SemMatcher: A Tool for Matching Ontology-based Schemas

Carlos Eduardo Pires[1], Damires Souza[2], Thiago Pachêco[1], Ana Carolina Salgado[1]

[1]Centro de Informática, Universidade Federal de Pernambuco, Recife, PE, Brazil

[2]Instituto Federal de Educação, Ciência e Tecnologia da Paraíba/IFPB, Brazil

`{cesp, tpap, acs}@cin.ufpe.br, damires@ifpb.edu.br`

Abstract. *In Peer Data Management Systems (PDMS), each peer is an autonomous source that makes available a local schema. Information exchange occurs through the establishment of schema mappings between local schemas. To help matters, ontologies have been considered as uniform representation of local schemas (i.e., peer ontologies). Consequently, ontology matching techniques have been used to determine schema mappings between peer ontologies. This work presents SemMatcher, a tool for matching ontology-based schemas. SemMatcher allows the identification of semantic correspondences between two peer ontologies using a domain ontology as background knowledge. Also, the tool determines a global similarity measure between the matching ontologies that can be used for peer clustering.*

1. Introduction

The development of communication infrastructures has led to a wide range of data sources being available through networks such as Peer Data Management Systems (PDMS) [Adjiman *et al*., 2007]. In these settings, each peer is an autonomous source that makes available a local schema. Schema mappings, i.e., correspondences between schema elements, are generated to allow information exchange between peers. To help matters, ontologies have been considered as uniform representation of peer schemas (referred here as peer ontologies) and, consequently, as a means for enhancing information integration [Xiao, 2006].

Peer ontologies are designed and developed autonomously, what entails several forms of heterogeneity between them, even between those on the same domain [Euzenat and Shvaiko, 2007]. Reconciling such ontologies and finding correspondences between their elements (concepts or properties) through ontology matching techniques is still a relevant research issue. In order to identify such correspondences, some works have used semantic additional descriptions, called background knowledge (e.g., domain ontologies) [Sabou *et al*., 2006]. Resulting correspondences between peer ontologies elements are usually associated with a similarity value which expresses the level of confidence on the correspondence. Particularly, in a PDMS, it is also important to have a global measure representing the overall similarity degree between two peer ontologies (and not only between their elements). For instance, the measure can be used to cluster semantically similar peers in the overlay network [Pires, 2009].

In this light, we present *SemMatcher*, a semantic-based ontology matching tool. Its main contributions are twofold: i) in order to identify semantic correspondences between two peer ontologies, *SemMatcher* takes into account a domain ontology as background knowledge; and ii) the tool determines a global similarity measure between the matching ontologies.

This paper is organized as follows: Section 2 explains the ontology matching process. Section 3 presents an instantiation of a process running in *SemMatcher*. Finally, in Section 4, we conclude our work giving some perspective about future works.

2. A Process for Matching Ontology-based Schemas

In our semantic-based approach, the process for matching peer ontologies brings together a combination of already defined matching strategies. In this process, a linguistic-structural matcher and a semantic matcher are executed in parallel. The obtained similarity values of both matchers are combined through a weighted average. Each matcher receives a particular weight according to its importance in the matching process. As shown in Figure 1, the process receives as input two ontologies (O_1 and O_2) and a domain ontology DO to be used as background knowledge. As output, it may produce one or two alignments (i.e., the semantic and final alignments, also called A_{SE} (Phase 1) and $A_{FI,}$ (Phase 2), respectively), and a global similarity measure between the involved ontologies.
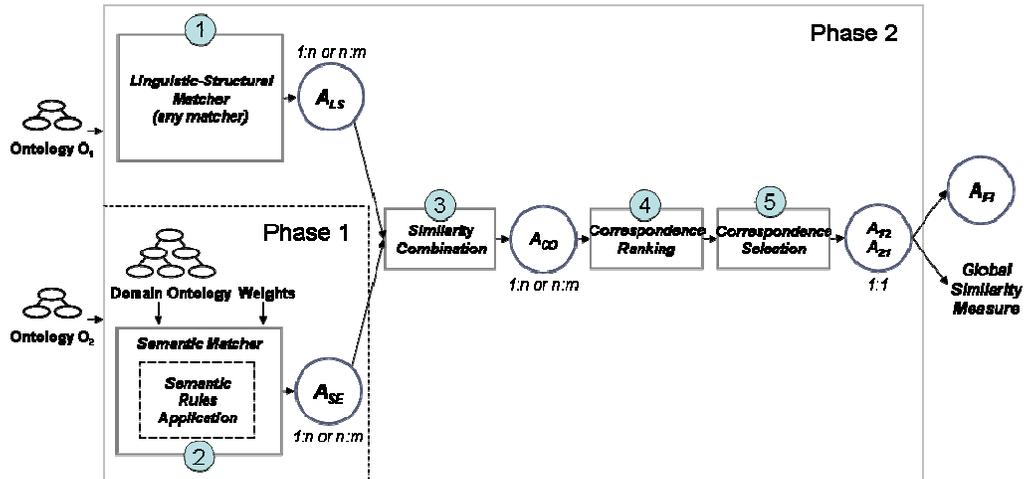
Figure 1. The overall process for matching ontology-based schemas

