A REFERENCE ONTOLOGY BASED APPROACH FOR SERVICE ORIENTED ONTOLOGY MANAGEMENT

Shuying Wang, Jinghui Lu and Miriam A. M. Capretz
Department of Electrical and Computer Engineering, Faculty of Engineering
University of Western Ontario, London, Ontario, N6A 5B9, Canada

Keywords: Service Oriented Architecture, Ontology Management.

Abstract: To establish effective information exchange among applications in a distributed B2B environment, the business participants are not only required to share their functions or service interfaces, but in many cases, they also need to exchange their data models. Ontology, as a popular semantic form of knowledge representation, can be used to represent data models, thus allowing applications to locate and integrate these models in a more intelligent way. In this paper, we introduce a reference ontology based approach for service oriented ontology management. Specifically, STAR, a domain specific reference ontology, is built and used for the experiments in a real life case. Furthermore, in order to validate and evaluate our approach and implementation, a prototype system is developed to provide ontology deploying, browsing and mapping operations on a service-oriented mechanism. Our experiments have provided promising results, which are consistent with our original ideas of managing ontologies and optimizing ontology mappings to facilitate data interoperability in a distributed environment.

1 INTRODUCTION

In Business-to-Business (B2B) applications, the interoperability of heterogeneous data sources is an important issue that is widely recognized in information technology intensive organizations. To establish effective information exchange among applications, the business participants are not only required to share their functions or service interfaces, but in many cases, they also need to exchange their data models. The traditional message-based approaches (Hohpe and Woolf, 2003) require developers to retrieve data models through messages and then to perform a one-to-one mapping in order to identify and characterize relationships between the models of two applications. However, it is a major challenge to create and maintain thousands of mappings for these models. Furthermore, in order to share their models, each application needs to publish its data model in a location where other applications can easily locate and retrieve the related models for information exchange.

As the core of the semantic web, ontology is the representation of knowledge in a certain domain. Representing data models by ontologies and mapping ontologies among the semantic resources is an important approach for achieving semantic data interoperability. At the same time, service oriented architecture (SOA) is a key technology for supporting interoperability among information and processing data model interoperability. Consequently, the significant potential of combining SOA and ontology provides a promising solution to improve semantic interoperability. For example, a well-defined mapping process can be considered as a component that provides a mapping service, which can be implemented with various applications.

In this paper, we propose a reference ontology based approach to support the interoperability of heterogeneous data sources. The main idea of our methodology is to make use of background knowledge in an industry domain to enhance the performance of alignment. Specifically, terms from the different data sources are first mapped to intermediate terms defined in the reference ontology, and then their mapping is deduced based on the semantic relation of the intermediate terms. Furthermore, in order to examine our approach, five experiments are designed to cover a comprehensive validation of the ontology mapping strategies. We examine these experiment results from generic

Wang S., Lu J. and A. M. Capretz M.
A REFERENCE ONTOLOGY BASED APPROACH FOR SERVICE ORIENTED ONTOLOGY MANAGEMENT.
DOI: 10.5220/0002787600690074
In *Proceedings of the 6th International Conference on Web Information Systems and Technology (WEBIST 2010)*, page
ISBN: 978-989-674-025-2

ontology mapping to specified case study and provide analysis of the proposed ontology mapping.

Section 2 introduces the related work on current research. In Section 3, it introduces reference ontology based mapping approach and the system architecture. In Section 4, we design five experiments aiming to evaluate the proposed approach. Lastly, Section 5 presents the conclusions and outlines a number of directions for future work.

2 RELATED WORK

Our work aims to extend the principles of the ontology mapping approach as well as the emerging Web Services standards in order to support the manageability and interoperability of heterogeneous data sources. A fundamental problem with ontology mapping involves the integration of heterogeneous data sources, which has been extensively researched in the last two decades (Rahm and Bernstein, 2001). Some research approaches (Sabou et al, 2006, Aleksovski et al, 2006) have considered the use of external background knowledge as a way of obtaining semantic mappings between syntactically dissimilar ontologies. WordNet is one of the most frequently used sources of background knowledge. The literature (Li et al, 1995) shows that WordNet has been used successfully for word sense disambiguation algorithms in other contexts, particularly in text. Moreover, SUMO (2009) has initially been created and further developed to facilitate data interoperability, information search and retrieval, and automated inference.

