

A Framework for Multilingual Ontology Mapping

Cássia Trojahn¹, Paulo Quaresma¹, Renata Vieira²

¹Universidade de Évora, ²Pontifícia Universidade Católica do Rio Grande do Sul
¹Rua Romão Ramalho 59 Évora Portugal, ²Av. Ipiranga, 6681, Porto Alegre, Brazil
cassia@di.uevora.pt, pq@di.uevora.pt, renata.vieira@puers.br

Abstract

In the field of ontology mapping, multilingual ontology mapping is an issue that is not well explored. This paper proposes a framework for mapping of multilingual Description Logics (DL) ontologies. First, the DL source ontology is translated to the target ontology language, using a lexical database or a dictionary, generating a DL translated ontology. The target and the translated ontologies are then used as input for the mapping process. The mappings are computed by specialized agents using different mapping approaches. Next, these agents use argumentation to exchange their local results, in order to agree on the obtained mappings. Based on their preferences and confidence of the arguments, the agents compute their preferred mapping sets. The arguments in such preferred sets are viewed as the set of globally acceptable arguments. A DL mapping ontology is generated as result of the mapping process. In this paper we focus on the process of generating the DL translated ontology.

1. Introduction

Open and dynamic systems, such as the Web and its extension, the Semantic Web, are by nature distributed and heterogeneous. Such characteristics implicate that the ontologies used to describe content and services can be represented using different formats and, more specifically, different languages. In this scenario, multilingual ontologies mapping is required. The mapping result can be used for ontology merging, agent communication, query answering, or for navigation on the Semantic Web. Examples of practical applications are distributed question answering systems (Pazienza et al., 2005) and comparison of ontologies describing legislations from different countries (Boer et al., 2003).

There are a lot of proposals to ontology mapping. Well-known approaches to the problem can be grouped into lexical, semantic, and structural ones, as terms may be mapped by a measure of lexical similarity, by semantic evaluation, usually on the basis of semantic oriented linguistic resources, or by considering the term positions in the ontology hierarchy. It is assumed that the approaches are complementary to each other and combining different ones reflect better solutions when compared to the solutions of the individual approaches. However, multilingual ontology mapping is an issue that is not well explored.

This paper proposes a framework for mapping of multilingual Description Logics (DL) ontologies. First, the *DL source ontology* is translated to the target ontology language, using a lexical database or a dictionary, generating a *DL translated ontology*. The *target* and the *translated* ontologies are then used as input for the mapping process. A *DL mapping ontology* is generated as result of this process. In this paper we focus on the process of generating the *DL translated ontology*.

The mappings are computed by specialized agents using different mapping approaches (lexical, semantic and structural). Next, these agents use argumentation to exchange their local results, in order to agree on the obtained mappings. Based on their preferences and confidence of the ar-

guments, the agents compute their preferred mapping sets. The arguments in such preferred sets are viewed as the set of globally acceptable arguments. See (Trojahn et al., 2008) for details of the argumentation process.

This paper is structured as follows. Section 2 comments on OWL (Ontology Web Language) Description Logics, the formalism used to represent our ontologies. Section 3 presents our architecture for multilingual ontology mapping. Section 4 presents an illustrative example. Section 5 comments on related work. Finally, section 6 presents the final remarks and the future work.

2. OWL-Description Logic

Description Logics (DLs) is the name for a family of knowledge representation (KR) formalisms that represent the knowledge of an application domain (“the world”) by first defining the relevant concepts of the domain (its terminology), and then using these concepts to specify properties of objects and individuals occurring in the domain (the world description) (Baader et al., 2003). A DL knowledge base (KB) comprises two components: TBox and ABox. The TBox contains the *terminology*, which specifies the vocabulary of an application domain. The ABox contains assertions about named individuals in terms of the TBox. The vocabulary consists of *concepts* and *roles*. *Concepts* denote set of individuals while *roles* denote binary relationship between individuals. Atomic concepts and roles can be used to build complex description of concepts and roles, using *constructors*. The language for building descriptions is a feature of different DLs, and different systems are distinguished by their description languages, i.e., the expressiveness of the language according with the *constructors* that they support.

