A Core Ontology for Brazilian Higher Education Institutions

Cleiton Silva¹ and Orlando Belo²

¹D.A.A., Federal Institute of Goiás, 75400-000 Inhumas-GO, Brazil
²Algoritmi R&D Centre – School of Engineering, University of Minho, 4710-057 Braga, Portugal

Keywords: Ontologies, Semantic Web, Web Science, Ontology Networks, Linked Data, Institutional Evaluation, Evaluation of Higher Education, SINAES.

Abstract: Within the past several years, Linked Open Data has reached a significant position in the Web, allowing for a significant increase of structured data in this domain. Although this, there are a lot to be done in order to ensure an effective way for supporting publication of connected open data. One of the strongest research and development line in this area appeals to the use of ontologies for structuring data and their relationships. Ontologies can be modeled for organizing knowledge about a domain, facilitating its sharing and reuse. As such, we have been developed an ontology especially oriented to support the organizational structure of higher education institutions in Brazil. This ontology was prepared for measuring and providing essential concepts and reference models that can be used by different stakeholders to develop and use ontologies and vocabularies for the National Higher Education Assessment System in Brazil. In this paper we present and discuss the most relevant aspects we approached during the conceptualization and development of the referred ontology.

1 INTRODUCTION

A Higher Education Institution (HEI) is a kind of organization that operates in the field of higher education. Like any other institution, a HEI exists for pursuing the objectives established in the different legal instruments, which in theory represent the expectations of a national or a supranational society that instituted them. Higher education in general, and in particular the results achieved by HEI, attracts the interest of different social actors (Bandeira et al., 2015; Dias Sobrinho, 2010; Firmino, 2013).

In recent years, especially after the approval of Law 12.527 of November 18, 2011 (Brasil, 2011), known as the Law on Access to Information (LAI), several documents and data related to the Brazilian higher education field are being published on the Web in different formats and without a reference conceptualization. Thus, the reuse of such resources for the production of new knowledge is quite limited, particularly when one wishes to process them using some kind of computational resource.

Problems such as these occur in different domains as more data are produced, published and reused on the Web. According to (O’Hara and Hall, 2011), the Web Science community has been studying and developed technologies in order to guarantee the sustainability of the Web.

The Semantic Web (Berners-Lee, 2000; Berners-Lee et al., 2006; Gruber, 2008; Sack et al., 2016; d’Aquin and Motta, 2016) is a set of technologies designed to make Web resources increasingly suitable for the consumption of people and machines. In (Silva and Belo, 2017) the authors highlight the main contributions of the Semantic Web to the implementation of the Data Web. In this work we are particularly interested in the technologies and tools used to add explicit semantics to datasets to be published on the Web.

Ontologies have been used in different applications domains, conferring explicit semantics and models for knowledge representation, both for the classic Web and for the new approaches where the Web is increasingly as a very promising data publishing platform (Silva and Belo, 2017). However, a careful literature review as well as the research carried out to date in the field did not identify any specific ontology developed with the objective of facilitating data interoperability in the domain of the National Higher Education Assessment System (SINAES).
Thus, we believe that is quite opportune and relevant to develop an initiative research where we could offer theoretical and empirical contributions, both in terms of knowledge advancement in ontology engineering and in terms of data treatment and exploration activities on the Web, as well as for the improvement of production processes and consumption of data in the field of higher education in Brazil.

In this paper, we present a proposal of an ontology especially oriented to support the organizational structure of HEI in Brazil. This ontology is being developed for measuring and providing essential concepts and reference models that can be used by different stakeholders to develop and use ontologies and vocabularies for the SINAES domain. The ontology proposed is integrated in a network of ontologies – OntoSINAES, which is under development in the course of a research project, in which we intend to implement an environment for the collaborative development of ontologies and vocabularies for the domain of the SINAES. Thus, following this brief introduction, in section 2, we present a literature review about ontologies in Web Science domain, highlighting some relevant works. In section 3, we present and discuss the ontology proposal, exposing its theoretical basis and the empirical results already achieved. Finally, in section 4, we present conclusions and point out some lines for future research and development.

