Using the AgreementMaker to Align Ontologies for the OAEI Campaign 2007*

William Sunna and Isabel Cruz

University of Illinois at Chicago, USA
wsunna@cs.uic.edu, ifc@cs.uic.edu

Abstract. In this paper, we present the AgreementMaker, an ontology alignment tool that incorporates the *Descendants Similarity Inheritance (DSI)* method. This method uses the structure of the ontology graphs for contextual information, thus providing the matching process with more semantics. We have tested our method on the ontologies included in the anatomy track of the OAEI 2007 campaign.

1 Presentation of the System

In distributed database applications with heterogeneous classification schemes that describe related domains, an ontology-driven approach to data sharing and interoperability relies on the alignment of concepts across different ontologies. Once the alignment is established, *agreements* that encode a variety of mappings between the concepts of the aligned ontologies are derived. In this way, users can potentially query the concepts of a given ontology in terms of other ontologies. To enable scalability both in the size and the number of the ontologies involved, the alignment method should be automatic. In order to achieve this, we have been working on a framework that supports the alignment of two ontologies. In our framework, we introduce an alignment approach that uses different matching techniques between the concepts of the aligned ontologies. Each matching technique is embedded in a what we refer to a mapping layer [2]. We have currently four layers in our framework with the possibility of adding more mapping layers in the future. The motivation behind our framework is to allow for the addition of as many mapping layers as possible in order to capture a wide range of relationships between concepts.

We have developed a tool, the AgreementMaker, which implements our approach. The user interface of our tool displays the two ontologies side by side as shown in Figure 1. We refer to the first ontology which is displayed on the left as the source ontology, and to the second ontology which is displayed on the right as the target ontology. After loading the ontologies, the domain expert can start the alignment process by mapping corresponding concepts manually or invoking procedures that map them automatically (or semi-automatically). The mapping information is displayed in the form of annotated lines connecting the matched nodes. Many choices were considered in the process of displaying the ontologies and their relationships [2].

* This research was supported in part by the National Science Foundation under Awards ITR IIS-0326284 and IIS-0513553.

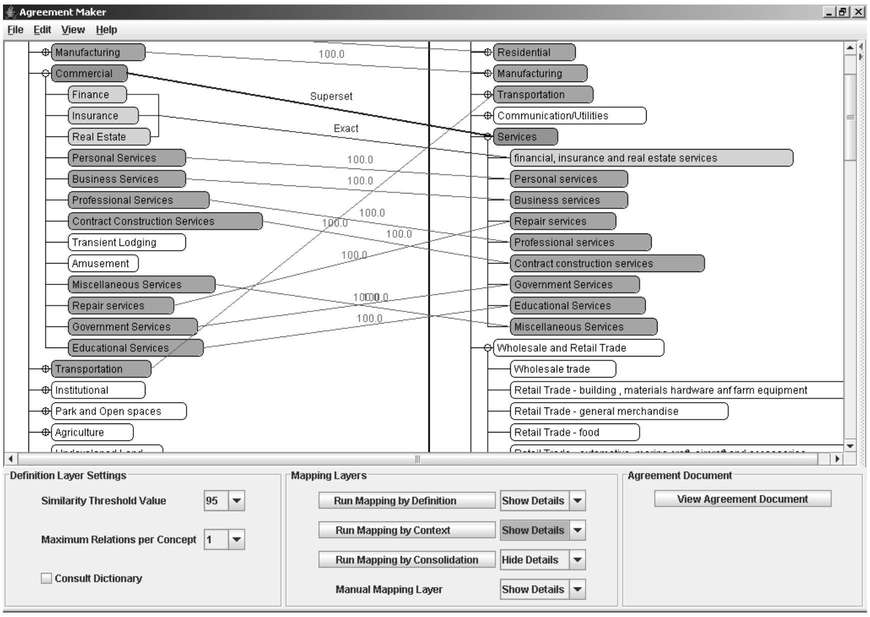

Fig. 1. Results of running three of the mapping layers.

1.1 Specific Techniques Used

In order to achieve a high level of confidence in performing the automatic alignment of two ontologies, a thorough understanding of the concepts in the ontologies is highly desired. To this end, we propose methods that investigate the ontology concepts prior to making a decision on how they should be mapped. We consider both the labels and the definitions of the ontology concepts and the relative positions of the concepts in the ontology tree. Our alignment method enables the user to select one of the following two matching methods: (1) applying the base similarity calculations only or (2) applying the base similarity calculations followed by the Descendant's Similarity Inheritance *(DSI)* method. The *DSI* method has been introduced to enhance the alignment results that were obtained from using the base similarity method previously proposed [2]. In what follows, we will present both our base similarity and our *DSI* methods.