The OWL-DL ontology language is a variant of the SHOIN(D) (Horrocks and Patel-Schneider, 2003) Description Logic, which provides constructors for full negation, disjunction, a restricted form of existential quantification, and reasoning with concrete datatypes. OWL-DL is the state-of-the-art to represent expressive ontologies.

The set of SHOIN(D) concepts is defined by the following syntactic rules, where A is an atomic concept, R is a role name, d is a concrete domain, c_i are individuals, and n is a non-negative integer:

$$\begin{aligned} C &\rightarrow A \mid \neg C \mid C_1 \sqcap C_2 \mid C_1 \sqcup C_2 \mid \exists R.C \mid \forall R.C \mid \\ &\quad n S \mid n S \mid \{a_1, \dots, a_n\} \mid n T \mid n T \mid \\ &\quad \exists T_1, \dots, T_n.D \mid \forall T_1, \dots, T_n.D \\ D &\rightarrow d \mid \{c_1, \dots, c_n\} \end{aligned}$$

We consider the semantic of a SHOIN(D) knowledge base KB by the mapping π proposed by (Haase and Motik, 2005) which transform KB axioms into a first-order formula. Each atomic concept is mapped into a unary predicate and each role is mapped into a binary predicate.

3. Architecture for Multilingual Ontology Mapping

Here, we focus in the process of translating ontologies in the context of multilingual ontology mapping. Figure 1 shows the proposed architecture. Initially, the agents interested in the mapping (*source* and *target* agents) send their ontologies to the *mediator agent*, which sends the ontologies to the *translation agent*. The *translation agent* is responsible for translating the source ontology into a *translated DL ontology*, in the target language, using a *dictionary*. The *translated*, target, and source ontologies are then sent to the *mapper agents*. The idea is to make available the *mediator agent* as a web service, in order to the translation process to be transparent to the *source* and *target* agents.

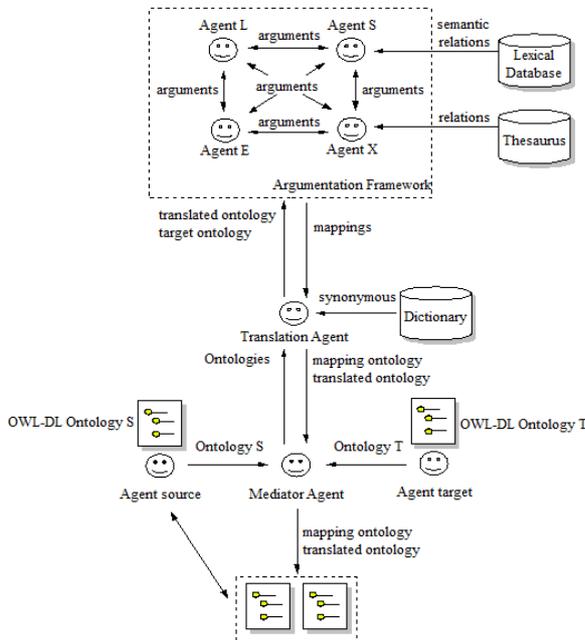


Figure 1: Architecture for multilingual ontology mapping.

In an initial setting of our framework, we have three *mapper agents*: lexical (L), semantic (S), and structural (E). However, our architecture is extensible for other kinds of agents. For instance, agents based on the similarity of instances, or agents based on the similarity of classes attributes, or

even other semantic agents relying on different semantic database or thesauri.

