Improving geoportal information search capabilities – an approach based on semantic similarity measurement

Miloš Bogdanović*, Aleksandar Stanimirović*, Leonid Stoimenov*
* Faculty of Electronic Engineering, University of Niš, 18000 Niš, Serbia
{milos.bogdanovic, aleksandar.stanimirovic, leonid.stoimenov}@elfak.ni.ac.rs

Abstract— **In this paper we will define and describe a novel approach for improving geoportal information search capabilities. The approach we present in this paper is a part of our ongoing research regarding the development of Web-based geographic information systems. Our approach is meant to be implemented within geoportals relying on federated geographic information systems (geo-information systems, GISs) as their spatial data infrastructure. Although it can be adapted for different meta-data, our approach was intended to rely on federated GISs which utilize ontological components (ontologies) for geospatial data integration purposes. Geoportal information search improvement is performed by taking advantage of existing ontological components, in particular the sense of the ontology concept names, and matching their sense with the sense of terms extracted from a natural language description of geo-information. In this way, our approach enables searching for geo-information in a heterogeneous and distributed GIS application environment.**

I. Introduction

Since the publication of the Infrastructure for Spatial Information in the European Community (INSPIRE) directive [1], we are witnessing a constant growth in a number of implemented spatial data infrastructures (SDI) [2][3][4][5]. Accessing and sharing geospatial data assembled from heterogeneous and distributed geospatial data sources are issues that necessarily emerge during the development of spatial data infrastructures. As stated in [1], "the problems regarding the availability, quality, organisation, accessibility and sharing of spatial information are common to a large number of policy and information themes and are experienced across the various levels of public authority". Therefore, development of a spatial data infrastructure (SDI) requires means which enable sharing, access and interoperability of spatial data and services.

Spatial data access is foreseen as one of the central components in a SDI. In most cases, this component is implemented as single points of discovery and access to information within SDI – a geoportal. The common characteristic of geoportal representatives is the usage of Web GIS applications in a distributed computing environment. Contemporary geoportals rely on technologies used for the development of distributed GIS solutions. The success of a single geoportal significantly depends on the quality of a Web GIS application that represents geoportal as seen from user˝s perspective. Therefore, it is evident that there is a strong dependency between the technologies used for the development of distributed GIS and geoportals.

The ability of each user to find (discover) what he/she is searching for, in terms of displayed geo-information and maps, is considered one of the primary geoportal objectives. At a glance, this request may seem rather straightforward but is in fact highly complex and directly affects geoportal usability as seen from user˝s perspective [6][7]. Individuals who do not belong to Geographic Information System (GIS) world, also referred to as "non GIS professionals", usually expect to be capable of discovering geo-information and maps using a natural language description of the information they are interested in. The ability such as this one raises geoportal usability issues and has proven to be a difficult one to achieve using current geoportal architecture [6].

Previously reported prominent examples of geoportal development, such as ones described in [8][9][10][11], indicate that contemporary geoportals relay on the usage of metadata catalogues. Metadata catalogue is usually implemented in accordance with the Open Geospatial Consortium (OGC) standard named OpenGIS Catalogue Services Specification [12]. The implementation of OGC standards improves the overall level of structural interoperability of a GIS solution. Still, in case of a geoportal implementation scenario, there are some important requests to be implemented or functionally improved: estimate relevance percentage of geo-information discovery results, classify the result using relevance percentage or analyse relationships between words used for geospatial data discovery [13]. The described state represents the motivation for the research we are continuously conducting. A part of this research will be presented in this paper. In particular, in this paper we will an approach for improving geoportal information search capabilities. The approach we will present in this paper enables users to discover geo-information within geoportals relying on federated geographic information systems (GISs) as their spatial data infrastructure. Our approach simplifies the discovery of geo-information within geoportals by enabling users to discover geo-information simply providing a natural language of geo-information they are interested in. The discovery process is based on semantic similarity measurement between natural language of geo-information and the description of ontological components used to describe the content of federated geographic information systems (GISs).

