Searching Web Resource Collections Using Multiple Ontologies

Abstract

The paper proposes an ontology mapping based framework that allows searching web resource collections annotated with various ontologies. Mappings among multiple ontologies we represent by the Simple Knowledge Organization System (SKOS) Mapping Vocabulary. Relaying on such mappings we developed a search algorithm that uses a concept of one ontology (source ontology) as the query argument, generates queries compliant with another ontology, and finally gets ranked search results relevant for the source ontology.

1 Introduction

In a past few years large collections of web resources became available either through the digital libraries (such as ACM Portal) or more importantly as widely dispersed web resources in many individual institutions. In order to have context relevant search engines the digital libraries and web resource collections usually employ explicit semantic information in the form of subject categories, taxonomies, or ideally richer ontologies. However, one can hardly find two different object repositories relaying on the same classification. Several interoperability initiatives are trying to solve the issue of searching across multiple collections.

2 Our approach

In order to address this problem we propose the use of ontology mapping techniques [Kalfoglou and Schorlemmer, 2003]. In fact, we use another ontology – mapping ontology – to define mapping relations among ontologies. In our approach we use the Mapping Vocabulary of the W3C RDF-based Simple Knowledge Organization System (SKOS)

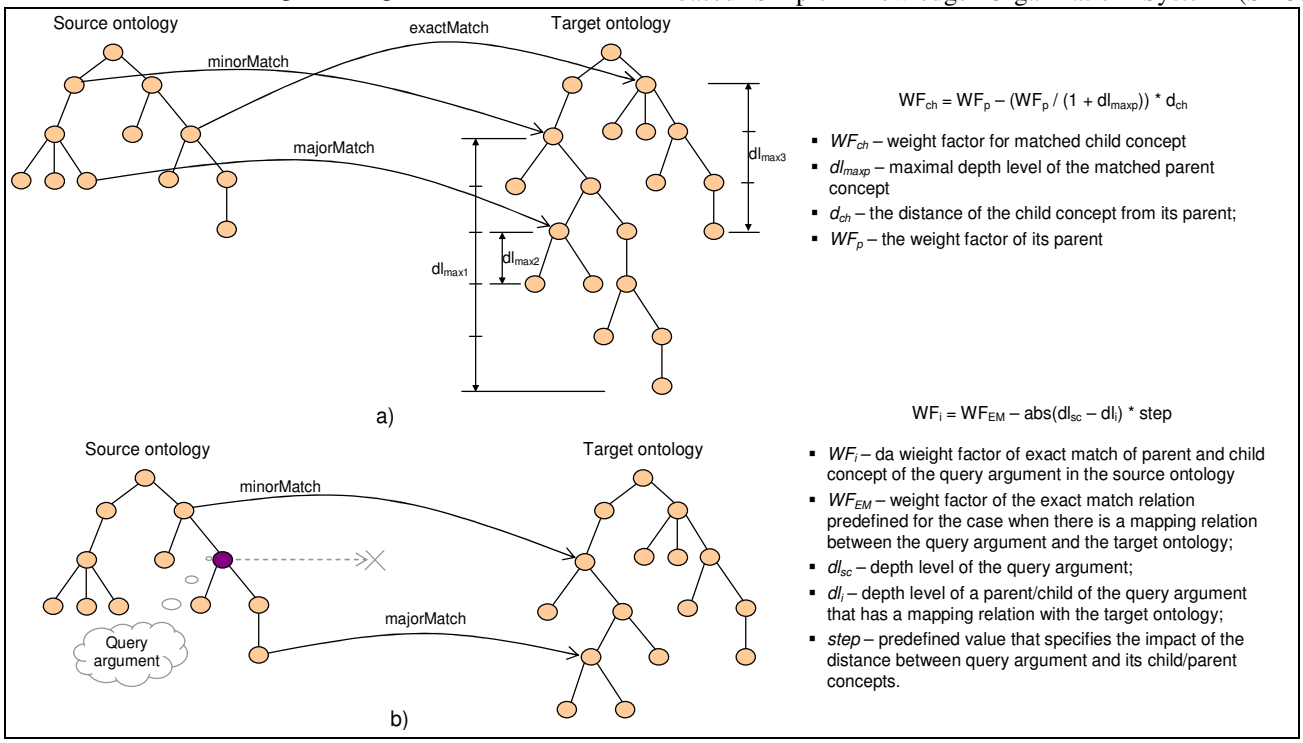

Figure 1. The search algorithm based on ontology mappings: a) all child nodes of the matched node in the target ontology are used; b) the case when there is no mapping between the query argument and concepts of the target ontology

[Miles and Brickley, 2005]. The *SKOS Mapping* contains a set of properties for specifying mapping relations between concepts from different domain ontologies (*broadMatch*, *narrowMatch*, *exactMatch*, *majorMatch*, *minorMatch*). Such a rich set of semantic relations for expressing ontology mapping is useful in ranking search results to reflect the weight of the mapping.