A substantial amount of literature has been published about the combination of SOA and ontology improving semantic interoperability (Staab and Studer, 2004). MAFRA (Silva and Rocha, 2003) is a toolkit used to maintain an ontology mapping system and provides support for ontology mapping tasks, such as the automatic specification of semantic relations, negotiation and evolution. Additionally, a distinct project proposed by Korotkiy and Top is known as Onto-SOA (Korotkiy and Top, 2006). Onto-SOA integrates ontologies and SOA to provide a mechanism for representing and exploiting both the conceptual and behavioral domain aspects. Specifically, it employs an ontology-based domain model as a direct input to a service and enables the exchange of messages between a service and its consumer.

3 A REFERENCE ONTOLOGY BASED APPROACH FOR SERVICE ORIENTED ONTOLOGY MANAGEMENT

3.1 Reference Ontology

In addressing the interoperability of heterogeneous data sources, our approach relies on the reference domain ontology as a semantic bridge between different data models. The basic process in this approach first aligns the concepts of the corresponding data models involved in the business process with the reference domain knowledge. Next, we use the semantic information from this reference knowledge to infer relationships between the models. Lastly, the relationships are utilized to induce an indirect set of mapping pairs and to generate the required correspondences between data models.

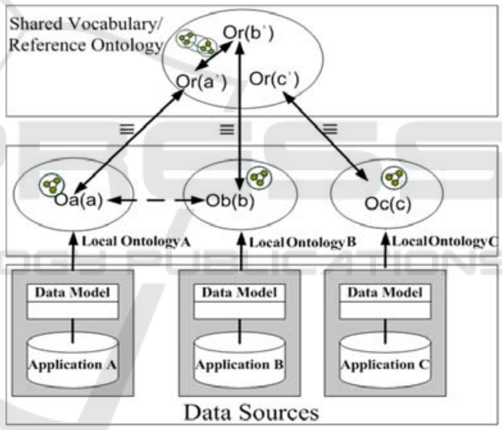


Figure 1: A Reference Ontology-Based Approach for Ontology Mapping.

In our approach, we develop local ontologies to represent different data models and reference ontology as the semantic bridge between local ontologies. The reference ontology represents the shared vocabulary of a domain and defines basic terms that can be combined to describe more complex semantics in the local ontologies. For example, in Figure 1, we show a local ontology Oa, for the data model of Application A and another local Ontology Ob, for the corresponding data model of Application B. Accordingly, these local ontologies are used for exchanging data between Application A and Application B. The reference ontology Or represents the set of basic domain terms that provide a semantic link between different data models. Consequently, for concept a in local ontology Oa and concept b in local ontology Ob, there are corresponding concepts a' and b' in the reference ontology.

The concrete steps of this approach are outlined below:

1. The reference ontology containing concepts, Or(a') and Or(b'), correspond to Oa(a) and Ob(b) in local ontologies;
2. For each pair of concepts:
   - If Or(a') is `equivalentClass` to Or(b') in the reference ontology, then Oa(a) from ontology A can be inferred as equivalent to Ob(b) from ontology B;
   - If Or(a') is `subClassOf` Or(b') in the reference ontology, then Oa(a) from ontology A can be inferred as child of Ob(b) from ontology B;
   - If Or(a') is `sameAs` Or(b') in the reference ontology, then Oa(a) from ontology A can be treated as synonym of Ob(b) from ontology B;
   - If Or(a') is `differentFrom` Or(b') in the reference ontology, then Oa(a) from ontology A can be inferred to be different from Ob(b) from ontology B.