2 RELATED WORK

The importance of the Web to today's society stems from advances in knowledge promoted by a community of researchers dedicated to making electronic information gradually less ambiguous for people and computers. This combination of efforts enhance the expression “Web Science”, which is used for representing a recent multidisciplinary area dedicated to advance the knowledge about the Web, both in terms of the technologies involved, and in terms of the various aspects involving its understanding as a specific phenomenon that affects different dimensions of the current society (O’Hara and Hall, 2011; Hendler et al., 2008; Hall and Tiropanis, 2012). In this work, we are interested in two different paradigms, namely: the Semantic Web and Linked Data (or Linked Open Data). The Semantic Web (Berners-Lee, 2000; Berners-Lee et al., 2001; Berners-Lee et al., 2006) has been proposed and is being developed with the aim to make the content of the Web more suitable for use both for people and computers. Ontologies occupy a central position in the conceptual model of the Semantic Web. In (Horrocks, 2002), Horrocks considers that ontologies can offer important contributions for a sustainable evolution of the Semantic Web, mainly by providing explicit semantics to the contents available on the Web, which favors their functioning from precisely defined sources of terms. Linked Data is a set of principles proposed by (Berners-Lee, 2006) for the publication of data on the Web. Linked Open Data is a movement within the World Wide Web Consortium (W3C) community, which aims to make the connected data available, free of charge. Combining these both ideas, we come to the Web paradigm as an open and connected data platform. Based on the main ideas of the article, we will use the terms “data connected” and “data open and connected” for referring Linked Data and Linked Open Data, respectively. According to the W3C (Anon, 2017), “Web of Data will become a reality when: (1) there is an enormous amount of data available, and (2) exists a standard format accessible and manageable by semantic Web tools.” In this sense, additionally to the Resource Description Framework (RDF), as a standard format, it is also necessary to ensure free access to data and to all the relationships that may exist among them, in order to have effectively open and connected data. In some perspectives, the RDF standard provides the basis for the publication of connected data, while ontologies and vocabularies provide the basis to formalize the different relationships between the data in a way that makes them quite useful. Many authors highlight the importance of ontologies, both in terms of guaranteeing higher levels of interoperability among data, and for providing semantic enrichment of knowledge bases.

The literature review has revealed several works (O’Leary, 2010; Reynolds, 2014a; Pereira and Almeida, 2014; Abramowicz et al., 2008) dealing with the use of ontologies for providing organizational conceptual models and in some cases especially oriented for the field of higher education (Falbo et al., 2014; Zemmouchi-Ghomari and Ghomari, 2013; Styles and Wallace, 2008; Pereira, 2015). The following is a brief summary of three papers, which deal specifically with organizational structures. For each of the cases described there, we try to point out an adequate interface to this work. The ontology published in (Reynolds, 2014) is a conceptual reference model for organizational structures recommended by the W3C, in order to support the publication of connected data. It should be noted that this organizational ontology was
designed to be generic, reusable and extended, and thus it is available for access in several formats, through the link https://www.w3.org/TR/vocabOrg/#organizational_structure. Considering our objective for developing a conceptual reference model for representing the organizational structure of Brazilian HEI, the ontology proposed by the W3C was used as a starting point for the development of the core ontology designed by us. Some authors emphasize that corporate ontologies are useful for many purposes, among which support the publication of connected data. According to (Falbo et al., 2014), an enterprise ontology should be: (1) flexible, to allow specialization in specific ontology projects; (2) broad, to the point of covering the whole application field; and (3) modular, so that only relevant fragments can be selected for each specific reality of the domain to be conceptualized. The authors of this work demonstrated their points of view through an ontology of the Brazilian Federal Universities.