The *lexical agent* (L) adopts a metric to compare string similarity, while the *semantic agent* (S) considers the semantic relations (i.e., synonym, hyponym, and hypernym) between concepts to measure the similarity between them, on the basis of a lexical database. The *structural agent* (E) considers the positions of the terms in the ontology hierarchy to verify if the terms can be mapped. An approach based on argumentation is used in order to the agents agree on the obtained mappings. An Extended Value-based Argumentation Framework (E-VAF) is used to represent arguments with *confidence degrees* (Trojahn et al., 2008). The E-VAF allows to determine which arguments are acceptable, with respect to the different *audiences* represented by different agents. To each argument is associated a *confidence degree*, representing how confident an agent is in the similarity of two ontology terms. Based on their preferences and confidence of the arguments, the agents compute their preferred mapping sets. The arguments in such preferred sets are viewed as the set of globally acceptable arguments. This set is assumed to be the result of the mapping process. The final mappings are then sent to the *translation agent*, which sends the *mapping* and the *translated ontologies* to the *mediator agent*. Finally, these ontologies are used to the *source agent*.

In the following we describe in detail the *translation agent*.

3.1. Translation Agent

The *translation agent* uses a dictionary (or a lexical database) to obtain the set of synonyms terms of each term from the source ontology. Each term is then transformed into a *concept description* in the *translated ontology*. The set of synonyms is used to compose the *concept description* of the corresponding term.

Formally, consider that T_{source} is the set of terms of the source ontology, in the source language; T_{trans} is the set of terms of the *translated ontology*. For each term $t_{source,i} \in T_{source}$, a set of synonyms T_{syn} (i.e., $\{t_{syn,1}, \dots, t_{syn,n}\}$) is generated in the target language. The term $t_{source,i}$ is then converted into a *concept description* $t_{trans,i} \equiv t_{syn,1} \sqcup \dots \sqcup t_{syn,n}$.

The number of terms of the *translated ontology* is equivalent to the sum of the terms in the source ontology with each set of synonyms of each term. Formally, $|T_{trans}| = \sum_{i=1, \dots, n} |T_{source,i}| \sum_{j=1, \dots, m} |T_{syn,j}|$, where $n = |T_{source}|$, and $m = |T_{syn}|$, for the term i .

Moreover, we can distinguish two kinds of terms: *single* (i.e., “Thesis”) and *compound* (i.e., “Master-Thesis”). The process of translating single terms is made according to the description above. A strategy to reduce the number of terms in the synonyms sets is to verify the terms which are super-classes of the compared terms, including these terms in the search. For instance, if “Publication” is the super-class of “Thesis”, the term “Publication” is used in the search of synonyms of the term “Thesis”. We call this search a *contextualized search*.

In the second kind, compound terms, first it is verified if the composite term is a synonyms entry in the dictionary. In this case, the process is made as commented above. Other-

wise, the terms are searched separately (i.e., the compound term is decomposed into single terms, such as a “tokenization” process). The *concept description* of a compound term is formed by the disjunction of the set of synonyms of each single term. For each synonyms set is generated a conjunction of the synonyms terms, as described above for the single terms. For instance, consider the Portuguese term “Tese-Mestrado” (“Master-Thesis” in English). First it is verified if the compound term is a direct entry in the dictionary. Considering that case, a description Tese-Mestrado \equiv (Master-Thesis \sqcup Thesis \sqcup ... \sqcup $t_{syn,n}$) could be generated. Otherwise, the terms are searched separately: “Tese” and “Mestrado”. In this case, “Tese” \equiv (Thesis \sqcup Dissertation \sqcup ... \sqcup $t_{syn,n}$), and “Mestrado” \equiv (Master \sqcap Master-Thesis \sqcap ... \sqcap $t_{syn,n}$). Then, the final description of “Tese-Mestrado” is the disjunction of the synonyms set of “Tese” and “Mestrado”: “Tese-Mestrado” \equiv (Thesis \sqcup Dissertation \sqcup ... \sqcup $t_{syn,n}$) \sqcap \equiv (Master \sqcap Master-Thesis \sqcap ... \sqcap $t_{syn,n}$).

4. An Illustrative Example

Consider the ontologies shown in Figures 2 and 3, in English and Portuguese languages, respectively. We exemplified as the terms “Topico” in the source ontology is mapped to the term “Subject” in the target ontology. These terms represent the research areas related to a publication.



Figure 2: Portuguese source ontology.



Figure 3: English target ontology.

First, the source term “Topico” is searched in a Portuguese-English lexical dictionary, and the set of synonyms {Subject, Topic, Matter, Issue} is generated. Each element from this set is added as a concept in the *translated ontology* (Figure 4). The term “Topico” is then described in terms of its synonyms set: “Topico” \equiv Subject \sqcup Topic \sqcup Matter \sqcup

Issue (Figure 4). The target (Figure 3) and the *translated* (Figure 4) ontologies are then used as input to the mapping process.

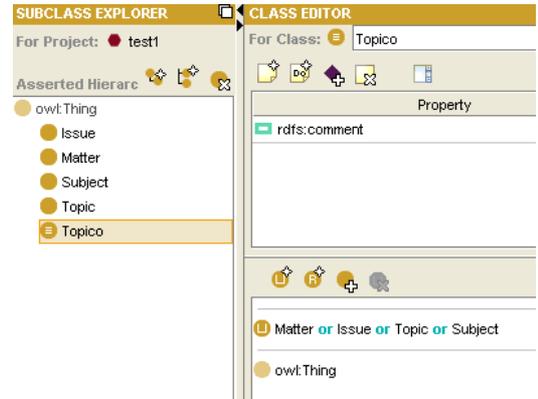


Figure 4: Translated ontology.

In this specific example, we consider the three mapper agents lexical (L), semantic (S) and structural (E). The lexical agent adopts the *lexical similarity* proposed by (Maedche and Staab, 2002). The semantic agent consider the semantic (i.e., synonym, hyponym, and hypernym) relations between concepts to measure the similarity between them, on the basis of WordNet¹ 2.1 database. The structural agent verified if the super-classes of the compared terms are lexically similar. If not, the semantic similarity between they is used. If the super-classes of the terms are lexically or semantically similar, the terms can be matched.

The agents are able to map the primary concepts from the *translated ontology* with the terms from the *target ontology*. In the example, “Topico” is not a primary concept, but a complex concept formed by the a conjunction of primary concepts. The primary concepts are “Issue”, “Matter”, “Subject”, and “Topic”.

For each pair (t_{trans}, t_{target}) the agents apply their algorithms and return two parameters (h, c): h , indicating that the mapping holds or not (+,-), and the *confidence degree* (c) of the mapping, which is related with the similarity of the terms, considering composite terms. In our argumentation framework, arguments with greater *confidence degree* are preferred. Here, we consider only the mapping between the primary terms and “Subject_{target}”. Table 1) shows the mappings.

Table 1: Mappings.

t_{trans}	t_{target}	L	S	E
Issue _{trans}	Subject _{target}	(-,0)	(+,1)	(+,0)
Matter _{trans}	Subject _{target}	(-,0)	(+,1)	(+,0)
Subject _{trans}	Subject _{target}	(+,1)	(+,1)	(+,0)
Topic _{trans}	Subject _{target}	(-,0)	(+,1)	(+,0)

From argumentation, the agents agree on the following arguments: Issue_{trans} \equiv Subject_{target} (+,1); Matter_{trans} \equiv Subject_{target} (+,1); Subject_{trans} \equiv Subject_{target} (+,1); and Topic_{trans} \equiv Subject_{target} (+,1).

¹<http://www.wordnet.princeton.edu>

Observing the case of a specific mapping, $\text{Subject}_{trans} \equiv \text{Subject}_{target}$, we can observe that Subject_{trans} is into the description of Topico_{trans} . This way, Subject_{target} is equivalent to the term Topico_{source} . Formally, “ $\text{Subject}_{target} \equiv$ ” “ Subject_{trans} ” and “ $\text{Subject}_{trans} \in$ ” “ Topico_{trans} ”(Subject \sqcup Topic \sqcup Matter \sqcup Issue), which is the translation of “ Topico_{source} ”. Then, “ Topico_{source} ” \equiv “ Subject_{target} ”.

