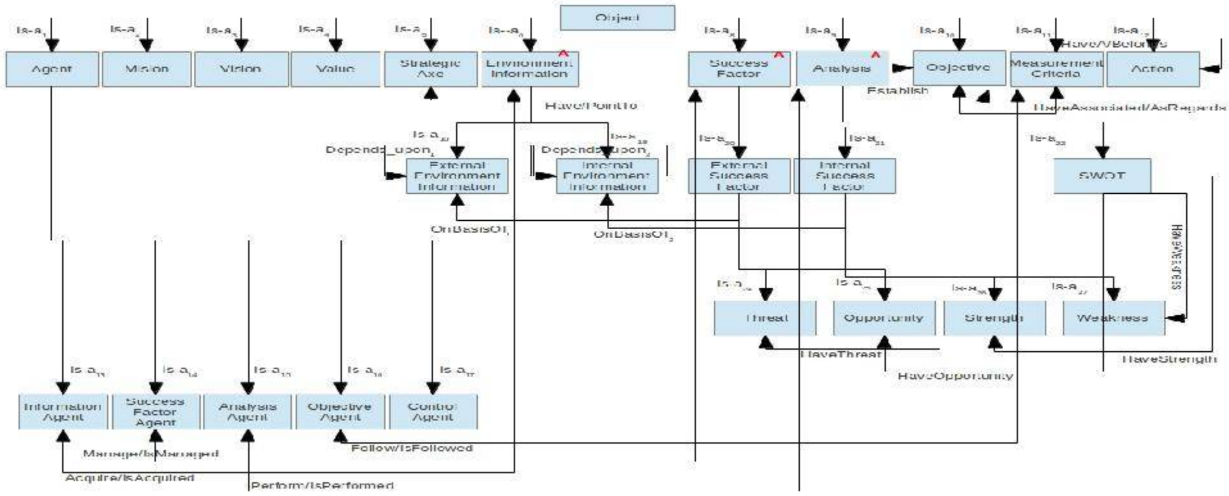


Figure 3. Ontology for SP process from [10]

III. THE ORS ONTOLOGY

The goal of ORS ontology is to ensure a formal, computer-readable representation of development priorities, specific goals as well as give answers to questions such as: *what* development priorities are (strategic elements), *how* they will be implemented (action plan elements), *who* is responsible for their implementation (action plan elements) as well as *where* and *why* they are being implemented in the first place (strategic elements).

As the basis for the ORS construction we used ontology structure shown on Figure 2 [10]. The semantic representation of the SWOT analysis from [10] as a part of the strategic planning process is extended with ontology concepts of the regional administrative bodies and development funds given in [12] as well as the SCORE ontology given in [13], which describes the development goals and socio-economic parameters of the region. The main concepts of the ORS ontology which semantically represent a SP process and planning strategy are shown on Figure 4.

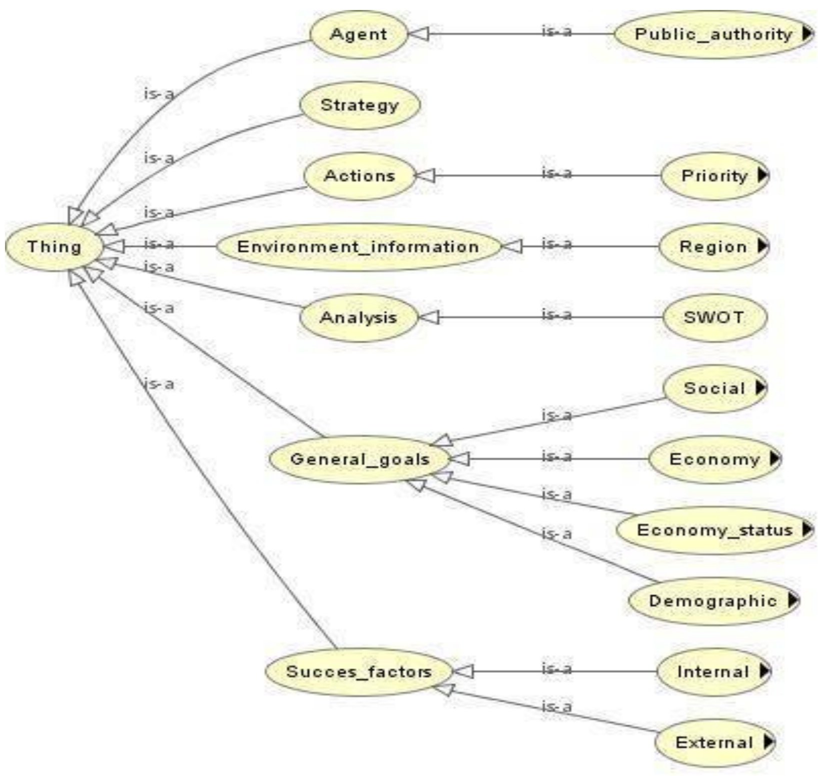


Figure 4. ORS ontology

As shown by Figure 4 the ontology proposed in [10] is extended in the following manner.

In the first step we create ontological concept that describes spatial aspect of the analyzed domain that includes populated areas and their administrative organization (county, municipality) and extend *Environment Information* concept from [10] with indentified instances as shown on the Figure 5.

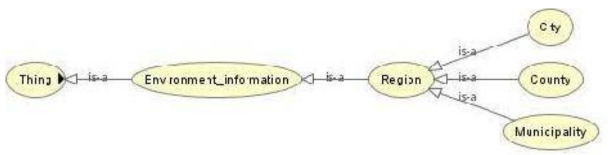


Figure 5. Region concept from ORS

In the second step we create *General Goals* concept. This concept describes the socio-economic parameters of the relevant region. It gives an insight into the current state and enables problems and limits of development to be identified. This concept is associated by relations *isThreat*, *IsOpportunity*, *IsStrenght* and *IsWeakness* with the concept *Success Factors* from [10].

In the third step we create *Public Authority* concept. This concept describes hierarchical links of regional administration, funds and other organizational bodies of state administration which have the task of implementing strategic plans. The *Agents* concept from [10] is extended with created *Public Authority* concept.

Extension of the concept *Action* from [10], with *Priority* concept described in [14] is a *fourth step*. *Priority* concept describes regional specifics, i.e. development priorities.

Once the ontology is formally defined and exactly specified (using an instances), and all information of the analyzed region are stored in ontology, ORS ontology can be reused for creation of a subsequent strategies.

IV. CONCLUSIONS

The proposed approach in creating ORS ontology ensures a formal, computer-readable description of elements of the process of strategic planning. The proposed methodological approach creates a common dictionary of terms used by all participants in the process of strategy-development.

The ontological representation of development strategies (the process of creating development strategies is specified by a group of concepts and concept instances) creates conditions for re-use of semantically described strategies in the process of creating new strategies for the relevant domain. Instances which populate the ontology in the strategy development process are machine-readable and may be used in the creation of a new strategy or redesign of a current strategy.

Creation of the described ontology is the first step towards a model of the software tool aimed to facilitate creation of development strategy. The semantic representation of SWOT analysis enables the creation of graphic tools that could simplify the process of strategy and action plan generation. Additionally, use of the ORS ontology in an ontology-driven information system could enable automation and unification of development strategies and action plans creation within the framework of designing strategic plans for the development of the analyzed region.

REFERENCES

[1] Brusa, G., Caliusco, M. L. and Chiotti, O., 2006. A Process for Building a Domain Ontology: an Experience in Developing a Government Budgetary Ontology. AOW ’06 Proceedings of the second Australasian workshop on advances in ontologies. 72. pp.7-15.

[2] Ali, A.,2010. Fareedi Ontology-based Model For The “Ward-round” Process in Health Care (OMWRP). LAP LAMBERT Academic Publishing.