Moreover, if no relationship is found, then no mapping is inferred.

3.2 STAR Ontology

In our case study, we are using the STAR ontology as the mediator, which contains rich reference knowledge for performing the intermediate mapping.

Table 1: The Metrics of the STAR Ontology.

Metrics	Count
Class	122
Object Property	117
Data Property	1168
Subclass Axioms	185
Object Property Domain Axioms	117
Object Property Range Axioms	115
Data Property Domain Axioms	1081
Data Property Range Axioms	1118

Standards for Technology in Automotive Retail (STAR, 2009) is a non-profit, unionized organization whose members include dealers, manufacturers, retail system providers and automotive-related industrial organizations. The goal of the STAR organization is to use non-proprietary information technology (IT) standards as a catalyst in fulfilling the business information needs of dealers and manufacturers. Using the STAR metadata, we have developed the STAR reference ontology in order to gain a high level of detailed knowledge from the automotive retail industry domain and therefore to facilitate the interoperability of automotive retail applications.

In general, the STAR ontology is based on the terminology in the automotive retail industry. Specifically, it is formalized in OWL-DL. Currently, it describes 1592 lexical terms; the ontology metrics are listed in Table 1. STAR ontology covers concepts in the automotive retail domain and, accordingly, it is structured in six different categories. These categories include General, Dealer, Customer, Parts Management, Vehicle Management and Sales and Vehicle Repair and Service. Together, these six categories represent the main organizational structure of STAR.

3.3 System Architecture

As depicted in Figure 2, our system is divided into three layers. These layers include the semantic layer, the service layer and the access layer. At the semantic layer, the semantic and expressive descriptions are used to describe the data models. Data model ontology sources, also known as local ontologies, are developed to represent different data models. Subsequently, the information corresponding to the ontologies, such as the business entity name and the business process name, are published using Web Services.

At the service layer, the regular Web Service technologies are utilized. In particular, we are using SOAP for messaging, WSDL for service description, and the service registry for publishing, discovering, and retrieving data models. Finally, a mapping engine is used to execute ontology mapping.

For implementing the service registry, we have also developed three key Web Services:

- The Publishing Service allows web users to submit their ontologies and other related information to the web. The submitted information includes the provider of the ontologies and the business process to which the ontologies are applied.
- The Discovery Service executes a search based on the information given by business processes or ontology providers. When a user submits a request for acquiring a certain ontology, the service returns the available ontology list as the search results.

- The Mapping Service provides the functionality of ontology mappings. The mapping engines are externally developed and imported into system.

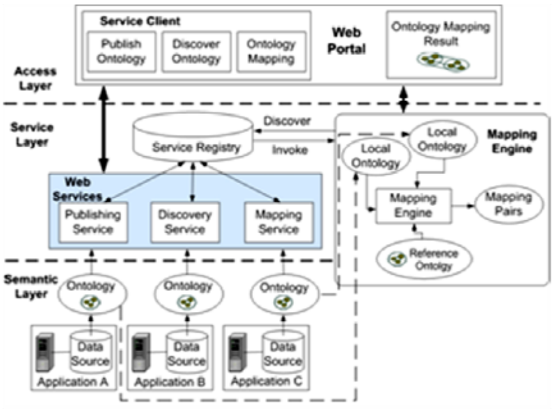


Figure 2: System Architecture.

The access layer contains the Web Portal, which enables users to access the available Web Services. Users can also execute specific functions, such as accessing the publishing ontology by using the publishing services, searching and retrieving ontologies via the discovery service, and performing ontology mapping through the mapping service.

4 EXPERIMENTS

4.1 Experiment Design

A web-based prototype called Service Oriented Ontology Management Framework (SOOMF) is implemented in order for the end user to manage ontologies. The implementation uses Struts and Spring architecture. Apache Tomcat is used as the container for the development and deployment in Web services. Also, Jena is used as the OWL parser.

