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.
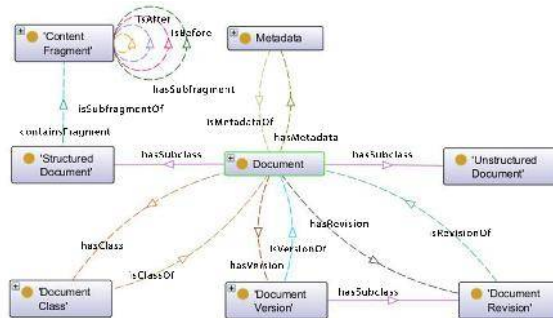

Figure 1. GDMO concepts

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.
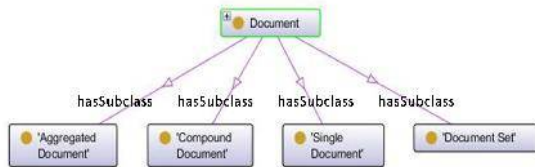

Figure 2. Document subclasses

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.
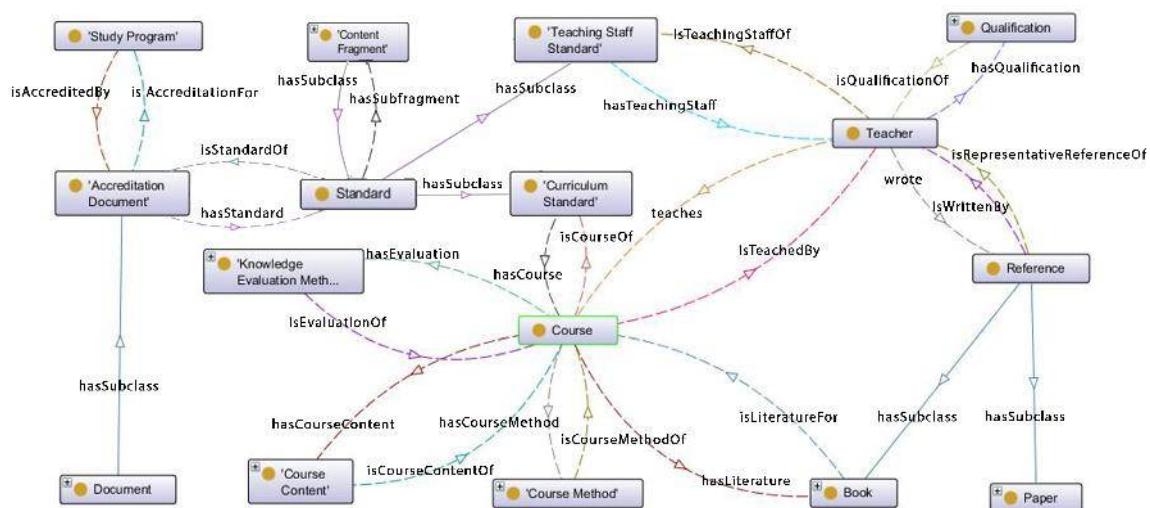
The graph shown in Figure 3 presents the proposed

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.

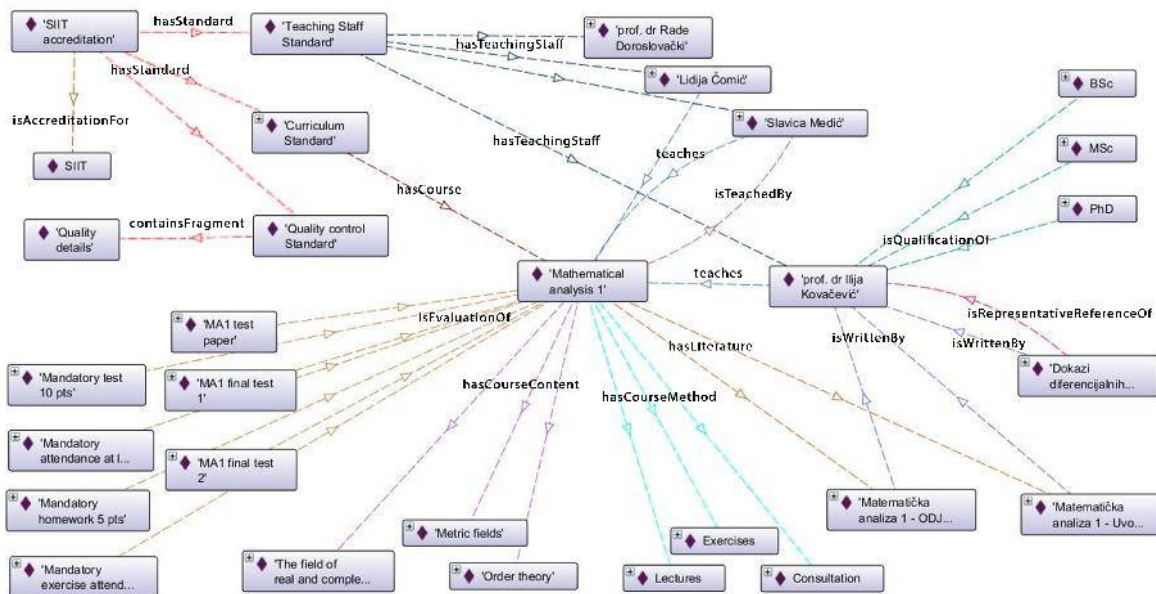Keeping in mind the complexity of the represented accreditation document, we have chosen a limited set of

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.

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