[3] Darr, T. P., Benjamin, P., and Mayer, R., 2009.Course of Action Planning Ontology. Ontology for the Intelligence Community.

[4] Houda, M., Khemaja, M., Oliveira, K., and Abed, M., 2010. A public transportation ontology to support user travel planning. In: International Conference on Research Challanges in Information Science (RCIS), pp. 127.

[5] J. F. Prescott and S. H. Miller. Proven Strategies in Competitive Intelligence: Lessons from the Trenches, Ed. Wiley, 2004

[6] C. Howson. Successful Business Intelligence: Unlock the Value of BI & Big Data, Ed. McGraw Hill Education, ISBN: 978-0071809184, 2013.

[7] N. Guarino. “Formal Ontology in Information Systems”. In Proceedings of FOIS'98, Trento, Italy, IOS Press, Amsterdam, 1998.

[8] P.Bolton. Vodič za Strateško planiranje za gradove i opštine, MIR2 2007.

[9] Timothy P. Darr, Perakath Benjamin, Richard Mayer. Course of Action Planning Ontology, Ontology for the Intelligence Community 2009 (OIC 2009).

[10] Juan Luis Dalmau-Espert, Faraón Llorens-Largo and Rafael Molina-Carmona. An Ontology for Formalizing and Automating the Strategic Planning Process. eKNOW 2015 : The Seventh International Conference on Information, Process, and Knowledge Management

[11] F. Lorens. Strategic Plan of the University of Alicante (Horizonte 2012) http://web.ua.es/es/peua/horizonte-2012.html, 2007. N. Pahl and A. Richter. Swot Analysis - Idea, Methodology and a Practical Approach, Ed. Books on Demand, 2009.

[12] S. Arsovski, B. Markoski, P. Pecev, N. Petrovački, D. Lacmanović (2014). Advantages of Using an Ontological Model of the State Development Funds, INT J COMPUT COMMUN, ISSN 1841-9836, 9(3):261-275.

[13] Laszlo Ratgeber, Saša Arsovski, Petar Čisar, Zdravko Ivanković, Predrag Pecev, Ontology driven decision support system for scoring clients in government credit funds, INTERNATIONAL Conference on Applied Internet and Information Technologies (2 ; 2013 ; Zrenjanin), pages (369-373), ISBN 978-86-7672-211-2

[14] CODEX , Coordinated Development and Knowledge Exchange,on Spatial Planning Methodology, Broj projekta: HUSRB/1203/213/151, 2013.

[15] Bryson, John M. and Farnum K. Alston, Creating and Implementing Your Strategic Plan. Jossey-Bass Publishers, San Francisco,1996.

Software System for Optical Recognition and Symbolic Differentiation of Mathematical Expressions

Ivan Perić, Marko Jocić, Đorđe Obradović
* University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
{ivanperic, m.jocic, obrad}@uns.ac.rs

Abstract – In this paper we propose a method for optical recognition and symbolic differentiation of printed mathematical expressions. Many areas of science use differential calculus as a mechanism to model and formally represent the dynamics of many processes. Function derivation process can be pretty exhausting, especially if we need to use it too often. Basically, it's based on the set of derivation rules which makes it suitable for automation and we will use that fact. Furthermore, to make the derivation process even faster, we will try to recognize the mathematical expression by taking a photograph of it, so the user will not have to enter complicated mathematical expressions. Maximum potential of this kind of system can be achieved by using it on a mobile platform where the user already has a mechanism to take photographs of mathematical expressions and differentiate them wherever he is. It could be used in scientific purposes to speed up the calculation process. On the other hand, it could be very helpful to students, who could check if their calculations are valid.

I. INTRODUCTION

Computer Vision is a field of artificial intelligence, widely used in robotics, applied medical sciences, physics, industrial processes and many more. It includes methods for acquiring, processing, analyzing, and understanding images in order to produce numerical or symbolical data about a real world. This way, machines are allowed to perceive and "understand" a world around them. Continuous enhancement in a field of computer hardware gives a boost to other fields of computer science, especially the ones that are computationally demanding and at the same time need to be executed in a real-time. That's the reason why computer vision has become very active field of research in previous years. One of the computer vision fields is the optical character recognition (OCR). OCR can be applied in many situations. Optical recognition of mathematical constructions is a special class of OCR problems.

The system proposed in this paper is divided into four subsystems, where each of them encapsulates logically similar operations.

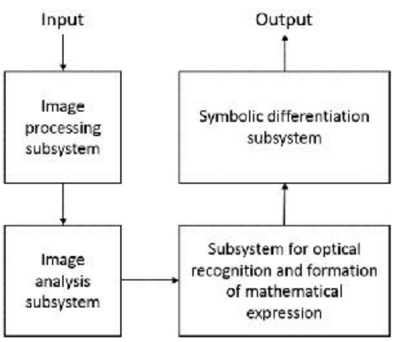


Figure 1 - System architecture

This paper contains formal specification and implementation steps of the proposed system. Computer vision techniques, machine learning, mathematical logic and numerical algorithms are used in the implementation.

II. RELATED WORK

There are many software systems for the optical character recognition. Mathematical expressions represent one problem type in this field of research. The first challenge in optical recognition and symbolic differentiation of mathematical expressions is the image segmentation process. Many different techniques can be used, but if the accuracy is important, those techniques tend to be more complicated and the speed of the system is be reduced. Researchers from Assam University and Manipur University in India proposed an approximation method for adaptive thresholding technique. [1] They managed to keep binarization process times around 200ms, while regular adaptive thresholding techniques went up to 13s for large window sizes. This means that their approach is more than 50 times faster. This is very important for systems that are executed in real-time environments or on mobile platforms, due to limited hardware resources. The second problem is the fact that mathematical expressions are not regular text. It's not easy to process spatial relationships between parts of the mathematical expression. Their complexity can be very high. Alan Sexton from the University of Birmingham in UK proposed one method for spatial relationships processing. [2] He did horizontal and vertical projections on the mathematical expression and tried to divide it. The problem he didn't solve was the fact that mathematical expressions can be rotated. In that case projections won't work. In the example of photographed mathematical expression, the probability of expression being rotated is very high. One possible solution to this problem will be presented in this paper. Commercial products for mathematical problem solving usually just solve mathematical equations, but don't incorporate OCR techniques. For example, there is Wolfram Alpha [3], which can solve equations and parse hand written symbols. This type of problem solving is very helpful to many people and that's why these commercial systems are constantly improving.

III. PRELIMINARIES

Approach proposed in this paper uses adaptive thresholding image segmentation techniques, morphological operators, least-squares approximation algorithm to make linear regression, artificial neural networks with back-propagation training algorithm and

symbolic logic incorporated into Prolog programming language.

A. *Image segmentation using adaptive threshold*

Threshold is the image segmentation method in which the specific *threshold* is calculated, and then every pixel value is compared to it. Usually there is one threshold value and all pixels are divided into two groups – below threshold value and above threshold value. Then the one group is classified as background, and the other one as a content of interest. This process is called **binarization**. Thresholding methods are usually applied to grayscale images, which is also the case in this paper. Threshold is a value that must be set or calculated. Most of the methods used to calculate threshold use **histogram** as a data source. Histogram of the grayscale image represents the number of pixels of the specific value on the image. Pixel values of the grayscale image are between [0,255], where 0 is black – zero light, 255 is white – maximum light. As it is shown in the Figure 2, threshold is calculated somewhere around 75. It can be observed that there is a larger number of pixels with low lightning, which means that the whole picture was in darker tone.