The rest of this paper is organized as follows. Section 2 discusses related work regarding semantic similarity measurement performed within federated geographic information systems for geo-information purposes. Also, this section discusses word sense disambiguation (WSD) methods that can be used to aid geo-information discovery process. In section 3, we describe a general architecture we envision our approach to be implemented in. Section 4 describes the approach we propose for improving geoportal information search capabilities. In section 5, we present our approach in practice by providing an example of semantic similarity determination between user-defined description and an ontology concept. Section 6 concludes with an outlook to future work.

II. Related work

In the environment of distributed geo-information sources, such as a spatial data infrastructure, users are usually provided with a single uniform access point over the refined data – a geoportal [13]. Geoportals commonly enable users discover geo-information they are interested in by implementing an information search capability. Geoportal users expect search results to be in a form of homogeneous data set(s). In order to be capable of providing users with results in such form, geoportals should rely on infrastructures capable of integrating data originating from several autonomous systems [14]. In the case we investigate in this paper, these autonomous systems are geographic information systems. Thus these infrastructures can also be observed as federated geographic information systems (federated GIS) [15].

Information integration is one of the core tasks performed by a federated GIS. To be able to perform such a task, federated GIS is in a need for a mechanism that can overcome the problem of semantic heterogeneity between different geospatial data sources [16][17]. Ontologies represent common means used to solve semantic heterogeneity problems within federated GISs [17][18]. If a federated GIS provides users with a single access point over integrated data and uses ontologies for resolving semantic heterogeneity problems, it has to solve two additional problems: perform mapping between ontologies and geospatial data sources used by individual GISs [19][20], and define discoverable Web interfaces which represent information source access points [21][22]. If a federated GIS implements means for overcoming these problems, in that case each GIS within a federated GIS can be discovered by utilizing its Web interface and mappings between ontologies and geospatial data sources can be used to retrieve integrated data. Thus, a federated GIS becomes an infrastructure which enables a geoportal to perform one of its core functionalities: provide users with ability to search through integrated data originating from heterogeneous sources.

It is our opinion that a geoportal information search capabilities can be improved in the described environment. The improvement we propose allows users to describe geo-information they are interested in using their own words, their own language. Our approach takes advantage of word sense disambiguation (WSD) algorithms as an intermediate task to aid semantic similarity measurement between the sense of the ontology concept names and the sense of terms extracted from a user-defined natural language description of geo-information.

A. *Semantic similarity measurement*

The notion of similarity originated in psychology. Similarity is considered one of the central theoretical constructs in psychology [23]. It can be used to perform grouping among entities and to determine if some entity categories are comparable to each other. Regarding semantic similarity measurement, it is usually performed between entity types whose representation can be highly complex. Entity type representation depends on the chosen representation language which in turn makes similarity measures difficult to compare [24].

Geographic Information Science (GIScience) has widely adopted semantic similarity measures over the past decade. One of the most widely adopted semantic similarity measures is called Matching Distance Similarity Measure (MDSM) [25]. This measure is based on Tversky's feature model [26]. It supports context theory, automatically determined weights and asymmetry. Raubal [27] suggested usage of conceptual spaces, as described in [28], to achieve cognitive semantic interoperability. Schwering and Raubal [29] proposed a method to extend current semantic similarity measures by accounting for the spatial relations between different geospatial concepts. Janowicz and Raubal argued that an affordance-based representation of the context in which similarity is measured, improves the quality of similarity measure [30]. Regarding ontologies described by description logics and used in geospatial domain, there is a number of prominent proposals suggested with aim to bring ontologies closer to existing similarity theories and semantic similarity measures [31][32][33].