Based on the prototype, we validate the accuracy improvement for the proposed mapping approach. We classify five experiments into two categories; the first category, which includes the first four tests, is known as generic experiments. These experiments are used to validate our reference ontology-based approach in comparison to other existing mapping methodologies (experiments 1 to 4). The second category includes the specified experiment, Experiment 5, where the newly developed STAR ontology will be examined for the improved utility and to a broader adoption of ontology mapping in the automotive retail domain.

In our study, we design two sets of experimental ontologies. First, in our generic experiments, we use the data set that has been used in the Information Interpretation and Integration Conference (I3CON, 2003) experimental ontologies. The second set of ontologies is created to present specific data models involved in certain business processes for specialized terms in the automotive retail industry. This data set is used in Experiment 5, which is treated as the specified experiment. In this set, the local ontologies are created to represent the data models in the automotive retail industry, while the SUMO (2009) and STAR ontologies are used as the reference ontologies.

Experiment 1 - Terminological Approach. In this experiment, we use the typical Terminological approach, which uses a combination of lexical and structural correspondence between source and target ontologies to compare strings. The correspondence of terms mapped from the two ontologies is generated directly by the terminological mapping, which results in a list of equivalent terms of pairs. We expect that the result of this experiment assist in constructing the evaluation baseline.

Experiment 2 - WordNet Hierarchical Distance. WordNet is a freely available English lexical database whose design is inspired by current psycholinguistic theories of human lexical memory (Li et al., 1995). This experiment is to implement the mapping by using WordNet as the thesaurus for calculating the hierarchical semantic distance.

Experiment 3 - The Combination of the Terminological Approach and WordNet Thesaurus. Experiment 3 combines the Terminological approach and the use of WordNet to refine the mapping results. We believe that combining the Terminological approach and the WordNet Thesaurus can overcome some of the lexical limitations and improve the mapping performance.

Experiment 4 - Reference Ontology-based Mapping using SUMO as the Generic Reference Knowledge. This experiment is to assess the value of using standard upper ontologies, thus we utilize the reference knowledge as an ontological bridge that indirectly infers mapping between local ontologies. SUMO is used as reference ontology.

Experiment 5 - Domain Specific Reference Ontology based Mapping. In this experiment, we utilize the reference ontology as a source of background information. Terms from the two local ontologies are first mapped to terms in STAR, and subsequently, their mapping is deduced on the basis of the semantic relation between the terms. The experiment includes three steps. In Step 1, we utilize the Terminological approach to generate a mapping baseline. Step 2 introduces SUMO as the reference ontology, showing the mapping result of this generic reference ontology. In Step 3, STAR is used to facilitate the attempt to capture correspondence in the automotive retail industry.

4.2 Measurement

Basically, a threshold score is a lower limit for the similarity score of two concepts that belong to the respective source and target ontologies and that will be treated as mapping pairs. For instance, if pairs of mapping results produced have a threshold score of 0.60, indicating that the two concepts are considered as mapping pairs if the similarity score between them is greater than or equal to 0.60.

The effectiveness of the mapping approaches can be measured by the precision of the mapping results. We define the mapping accuracy as the ratio of correct mappings N to the number of discovered mappings M. The formula in Equation 1 shows the mapping precision as a percentage value.

$$p(MappingPrecision) = \frac{N(CorrectMapping)}{M(MappingPairs)} \times 100\%$$

Equation 1: Mapping Precision.

4.3 Generic Experiment Analysis

The generic experiments consist of Experiments 1, 2, 3, and 4. Accordingly, Figure 3(a) demonstrates the mapping precision arranged by the threshold value of similarity, where the left side of the graph displays the lowest value; the right side shows the highest value; and the vertical line illustrates the precision. The Terminological approach in Experiment 1 is presented as the baseline experiment in order to evaluate the other mapping results. Specifically, we are attempting to directly discover mapping pairs between the local ontologies.

The results of the four generic experiments are presented in Figure 3(b). In comparison to the Terminological approach used in Experiment 1, the WordNet Distance approach in the second experiment generates more mapping pairs.

For example, when the threshold is 0.8, the WordNet Distance method (Experiment 2) returns 68 mapping pairs and among them, 16 are correct as the precision is only 24%, whereas in the Terminological approach (Experiment 1), the result is 29 mapping pairs with 16 correct mappings, and the mapping precision increases as 55%. Overall, the results indicate that the mapping precision of the WordNet Distance approach is not necessarily more effective than the Terminological approach.

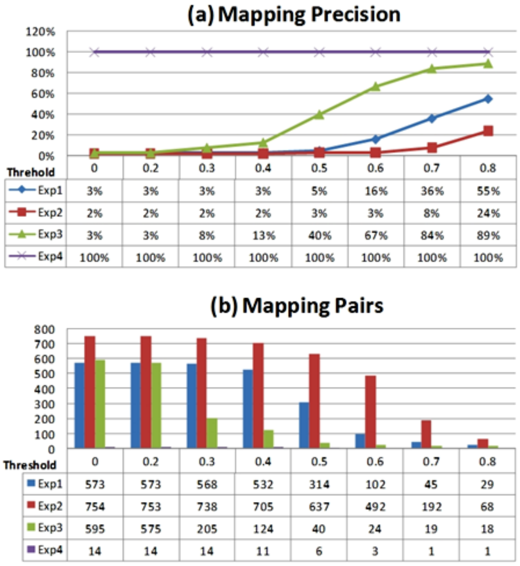
Furthermore, the federated mapping approach of Terminological and WordNet in Experiment 3 demonstrates that the mapping results are more effective than they are in the two approaches used in Experiments 1 and 2 respectively, and moreover, a lot of incorrect mapping results are eliminated in Experiment 3. For example, when the threshold level is 0.6, the mapping precision of the federated approach is 67%; in comparison to the 16% of the Terminological approach.

We also obtain a high mapping precision when using the reference ontology based approach in Experiment 4. However, we also observe while the similarity threshold is increased, the mapping pairs are reduced. For example, the number of mapping pairs is 14 at threshold 0.2, while this number is reduced to 1 at a threshold of 0.7. This drastic decrease occurs because the terms used in experimental ontologies do not have corresponding definitions in SUMO, and therefore, most terms in local ontologies cannot be bridged by the reference ontology. Therefore, in order to overcome this disadvantage, the appropriate reference ontology needs to be selected prior to mapping in a specific domain. Moreover, in order to obtain a more effective mapping result, the pre-selected reference ontology should include as many terms as possible.

4.4 Specified Experiment Analysis

In comparison to the traditional Terminological approach, the domain specific reference ontology results in improved mapping precision at the same threshold level. For example, Figure 4(a) shows that at threshold level 0.2, the mapping precision of using STAR is 90%. Compared to the 12% achieved in the Terminological approach, the precision with STAR is increased by 78%.

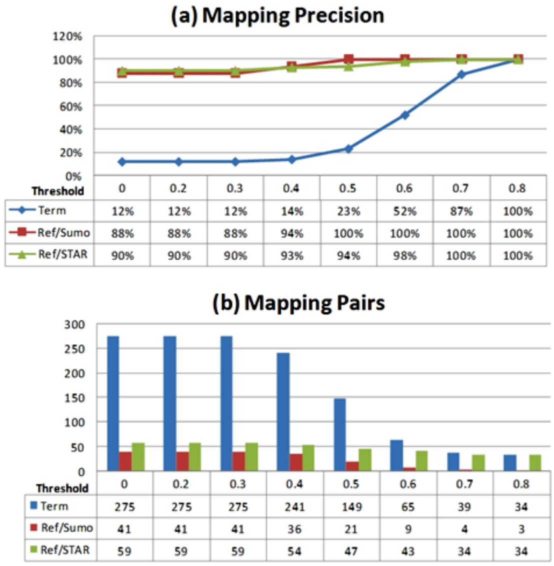
The experiment also proves that domain-specific information can help to improve the mapping results for the reference ontology based approach. For example, when using either SUMO or STAR, the mapping precision curves are very similar; however, when the threshold level is increased from 0.2 to 0.8, STAR is more effective than SUMO. In particular, as shown in Figure 4(b), using SUMO reduces the mapping pairs from 41 to 3, whereas using STAR only decreases the pairs from 59 to 34. This discrepancy occurs since the increased inclusion of terms in the STAR, has a significant effect on the mappings. In particular, the resulting similarity scores of correct concept pairs are increased, therefore demonstrating that the domain specific specific reference ontology can increase the amount mapped terms and thus lead to more meaningful mappings.



