Aligning Top-level and Domain Ontologies - Expected date for defense: March/2019

Daniela Schmidt*, Cassia Trojahn†, Renata Vieira*

[1*]Pontifical Catholic University of Rio Grande do Sul (Brazil)
daniela.schmidt@acad.pucrs.br, renata.vieira@pucrs.br
† Institut de Recherche en Informatique de Toulouse (France)
cassia.trojahn@irit.fr

Abstract. *Many efforts in the ontology matching field have been particularly dedicated to domain ontologies, however the problem of matching domain and top-level ontologies has been addressed to a lesser extent, particularly due to their different levels of abstraction. This work aims at filling the gap in this area. We intend to propose an approach to align top-level and domain ontologies. The use of general lexical databases as an intermediary layer is a direction.*

1. Introduction

Ontologies have been applied in many areas motivated mainly by the need to create, to share and to reuse knowledge. The rich semantics and formalization of top-level ontologies are important requirements for ontology design [Mika et al. 2004], they act as well as semantic bridges supporting very broad semantic interoperability between ontologies [Mascardi et al. 2007, Mascardi et al. 2010]. In that sense, they play a key role in ontology matching, which is the process of finding correspondences between entities from different ontologies.

The advantage of top-level ontologies is to gather lots of available knowledge and create super structures for information that provide interoperability for many applications. However, most efforts in ontology matching have been particularly dedicated to domain ontologies and the problem of matching domain and top-level ontologies has been addressed to a lesser extent. This problem poses different challenges in this field, particularlly due to the different levels of abstraction of these ontologies. This is a complex task, that requires knowledge about the semantic context of concepts, which goes beyond the frontiers of what is encoded in the ontology.

In particular the differences in the abstraction levels of domain and foundational ontologies will require a change of focus from finding equivalence relation to the identification of subsumption relations. In fact, when having different levels of abstraction it might be the case that the matching process should focus in finding subsumption rather than equivalence correspondences, since the top-level ontology has concepts at a higher level. This is largely neglected by most matching systems. Approaches dealing with this task are mostly based on manual matching [Brodaric and Probst 2008, Mika et al. 2004].

In order to evaluate the quality and correctness of the generated alignments in the process of top-level and domain ontology matching, reference alignments (also called gold standard) are required. Reference alignments could be developed manually by experts or in a semi-automatic way, where the resultant alignment from matching systems is used against the manual analysis.

Considering the discussion above, this thesis aims to contribute to the problem of matching domain and top-level ontologies. We are proposing an approach to automatically align domain and top-level ontologies. This paper summarizes the thesis proposal which will be estimated to be developed till march/2019.

The remaining of this paper is organized as follows. Section 2 introduces the theoretical background on top-level ontology and ontology alignment. Then, we discuss on available state-of-the-art matching systems and the lexical database WordNet. Section 3 presents the related work. Section 4 describes our initial experiments. Section 5 presents our thesis proposal including research hypothesis, and research goals. Section 6 concludes this paper.

2. Background

2.1. Top-level ontologies

A top-level ontology is a high-level and domain independent ontology. The concepts expressed are intended to be basic and universal to ensure generality and expressivity for a wide range of domains. It is often characterized as representing common sense concepts and concerns concepts which are meta, generic, abstract and philosophical. Some examples of well known top-level ontologies are BFO [Grenon et al. 2004], DOLCE [Gangemi et al. 2002], GFO [Herre et al. 2007], SUMO [Niles and Pease 2001] and UFO [Guizzardi 2005]. A review of them is presented in [Mascardi et al. 2007].