B. *Word Sense Disambiguation*

Natural Language Processing (NLP) proclaims Word Sense Disambiguation (WSD) as one of its core tasks [34]. Word Sense Disambiguation (WSD) algorithm assigns an appropriate sense(s) to each word of a given text. WSD algorithms are divided into two classes of algorithms: supervised and unsupervised methods. A supervised WSD algorithm compares information by taking advantage of labeled training data, whereas the unsupervised method does not. According to [35], WSD methods can be classified into: path-based, information content based, gloss based and vector based methods.

In a majority of cases, external knowledge sources are considered fundamental components to perform WSD. Ontologies, glossaries, thesauri, computational lexicons, corpora of texts and other sources are often utilized as external knowledge sources. Regarding WSD methods, probably the most employed external knowledge source is WordNet [36]. “Synonym sets” (synsets) represent the main building blocks used to organize WordNet as a computational lexicon. The latest version of WordNet (3.1) is available online and it contains over 155000 terms for 117000 synsets. A “synonym set” (synset) within WordNet is a structure built up of the following components: a term (word), its class (verb, noun, adjective etc.) and connections to all semantically related words along with a brief definition („gloss“) illustrating the use of the synset members. Aside from the previously enumerated components, each synset may have semantic relations defined. A semantic relation defined for a synset can be applied to all its members. WSD methods mostly utilize the following semantic relations: hypernymy (also called kind-of or is-a), hyponymy (the inverse relations of hypernymy), meronymy (also called part-of) and holonymy (the inverse of meronymy).

Unsupervised WSD methods can utilize semantic similarity measures used to perform word disambiguation. WordNet can be effectively used within a majority of these WSD methods. In such case, similarity between terms (synsets) is determined by utilizing semantic relations defined within WordNet. For example, path-based methods measure the length of the path between two words in a graph-like structure. In this case, WordNet can be used as a graph-like structure which provides the paths. Approaches such ones described in [37], [38] and [39] have successfully utilized WordNet as a graph-like structure to perform word similarity measurement.

III. Geoportal environment

The purpose of this paper is to present an approach we have developed as a part of our ongoing research regarding the development of Web-based geographic information systems. Our approach targets geoportals relying on federated geographic information systems (GISs) as their spatial data infrastructure. In particular, the approach we present in this paper relies on federated GISs which utilize ontological components (ontologies) for geospatial data integration purposes. Geoportal information search improvement is performed by taking advantage of existing ontological components. We take advantage of the sense of the ontology concept names, and match their sense with the sense of terms extracted from a natural language description of geo-information. In this way, our approach enables searching for information in a federated GIS environment, as shown in Fig. 1.

Fig. 1 represents a visualization of an environment we envision our approach to be implemented in. The main facilities of the environment are the following:

- Natural language description – users define a natural language description of geo-information they are interested in.
- Federated geo-information systems (Federated GISs) – in the context of our approach, federated GISs utilize ontologies to overcome semantic heterogeneity problems
- Ontologies – a content description of different geospatial data sources used by different GISs in a federated GIS; each federated GIS maintains its ontology within a central ontology repository.
- Ontology repository – a repository capable of storing ontologies for each GIS within a federated GIS; this component can be omitted if each part of federated GIS stores ontology locally.
- WordNet computational lexicon – a computational lexicon used as a provider of a taxonomy of terms; it is used as a knowledge source to associate the most appropriate senses with terms (words) given by the user; also, it is used as a knowledge source for semantic similarity measurement purposes.
- Search Engine – this module implements information search capability by matching federated GIS ontology concepts with user-defined geospatial data description; for this purpose, search engine utilizes WordNet computational lexicon as a knowledge source.