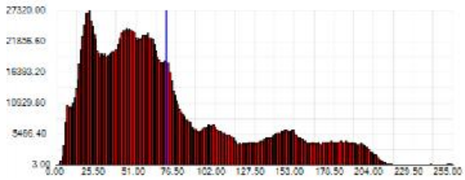


Figure 2 - Histogram for threshold calculation

The main problem with the thresholding methods can be seen if the picture is not uniformly illuminated. When you make a photograph with a flash turned on, there is a possibility that this will happen. One part of the image will be very bright, and the other one will be very dark (Figure 3).

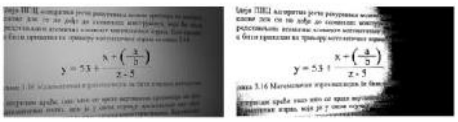


Figure 3 - Global threshold problem

This problem can be fixed by using **adaptive threshold** techniques. Adaptive threshold uses a locality principle, which means that the threshold is calculated for many image segments independently. By this approach results can be improved, but the binarization process will be more complicated and more time-consumable. Threshold calculation needs many statistical characteristics of the input image, which are time-consumable by themselves, and they have to be calculated many times – for each local threshold. An approximation will be used in this paper, and the performance problem will be minimized.

B. *Morphological operators – dilation and erosion*

Definition. Dilation of the set ***X*** by structure element ***S*** is defined as a binary image which represents the set of points [m,n] such that after translation of the structured element to point [m,n] intersection of the sets $\boldsymbol{S_{mn}}$ and ***X*** is not an empty set:

$$S \oplus X = \{m, n: S_{mn} \cap X \neq \emptyset\}. \ [4]$$

Dilation is a morphological operation based on a logical operator OR between the neighboring pixels. If at least one point the neighborhood is black, the current point will be black. [5]


Figure 4 - Dilation operator

Definition. Erosion of the set ***X*** by structure element ***S*** is defined as a binary image which represents the set of points [m,n] such that after translation of the structured element to point [m,n] the whole structured element $\boldsymbol{S_{mn}}$ is included in the set ***X***:

$$X - S = \{m, n: S_{mn} \subseteq X\}. \ [4]$$

Erosion is a morphological operation based on a logical operator AND between neighboring pixels. If all the points in the neighborhood are black, the current point will be black. [5]


Figure 5 - Erosion operator

C. *Artificial neural networks and Back-Propagation algorithm*

Definition. Artificial neural network can be observed as an artificial replica of the human brain and tends to simulate the learning process. However, they are just a rough approximation of biological neural networks. Artificial neural networks are designed to do a nonlinear mapping of input data to some output data. Formally, artificial neural network can be represented as:

$$f_{ANN}: R^l \rightarrow R^k.$$

Artificial neural network consists of a set of simple process elements, called neurons. Those neurons represent a model of biological neurons. Neural network stores the knowledge in connections between neurons. This knowledge is calculated in a neural network learning process (training).

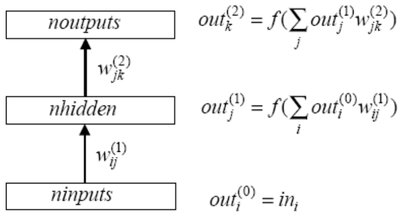


Figure 6 - Multilayer perceptron

Neural network proposed in this paper is trained by a back-propagation algorithm. This algorithm propagates an error value backwards in a neural network topology. Neural network used here is a feed forward neural network and consists of 5 layers. Input layer has 64 neurons, three hidden layers consists of 100 neurons each, and the output layer has 49 neurons. The parameter μ, which represents the speed of learning, is set to 0.05, and the learning process is completed after 10.000 iterations or if the error value drops below 10^{-4}.

The main problem in multilayer perceptron training process in the fact that the target value is known only in the output layer. Algorithm must be able to correct all weights by the same minimization criteria.

Multilayer perceptron training tends to minimize error function

$$E\left(w_{ij}^{(n)}\right) = \frac{1}{2}\sum_{p}\sum_{j}(targ_j^p - out_j^{(N)}(in_i^p))^2$$

and corrections given by the gradient formula are used:

$$\Delta w_{kl}^{(m)} = -\eta \frac{\partial E(w_{ij}^{(n)})}{\partial w_{kl}^{(m)}}$$

In the first step, training set is formed and the neural network architecture is defined. After neural network formation, initial values of weights between neurons must be set. Values are usually defined in a random manner. Also, error function $E(w_{ij})$ must be chosen, and also a training speed η. In the second step, the input data vector is sent to the neurons in neural network's input layer. In the third step the neural network output is calculated. The whole weight correction process is represented as an optimization process of the error function.

Output value formula for the neural network with multiple hidden layers looks like this:

$$delta_k^{(N)} = \left(targ_k - out_k^{(N)}\right) \cdot f'\left(\sum_j out_j^{(1)} w_{jk}^{(N)}\right) = (targ_k - out_k^{(N)}) \cdot out_k^{(N)} \cdot (1 - out_k^{(N)})$$

formula for the hidden layers:

$$delta_k^{(n)} = \left(\sum_k delta_k^{(n+1)} w_{lk}^{(n+1)}\right) \cdot f'\left(\sum_j out_j^{(n-1)} w_{jk}^{(n)}\right) = \left(\sum_k delta_k^{(n+1)} w_{lk}^{(n+1)}\right) \cdot out_k^{(n)} \cdot \left(1 - out_k^{(n)}\right)$$

and a formula for weights correction:

$$\Delta w_{hl}^{(n)} = \eta \sum_p delta_l^{(n)} out_h^{(n-1)}$$

Algorithm continues by correcting all weights, for every input pattern from the training set. This step is repeated until the error is not low enough. When error drops below certain value, training process is finished. [6] [5]

D. Prolog

Many problems in the real world are calculated much faster by machines than by the human brain, but some of them are not. Sometimes it's easier to represent a problem by a set of rules and facts, which is more similar to a human perception. Prolog is a declarative programming language that has a knowledge base, which consists of a set of **facts** and a set of **rules**. Fundamentals of Prolog are based on a predictive logic. Prolog is very suitable for problem solving where problem consists of objects and relations between them.

In the first step, the knowledge base is formed. When a query is passed to Prolog interpreter, decision tree is made by using facts and rules from knowledge base. Then the result is calculated by following the tree branches.

IV. PROPOSED ALGORITHM

Software system proposed in this paper is divided into four subsystems (image processing subsystem, image analysis subsystem, subsystem for optical recognition and formation of the mathematical expression and a subsystem for symbolic differentiation). Every subsystem does a special set of tasks. Input to the system is a digital image of the mathematical expression, and an output is a derivative of that mathematical expression in a form of text.

1. Image processing subsystem

This subsystem is the first in the series of subsystems, and it's input is also an input of the whole system.

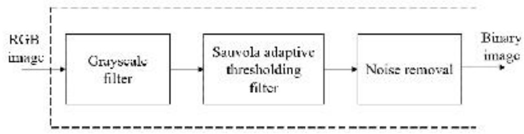


Figure 7 - Image processing subsystem block diagram

Color image, represented in a RGB color system, first needs to be converted into grayscale color image. Mathematical expressions are usually printed on a very bright surfaces (white paper), and the text is usually black, so it's enough to look every pixel's luminosity level. Grayscale color space shows the luminosity level of every pixel and that's why this approximation is applied. In the next step, grayscale image is binarized and every pixel is classified as a background or a content of interest. This is done by adaptive thresholding technique. The main problem of adaptive threshold techniques is a performance time. Adaptive threshold approximation will be proposed in this paper, by using **integral images**.