2.2. Ontology matching

The process of finding correspondences between ontology entities is known as ontology matching. It takes as input two ontologies o_s (source) and o_t (target) and an (possibly empty) alignment A to be completed, and determines as output an alignment A', i.e., a set of correspondences. Here, we borrow the definition of correspondence from [Euzenat and Shvaiko 2007]:

Definition 1 (Correspondence) *A correspondence can be defined as $<e_s,e_t,r,n>$, such that: e_s and e_t are entities (e.g., concepts, properties, instances) of o_s and o_t, respectively; r is a relation holding between two entities e_s and e_t, (for instance,* equivalence*,* subsumption*,* disjointness*,* overlapping*); and n is a confidence measure number in the [0;1] range. The confidence assigns a degree of trust on the correspondence from the matcher.*

2.3. WordNet

WordNet [Miller 1995] is a general-purpose large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. A synset denotes a concept or a sense of a group of terms. WordNet also provides textual descriptions of the concepts (gloss) containing definitions and examples. For instance, for the concept "Poster", one of the associated WordNet synsets (SID-06793426-N) groups the synonyms "poster, posting, placard, notice, bill, card", together with a gloss "a sign posted in a public place as an advertisement; a poster advertised the coming attractions".

3. Related Work

In the literature it is possible to identify different uses of top-level ontologies. In the ontology matching field, top-level ontologies could be seen as a resource to obtain or improve the alignment between domain ontologies. In this way, [Padilha et al. 2012] propose an approach to explore alignment patterns based on the Unified Foundational Ontology (UFO). In [Mascardi et al. 2010], a set of algorithms is developed to exploit top-level ontologies as semantic bridges to solve heterogeneity problems of domain ontology alignments. One algorithm that helps the developer to choose a more suitable top-level ontology for use together with domain ontologies is presented in [Khan and Keet 2012]. The top-level ontologies which are able to be recommended are DOLCE, BFO, GFO, and SUMO. The same authors present a repository called ROMULUS with the aim of improving semantic interoperability of top-level ontologies [Khan and Keet 2013]. In ROMULUS, there are alignments among three top-level ontologies (DOLCE, BFO and GFO) with each other manually and using automatic matching tools [Khan and Keet 2013]

We are particularly interested in aligning top-level and domain ontologies. Therefore, we investigated the existing alignments regarding top-level ontologies and external resources such as WordNet, once, background knowledge from external resources such as WordNet has been largely exploited in matching domain ontologies, as a way for improving similarity measures [Lin and Sandkuhl 2008]. In this way, Gangemi et al. [Gangemi et al. 2002] present an effort to align WordNet nouns with DOLCE top-level ontology. More recently Silva et al. [Silva et al. 2016] present an extension of the previous alignment between DOLCE and WordNet nouns [Gangemi et al. 2002] to include verbs. In the same way, it is possible to see alignments between WordNet and other top-level ontologies as BFO [Seppälä 2015], Cyc [Reed and Lenat 2002], and SUMO [Niles and Pease 2003].

During the investigation, we observe the importance of using top-level ontologies to aggregate semantics and reduce heterogeneity problems between ontologies. Moreover, we can observe that lexical databases such as WordNet are used in the ontology alignment task as an external resource to identify domain concepts correspondence. However, we identify less efforts to automatize the process of alignment. In fact, we do not identify an application developed specifically to align domain and top-level ontologies. The available matching systems were developed to align ontologies in a same domain and there is no evaluation of them in the task of align domain and top-level ontologies. Therefore, we are proposing a matching approach and a reference alignment for its evaluation. We start our investigation evaluating the output of available matching systems [Schmidt et al. 2016a]. We also investigate the use of WordNet as a validator for the alignments [Schmidt et al. 2016b]. These studies are described in the next section.

4. Initial Experiments

In order to define our proposal, we develop two initial experiments towards the problems we will address in our thesis. First, we made an analysis of some available matching systems participating in previous OAEI campaigns. We use these matchers to align domain and top-level ontologies. Even though they were not developed specifically for that purpose, they are the currently available tools. Also they present many different approaches for the alignment problem and their output might help us to investigate the problem. This

study has been published in the International Workshop on Ontology Matching (OM-2016) [Schmidt et al. 2016a] and discusses the challenges observed in the task of align domain and top-level ontologies.