Using SKOS Mapping relations between ontologies we developed a search algorithm that takes arguments from one ontology and generates queries compliant with other ontologies. The algorithm uses all the children of matched concepts in the target ontology (see Figure 1a). It also calculates the weight factor to determine ranks of both matched concepts and their children in the resulting list of concepts (see see Figure 1a). The referent weight factor is for the *exactMatch* relation, while others are calculated relatively to it. The search algorithm also captures the case when mapping is not defined between the query argument and the target ontology (see Figure 1b). In such a case, the algorithm looks for both child and parent concepts in the source ontology that have specified mappings with concepts of the target ontology when the query argument does not have any defined mappings. The algorithm takes into account the fact that the distance between the query arguments and all its child/parent concepts with defined matching relations is not the same. Accordingly, it calculates the weight factor (WF_i) for exact match for each parent and child concept of the query argument in the source ontology (Figure 1b).

3 Evaluation

In order to evaluate the search algorithm we developed a web-based application in the domain of e-learning for an information management course (Figure 2). The application uses a course ontology to search two different collections of web resources based on their own classifications: ACM Digital Library (DL) (http://www.acm.org/dl) and Merlot learning object repository (http://www.merlot.org).

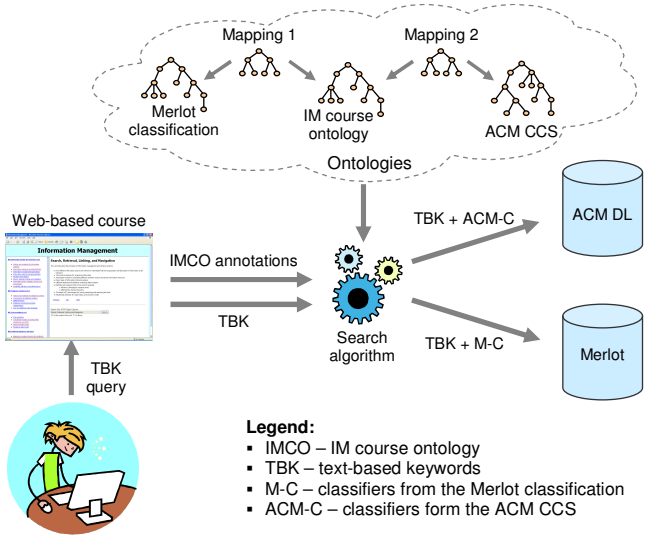

Figure 2. An evaluation system: a web-based application in the domain of e-learning. Students define text-based keyword queries. The application collects annotation of the current page within the course ontology (embedded in the web page as RDF) and sends it with the keyword to our search algorithm implemented in Jess. The search algorithm maps it to the target classification systems, and sends an expanded query to the chosen collection of web resources

The analysis of evaluation results showed that the combination of text-based keywords and repositories' classifiers reduces the number of found objects, and improves looking for more suitable learning resources. However, we noticed several peculiarities of the ACM DL Advanced Search and its underlying search engine (http://www.verity.com). We found out the total number of results was decreased when we increased the number of classifiers in a query. It was completely opposite to expectations since those query arguments are connected with the OR logical operator. In fact, the ACM search engine has a threshold for every found resource. Increasing the number of classification parameters also increases the threshold and therefore eliminates some of the objects. However, this does not affect the best results.

In the experiment with Merlot we obtained results according to the expectations: the higher number of target ontology classes defined for a concept in the source ontology the higher number of found learning objects. Analyzing the results, we saw that the combination of classification-based search and text-based keyword search reduced the number of found learning object to 1.65% on average.

4 Conclusion

The paper presented a way to achieve semantic interoperability when searching for web resources in web sources annotated by different domain ontologies. The evaluation example showed benefit to have a combined text-based keyword search with ontology annotated content in order to provide more relevant web resources.

In the future we plan to integrate the developed search algorithm into the *eduSource Communication Layer (ECL)* as a part of its federated search engine. We also plan to research how we can automatically generate ontology mapping relations the search algorithm relays on. The idea is to employ algorithms from our on-going project for generating semantic signatures based on both the WordNet [Choi and Hatala, 2005] and content of web resources.

Acknowledgement

This work was funded by Canada's LORNET NSERC Research Network.

References

[Choi and Hatala, 2005] A. Choi, and M. Hatala. Towards Browsing Distant Metadata Using Semantic Signature, *KCAP2005 Workshop on Integrating Ontologies*, submitted, Banff, Canada, 2005.

[Kalfoglou and Schorlemmer, 2003] Y. Kalfoglou and M. Schorlemmer. Ontology mapping: the state of the art, *The Knowledge Engineering Review*, 18(1):1-31, 2003.

[Miles and Brickley, 2005] A. Miles, and D. Brickley, SKOS Core Vocabulary Spec., W3C Working Draft, http://www.w3.org/TR/swbp-skos-core-guide, 2005