Base similarity calculations The very first step in our approach is to establish initial mappings between the concepts of the source ontology and the concepts of the target ontology. These initial mappings will be a starting point for the *DSI* method. We try to find matching concepts in the target ontology for each concept in the source ontology. This is achieved by defining a similarity function that takes a concept in the source ontology and a concept in the target ontology and returns a similarity measure between them. If the similarity measure is equal or above a certain threshold decided by the domain expert, then the two concepts match each other. In order to find the base similarity measure between two concepts, we utilize the concepts' labels and in some cases their definitions as provided by a dictionary [2]. In what follows, we present the details of finding the base similarity between a concept in the source ontology and a concept in the local ontology:

– Let *S* be the source ontology and *T* be the target ontology.
– Let *C* be a concept in *S* and C' be a concept in *T*.
– We use function *base_sim*(C, C') that yields a similarity measure *M*, such that $0 \leq M \leq 1$.
– Parameter *TH* is a threshold value such that C' is matched with C when *base_sim*$(C, C') \geq TH$.
– For every concept *C* in *S*, we define the mapping set of C, denoted *MS*(C), as the set of concepts C' in T that are matched with C (i.e., *base_sim*$(C, C') \geq TH$).

Establishing base similarities between concepts of the source ontology and concepts of the target ontology may not be sufficient to achieve a high degree of precision in relating concepts in the two ontologies. To exemplify this point, we give an example in the geospatial domain, in particular, we align two ontologies describing wetlands. The first ontology describes the "Cowardin" wetland classification system [1] which is adopted in the United States. The second ontology describes the South African wetland classification system [3]. Figure 2 shows part of the "Cowardin" classification on the left, which is the source ontology, and part of the South African classification on the right, which is the target ontology. When calculating the base similarities between concepts of the two ontologies, the concept *Reef* that belongs to the *Intertidal* wetland subsystem in the source ontology will yield a base similarity measure of 100% with the concept *Reef* that belongs to the *Intertidal* wetland subsystem in the target ontology. Furthermore, it will also yield a base similarity measure of 100% with the concept *Reef* that belongs to the *Subtidal* wetland subsystem in the target ontology. This example shows that the base similarity measure is misleading because it does not correctly express the true meaning of the relationship between the two concepts, which should not be related because they belong to different wetland subsystems.

In order to eliminate such situations, we propose the Descendant's Similarity Inheritance *(DSI)* method, which reconfigures the base similarity between the concepts based on the similarity of their parent concepts.

Descendant's Similarity Inheritance (DSI) method We define the *DSI* reconfigured similarity between a concept *C* in *S* and a concept C' in *T* as *DSI_sim*(C, C'). In what follows, we present the details on how to determine *DSI_sim*(C, C'):

– Let *path_len_root*(C) be the number of edges between the concept *C* in *S* and the root of the ontology tree *S*. For example, in Figure 3, *path_len_root*$(C) = 2$. Similarly, we define *path_len_root*(C') with respect to T. For example, in Figure 3, *path_len_root*$(C') = 2$.
– Let *parent*$_i(C)$ be the *ith* concept from the concept *C* to the root of the source ontology *S*, where $0 \leq i \leq$ *path_len_root*(C). Similarly define *parent*$_i(C')$ with respect to T. For example, in Figure 3, *parent*$_1(C) = B$ and *parent*$_1(C') = B'$.

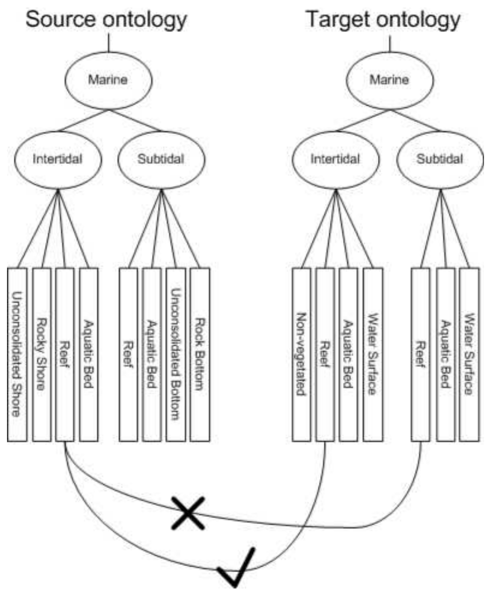

Fig. 2. An example of a case where misleading mappings may occur when two concepts have the same label.