In a second analysis, we used WordNet to automatically validate the generated alignments from the first study. This study was concerned with the automatic validation of candidate alignments between top-level and domain ontologies, exploiting WordNet background knowledge and the notion of *context*. We apply our approach for validating alignments generated by the matching tools and discuss the results against the manual validation presented in [Schmidt et al. 2016a]. This study has been published in the Brazilian Ontology Research Seminar (ONTOBRAS-2016) [Schmidt et al. 2016b] and discusses the obtained results and benefits to adopt external resources to validate generated correspondences.

5. Research hypothesis, Goals and Methodology

Considering the depicted discussed challenges of aligning top-level and domain ontologies, we define a research hypothesis which will be a guide of our research work as follow: The lexical database WordNet could be a way to improve and obtain gains in relation to the state-of-the-art matching systems in the task of align general domain and top-level ontologies, once, there are previous alignments between top-level ontologies and WordNet. Also, WordNet is a source of knowledge regarding subsumption relation between terms, which is essential for this kind of alignment.

We aim at proposing an approach for automatic ontology alignment between top-level and domain ontologies. To explore this field, we developed some preliminary studies which made it possible to delimit the scope of our work. In the experiments, we adopted the conference domain which is one of the participant domains of OAEI evaluation campaigns. For this domain, there is no reference alignments involving top-level ontologies. Reference alignments are important sources to compare and evaluate automatic matching approaches. Hence, we intend to build a reference alignment between ontologies of the conference domain and top-level ontologies.

Considering our preliminary studies, we intend to adopt the previous alignments between the lexical database WordNet and some well known top-level ontologies such as DOLCE and SUMO as an intermediary step to align domain ontologies with some of the top-level ontologies. The option of the WordNet regarding the other databases is because there are previous alignments of this base with several top-level ontologies. Moreover, WordNet presents common sense terms so it can help us in the identification of correspondences between domain and top-level concepts, where the domain if of a more general kind such as the conferences domain.

Our approach consists in, given a domain ontology as input, identifying the most appropriate WordNet synset for each domain concept. The synset identification is based on the concept context. A context is constructed from all information available about an ontology entity, including entity naming (ID), annotation properties (usually labels and comments) and information on the neighbors (super and sub-concepts). Next, the selected synset has to be found in the alignment between WordNet and top-level ontology. Since the synset is previously mapped to the top-level ontology, we will track the top-level concept to map them with the domain concept.

Hence, our process could be divided in four main steps: (i) extraction of the ontology concepts; (ii) processing of the context related to each concept; (iii) use of the context to retrieve the appropriate synset in WordNet; and (iv) the track of the top-level concept based on the alignment of the WordNet synset to top-level ontology concept.

We intend to evaluate our approach through a prototype which can be compared with other available matching tools. We are aware that they are not developed specifically to align domain and top-level ontologies, however that is currently the only comparison possible. In addition, we intend to develop a reference alignment which will allow us to evaluate our approach in terms of precision recall and F-measure. Hence, in the end of our work, we intend to contribute with *(i)* the creation of a reference alignment for domain and top-level ontologies; The chosen domain is conferences adopted in our previous experiments regarding DOLCE and SUMO top ontologies which we intend to adopt in our work. *(ii)* an approach to align top-level and domain ontologies automatically; and *(iii)* a prototype implementing our alignment approach and its evaluation.

6. Final Remarks

In this paper, we present a summary of our thesis proposal. We described the theoretical background adopted as a basis to our work and current related work. We developed some initial analysis which were useful to identify the current situation of available tools in the task of align top-level and domain ontologies. In the same way, we investigated the use of external resources as a way to validate the resulting alignments. Hence, we consider that the lexical database WordNet could be a useful external resource in the task of aligning top-level and domain ontologies, specially because there are alignments between some of the most common top-level ontologies and WordNet.

We are aware of some limitations regarding the use of WordNet, since it usually lacks very specialized terms and it does not contain compound terms. Also, regarding the previous alignment involving WordNet and top-level ontologies we are limited to the version of WordNet used in the alignment and in some top-level ontologies, the alignment could be not complete and cover only nouns, or other specific group of terms. Besides these limitations, considered the lack of options for dealing with this problem and, on the other hand, alignments already given for WorNet and top ontologies, we believe that such proposal may improve the current status in this research area.