Our work extends the domain described by the ontology proposed in (Falbo et al., 2014), but dealing also with the other types of HEI. Finally, we discovered in (Pereira, 2015) a proposal of an organizational reference ontology, specified in a meta-model and conceived in accordance with the ontological distinctions of the Unified Ontology Fundamental (UFO), which extends the social concepts of UFO-C (Guizzardi, 2005). This core ontology, called OntoUML Organizational Ontology (O3), was designed for using as a reference model in the definition of organizational structures. One of the objectives of O3 is to support the creation of domain ontologies through the specialization of its concepts and relations (Pereira, 2015). In this work – similar to what we are developing –, the author presents a model of the domain of the active structure of organizations. As so, the conceptual model of O3 has become an ontological resource quite relevant to the development of the ontology we present and discuss in this paper. According to what was possible to note in our literature review, we may conclude that the use of ontologies for representing organizational structures, besides being feasible, can offer relevant contributions for producing and consuming open and connected data on the Web.

3 THE ONTOLOGY

Ontologies have been the object of study of several works since the antiquity. In the last few years, many concepts of ontologies have been presented and defended by several authors within the Computer Science (Guizzardi, 2005; Gruber, 1993; Guarino, 1995; Noy and McGuinness, 2001; Almeida, 2013; Guarino, 1998). However, we can see an ontology as a logical theory that explains the intended meaning of a formal vocabulary, which is its ontological commitment to a particular conceptualization of the world (Guarino, 1998). Take this into consideration we develop the main theoretical foundations and a meta-model for the core ontology we present and discuss in this paper, with the objective of providing a minimum conceptualization of references for the organizational structure of Brazilian HEI. Core ontologies occupy an intermediate position between the superior (or foundation) ontologies and domain ontologies. In general, core ontologies rely on foundation ontologies to add real world semantics to conceptual models, avoiding ambiguities and making them more independent of the domain. The development of this core ontology is being based on Methodology for Building Ontology Networks (NeOn) (Suárez-Figueroa, 2010).

3.1 Requirements Specification

The core ontology we are presenting here aims to provide a semantic model to represent formally the IES organizational structure, covering all the general concepts referenced in the legislation that regulates the SINAES – as reference ontology, it must be formal and implemented in OWL DL version 2. The ontology was designed in a way that any intelligent agent interested in the production, use or reuse of information about higher education in Brazil could be use it, involving knowledge about the organizational structures of HEI. At this stage – requirements specification – we prepared and generated the correspondent Ontology Requirements Specification Document (ORSD) for the ontology. In this document we included a set of non-functional requirements (general aspects not related directly to the content of the ontology) and a set of functional requirements. The non-functional requirements were defined taking into account the motivation and the scope of this work. They are:

- RNF-01 – the ontology should be described and documented in Portuguese.
- RNF-02 – all the concepts related to the ontology will be classified according to the ontological categories defined by the Unified Foundational Ontology (UFO) (Guizzardi, 2005), (Guizzardi et al., 2008), using the OntoUML language for development, verification and validation of the model (Albuquerque and Guizzardi, 2013).
RNF-03 – the ontology should be implemented in OWL DL version 2;
RNF-04 – the ontology should formally integrate the OntoSINAES ontology network.

Table 1: Competence questions for the initial iteration.

<table>
<thead>
<tr>
<th>Competence Questions (CQ)</th>
<th>CQ Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQ-001: What are the Brazilian HEI classified as university institution?</td>
<td>Listing of all the institutions that act in the offer of higher education in Brazil, which have the legal prerogatives of university and university center.</td>
</tr>
<tr>
<td>CQ-002: What are the Brazilian HEI classified as university?</td>
<td>Listing of all the institutions that act in the offer of higher education in Brazil, which have the legal prerogatives of university.</td>
</tr>
<tr>
<td>CQ-003: What are the Brazilian HEI classified as university center?</td>
<td>Listing of all the institutions that act in the offer of higher education in Brazil, which have the legal prerogatives of university center.</td>
</tr>
<tr>
<td>CQ-004: What are the Brazilian HEI classified as non-university?</td>
<td>Listing of all the institutions that act in the offer of higher education, which according to the Brazilian legislation are considered as non-university HEI.</td>
</tr>
<tr>
<td>CQ-005: Who is the sponsor of the HEI?</td>
<td>An individual or legal entity that provides the necessary resources for the operation of the HEI.</td>
</tr>
<tr>
<td>CQ-006: Which are the regulatory acts that classify HEI as university, university center and non-university HEI?</td>
<td>Relation of regulatory acts of HEI accreditation and re-accreditation.</td>
</tr>
</tbody>
</table>