The main steps carried out by the semantic-based ontology matching process are:

(1) *Linguistic-Structural Matching*

In this step, any existing ontology matching tool including linguistic and/or structural matchers can be used. Such linguistic and structural matchers are handled as a hybrid matcher, i.e., as a fixed combination of simple matchers. The combination of their similarity values depends on the composition strategy of the ontology matching tool that is used. The alignment produced by the hybrid matcher is denoted by A_{LS}.

(2) *Semantic Matching*

Using a DO, the semantic matcher applies a set of semantic rules described in Section 2.1 to derive the type of semantic correspondences between O_1 and O_2. The resulting alignment is denoted by A_{SE}.

(3) *Similarity Combination*

The individual similarity values of the correspondences produced by the hybrid matcher and the semantic matcher are associated in a combined similarity value, through a weighted average of the values generated by the individual matchers. A weighted average is used because matchers may produce opposing similarity values.

(4) *Correspondence Ranking*

Correspondences containing the elements of O_1 are ranked (in descending order) according to the associated similarity values.

(5) *Correspondence Selection*

A filter strategy is applied to choose the most suitable correspondence for each O_1 element. The strategy consists in selecting the correspondence with the highest combined similarity. As a result, an alignment A_{12} is generated.

Steps 4 and 5 are also executed in the opposite direction. Correspondences containing the elements of O_2 are ranked according to the associated similarity values and the same filter strategy is applied. An alignment A_{21} is then produced. The final alignment (A_{FI}) is obtained as the union between the alignments A_{12} and A_{21}. At the end of the matching process, three goals are possible to be achieved:

A. Generating only the semantic alignment (A_{SE}): in this option (Phase 1), only the alignment produced by the semantic matcher is shown. The alignment can be used for query reformulation purposes [Souza *et al*., 2009];

B. Generating the final alignment (A_{FI}): in this option (Phases 1 and 2), the resulting set of correspondences identified by the linguistic-structural matcher is combined with the correspondences produced by the semantic matcher. The final alignment can be used as a schema mapping between two peer ontologies [Pires, 2009];

C. Calculating the global similarity measure: in this case (Phases 1 and 2), the global measure between O_1 and O_2 is generated. The measure can be used to cluster semantically similar peers in a PDMS [Pires, 2009].

2.1 Semantic Matching

In our work, we consider domain ontologies (DO) as reliable references that are made available on the Web. We use them in order to bridge the conceptual differences or similarities between two peer ontologies. In this sense, first concepts and properties from the two peer ontologies are mapped to equivalent concepts/properties in the DO and then their semantic correspondence is inferred based on the existing semantic relationship between the DO elements. To specify the correspondences, we take into account four aspects: i) the semantic knowledge found in the DO; ii) if the peer ontology concepts share super-concepts in the DO; iii) if these super-concepts are different from the root concept; and iv) the depth of concepts measured in number of nodes.

We have defined seven types of semantic correspondences [Souza *et al.*, 2009], each one associated with a particular weight which corresponds to the level of confidence on such correspondence, as follows: *isEquivalentTo* (1.0), *isSubConceptOf* (0.8), *isSuperConceptOf* (0.8), *isCloseTo* (0.7), *isPartOf* (0.3), *isWholeOf* (0.3), and *isDisjointWith* (0.0). Particularly, the *isCloseTo* correspondence denotes a strong semantic relationship holding between two concepts. We added this correspondence type as a way to enrich queries and provide users with approximate answers.

