Semantic Matching of Ontologies

Christoph Quix, Marko Pascan, Pratanu Roy, David Kensche

Informatik 5 (Information Systems), RWTH Aachen University, Germany
http://www.dbis.rwth-aachen.de

Introduction The discovery of semantic relationships such as subsumption and disjointness is still a challenge in ontology matching [6]. Existing methods use logical reasoning over computed equivalence relationships or machine learning based on lexical and structural features of ontology elements [7, 3]. While these methods deliver good results for some cases, they are limited to the information contained in the input ontologies to be matched. Therefore, background knowledge in form of an additional ontology may be useful to detect semantic relationships. In existing approaches, the identification of an appropriate ontology as background knowledge is often a task left for the user. We present two enhanced approaches for identifying semantic relationships. The first one is based on background knowledge; in contrast to other approaches, it is able to identify a background ontology automatically. The second approach builds on existing machine learning methods for identifying semantic relationships. First evaluation results for these methods and combined approaches show that the integration of these methods is reasonable as more semantic relationships are identified.

Semantic Matching using Background Knowledge It has been shown in previous works that using an ontology as background knowledge can improve the match result [1]. The selection of the background ontology is obviously an important step in such an approach. While earlier works either relied on the user to provide such an ontology [1], or used very general upper ontologies (e.g. SUMO-OWL, [5]), our approach is able to select the background ontology automatically. The idea is illustrated in fig. 1 and 2. For the input ontologies S and T, we generate keyword queries for a web search engine (e.g., Google or Swoogle), and for our local ontology repository. The external search engine is only used if the local repository does not contain an appropriate ontology. When a background ontology O is found it can be used for matching. In addition to the direct alignment A_{dir}, two alignments $A_{O,S}$ and $A_{O,T}$, between the input ontologies and the background ontology, are computed. Then, for each pair of correspondences from $A_{O,S}$ and $A_{O,T}$, existence of a relationship (i.e., equivalence, subsumption) between the model elements from O is determined. If that is the case, a new correspondence between the concepts from S and T can be inferred. All the correspondences found in this way are called *semantic matches* (A_{sem}). Eventually, the final result is created by building the union of A_{dir} and A_{sem}.

Using Machine Learning for Semantic Matching Our second, complementary method uses machine learning to identify semantic relationships. We implemented an approach similar to the method presented in [7]. The computation of subsumption relationships is considered as a binary classification task, i.e., a concept pair is classified into two possible classes: subsumption and not-subsumption relationship.

Because ontologies are usually hierarchical structures, the subsumption relationships found in input ontologies can be used as training examples, making the process of classifier training independent of alignment computation. Each training example is a

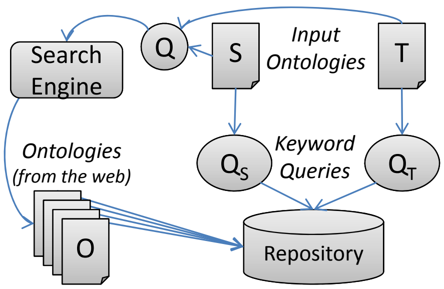

Fig. 1. Selection of the background ontology

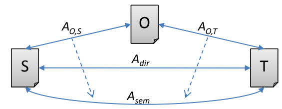

Fig. 2. Matching using a background ontology

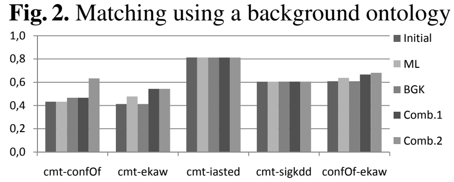

Fig. 3. Recall for OAEI conference track

concept pair (c_i, c_j), where $c_i \sqsubseteq c_j$ and both concepts belong to a single ontology, i.e., the source or target ontology of a matching task. We use distinct words, extracted from both source and target ontologies, as features for the machine learning method. In order to represent the concept pairs in the feature space, each concept c_i of a concept pair is described by a set of feature space words that can be found in its neighborhood, constituting the concept's *context document* D_{c_i}. The notion of the concept's neighborhood can be defined in various ways. In our implementation D_{c_i} is created from words found in: name, label, comment, instances, data and object properties of c_i, direct sub or super concepts of c_i, concepts in union or intersection definitions of c_i and equivalent concepts of c_i. The context documents of concepts are translated to feature vectors which are used as input for the machine learning method (currently, C4.5 decision tree). We use an optimization to avoid the classification of all concept pairs.

Conclusion The presented approaches have been integrated into our generic matching system *GeRoMeSuite* [4]. To evaluate our approach, we used *semantic* precision and recall as measures [2] and data sets from the oriented track of OAEI 2009. We created also two combined match configurations using both approaches. The results in fig. 3 show very good results for the combined approaches.

Acknowledgments: This work is supported by the Umbrella Cooperation Programme (`http://www.umbrella-coop.org/`).

References

1. Z. Aleksovski, M. Klein, W. ten Kate, F. van Harmelen. Matching Unstructured Vocabularies Using a Background Ontology. *Proc. EKAW*. 2006.
2. J. Euzenat. Semantic precision and recall for ontology alignment evaluation. *Proc. IJCAI*, pp. 348–353. 2007.
3. Y. R. Jean-Mary, E. P. Shironoshita, M. R. Kabuka. Ontology matching with semantic verification. *Journal of Web Semantics*, **7**(3):235–251, 2009.
4. D. Kensche, C. Quix, X. Li, Y. Li. *GeRoMeSuite*: A System for Holistic Generic Model Management. *Proc. VLDB*, pp. 1322–1325. 2007.
5. V. Mascardi, A. Locoro, P. Rosso. Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation. *IEEE Trans. on Knowl. & Data Eng.*, **22**:609–623, 2010.
6. P. Shvaiko, J. Euzenat. Ten Challenges for Ontology Matching. *Proc. ODBASE*, pp. 1164–1182. 2008.
7. V. Spiliopoulos, G. A. Vouros, V. Karkaletsis. On the discovery of subsumption relations for the alignment of ontologies. *Web Semantics*, **8**(1):69–88, 2010.