The functional requirements were specified based on a set of competency questions (CQ), formulated based on a thorough analysis of the concepts and relationships between them that were verified in legal texts regulating higher education in Brazil and directly related to the structure of HEI. The initial set of functional requirements was considered stable and adequate for the purposes of a first iteration of the core ontology lifecycle we designed. The initial CQ set was obtained essentially from the analysis of the Law 10.861, from April 14, 2004 (Brasil, 2004), which is known as the SINAES Law. Then all the CQ were grouped into several categories, each one representing different aspects of the organizational structure of HEI. In order to facilitate the understanding of the domain, these categories represent ontology modules that may guide the development of specific ontologies in the future. The set of functional requirements, in the form of CQ, was then submitted to a validation process, based on the verification of the criteria suggested by the NeOn methodology, namely concision, correction, consistency and lack of ambiguities. Domain experts attended the validation process. Finally, the requirements were prioritized in order to be satisfied in two iterations. In Table 1, we can find the CQ defined to be answered at the end of the first iteration, grouped into categories. The last step of the ontology specification process was the organization of a glossary of terms directly related to CQ, their responses and named entities.

### 3.2 Project Planning

The design of the ontology project was based on ORSD. At this stage we tried to organize in a timely manner the different processes and activities to be executed during the development process of the ontology. The lifecycle model chosen is based by an incremental iterative approach. Thus, this work takes as reference the execution of the first iteration, which included several scenarios of NeOn, namely the scenarios 1, 2 and 3 (Suárez-Figueroa, 2010, p.83).

To choose the life cycle model and the respective scenarios, we considered a specific set of CQ and non-functional requirements. Given the complexity of the domain of SINAES and the diversity of the organizational structures that are established among the Brazilian HEI, we decide for reusing non-ontological resources, which can be found in official information sources such as government websites, particularly the Brazilian Portal of Open Data (Brasil, 2017), and the laws and other official documents related to SINAES. As for the ontological resources, in this first iteration we had by reference the ontologies proposed in (Reynolds, 2014; Pereira, 2015). In Figure 1, presents a general overview of the main phases of the first iteration of the ontology development project life cycle.
The NeOn methodology prescribes for the life cycle based on an incremental iterative model, an initial phase that is oriented to the elaboration of a global development plan, containing a set of basic ontology requirements. Subsequently, during the initiation phase of each iteration, the detailed iteration planning, a review of the set of initial requirements, and the overall development plan are carried out. Each iteration can be regulated for one of the five versions that the cascade model provides in the NeOn methodology. For the first iteration of this project we chose the cascade model with six phases, taking into account the need for reengineering the ontological and non-ontological resources.

3.3 Conceptualization and Formalization

In these phases we produced the conceptual and formal models of the ontology. For the production and validation of the models, we used the OntoUML Lightweight Editor – OLED tool (Ufes, 2017). The OntoUML language is a profile of the Unified Modeling Language (UML), with an ontological basis based on the UFO ontology. The tool allows for constructing and validating the conceptual models in OntoUML, and generating automatically their implementations in RDF, OWL and UML.

In Figure 2, we can see a fragment of the conceptual model of the ontology we developed.