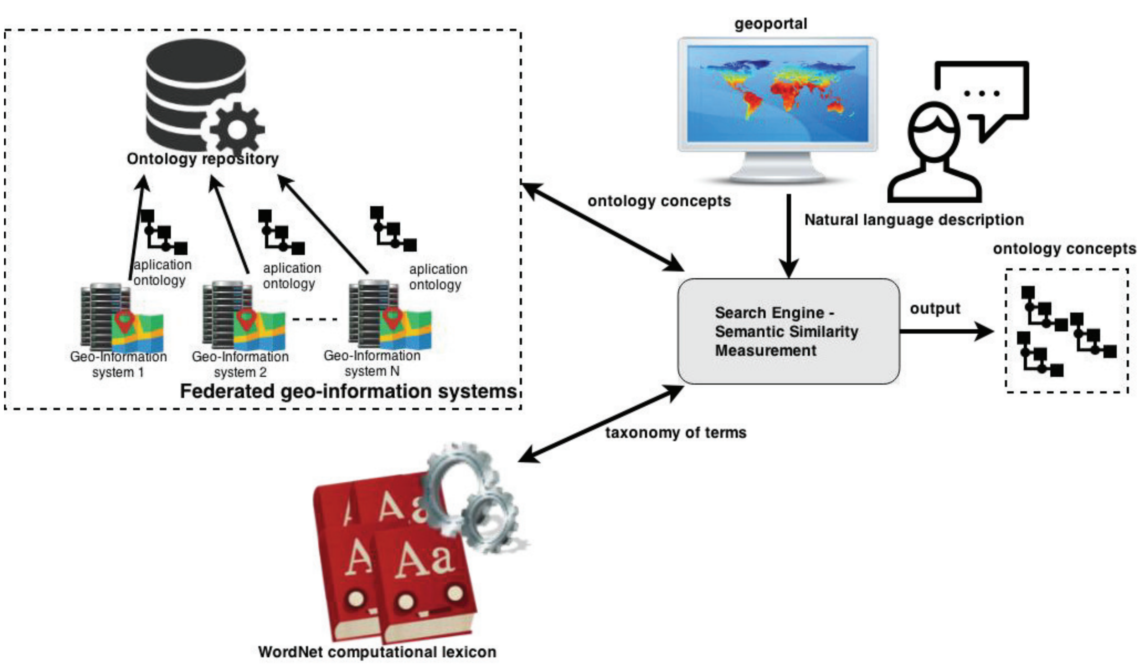

Figure 1: Geoportal in a federated GIS environment

IV. An Approach for Geoportal Information Search Improvement Based on Semantic Similarity Measurement

In this section, we will describe the approach we have developed for the purpose of improving geoportal information search capabilities. Our approach will be presented in the form of an algorithm with four sequential steps. In starts by utilizing a natural language description of geo-information given by user(s) and determines a set of concepts that belong to ontology used to describe the content of federated GIS components.

A. *Disambiguate user-defined description of geospatial data*

Our algorithm starts with tokenizing natural language description of geospatial information given by user(s) into a list of words. For these purposes, regular expressions are used. Afterwards, WordNet computational lexicon is utilized to identify a correct part of speech for each of the tokenized words, as shown in Fig. 2. The identification process includes the stripping of suffixes from words, as shown in Fig. 2. Suffix stripping is implemented according to algorithm defined in [40]. After the identification process is complete, words identified as nouns are extracted to a separate term set, which will be referred to as *„natural language description term set"*.

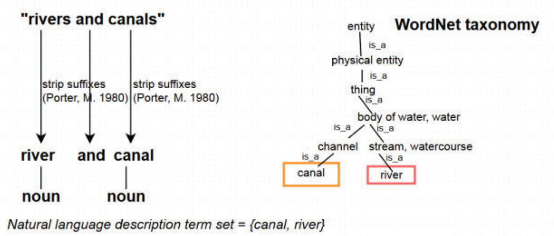

Figure 2. Disambiguation process for a user-defined description

B. *Acquire and disambiguate description of ontology concepts*

In most cases, ontologies concept descriptions can be acquired from *rdfs:comment* or *rdf:label* tags contained within ontology concept definitions (*rdfs:Class*). Because of performance issues, our proposal currently prefers *rdf:label* over *rdfs:comment*, since *rdf:label* contains a brief description of concept (single sentence) as opposed to *rdfs:comment* which can contain a wider description (a few sentences). The description contained within *rdfs:comment* or *rdf:label* tags is disambiguated in a way described in *Step A*. As a result of this process, a term set referred to as *„concept description term set"* is created for each concept of an ontology. As an example, Fig. 3 illustrates a disambiguation process for a description of the concept "*Word*" as defined in Suggested Upper Merged Ontology (SUMO) [41]. In this case, *rdfs:comment* was used to demonstrate the disambiguation process.