References

Brodaric, B. and Probst, F. (2008). DOLCE ROCKS: Integrating Geoscience Ontologies with DOLCE. In *Semantic Scientific Knowledge Integration*, pages 3–8.

Euzenat, J. and Shvaiko, P. (2007). *Ontology Matching*. Springer-Verlag, 1st edition.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. (2002). Sweetening Ontologies with DOLCE. In *13th Conf. on Knowledge Engineering and Knowledge Management*, pages 166–181.

Grenon, P., Smith, B., and Goldberg, L. (2004). Biodynamic Ontology: Applying BFO in the Biomedical Domain. In *Stud. Health Technol. Inform.*

Guizzardi, G. (2005). *Ontological foundations for structural conceptual models*. PhD thesis, University of Twente, Enschede, The Netherlands, Enschede.

Herre, H., Heller, B., Burek, P., Hoehndorf, R., Loebe, F., and Michalek, H. (2007). General Formal Ontology (GFO): A Foundational Ontology Integrating Objects and Processes. In *Basic Principles, Research Group Ontologies in Medicine)*.

Khan, Z. C. and Keet, C. M. (2012). ONSET: Automated Foundational Ontology Selection and Explanation. In *Proc. of the 18th Intern. Conf. on Knowledge Engineering and Knowledge Management*, pages 237–251.

Khan, Z. C. and Keet, C. M. (2013). The Foundational Ontology Library ROMULUS. In *Proc. of. the 3rd Intern. Conf. on Model and Data Engineering*, pages 200–211.

Lin, F. and Sandkuhl, K. (2008). A Survey of Exploiting WordNet in Ontology Matching. In *IFIP 20th World Computer Congress*, pages 341–350.

Mascardi, V., Cordì, V., and Rosso, P. (2007). A Comparison of Upper Ontologies. In *8th AI*IA/TABOO Joint Workshop on Agents and Industry*, pages 55–64.

Mascardi, V., Locoro, A., and Rosso, P. (2010). Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation. *IEEE Trans. on Knowl. and Data Eng.*, 22(5):609–623.

Mika, P., Oberle, D., Gangemi, A., and Sabou, M. (2004). Foundations for Service Ontologies: Aligning OWL-S to Dolce. In *13th Conf. on World Wide Web*, pages 563–572.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Commun. ACM*, 38(11):39–41.

Niles, I. and Pease, A. (2001). Towards a Standard Upper Ontology. In *Conf. on Formal Ontology in Information Systems*, pages 2–9.

Niles, I. and Pease, A. (2003). Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proc. of the Intern. Conf. on Information and Knowledge Engineering*, pages 412–416.

Padilha, N. F., Baião, F., and Revoredo, K. (2012). Alignment Patterns based on Unified Foundational Ontology. In *Proc. of the Brazilian Ontology Research Seminar*, pages 48–59.

Reed, S. and Lenat, D. (2002). Mapping Ontologies into Cyc. In *Proc. of the Workshop on Ontologies For The Semantic Web*, pages 1–6.

Schmidt, D., Trojahn, C., and Vieira, R. (2016a). Analysing Top-level and Domain Ontology Alignments from Matching Systems. In *Proc. of the 11th Intern. Workshop on Ontology Matching*, pages 1–12.

Schmidt, D., Trojahn, C., Vieira, R., and Kamel, M. (2016b). Validating Top-level and Domain Ontology Alignments using WordNet. In *Proc. of the Brazilian Ontology Research Seminar*, pages 1–12.

Seppälä, S. (2015). Mapping WordNet to Basic Formal Ontology using the KYOTO ontology. In *Proc. of the Intern. Conf. on Biomedical Ontology*, pages 1–2.

Silva, V. S., Freitas, A., and Handschuh, S. (2016). *Word Tagging with Foundational Ontology Classes: Extending the WordNet-DOLCE Mapping to Verbs*, pages 593–605. Springer International Publishing.