The fragment presented highlights some of the most essential concepts of the organizational structure and organizational roles, which together make up the abstract concept of a Brazilian HEI. The concept of HEI specializes the concept of organization (Firmino, 2013; Brasil, 2011). Like any other kind of organization, HEI are social agents explicitly instituted to develop higher education in Brazil. The HEI concept was specialized in institution and academic units. An institution is a type of formal organization, recognized by the external environment, while an academic unit is an organizational unit, recognized in the internal context of an institution. Academic units represent the working groups of an institution. The model also indicates that an HEI is accredited at the moment of its constitution, as a university, university center or as a non-university. At the time of its existence, regulatory acts will determine the maintenance or change of its classification.

3.4 COrg Ontology: Implementation

The portion of knowledge of the domain of SINAES that in the previous stage was organized in a formal and well-founded conceptual model, in this phase was transformed into a computable model, using the Protégé tool (Musen and Team, 2015). Based on the ORSD, the ontology language used was the OWL DL version 2.

Because it is a core ontology, the relationships between concepts and formal properties in the conceptual model were enriched with new constraints and axioms, aiming to increase the ontology's expressiveness and inference capacity. The ontology implemented in this work is the first version of the Core Ontology of the HEI organizational structure in Brazil (COrg).

In Figure 3, we present the main classes that are already part of the ontology.
The COrg ontology is an extension of the ontology of organizations (ORG) (Reynolds, 2014b) recommended by the W3C. The ontology ORG has a reference taxonomy, on which there already exists a consensus established by the use in different projects and domains.

The reuse of ontological resources is a good practice already consolidated by the ontology engineering community. In this sense, during the last iteration, the "pruning" of the ontology will be performed, aiming at the removal of elements considered irrelevant for the scope of COrg. This activity aims to make the resulting ontology suitable for reuse, extension or specialization in other projects and applications.

4 CONCLUSION AND FUTURE WORK

The environment of an organization usually involves several activity domains, some geographic dispersion and different social actors. It is increasingly challenging to organize the knowledge necessary for the healthy and productive development of activities in different organizational contexts.

However, the scientific literature demonstrates that corporate ontologies can be very useful for such purposes. Additionally, the literature also shows that large ontologies, particularly when involving multiple domains, can be developed as a network of ontologies containing different types of ontologies smaller, simpler, and mainly more suited to more specific and dynamic requirements. It was also verified that a network of ontologies it will better developed through intense collaboration, especially among general stakeholders, domain experts and ontology engineers. In this case, the "added value" of collaboration would be a conceptualization of consensus, from which new concepts could be formalized and reused.

Therefore, in our opinion a network of ontologies in which one intends to contribute to the organization of corporate knowledge should provide a "beginning point", a set of well-founded general conceptual categories, organized in networks of core ontologies or ontology patterns. The development of the ontology has demonstrated that the identification of the fundamental concepts must be gradual and very well founded, so that they can effectively be specialized in different ontologies and preserving the conceptual reference that is necessary to guarantee the semantic interoperability of the data. Regarding the conceptualization aspects for the development of the ontology, the legal framework was the main source of conceptual categories, by the fact that being in force the laws can be considered as social "consensuses". The analysis of the legislation was also relevant to ensure the existence of conceptual categories that covered the entire domain of SINAES.

It is our opinion that in projects developed for contexts similar to SINAES, the laws are unavoidable non-ontological resources, particularly at the beginning of projects.

As future work, we intend to integrate COrg into the network of ontologies we are developing within the OntoSINAES project. The project also foresees the implementation of a Web environment to support the collaborative development of knowledge representation models for the SINAES domain.

ACKNOWLEDGEMENTS

This work has been supported by COMPETE: POCI-01-0145-FEDER-007043, by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013 and by CAPES Foundation, Ministry of Education of Brazil.

REFERENCE


Anon, 2017. Data - W3C. Available at: https://www.w3.org/standards/semanticweb/data [Accessed January 18, 2017].


Bermers-Lee, T., 2006. Linked Data - w3.org, p.https://www.w3.org/DesignIssues/LinkedData.html. Available at: