A Framework for Multilingual Ontology Mapping

Cássia Trojahn
Universidade de Évora
Rua Romão Ramalho, 59
Évora, Portugal
cassia@di.uevora.pt

Paulo Quaresma
Universidade de Évora
Rua Romão Ramalho, 59
Évora, Portugal
pq@di.uevora.pt

Renata Vieira
Universidade do Vale do Rio dos Sinos
Av. Unisinos, 950
São Leopoldo, Brazil
renatav@unisinos.br

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 [7] and comparison of ontologies describing legislations from different countries [2].

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, or they can be evaluated semantically, usually on the basis of semantic oriented linguistic resources, or 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 what 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 langue, 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 arguments, the agents compute their preferred mapping sets. The arguments in such preferred sets are viewed as the set of globally acceptable arguments. See [10] for details of the argumentation process.

This paper is structured as follows. Section 2 comments on OWL 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) [1]. 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) [5] 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 os an role name, d is a concrete domain, c_i are individuals, and n is a non-negative integer:

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

We consider the semantic of a SHOIN(D) knowledge base KB by the mapping π proposed by [4] which transform KB axioms into a first-order formula. Each atomic concept is

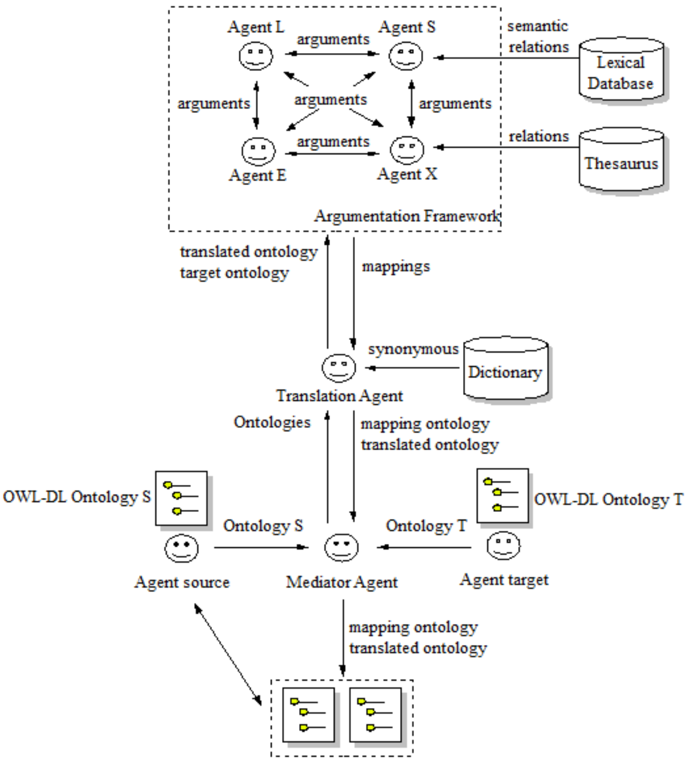

Figure 1: Architecture for multilingual ontology mapping.

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 langue, 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.

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* [10]. 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 be the result of the mapping process.

The final mappings are the 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 synonymous terms of each term from the source ontology. Each term is then transformed into a *concept description* in the *translated ontology*. The set of synonymous 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 langue; T_{trans} is the set of terms of the *translated ontology*. For each term $t_{source,i} \in T_{source}$, a set of synonymous T_{syn} (i.e., $\{t_{syn,1}, ..., t_{syn,n}\}$) is generated in the target langue. The term $t_{source,i}$ is then converted into a *concept description* $t_{trans,i} \equiv t_{syn,1} \sqcup ... \sqcup t_{syn,n}$.

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

Moreover, we can distinguished 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 synonymous 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 synonymous 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 synonymous entry in the dictionary. In this case, the process is made as commented above. Otherwise, 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 terms is formed by the disjunction of the synonymous set of each single term. For each synonymous set is generated a conjunction of the synonymous terms, as described above for the single terms. For instance, consider the Portuguese term "Tese-Mestrado" ("Master-Thesis"m 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 gener-

Figure 2: Portuguese source ontology.

Figure 3: English target ontology.

ated. 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 $\sqcap$ ...$\sqcup$ $t_{syn,n}$). Then, the final description of "Tese-Mestrado" is the disjunction of the synonymous set of "Tese" and "Mestrado": "Tese-Mestrado" $\equiv$ (Thesis $\sqcup$ Dissertation $\sqcup$ ... $\sqcup$ $t_{syn,n}$) $\sqcap \equiv$ (Master $\sqcap$ Master-Thesis $\sqcap$ ...$\sqcup$ $t_{syn,n}$).