The following steps, Step C and Step D, must be repeated for each *„concept description term set"* created for each ontology concept. The same *„natural language description term set"* is used in all iterations.

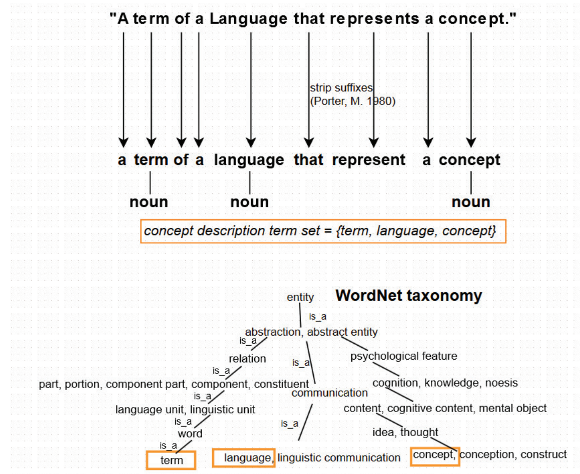

Figure 3: Disambiguated description of concept "Word" – SUMO ontology

C. *Initialize and populate semantic similarity matrix*

Semantic similarity matrix *S[m, n]* is created for a pair of term sets whereas each pair consists of a *„natural language description term set"* and a *„concept description term set"* created for the ontology concept whose similarity is currently being measured. *m* represents the number of terms within *„natural language description term set"* while *n* represents the number of terms within *„concept description term set"*.

In order to populate semantic similarity matrix *S[m, n]*, semantic similarity measurement is performed for each pair of terms T_{NL} and T_{CD} from *„natural language description term set"* and a *„concept description term set"*, respectively. Semantic similarity between terms T_{NL} and T_{CD} is computed according to algorithm described in [39]. Algorithm described in [39] measures the path length to the root node from the least common subsumer (LCS) of the two terms compared within a graph-based structure. In this case, WordNet computational lexicon provides paths between the observed terms. In case one of the terms or both of them do not exist in the WordNet lexicon, semantic similarity is determined according to Levenshtein distance [42]. Computed semantic similarity represents *S[i, j]*, whereas *i* represents the index of term T_{NL} within *„natural language description term set"* and j represents the index of term T_{CD} within *„concept description term set"*.

D. *Calculate semantic similarity between user-defined description of geospatial data and ontology concept*

The overall similarity between user-defined description of geospatial data and ontology concept is computed as an average value of similarities for each pair of terms from *„natural language description term set"* and a *„concept description term set"*, respectively (equation 1). Similarity value ranges from 0.0 to 1.0.

$$sim = \sum_{i=0}^{m-1}\sum_{j=0}^{n-1} S[i,j] \Big/ m*n \qquad (1)$$

V. An Example of Semantic Similarity Determination Between User-Defined Description and an Ontology Concept

To demonstrate the algorithm described in section IV, this section will demonstrate a brief example semantic similarity computation. Let us suppose that the user is interested in finding all available geo-information regarding canals and rivers using a geoportal instance. Thus, a natural language description of geo-information given by the end user in this case could be *"rivers and canals"*. In that case, the output of the first algorithm step would be the one presented on Fig. 2. Natural language description term set consists of two terms: *"river"* and *"canal"*.