Definition. An integral sum image **g** of an input image *I* is defined as an image in which the intensity at a pixel position is equal to the sum of the intensities of all the pixels above and to the left of that position in the original image. [1]

Integral image intensity at *(x,y)* can be calculated as:

$$g(x,y) = \sum_{i=1}^{x}\sum_{j=1}^{y} I(i,j)$$

The integral sum image of any grayscale image can be efficiently computed in a single pass as:

$$g(1,y) = I(1,y) + g(1,y-1), \quad y = 2..n$$

$$g(x,1) = I(x,1) + g(x-1,1), \quad x = 2..m$$

$$g(x,y) = I(x,y) + g(x,y-1) + g(x-1,y) - g(x-1,y-1), \quad x = 2..m, \ y = 2..n$$

Definition. In Sauvola's binarization method [1], the threshold ***T(x,y)*** is calculated using the mean ***m(x,y)*** and standard deviation ***δ(x,y)*** of the pixels within a window of size ***ω×ω*** as:

$$T(x,y) = m(x,y)\left[1 + k\left(\frac{\delta(x,y)}{R} - 1\right)\right]$$

where R is the maximum value of the standard deviation (R = 128 for a grayscale document), and k is a bias, which takes positive values in the range [0.2, 0.5]. [1]

However, standard deviation is calculated for every local thresholding window and that consumes a lot of time. That's why mean standard deviation is used, and the process is much faster. After this approximation, threshold formula looks like this:

$$T(x,y) = m(x,y)\left[1 + k\left(\frac{\partial(x,y)}{1-\partial(x,y)} - 1\right)\right]$$

where $\boldsymbol{\partial(x,y) = I(x,y) - m(x,y)}$ is the mean deviation, and **k** is a bias which can control the level of adaptation varying threshold value. Also $k \in [0,1]$. [1]

Calculation of the *m(x,y)* requires that the local sum ***s(x,y)*** is calculated before.

Definition. The local sum ***s(x,y)***, which is the center of the local window of size *ω×ω* of an image *I* is the sum of all the pixel intensities within the local window. It can be calculated in two passes as:

$$s(x,y) = \sum_{i=x-c}^{x+c} \sum_{j=y-c}^{y+c} I(i,j)$$

where $c = \frac{\omega-1}{2}$, since *ω* is an odd number. [1]

Once we have the integral sum image **g**, the local sum ***s(x,y)*** of any window size can be computed simply by using two addition and one subtraction operations without depending on window size as:

$$s(x,y) = [g(x+d-1, y+d-1) + g(x-d, y-d)] - [g(x-d, y+d-1) + g(x+d-1, y-d)]$$

where $d = round\left(\frac{\omega}{2}\right)$.

Definition. The local arithmetic mean ***m(x,y)*** at *(x,y)* is the average of the pixels within the window of *ω×ω* of the image I. It can be calculated as:

$$m(x,y) = \frac{s(x,y)}{\omega^2}$$

In this way the local mean can be calculated efficiently in a single pass without depending on local window size using integral sum image. [1]

Adaptive thresholding gives very good results without a lot of noise after binarization process. However, there is a noise removal filter to handle those situations. In this case it was enough to use just morphological operators – dilation and erosion.

2. *Image analysis subsystem*

When binary image is filtered, regions of interest can be found. Every region of interest is a list of black pixels.

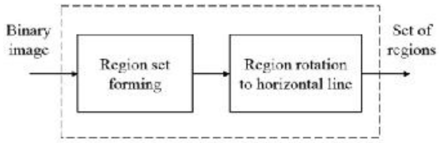


Figure 8 - Image analysis subsystem block diagram

Region formation starts with the one pixel that belongs to it, and then a ***Depth first search – DFS*** is performed to collect all the other points of that region. The process continues until all pixels are classified to their corresponding regions.

If the image is rotated, regions will be rotated too. This is going to be a problem in the next subsystem, and that's why a rotation angle of the whole expression needs to be calculated. Then, every region will be rotated and placed on a horizontal line.

First we need to find a rotation angle **φ**. Region centers will be taken, and then linear regression is used to find a line that best fits an angle of the expression.

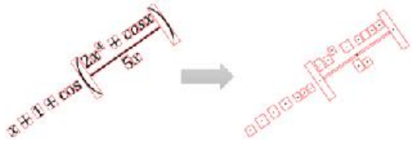


Figure 9 - Preparation for the approximation

When regression line is found, it's easy to find a rotation angle φ.

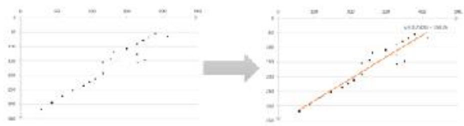


Figure 10 - Angle approximation process

$$\varphi = \tan^{-1}(k).$$

Now it's possible to rotate the expression. Point of the rotation is going to be an intersection of a y-axis and a

Figure 11 - Rotation of the expression

regression line. If a negative value is calculated, all the regions are translated to the positive part of y-axis so the digital image coordination system is preserved. [7]

Every point of every region is rotated around rotation point. Original point with coordinates ***(x,y)*** is transformed into point with coordinates (r_x, r_y), where

$$r_x = \cos(\varphi) * (x - m_x) - \sin(\varphi) * (y - m_y) + m_x$$

$$r_y = \sin(\varphi) * (x - m_x) + \cos(\varphi) * (y - m_y) + m_y$$

3. *Subsystem for optical recognition and formation of mathematical expression*

This subsystem is very important, because it will do the optical recognition of every symbol in the mathematical expression, and it will do the processing of spatial relationships between regions. These two functionalities will be placed into two different modules.

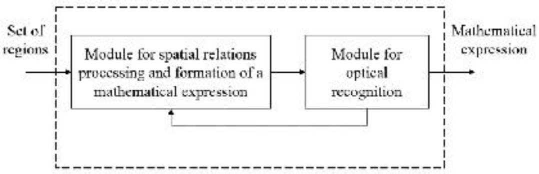


Figure 12 - Block diagram of the subsystem for optical recognition and formation of mathematical expression

Module for optical recognition consists of a feed-forward neural network which is trained by back-propagation algorithm. This module functionalities will be used by the other module.

Module for spatial relationships processing and formation of a mathematical expression makes a tree that represents a mathematical expression, and then transforms it to plain text. The projection profile cutting algorithm – PPC, is used in this module to process spatial relationships of regions.

PPC algorithm recursively divides the image space and tries to get atomic elements of the mathematical expression. Along all the way, a tree that represents a mathematical expression is created. PPC algorithm does a vertical projection and divides a space into vertical segments. Then, every vertical segment is divided horizontally. Now every horizontal subsegment is divided vertically and so on. The algorithm is finished when every subsegment has only one element.

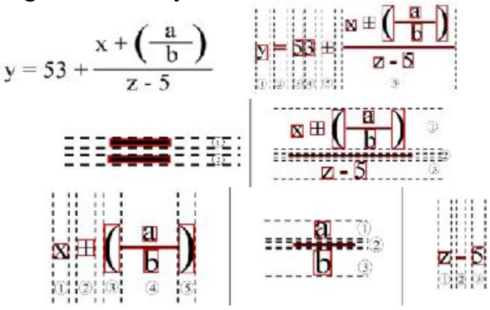


Figure 13 - PPC algorithm

PPC algorithm creates a parsing tree that represents the mathematical expression. A parsing tree for the expression shown above is shown in a Figure 14.

This tree has to be transformed into text in order to be sent to the Prolog interpreter. Here we propose a method in which the parsing tree is modeled by Composite design pattern.

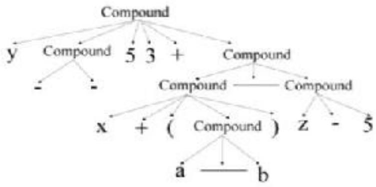


Figure 14 – A parsing tree of mathematical expression

The instance of a parsing tree is a Singleton object which has a reference to the root node. Every node, including the root, has references to all it's successors. Every node can be an instance of CompoundExpression, which represents a complex structure in a tree, or an instance of AtomicExpression, which represents just one element in the mathematical expression. Atomic expressions are leafs in the parsing tree, and compound expressions can be divided into other tree nodes. Compound expression can encapsulate many other compound or atomic expressions. This recursive structure allows this parsing tree to fit any complexity of mathematical expression. Atomic expressions are transformed into text just by returning the symbol that represents that tree node. In the compound expression, every child node is transformed into text independently, surrounded by parenthesis to preserve operator priorities is case of fractions and functions with complex arguments.

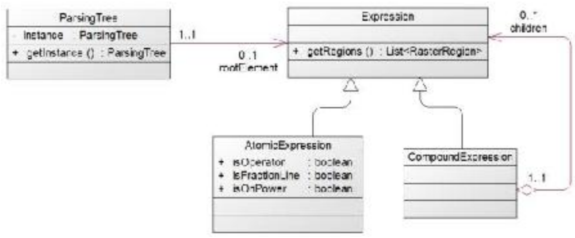


Figure 15 - Parsing tree model by Composite design pattern

4. *Symbolic differentiation subsystem*

This subsystem is designed to do the symbolic differentiation. This functionality is achieved by using Prolog programming language. Knowledge base consists of a differentiation rules for main function types, as well as rules for differentiation of complex functions and differentiation of the sum, subtraction, division and multiplication of two functions. This subsystem receives a mathematical expression in a form of text, and returns differentiated mathematical expression, also as text.

V. RESULTS

Software system proposed in this paper is tested from two different perspectives – accuracy and performance. Performance testing is done on two different machines.

	Machine 1	Machine 2
Processor	Intel Core i7 (QuadCore) @ 2.1GHz	Intel Celeron (SingleCore) @ 2.1GHz
RAM	8 GB	2GB
Graphic card	AMD Radeon 7670	Intel HD Graphics 4000

Table 1 - Testing machines specifications

a) Accuracy evaluation

Evaluation of the accuracy is done in two phases. In the first one, simple mathematical expressions are evaluated (simple fractions, no rotation, no square roots, etc). These expressions are simple and system should work accurately.

Mathematical expression	**Accuracy**
$\mathbf{sinx + cos5x}$	100%
$\mathbf{(x+5)*(60*x)+50}$	100%
$\mathbf{x+7*\left(\frac{4}{5}+3*x\right)}$	100%

Table 2 - Simple mathematical expressions accuracy evaluation

Now more complex mathematical expressions will be evaluated (complex fractions, rotated expressions, complex functions, etc). Some expressions are rotated, so the rotation algorithm will be evaluated.

Mathematical expression	**Accuracy**
$\sin x + \cos 5x$	100%
$\sin\frac{180x}{3} - 5*x+5$	100%
$(x+5)*(60*x)+50$	100%
$x+7*\left(\frac{4}{5}+3*x\right)$	100%
$\sqrt{2x+1}-\frac{2x}{7}$	90% (1)
$x^3+2x^2+x+2+2\tan x$	100%
$x+1+\sin\left(\frac{2x^3+\sin x}{5x}\right)$	100%
$x^2+2x+\frac{5+x^3}{x+5x}$	94% (2)
$\sqrt{x+3}+\frac{x}{y}-\sqrt{x^2+2}$	100%

Table 3 - Complex mathematical expressions accuracy evaluation

(1) Subtraction sign (-) recognized as a letter „i“.

(2) First element x^2 recognized as x^x.

After accuracy evaluation, it can be observed that system accuracy is very high, and somewhere around 98%. All errors were made by neural network, and other subsystems worked well.

b) Performance evaluation

Performance evaluation is done on the expression set from the second phase of the accuracy testing (complex mathematical expressions) in the same order they appeared in the Table 3. The slowest performance time on the old generation machine was around 250ms, and around 75ms for a modern machine. System can be classified as really fast. Performance time of 250ms average user sees as almost instant response from the system. All performance times are shown in a Figure 16.

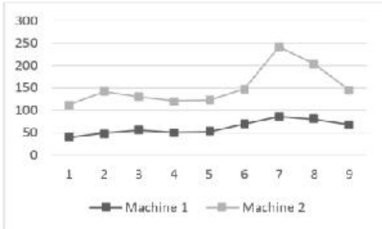


Figure 16 - Performance test results

VI. CONCLUSION

In this paper we proposed a system for optical recognition and symbolic differentiation of mathematical expressions. Many algorithms from computer vision, artificial intelligence, calculus and mathematical logic were used and system resulted to be fast and accurate. This type of system would certainly be a great help to people who do calculations every day, and also to students who could check their calculations. Future research would provide the proposed algorithm of handling images made from perspective, where parts of the mathematical expression are skewed. Furthermore, Prolog segment could be improved to be able to return a step by step solution of the symbolic differentiation process.

REFERENCES

[1] R. T. Singh, S. Roy, T. Sinam and M. K. Singh, "A new local adaptive thresholding technique in binarization," in *Interantional Journal of Computer Science Issues, Vol.8, Issue 6, No2, ISSN(Online): 1694-0814*, 2011.

[2] A. Sexton, A. Raja, M. Rayner and V. Sorge, "Towards a Parser for Mathematical Formula Recognition," University of Birmingham, Birmingham, United Kingdom, 2014.

[3] "Wolfram Alpha," [Online]. Available: (http://www.wolframalpha.com/).

[4] S. Graovac and V. Papić, "Digitalna obrada slike," Elektrotehnički fakultet, Univerzitet u Beogradu, Beograd, 2013.

[5] M. Jocić, Đ. Obradović, Z. Konjović and E. Pap, "2D fuzzy spatial relations and their applications to DICOM medical images," in *Intelligent Systems and Informatics (SISY), 2013 IEEE 11th International Symposium on Intelligent Systems and Informatics*, 2013.

[6] J. Freeman and D. M. Skapura, "Neural networks: algorithms, applications, and programming techniques," MA: Addison-Wesley, Reading, 1991.

[7] S. Weisberg, Applied Linear Regression, Minneapolis, Minnesota: Wiley, 2005.

SOFTWARE TOOL FOR RADON MAPPING AND DOSIMETRIC MODELING

Mladen D. Nikolić[1], Svetlana M. Stevović[2], Jovana Đ. Jovanović[3]
[1]*College of Applied Sciences in Technical Studies, Kosančićeva 36, 37000 Kruševac*
[2, 3]*Union Nikola Tesla University, Faculty Of Ecology And Environmental Protection, Cara Dušana 62-64, Belgrade, Serbia*

***Abstract** -The effects of exposure to radon on workers and members of the public have been examined for many years. Recent advances have been made in evaluating the risk associated with radon exposure and in implementing remediation programs in dwellings. However, decisions about whether to implement countermeasures to reduce radon exposures may benefit from an enhanced capability to evaluate and understand the associated health risk.*

This research presents the development of a user friendly software package based upon current information on radon epidemiology, radon dissymmetry, demography, and countermeasure efficiency. The software has been designed to perform lung cancer risk calculations specific to populations for various exposure profiles and to evaluate, in terms of risk reduction, the efficiency of various countermeasures in dwellings. This paper presents an overview of the general structure of the software and outlines its most important modeling approaches.

In this context, software has been developedbased upon current information on radon epidemiology, radon dissymmetry, radon exhalation rates, radon concentration and indoor ventilation rates. The software has been designed to perform lung cancer risk calculations for various exposure profiles and to evaluate, in terms of risk reduction, the efficiency of ventilation in dwellings. This paper presents an overview of the general structure of the software and outlines its most important modeling approaches.

Key words: software, , building, radon, dwelling, health risk, lung cancer.

1. INTRODUCTION

Radon (^{222}Rn) and its progeny are present in all dwellings, because radium is present in building materials as well as in the soil. It is important to understand the generation and migration process of radon from building materials, which contributes to 55% of total radiation dose received by the population from the environment (UNSCEAR, 2000) [1]. A research dealing with the radiation exposure produced by radon decay products in homes has shown that the cause of increased radiation may as well be the construction material of the building[2]. When estimating an average annual radiation exposure one should also take into account life at home, because the level of radiation can be higher within homes than without, bearing in mind that walls themselves can contain and emit radionuclides. Depending on the type of the construction material and ventilation efficiency [2,3,4], radiation indoors can be many times higher than outdoors, and as such could represent a serious health hazard.

The average annual dose in Europe from radon and its progeny in homes and workplaces is estimated to be 1,6 mSv.

Radon gas is formed inside building materials by decay of the parent nuclide ^{226}Ra. However, it is not possible to determine the radon exhalation rate simply from the activity concentration of ^{226}Ra. Instead one must measure radon exhalation rates directly from the surface of the material, or measure activity concentration of radon inside premises by long term or short term measurements as shown by many studies in the literature. [3,4].

When radon activity concentrations are calculated from exhalation rates or measured it is possible to estimate effective dose and risk of lung cancer from radon.

Based on these requirements relational database and Web technology were selected provided by its access and update. In this way, the data that is mutually correlated is optimally connected, and access and update could be provided from different locations and with more clients simultaneously. For the realization MySql database and web programming language PHP and JavaScript were used.

2. METHOD OFCALCULATION

2.1. Radon activity concentration

Data stored in the database are subject of further analysis and processing.

Different tasks can be applied, such as the following general tasks from the domain of low-level user activities:

- value retrieval – finding of specific attribute values for data cases (e.g. find measurements for specific date and location),
- filtering – finding data cases which satisfy the given conditions in relation to their attributes (e.g. find locations that have radon activity concentrations higher than 200 Bqm^{-3}),
- compute derived value – finding the appropriate aggregate numerical value for data cases (e.g. find the

average radon activity concentration for the given location in the given time),
- find extreme values – finding data cases with extreme attribute values (e.g. find the location with highest average radon activity concentration),
- determine range – finding the range of attribute values of interest for data cases (e.g. find the span of two radon activity concentrations for a given area),
- characterize distribution – finding the distribution of the quantitative attribute of interest for data cases (e.g. find the distribution of radon activity concentrations for a given time),
- find anomalies – identification of any anomalies with respect to a given relation or expectation, i.e. statistical outliers (e.g. sudden concentration changes),
- clustering – finding clusters of similar attribute values for data cases (e.g. divide the area into two groups according to their levels of radon activity concentration),
- correlation – finding useful relations between values of two given attributes for data cases (e.g. find the correlation between sudden increase of radon activity and earthquakes).

As can be seen, the suggested starting framework for the data analysis offers a very broad spectrum of useful applications which should certainly contribute to the environmental radon monitoring.

Advanced techniques of data analysis can be relatively easily embedded into the existing general framework of the software, if there is a specific demand.

Various techniques from the domain of artificial intelligence aimed at expert systems are available, such as pattern recognition, neural networks, fuzzy logic application, etc. [6-8]

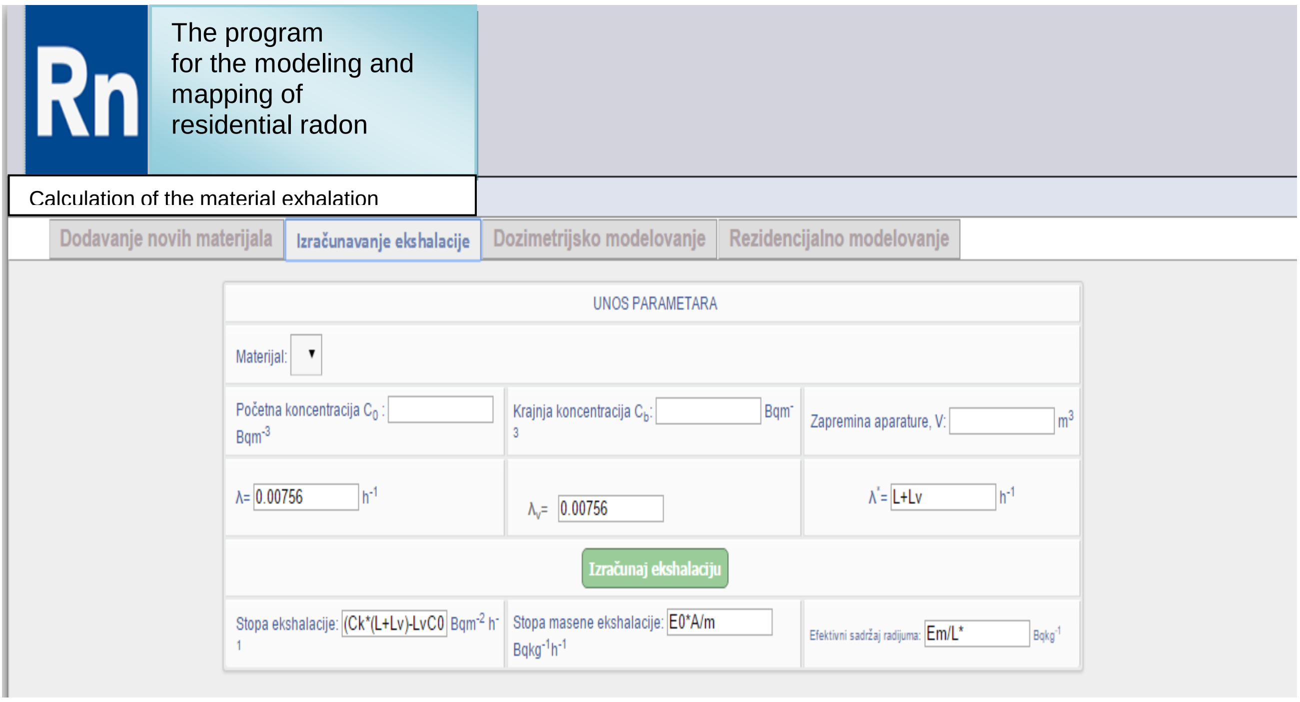


Figure 1. Appearance of the panel for exhalation rate calculation

Figure 1 shows panel for exhalation rate calculations where the radon exhalation rate E_0 is defined as the liberation quantity of radon activity from the surface area of building materials per unit time (Bq m^{-2} h^{-1}). Using this value and the area of the indoor surface, radon emanated per unit time could easily be calculated and used to estimate the radon concentration in the indoor environment.

2.2. Dosimetric model

The dosimetric model selected for the software is a radon specific model on the basis of the model used to describe and validate the human respiratory tract model recommended in 1994 by the International Commission on Radiological Protection (ICRP) in its publication 66 (1). This model relies on various parameters specific to the exposed individual (age, sex, activity level, nose or mouth breathing) together with parameters describing the radon progeny aerosol. It also takes into account the radiation weighting factor for alpha particles and factors describing the relative radio sensitivity of the three lung compartments.

2.3. Radon exposure model

The collective exposure of the occupants of a dwelling is determined by an exposure model that uses information on the age and sex of the occupants and also the number and type of rooms (kitchen, living room, and bedroom) in the dwelling. The room type is used to determine the equilibrium factor and the

aerosol characteristics. Additionally, for each room, the average concentration of radon-gas is required (this is adjusted using an ad-hoc factor in case of a measurement of a below one-year duration), and the rooms occupancy (also the time spent in each activity level in case of the use of the dosimetric approach).Figure 2. shows application for dosimetric modeling. In order to calculate annual effective dose certain residential scenario must be made and it is explained in this paper.