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.

First, the source term "Topico" is searched in a Portuguese-English lexical dictionary, and the set of synonymous {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 synonymous 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.

In this specific example, we consider the three mapper agents lexical (L), semantic (S) and structural (E). The lexical agent adopts the the *lexical similarity* proposed by [6]. 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 WordNet[1] 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.

[1]http://www.wordnet.princeton.edu

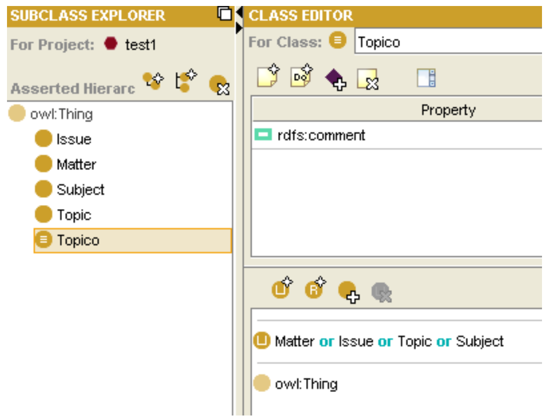

Figure 4: Translated ontology.

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: $\text{Issue}_{trans} \equiv \text{Subject}_{target}$ (+,1); $\text{Matter}_{trans} \equiv \text{Subject}_{target}$ (+,1); $\text{Subject}_{trans} \equiv \text{Subject}_{target}$ (+,1); and $\text{Topic}_{trans} \equiv \text{Subject}_{target}$ (+,1).

Looking for 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, "$\text{Subject}_{target} \equiv$" "$\text{Subject}_{trans}$" 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, [8] and [9] 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 proposal for using multilingual ontologies in Information Retrieval field. [3]

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, [7] use a multilingual ontology to search across multilingual source of data, where the multilingual ontologies are manually defined. An initiative to create a extension of the well-known WordNet English database is the EuroWordNet[2], 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 LOIS[3] 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 proposed a framework for automatic mapping of multilingual Description Logics (DL) ontologies. In the ontology mapping field, multilingual ontology mapping is still not explored. Several applications, such as web question answering, require automatic tools to map the description of their content and services.

First, the *source ontology* is translated to the target ontology langue, using a lexical database or a dictionary, generating a *DL translated ontology.* The *target* and *tranlated* ontologies are used as input for the mapping process. A *DL mapping ontology* is generated as result of this process. The mappings are computed by specialized agents using different mapping approaches (lexical, semantic and structural). These agents use argumentation to exchange their local results, in order to agree on the obtained mappings. In this paper we focused in the translation agent.

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

7. REFERENCES

[1] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications.* Cambridge University Press, 2003.

[2] A. Boer, T. van Engers, and R. Winkels. 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, 2003. ACM Press.

[3] J. Guyot, S. Radhouani, and G. Falquet. Ontology-based multilingual information retrieval. In *CLEF*, 2005.

[4] P. Haase and B. Motik. A mapping system for the integration of owl-dl ontologies. In *IHIS*, pages 9–16, 2005.

[5] I. Horrocks and P. Patel-Schneider. Reducing owl entailment to description logic satisfiability. In *Proceedings of the 2nd International Semantic Web Conference*, 2003.

[2]http://www.illc.uva.nl/EuroWordNet/

[3]http://www.loisproject.org/

[6] A. Maedche and S. Staab. Measuring similarity between ontologies. In *Proceedings of the European Conference on Knowledge Acquisition and Management*, pages 251–263, 2002.

[7] M. Pazienza, A. Stellato, F. Zanzotto, L. Henriksen, and P. Paggio. Ontology mapping to support ontology based question answering. In *Proceedings of the Meaning 05 Workshop*, 2005.

[8] E. Rahm and P. A. Bernstein. A survey of approaches to automatic schema matching. *VLDB*, 10:334–350, 2001.

[9] P. Shvaiko and J. Euzenat. A survey of schema-based matching approaches. Technical report, Informatica e Telecomunicazioni, University of Trento, 2004.

[10] C. Trojahn, P. Quaresma, and R. Vieira. A cooperative approach for composite ontology mapping. *LNCS Journal of Data Semantic (to appear)*, 2007.