Learning Ontologies to Improve the Quality of Automatic Web Service Matching

Hui Guo
Department of Computer Science
Stony Brook University
Stony Brook, NY 10532
huguo@cs.sunysb.edu

Anca Ivan
IBM TJ Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
ancaivan@us.ibm.com

Rama Akkiraju
IBM TJ Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
akkiraju@us.ibm.com

ABSTRACT

This paper presents a novel technique that significantly improves the quality of semantic Web service matching by (1) automatically generating ontologies based on Web service descriptions and (2) using these ontologies to guide the mapping between Web services. The experimental results indicate that with our unsupervised approach we can eliminate up to 70% of incorrect matches that are made by dictionary-based approaches.

Categories and Subject Descriptors

C.2.0 [**Computer Systems Organization**]: COMPUTER-COMMUNICATION NETWORKS—*Data communications*

General Terms

Algorithms, Standardization

Keywords

Semantics, Web services

1. INTRODUCTION

Semantic Web services represent the next generation of Web services, designed to support automatic discovery, matching and composition of Web services. This paper describes a novel ontology-learning approach that matches a source and a target set of Web services by refining the classic dictionary-based approach by capturing the relationships between tokens inside service tags. The idea is to disambiguate such tag matches by capturing the word relationships in an ontology, and use the ontology to guide the matching process.

2. DICTIONARY-BASED APPROACH

The idea of the dictionary-based approach is to decide whether two web services match by extracting all words from the service descriptions and computing the best matches based on dictionaries. For example, the tags *customerName* and *clientId* might match because *customer* is a synonym for *client*, and *name* is a synonym for *ID*. In general, the dictionary-based approach contains three stages.

1. Tag Processing. After parsing the Web services and extracting all tags, each tag is divided into multiple tokens (*clientName* splits into *client* and *Name*), expanded from abbreviations into full words (*ClientInfo* is expanded to *ClientInformation*), and associated with a list of its synonyms, which is built using a domain-independent dictionary (e.g., WordNet).

2. Finding Matches. The "best" matches for a given word are given by a "similarity score": *Score(A, B) = synNum / max(len(A), len(B))* where *synNum* is the number of matching words (i.e., the tokens that are synonyms), and $len(X)$ is the number of tokens in the tag X. For example, the similarity score between *CustomerCare* and *ClientSearch* is 0.5, because *Customer* and *Client* are synonyms, but *Care* and *Search* are not.

3. Filtering The filtering step reduces the set of all possible matches to a smaller set, based on their significance. However, many of the final matches are incorrect: The tags *CompanyID* and *CountryID* should not be matched since *CompanyID* is about a company, while *CountryID* is about a country. The tag *LegalContractID* should be matched to *Contract* instead of *LegalAccountID*. The reason for the false positives is that dictionary-based approach treats each tag as a bag of words and ignores the relationships between the words in a tag.

3. ONTOLOGY-LEARNING APPROACH

The main idea of the ontology-learning approach is to capture the relationships between the words contained in a tag, and match tags if both the words are similar (dictionary-based approach) and the relationships are equivalent. This approach started from the observation that people tend to use same simple patterns if they have space constraints. In order to discover the patterns and confirm our theory, all 27,026 tags from 919 WSDL files pulled off the Internet were categorized into several patterns.

No. tags	Pattern	Example
22,126	*Prefix+Noun1 +Noun2*	*LegalContractID*
2,585	*Property+Noun*	*ClientName*
17,885	*Verb+Noun*	*selectGift*

The ontology-learning approach starts from these observations and refines the dictionary-based approach as described below.

Step 1. Ontology Learning The first step in the ontology-learning approach is capturing the relationships between the words in a tag, as given by the observed patterns, and save them in an ontology (see Table 1). Using these rules, a source ontology is generated from the source WS collection,

WWW 2007, May 8–12, 2007, Banff, Alberta, Canada.
ACM 978-1-59593-654-7/07/0005.

Table 1: Transformation rules

Rule Pattern	Relationships
Prefix+Noun1+Noun2	*Prefix+Noun1* **hasProperty** *Tag* *Tag* **subclassOf** *Noun2*
Adjective+Noun	*Tag* **subclassOf** *Noun*
Verb+Noun	*Tag* **hasObject** *Noun* *Tag* **hasVerb** *Verb*

while a target ontology from the target WS collection. The following step in the ontology-learning matching is to match the two source and target ontologies.