Threhold	0	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Exp1	3%	3%	3%	3%	5%	16%	36%	55%
Exp2	2%	2%	2%	2%	3%	3%	8%	24%
Exp3	3%	3%	8%	13%	40%	67%	84%	89%
Exp4	100%	100%	100%	100%	100%	100%	100%	100%

Threshold	0	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Exp1	573	573	568	532	314	102	45	29
Exp2	754	753	738	705	637	492	192	68
Exp3	595	575	205	124	40	24	19	18
Exp4	14	14	14	11	6	3	1	1

Figure 3: The Mapping Results of Generic Experiments.



Threshold	0	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Term	12%	12%	12%	14%	23%	52%	87%	100%
Ref/Sumo	88%	88%	88%	94%	100%	100%	100%	100%
Ref/STAR	90%	90%	90%	93%	94%	98%	100%	100%

Threshold	0	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Term	275	275	275	241	149	65	39	34
Ref/Sumo	41	41	41	36	21	9	4	3
Ref/STAR	59	59	59	54	47	43	34	34

Figure 4: The Mapping Result of Specified Experiments.

5 CONCLUSIONS

The goal of our research is to show the feasibility and potential advantages of using a service-oriented mechanism to build an ontology management framework and using reference ontology as background knowledge for ontology mapping. The implementation and experiments based on real world case have provided positive results, which are consistent with our original ideas of ontology management and mapping in the distributed environment. Our future work includes building a new mechanism to facilitate reference ontology to discover more mapping pairs.

REFERENCES

Aleksovski, Z., Klein, M., ten Kate, W., Harmelen, F. van., (2006). "Matching Unstructured Vocabularies using a Background Ontology", in Managing Knowledge in a World of Networks, Springer-Verlag.

Hohpe, G., Woolf, B., (2003). Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions, Addison Wesley.

I3CON, (2003). http://www.atl.external.lmco.com/projects/ontology/i3con.html

Korotkiy, M., Top, J., (2006). "Onto-SOA: From Ontology-enabled SOA to Service-enabled Ontologies", Proceedings of the Advanced Int'l Conference on Telecommunications and Int'l Conference on Internet and Web Applications and Services, IEEE Computer Society, Washington, DC, USA, pp.124-124.

Li, X., Szpakowicz, S., Matwin, S., (1995). "A WordNet-based Algorithm for Word Sense Disambiguation", Proceedings of IJCAI-95. Montréal, Canada.

Rahm, E., Bernstein, (2001). P. A., "A survey of approaches to automatic schema matching", The VLDB Journal, 10, pp. 334–350.

Sabou, M., d'Aquin M., Motta, E., (2006). "Using the Semantic Web as Background Knowledge for Ontology Mapping", Proceedings of the 3rd international conference on Knowledge capture, ACM, New York, USA, 2006, pp.175-176.

Silva, N., Rocha, J., (2003). "Service-Oriented Ontology Mapping System", Proceedings of the Workshop on Semantic Integration of the International Semantic Web Conference, Sanibel Island, FL, USA.

Staab, S., Studer, R., (2004). Handbook on Ontologies, Springer.

STAR, Standards for Technology in Automotive Retail, (2009). http://www.starstandard.org/.

SUMO, Suggested Upper Merged Ontology, (2009). http://www.ontologyportal.org/.