Also, let us suppose that the ontology used to describe the content of underlying geoportal infrastructure is Suggested Upper Merged Ontology (SUMO) [41]. Algorithm steps B, C and D will be demonstrated using SUMO ontology concept *WaterArea (rdfs:Class rdf:ID= "WaterArea")*. SUMO ontology describes water area as "*A body which is made up predominantly of water, e.g. rivers, lakes, oceans, etc."*. The disambiguation process regarding this description is shown on Fig. 4. As a result, concept description term set consists of five terms: *"body"*, *"water"*, *"river"*, *"lake"* and *"ocean"*.

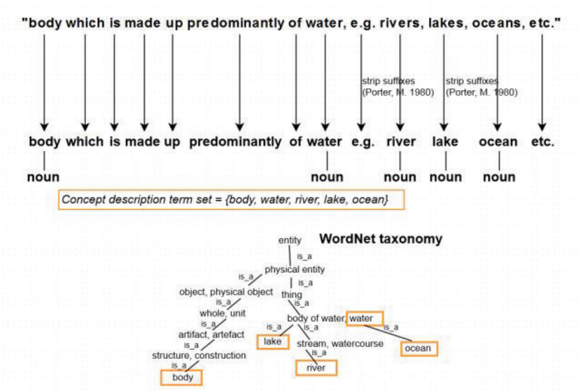

Figure 4. Disambiguation process for SUMO ontology concept *"WaterArea"*

According to algorithm Step C, semantic similarity matrix *S[2, 5]* is created and populated with semantic similarity measurement results for each pair of terms from natural language description term set and concept description term set. The resulting matrix is shown on Fig. 5.

	body	*water*	*river*	*lake*	*ocean*
river	0.5	0.8	1.0	0.73	0.73
canal	0.71	0.8	0.67	0.73	0.73

Figure 5. Similarity matrix S

The overall similarity between user-defined description of geospatial data and SUMO ontology concept named *WaterArea* is computed to 0.74 which indicates a high level of possibility that *WaterArea* ontology concept is related to geo-information the user is searching for.

VI. Conclusion

For the purpose of improving geoportal information search capabilities, this paper proposes an approach based on semantic similarity measurement which utilizes existing ontological components of federated GISs. The approach core is a process of computing semantic similarity between terms extracted from a natural language description of geo-information, defined by the end users, with the terms extracted from the description of ontology concepts. This process uses WordNet computational lexicon as a knowledge base for semantic similarity measurement. As a result, the approach presented in this paper outputs a set of ontology concepts that describe geospatial content within federated GISs.

Core benefits we envisioned to bring through this proposal can be summarized as follows:

- Simplified geoportal information search capabilities using natural language description of geo-information.
- The described algorithm can be implemented as a part of existing software components or as a separate Web service. Thus, it becomes independent of federated GIS"s tier that implements geo-information access capability.
- Ontologies are used in their original form, as implemented by federated GIS developers. Therefore, our approach retains the possibility to perform reasoning over existing ontologies.

Although presented approach can be easily implemented, additional efforts should be made to make it omni-implementable. Special attention should be devoted to processing complex text expressions along with ability to utilize other types of semantic descriptions instead of ontologies.

Acknowledgment

Research presented in this paper was funded by the Ministry of Science of the Republic of Serbia, within the project "Technology Enhanced Learning", No. III 47003.

References

[1] The European Parliament and The Council of the European Union, Directive 2007/2/EC of The European Parliament and of The Council, of 14 March 2007: establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), 2007

[2] Department of Environment, Community And Local Government, "Irish Spatial Data Infrastructure – Implementing the INSPIRE Directive regulations in public bodies", 2011

[3] D. Vandenbroucke and D. Biliouris, "Spatial Data Infrastructures in Slovak Republic: State of play 2010", University of Leuven, 2010

[4] Lj. Živković, "Approach to Spatial Data Infrastructure Development for Spatial Planning in Serbia", REAL CORP 2013, the 18th International Conference on Urban Planning, ISBN:978-3-9503110-5-1 Rome, Italy, May 2013

[5] D. Vandenbroucke and D. Biliouris, "Spatial Data Infrastructures in Czech Republic: State of play 2010", University of Leuven, 2010

[6] T. Aditya and M-J. Kraak, "Reengineering Geoportal: Applying HCI and Geovisualization Disciplines", In Proceedings of the 11th EC-GI & GIS Workshop, Alghero, Italy, 29 June–1 July 2005.