Step2. Ontology Matching Similar to the process described in Section 2, the ontology-learning approach finds all the matches that contain related words, and uses a filter to extract the final matches. The main difference is that tags are matched based on word similarity and on the relationships between words. Two ontological concepts are matched if and only if one of the following is true:

The concepts are synonyms of each other, as given by a dictionary (e.g., *Client* matches *Customer*)
One concept (or its synonym) is the property of the other concept (e.g., *StateCode* is a property of *State*)
One concept (or its synonym) is the subclass of the other concept (e.g., *OrderID* is a subclass of *ID*)
The concepts are matched only if all their relationships are matched (e.g., *ClientLocation* matches *CustomerAddress* because *Client* is a synonym of *Customer*, *Location* is a synonym of *Address*, and the relationship between *ClientLocation* and *Location* is identical with the relationship between *CustomerAddress* and *Address*).

Based on the above rules, the refined ontology-learning technique defines which tags are really connected to each other, and avoids generating matches between irrelevant concepts (i.e., false positives).

Step 3. Generating Matches The next step is finding the tag matches given the matches between concepts at the ontology level. Similar to the dictionary-based approach, the quality of a match is reflected by its similarity score.

4. EVALUATION

The goal of the evaluation is to quantify the improvement brought by the ontology-learning approach when compared to the dictionary-based approach. As Figure 1 shows, the ontology-learning approach generates 65% correct matches, while the dictionary-based approach generates 68%. However, the ontology-learning approach generates only 14% incorrect matches, which is considerable lower than the number generated by the dictionary-based approach 58%.

5. RELATED WORK

Researchers have done a considerable amount of work on the problem of Web service matching [4, 2]. However, they treat all terms from a WSDL document as a bag of words and do not guarantee a match of operations to operations, messages to messages, etc. Knowledge Representation (KR) community [6] learns ontologies from glossaries and free form text based on documents with a lot of contextual information; Web services often do not have any documentation beyond a set of words/terms used to describe various parameters. Work on totally or mostly automatic ontology extraction includes [1, 3, 5]. However, most of them require big data sets or context information (e.g., corpus, dictionary and relation schema).

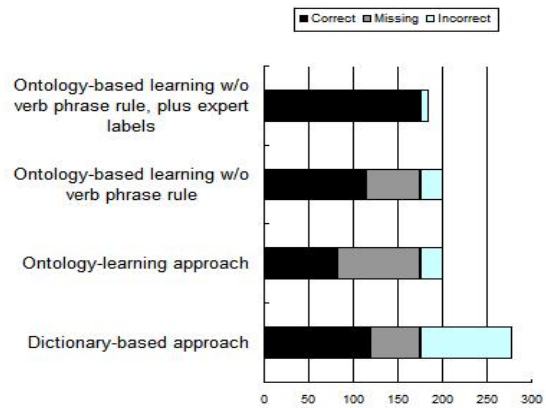

Figure 1: Quality of results

6. CONCLUSION AND FUTURE WORK

This paper described an ontology-learning approach to Web service matching. Its main contributions are the capability to automatically learn ontologies based on Web service descriptions and use these ontologies to guide the Web service matching process.

7. ADDITIONAL AUTHORS

Additional authors: Richard Goodwin (IBM TJ Watson Research Center, email: `rgoodwin@us.ibm.com`).

8. REFERENCES

[1] E. Agirre, O. Ansa, E. Hovy, and D. Martinez. Enriching very large ontologies using the WWW. In *Proceedings of the Workshop on Ontology Construction of the European Conference of AI (ECAI-00)*, 2000.

[2] X. Dong, A. Halevy, J. Madhavan, E. Nemes, and J. Zhang. Similarity search for web services. In *Proceedings of VLDB*, 2004.

[3] A. Faatz and R. Steinmetz. Ontology enrichment with texts from the WWW. In *Semantic Web Mining 2nd Workshop at ECML/PKDD-2002*, 2002.

[4] A. Hess and N. Kushmerick. Learning to attach semantic metadata to web services. In *Proceedings of 2nd International Semantic Web Conference (ISWC2003)*, 2003.

[5] J. Jannink. Thesaurus entry extraction from an on-line dictionary. In *Proceedings of Fusion '99*, 1999.

[6] P. Y. Glossont: A concept focussed ontology building tool. In *Knowledge Representation Conference*. 2004.