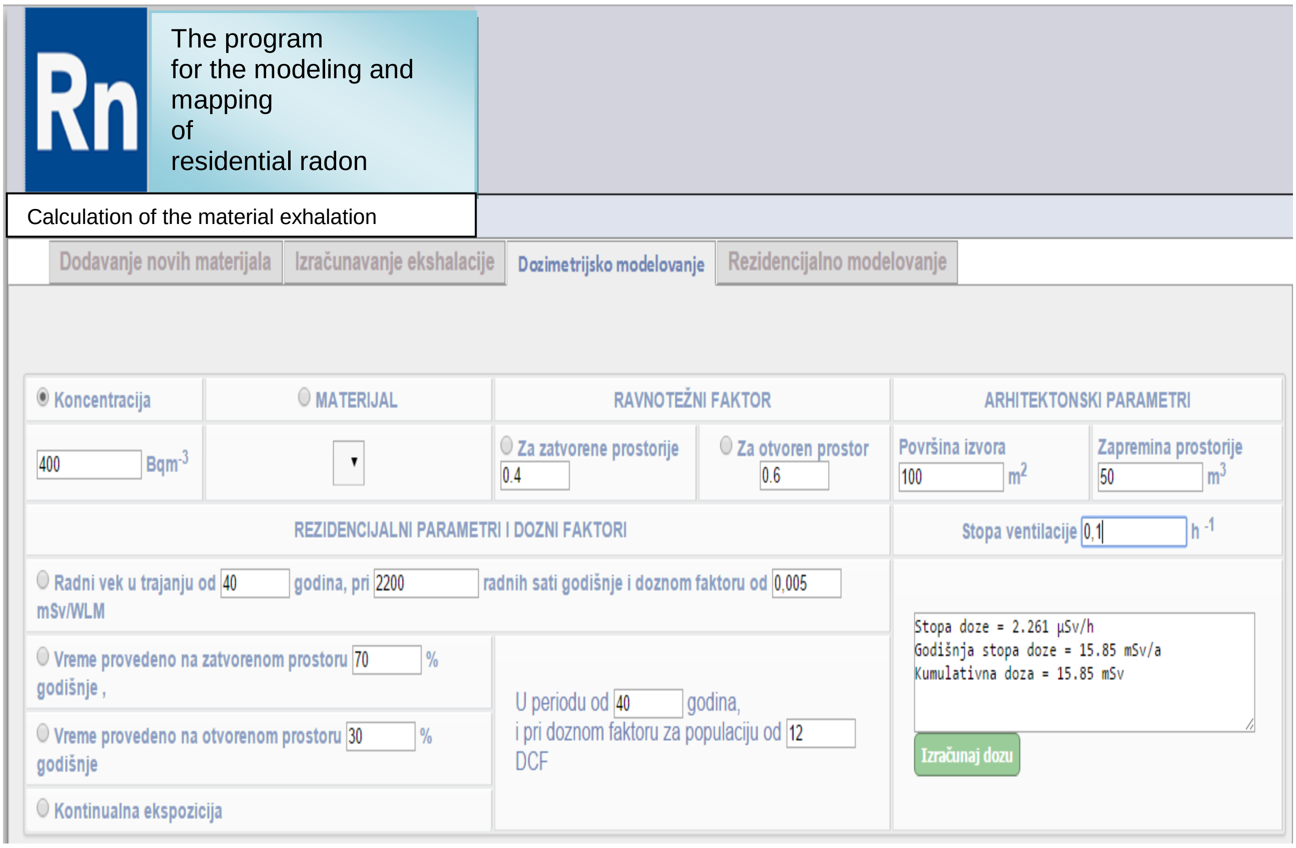


Figure 2. Application for dosimetric modeling

This paper presents a simplified model of a home with basis of 48 m^2, whose walls are built of siporex blocks 0.25 m thick and 0.6 g/cm^3 dense while floors and ceilings are made of a 2.3 g/cm^3 dense concrete.

The assumption is that the home consists of four rooms: bedroom, livingroom, vestibule, and bathroom and the height of the apartment is 2.2 m. In the center of each room there is a recipient (tenant) who spends some required time in it. Table 1 shows exact room dimensions as well as total walls surface and volumes of rooms.

Table 1 shows architecture parameters necesery for the calculations while Figure 3 shows floor plans of the apartment modelwith dimensions of residential premisses in spaces.

Table 1. Dimensions of the rooms

Room type	Room dimensions [m]	Total windows and doors surface [m^2]	Total walls surface made of siporex [m^2]	Total floor and ceilings surface made of concrete [m^2]	Room volume [m^3]
Living room	4x5	4.39	35.21	40	44
Bedroom	4x3	4.92	25.88	24	26.4
Vestibule	2x5	5.54	25.26	20	22
Bathroom	2x3	2.37	19.63	12	13.2

Contribution to total radon activity concentration of floors and ceilings made out from concrete for each room are taken into the consideration and they are calculated in the same manner and added to final radon concentration activity assasment, which is visable through all tables and figures, selected from the research and presented here.