[7] B. Resch and B. Zimmer, "User Experience Design in Professional Map-Based Geo-Portals", ISPRS International Journal of Geo-Information, 2013, 2, 1015-1037; doi:10.3390/ijgi2041015

[8] V. Bernhard, A. Richter and M. Mittlböck, "From Geoportals to Geographic Knowledge Portals", ISPRS International Journal of Geo-Information 2.2 (2013): 256-275, 2013

[9] E. Sakkopoulos, T. Mildorf, K. Charvat, I. Berzina and K. U. Krause, "Plan4All GeoPortal: web of spatial data", In Proceedings

of the 21st international conference companion on World Wide Web (pp. 279-282). ACM, 2012

[10] F. R. Salas, E. Boldrini, D. R. Maidment, S. Nativi and B. Domenico, "Crossing the digital divide: an interoperable solution for sharing time series and coverages in Earth sciences", Natural Hazards and Earth System Sciences, 12, 3013–3029, 2012, doi:10.5194/nhess-12-3013-2012

[11] L. Bernard, I. Kanellopoulos, A. Annoni and P. Smits, "The European geoportal – one step towards the establishment of a European Spatial Data Infrastructure", Computers, Environment and Urban Systems, 29 (2005) 15–31, 2005

[12] Open Geospatial Consortium, D. Nebert, A. Whiteside and P. Vretanos Eds., OpenGIS Catalogue Services Specification, v2.0.2, OGC 07-006r1, 2007, Available from: http://portal.opengeospatial.org/files/?artifact_id=20555

[13] A. Tellez-Arenas, "Best Practice Report on Geoportals", ECP-2007-GEO-317001, OneGeology-Europe, 2009

[14] D. Calvanese and G. D. Giacomo, "Data integration: a logic-based perspective", Artificial Intelligence Magazine 26 (1) , 59–70, 2005

[15] A. Sayar, "High Performance, Federated, Service-Oriented Geographic Information Systems", PhD Thesis, Department of Computer Science, Indiana University, 2009

[16] F. Hakimpour, "Using ontologies to resolve semantic heterogeneity for integrating spatial database schemata", Ph.D. Dissertation, Zurich University, Switzerland, 191pp., 2003

[17] A. Buccella, A. Cechich, and P. Fillottrani, "Ontology-driven geographic information integration: A survey of current approaches", Computers & Geosciences, 35 (4), 2009

[18] T. Gruber, "A translation approach to portable ontology specifications", Knowledge Acquisition 5 (2), 199–220, 1993

[19] W3C RDB2RDF Incubator Group, A Survey of Current Approaches for Mapping of Relational Databases to RDF, 2009, Available from: http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf [03 December 2014].

[20] W3C RDB2RDF Working Group, RDB2RDF, 2012, Available from: http://www.w3.org/2001/sw/wiki/RDB2RDF [03 December 2014]

[21] G. Meditskos and N. Bassiliades, "A combinatory framework of Web 2.0 mashup tools, OWL-S and UDDI", Expert Systems with Applications, 38, pp. 6657–6668, 2011

[22] Y. Tian, M. Huang, "Enhance discovery and retrieval of geospatial data using SOA and Semantic Web technologies", Expert Systems with Applications, Vol. 39 (16), pp. 12522–12535, Elsevier Ltd, 2012

[23] D. Medin, R. Goldstone and D. Gentner, "Respects for similarity", Psychological Review 100(2), 254-278, 1993