2.2. Calculating the Global Similarity Measure

The evaluation of the overall similarity between the ontologies O_1 and O_2 is an additional step in the proposed ontology matching process. Such step takes as input the alignments A_{12} and A_{21} (Step 5) to calculate the similarity value. The value indicates the global similarity degree between the ontologies. Existing similarity measures such as *dice* [Aumüller *et al.*, 2005], *weighted* [Castano *et al.*, 1998] and *overlap* [Rijsbergen, 1979] can be adapted to calculate the global similarity degree (Figure 2). They consider the size of the input ontologies. In this work, the size of an ontology is determined by the number of its elements and is denoted by |O|.

$$Dice(O_1, O_2) = \frac{|A_{12}| + |A_{21}|}{|O_1| + |O_2|} \quad Weighted(O_1, O_2) = \frac{\sum_{i=1}^{|A_{12}|} n + \sum_{i=1}^{|A_{21}|} n}{|O_1| + |O_2|} \quad Overlap(O_1, O_2) = \frac{|A_{12} \cap A_{21}|}{\min(|O_1|, |O_2|)}$$

Figure 2. The measures used to determine the overall similarity between ontologies

3. *SemMatcher*

In the following, we describe the tool in practice throughout an example. In such example, we use ontologies belonging to Education's knowledge domain: *Semiport.owl* (O_1), *UnivBench.owl* (O_2), and *UnivCsCMO.owl* (DO). All ontologies can be found in the tool's web site[1] as well as some experimental results obtained with *SemMatcher*.

The *SemMatcher* interface is presented in Figure 3. First, the user must choose the two ontologies (O_1 and O_2) to be matched as well as the domain ontology (DO) which will be used as background knowledge. After that, the user must press the "Run" button and options regarding the possible outputs (illustrated in Figure 4) are presented (semantic alignment, global similarity measure, and final alignment). In the example, all three options have been selected indicating that the whole matching process (i.e., Phases 1 and 2) must be executed.

Once the output is indicated, the matching process is executed and the selected outputs are shown. In this version, the user is allowed to use H-Match[2] and Alignment API[3] to generate a structural-linguistic alignment. However, s/he can also use an alignment generated by any other ontology matching tool that uses the alignment format described in [OAEI, 2009]. In the example, H-Match has been chosen. The semantic alignment is shown in Figure 3 while the final alignment and the global measure are depicted in Figure 5. For the sake of space, not all correspondences are shown.
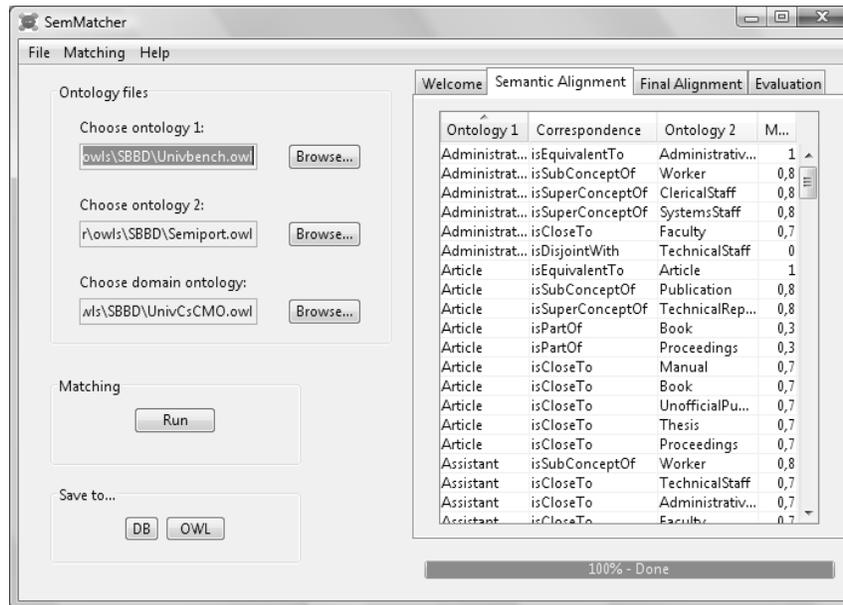
Figure 3. SemMatcher's main screen
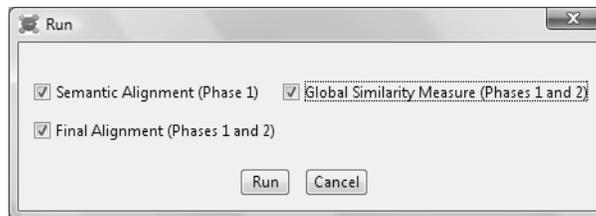
Figure 4. SemMatcher's output options
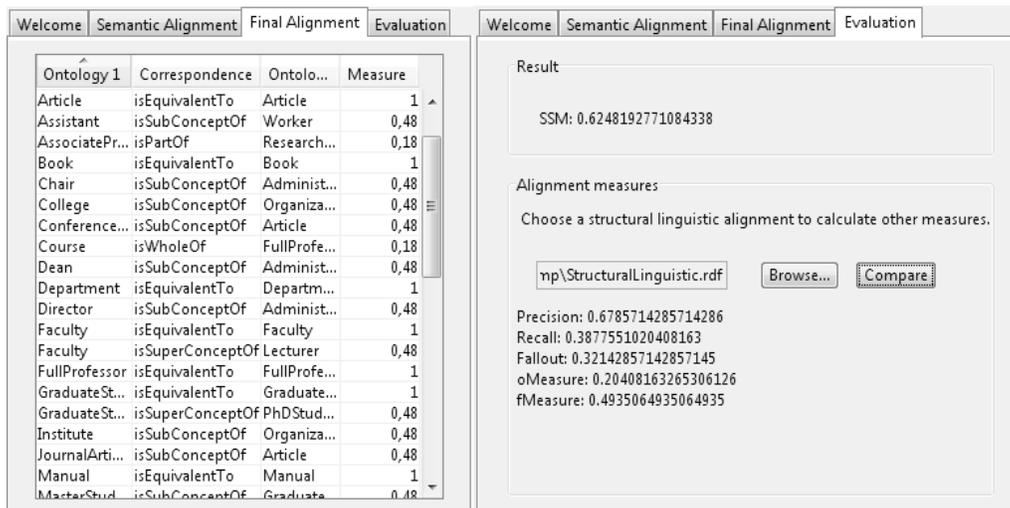
Figure 5. SemMatcher's results (Phase 2): final alignment and global similarity measure