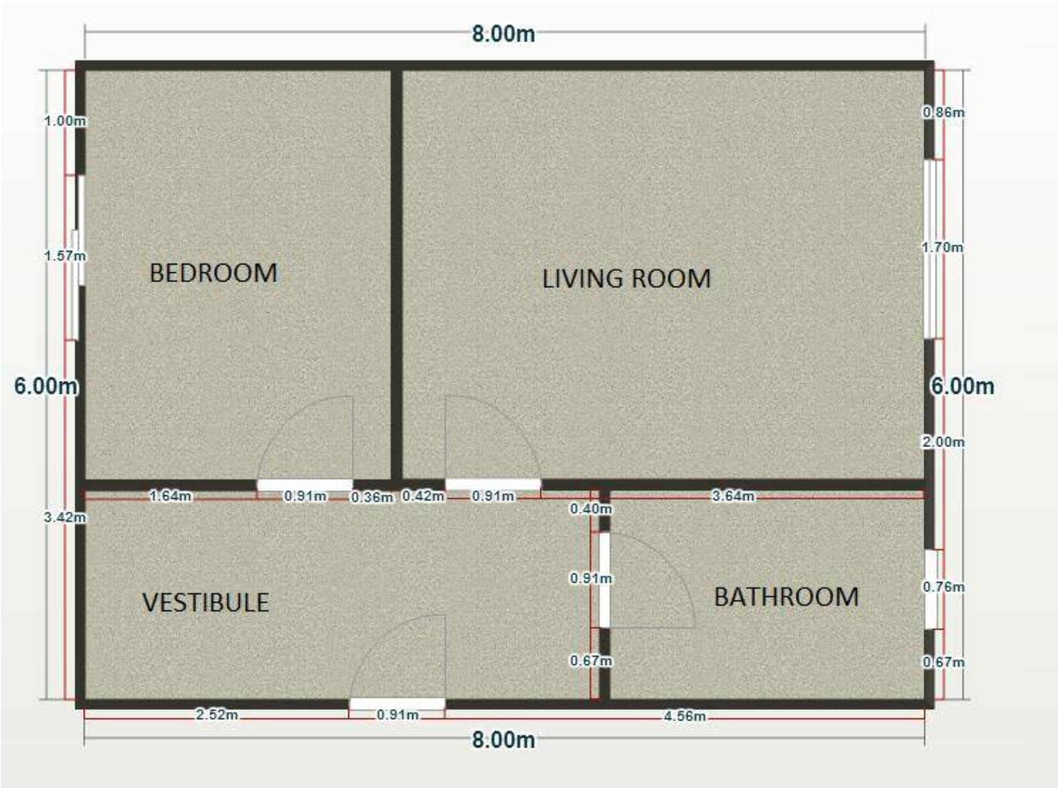


Figure 3. Floor plans with dimensions of residential premisses

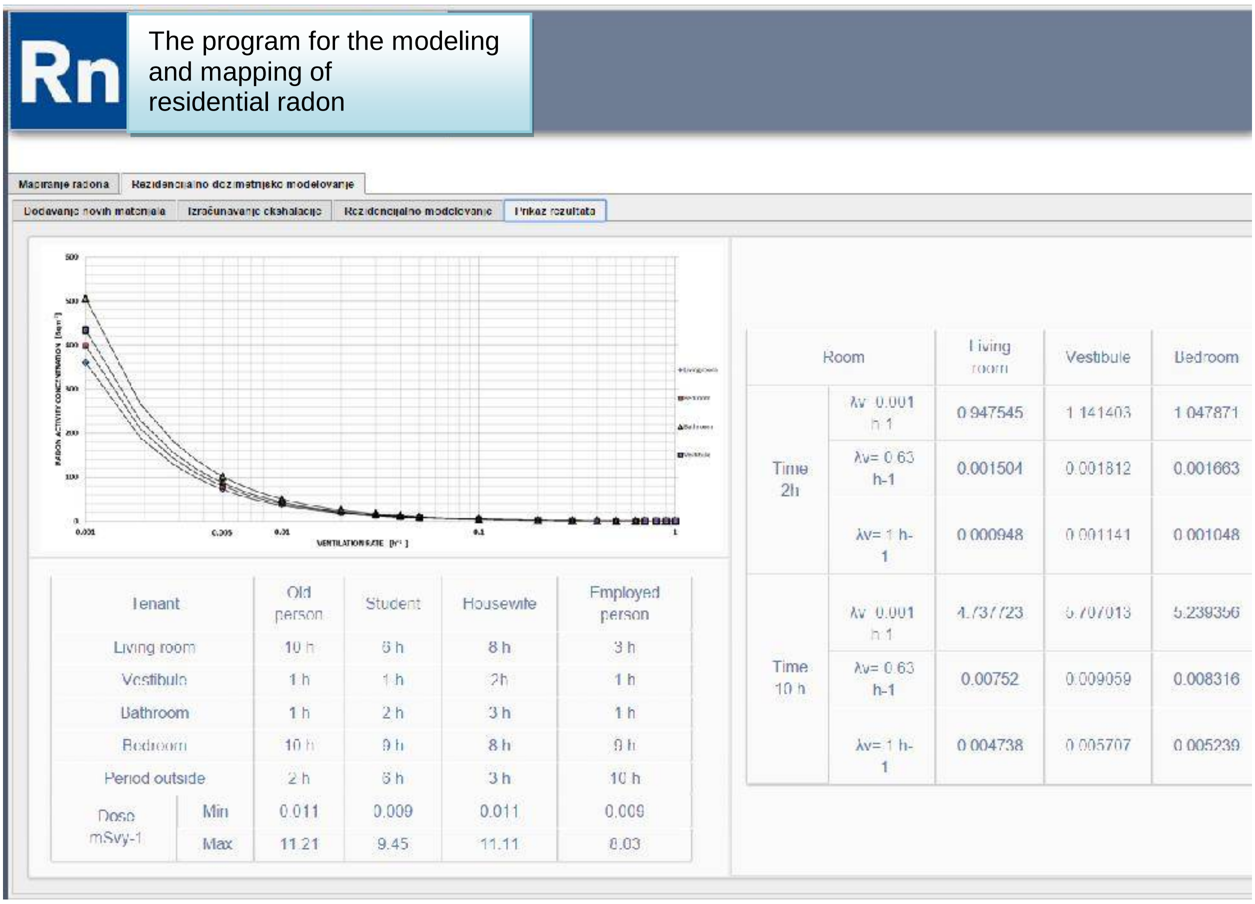


Figure 4. Radon activity concentration in the function of ventilation as countermeasure efficiency

Radon activity concentration can be either calculated through exhalation rate or entered as value into related textbox.

3. THE RESULTS

Figure 4. shows application with the results of residential scenario. Other software features are:

All the data (input data and results) may be viewed and edited as tables via the built-in spreadsheet system. Data may also be presented graphically, using a built-in system (see Figure 4). Data may be exported in standard tabular or graphical formats. Multidimensional data may be presented under any user-defined order of dimensions.

The results of various calculations may be compared in a single tabular or graphical output in order to facilitate the identification of the most cost-effective countermeasure as well as the analysis of the impact on the results of the simultaneous variation of one or several input parameters.

4. CONCLUSION

Software tool for radon mapping anddosimetric modeling for the evaluation of risks associated with radon exposure in dwellings, is user friendly and was developed on the basis of the most recent data on radon epidemiology, radon dissymmetry, demography, and countermeasure efficiency. This software could play a role in general training, the provision of information to the public and in investigating the effect on risk of different dose reduction strategies. Finally, its sensitivity analysis capabilities and its database system together with its user friendly configuration capabilities should make it an easy to use tool for the risk evaluation experts of various countries to perform useful calculations, appropriate to their situations with regard to local radon and smoking patterns.

REFERENCES:

1. United Nations Scientific Committee on the Effects of Atomic Radiation, UNSCEAR 2000 Report to the General Assembly, VOL I,UNSCEAR secretariat Vienna International Centre
2. Mladen Nikolić, Milena Nikolić, Anastasija Matić, A Resrad Build Programme Based Assessment of Radiation Exposure as a Serious Health Hazard in Homes Built of Siporex Blocks, International Conference of Biomedicine and Geosciences, June 2014, Belgrade, Serbia
3. Svetlana Stevovic,Dragana Vasilski,*Smart materials in architecture as a function of sustainable development*, Proceeding of theConference’’Contemporary theory and practice in building development’’,Publisher: Institute for Construction Banja Luka,pp 87-97,B. Luka, 2013.
4. Hranislav Milošević, Dragana Vasilski,Svetlana Stevović,Full Length Research Paper: *Sustainable technology for existing buildings renovation in the function of energy efficiency*,Techniques Technologies Education Management Journal, vol 8 no 1, 2013
5. Cavallo, A., Gadsby, K., Reddy, T.A. (1996) Comparison of natural and forced ventilation for radon mitigation in houses. Environment International,Volume 22, Supplement 1, 1996, Pages 1073–1078,The Natural Radiation Environment VI
6. Mladen Nikolic, Predrag Sibinovic, Nemanja Ilic, Information System in The Function of a Quality System Development Illustrated in the Example of a Serbian Academy of Applied Studies, International Conference on Quallity Assurance for Successful Business and Competitiveness, December 2014, Kopaonik, Serbia
7. Svetlana Stevović, Predrag Popović, *Fuzzy methodological approach to technical and environmental quality management,* 6th International Working Conference ‘Total quality management advanced and intellegent approaches’, 1st – 4th June, 2009, Belgrade,
8. Danijela Tadić, Svetlana Stevović, Dragan Milanović, *Improvement of Quality Tools and Methods by Applying Fuzzy Sets Theory*, 6th International Working Conference ‘Total quality management advanced and intellegent approaches’, 1st – 4th June, 2009, Belgrade
9. N.P Petropoulos, , M.J Anagnostakis, S.E Simopoulos, Building materials radon exhalation rate: ERRICCA intercomparison exercise results, Science of The Total Environment, Volume 272, Issues 1–3, 14 May 2001, Pages 109–118
10. Samuelson C. , The closed-can exhalation method for measuring radon. Journal of Research of the National Institute of Standards and Technology, 95, 167-169, 1990.
11. Chao,C.Y.H., et al., Determination of Radon Emanation and Back Difussion Characteristics of Building Material in Small Chamber Tests, Building Environment, 32, (1997),4, pp. 355-352
12. G Keller, B Hoffmann, T Feigenspan, Radon permeability and radon exhalation of building materials, Science of The Total Environment, Volume 272, Issues 1–3, 14 May 2001, Pages 85–89