[24] K. Janowicz, C. Keßler, M. Schwarz, M. Wilkes, I. Panov, M. Espeter and B. Baeumer, "Algorithm, implementation and application of the SIM-DL similarity server", In F.T. Fonseca, A. Rodriguez and S. Levashkin (Eds), Proceedings of the Second International Conference on GeoSpatial Semantics (GeoS 2007). Berlin, Springer Lecture Notes in Computer Science Vol. 4853: 128–45, 2007

[25] A. Rodriguez and M. Egenhofer, "Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure" International Journal of Geographical Information Science, 18(3) 229-256, 2004

[26] A. Tversky, "Features of similarity", Psychological Review, 84(4) 327-352, 1977

[27] M. Raubal, "Formalizing conceptual spaces", In A. Varzi, L. Vieu (Eds.): Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004). Volume 114 of Frontiers in Articial Intelligence and Applications. IOS Press, Amsterdam, 153-164, 2004

[28] P. Gaerdenfors, "Conceptual Spaces - The Geometry of Thought", Bradford Books, MIT Press, Cambridge, MA, 2000

[29] A. Schwering and M. Raubal, "Spatial relations for semantic similarity measurement", In J. Akoka, S. Liddle, I.Y. Song, M. Bertolotto, I. Comyn-Wattiau, W.J. vanden Heuvel, M. Kolp, J. Trujillo, C. Kop and H. Mayr, (Eds.): Perspectives in Conceptual Modeling: ER 2005 CoMoGIS Workshop, Klagenfurt, Austria. Volume 3770 of Lecture Notes in Computer Science. Springer, Berlin, 259-269, 2005

[30] K. Janowicz and M. Raubal, "Affordance-based similarity measurement for entity types", In: 5th Conference on Spatial Information Theory (COSIT 2007). Lecture Notes in Computer Science, Springer, 2007

[31] A. Borgida, T. Walsh and H. Hirsh, "Towards measuring similarity in description logics", In: Proceedings of the 2005 International Workshop on Description Logics (DL2005). Volume 147 of CEUR Workshop Proceedings, CEUR, Edinburgh, Scotland, UK 2005

[32] C. d'Amato, N. Fanizzi and F. Esposito, "A dissimilarity measure for ALC concept descriptions", In: Proceedings of the 2006 ACM Symposium on Applied Computing (SAC), Dijon, France, 2006

[33] K. Janowicz, "Sim-dl: Towards a semantic similarity measurement theory for the description logic *ALCNR* in geographic information retrieval", In R. Meersman, Z. Tari, P. Herrero, al., e., (Eds.), SeBGIS 2006, OTM Workshops 2006. Volume 4278 of Lecture Notes in Computer Science. Springer, Berlin (2006)

[34] R. Navigli, "Word sense disambiguation: A survey", ACM Computing Surveys (CSUR) Surveys, Volume 41 Issue 2, Article No. 10, ACM New York, NY, USA, 2009

[35] T. Zesch, and I. Gurevich, "Wisdom of Crowds Versus Wisdom of Linguists: Measuring the Semantic Relatedness of Words", Natural Language Engineering, 16(01), 1351-3249, 2010

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

[37] R. Rada, H. Mili, E. Bicknell, and M. Blettner, "Development and Application of a Metric on Semantic Nets. Systems, Man and Cybernetics", IEEE Transactions, 19(1), 17-30, 1989

[38] C. Leacock and M. Chodorow, "Combining Local Context and WordNet Similarity for Word Sense Identification", In C. Fellbaum (Ed.), WordNet: An Electronic Lexical Database (pp. 305-332). MIT Press, 1998

[39] 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

[40] M. Porter, "An algorithm for suffix stripping", Program (Automated Library and Information Systems), 14(3):130-137, 1980

[41] Institute of Electrical and Electronics Engineers (IEEE), Suggested Upper Merged Ontology (SUMO), Available from: http://www.adampease.org/OP/ [03 December 2014]

[42] V. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals", Soviet Physics-Doklady, Vol. 10, No. 8, 707–710. Original in Russian in Dokl. Akad. Nauk SSSR 163, 4, 845–848, 1965.