5. Related Work

In the field of ontology mapping, (Rahm and Bernstein, 2001) and (Shvaiko and Euzenat, 2004) present a broad overview of the various approaches on automated ontology matching. Multilingual ontology is an issue what is not explored in these works. We find some proposals for using multilingual ontologies in Information Retrieval field. (Guyot et al., 2005) explore a translation-free technique for multilingual information retrieval. Multilingual ontologies are used for documents/queries representation to map a term to its corresponding concept. The ontologies are supposed to be created manually. In multilingual question-answering systems, (Pazienza et al., 2005) use a multilingual ontology to search across multilingual source of data, where the multilingual ontologies are manually defined. An initiative to create an extension of the well-known WordNet English database is the EuroWordNet², a multilingual database with wordnets for several European languages (Dutch, Italian, Spanish, German, French, Czech and Estonian). In the juridical domain, a proposal of a multilingual lexical database LOIS³ had supported the creation of multilingual ontologies of terms from juridical domain, using European languages. The LOIS database can be used also as a dictionary to translate terms from different languages.

6. Final Remarks and Future Work

In this paper we presented a framework for automatic mapping of multilingual Description Logics (DL) ontologies. We focused in the process of translating ontologies. In the presented framework, the *source ontology* is translated to the target ontology language, using a dictionary, generating a *DL translated ontology*. The *target* and *translated* ontologies are used as input for the mapping process. We had used three agents representing different mapping approaches (lexical, semantic, and structural). These agents use argumentation in order to agree on the obtained mappings. A *DL mapping ontology* is generated as result of this process. Using an illustrative example, we presented in details the mapping process of two ontologies in English and Portuguese languages.

In the future, we intend to evaluate our approach in a practical use; develop further tests using ontologies from different languages; evaluate our automatic systems against manual mappings; and use mapping in a multilingual question-answering system.

7. References

- Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.
- Alexander Boer, Tom van Engers, and Radboud Winkels. 2003. Using ontologies for comparing and harmonizing legislation. In *ICAIL '03: Proceedings of the 9th international conference on Artificial intelligence and law*, pages 60–69, New York, NY, USA. ACM Press.
- Jacques Guyot, Saiid Radhouani, and Gilles Falquet. 2005. Ontology-based multilingual information retrieval. In *CLEF*.
- Peter Haase and Boris Motik. 2005. A mapping system for the integration of owl-dl ontologies. In *IHIS*, pages 9–16.
- I. Horrocks and P. Patel-Schneider. 2003. Reducing owl entailment to description logic satisfiability. In *Proceedings of the 2nd International Semantic Web Conference*.
- A. Maedche and Steffen Staab. 2002. Measuring similarity between ontologies. In *Proceedings of the European Conference on Knowledge Acquisition and Management*, pages 251–263.
- M. Pazienza, A. Stellato, F. Zanzotto, L. Henriksen, and P. Paggio. 2005. Ontology mapping to support ontology based question answering. In *Proceedings of the Meaning 05 Workshop*.
- E. Rahm and P. A. Bernstein. 2001. A survey of approaches to automatic schema matching. *VLDB*, 10:334–350.
- P. Shvaiko and J. Euzenat. 2004. A survey of schema-based matching approaches. Technical report, Informatica e Telecomunicazioni, University of Trento.
- C. Trojahn, P. Quaresma, and R. Vieira. 2008. A cooperative approach for composite ontology mapping. *LNCIS Journal of Data Semantic*, Volume 4900/2008:237–263.

²<http://www.illc.uva.nl/EuroWordNet/>

³<http://www.loisproject.org/>