All alignments generated during the matching process are saved in an operating system's folder that can be configured in the "File → Preferences → Directory Options" menu. The alignments can also be stored in a MySQL database. Other parameters (e.g. the semantic correspondence's weight, database credentials, and linguistic-structural

matcher) can also be configured in the Preferences menu. After the matching process is finished, the user can compare the final alignment against a reference alignment. The reference alignment must be manually generated by expert users which are knowledgeable about the corresponding domain and must be represented in the format defined by [OAEI, 2009]. The final and reference alignments are compared using the measures suggested by OAEI: *Precision*, *Recall*, *Fallout*, *oMeasure*, and *fMeasure*. Figure 5 depicts the results obtained for the two ontologies of our example. *SemMatcher* has been implemented in Java. In order to provide ontology manipulation and reasoning, we have used Jena[4] and OWL API[5].

4. Conclusions

In highly dynamic environments such as PDMS, the semantics surrounding correspondences among ontologies representing peer schemas is rather important for tasks such as query answering or peer clustering. This work has presented a tool for matching ontology-based peer schemas, combining different matching strategies (e.g., linguistic, structural, and semantic). Furthermore, as a result of the overall process, we have introduced the determination of a global similarity measure between the matching ontologies. Such measure is used to semantically cluster peers in our PDMS. Currently, we are extending *SemMatcher* by considering properties both in the correspondences identification and in the determination of the global similarity measure.

References

Adjiman, P., Goasdoué, F., Rousset, M.-C. (2007) "SomeRDFS in the Semantic Web", In: Journal on Data Semantics, LNCS, Vol. 8, pp. 158-181.

Aumüller, D., Do, H. H., Massmann, S., Rahm, E. (2005) "Schema and Ontology Matching with COMA++". In: International Conference on Management of Data (SIGMOD), Software Demonstration.

Castano, S., Antonellis, V., Fugini, M. G., Pernici, B. (1998) "Conceptual Schema Analysis: Techniques and Applications". In: ACM Transactions on Database Systems, Vol. 23, No. 3, pp. 286-333.

Euzenat, J., Shvaiko, P. (2007) "Ontology Matching". Springer-Verlag.

OAEI – Ontology Alignment Evaluation Initiative. (2009) http://oaei.ontologymatching.org/

Pires, C. E. S. (2009) "Ontology-based Clustering in a Peer Data Management System". PhD Thesis. Federal University of Pernambuco (UFPE), Recife, PE, Brazil.

Rijsbergen, C. J. (1979) "Information Retrieval", 2nd Edition, Stoneham, MA: Butterworths. http://www.dcs.gla.ac.uk/Keith/Preface.html.

Sabou M., D'Aquin M., Motta E. (2006) "Using the Semantic Web as Background Knowledge for Ontology Mapping". In: ISWC'06 Ontology Matching WS.

Souza D., Arruda T., Salgado A. C., Tedesco P., Kedad, Z. (2009) "Using Semantics to Enhance Query Reformulation in Dynamic Environments". To appear in the 13th East European Conference on Advances in Databases and Information Systems, Riga, Latvia.

Xiao, H. (2006) "Query Processing for Heterogeneous Data Integration using Ontologies". Ph.D. Thesis. University of Illinois at Chicago, Chicago, USA.

---

[4] Jena, http://jena.sourceforge.net/
[5] OWL API, http://owlapi.sourceforge.net/