- Define *MCP* as the *main contribution percentage*, which is the fraction of the similarity measure between *C* and C' that will be used in determining the overall *DSI_sim(C,C')*.
- We compute *DSI_sim*(C, C') as follows:

$$\mathit{MCP}\cdot \mathit{base\_sim}(C,C')+\frac{2(1-\mathit{MCP})}{n(n+1)}\sum_{i=1}^{n}(n+1-i)\mathit{base\_sim}(\mathit{parent}_i(C),\mathit{parent}_i(C')))$$

where $n = \min(\mathit{path\_len\_root}(C), \mathit{path\_len\_root}(C'))$

The main characteristic of the *DSI* method is that it allows for the parent and in general for any ancestor of a concept to play a role in the identification of the concept. Intuitively, the parent of a concept should contribute more to the identity of the concept than its grandparent. This is achieved by setting a relatively high value to *MCP*. The grandparent concept contributes more than the great grandparent, and so on, until the root is reached. This can be demonstrated by considering the example in Figure 3. In the figure, we show how the *DSI* similarity is determined between the concept *C* in the source ontology *S* (shown left) and the concept C' in the target ontology *T* (shown right) when applying the *DSI* method using an *MCP* value of 75%. The *DSI* similarity is determined by adding 75% of the base similarity between *C* and C' to 17% of the base similarity of their immediate parents (*B* and B') and finally to 8% of the base similarity of their grandparents (*A* and A'). Experiments have shown that 75% for the value of the

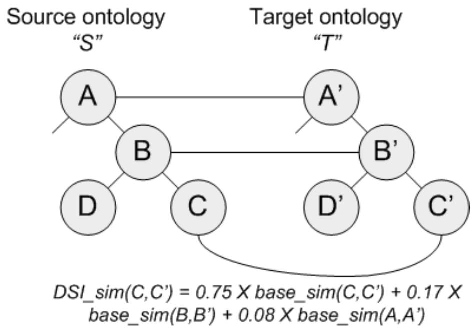

Fig. 3. Applying the DSI method to calculate the similarity between C and C'

.

MCP factor works well (in fact, any values in that neighborhood performed similarly). The following example illustrates just one such case.

Considering the case of Figure 2, the base similarity between the concepts *Intertidal* in the source ontology and the concept *Subtidal* in the target ontology is 37%. The base similarity between the concepts *Marine* in the source ontology and the concept *Marine* in the target ontology is 100%. When applying the *DSI* method with an *MCP* value of 75%, the *DSI* similarity between the concept *Reef* that belongs to the *Intertidal* wetland subsystem in the source ontology and the concept *Reef* that belongs to the *Subtidal* wetland subsystem in the target ontology will be 88%. Applying the *DSI* method again between the concept *Reef* that belongs to the *Intertidal* wetland subsystem in the source ontology and the concept *Reef* that belongs to the *Intertidal* wetland subsystem in the target ontology will yield a similarity of 100%. Therefore, we conclude that the last match is the best one (in fact the optimal one). This is just one example that shows how the *DSI* method can be useful in determining more accurate similarity measures between concepts.

1.2 Link to the system and parameters file

http://www.cs.uic.edu/˜advis/OAEI2007/align-code.zip

1.3 Link to the set of provided alignments (in align format)

The results of the three tasks for the anatomy track can be found at:
http://www.cs.uic.edu/˜advis/OAEI2007/sunna-cruz.zip

2 Anatomy Track Results

We have focused on the "Anatomy" track of the 2007 campaign. The purpose of this track is to find alignments between the ontology of the Adult Mouse Anatomy and the

NCI Thesaurus, which describes the human anatomy. The ontology of the Adult Mouse Anatomy has 2744 classes and the NCI Thesaurus has 3304 classes. Since the class IDs of all the classes of the ontologies do not describe what they refer to, a lookup file which contains the IDs of the classes and their labels has been produced. The lookup file has been used in the alignment process of the ontologies using the *DSI* method. The alignment process of the anatomy ontologies took around 9 minutes on an 1.6 GHz Intel Centrino Duo CPU with 1GB of RAM, running Windows XP.

3 Conclusions

We have presented the *Descendant's Similarity Inheritance (DSI)* method, that enhances our Base similarity method. The *DSI* method uses the structure of the ontology graph by utilizing the information associated with the descendants of each concept for contextual information thus providing the matching process with more semantics. We have applied our *DSI* method on the ontologies in the anatomy track of the OAEI 2007 campaign.

In addition to the *DSI* method, we have proposed the *Sibling's Similarity Contribution (SSC)* method [4], which uses the relationships between sibling concepts to further enhance the process of the alignment. For the purposes of this campaign, we decided to only apply the *DSI* method which performs better than the *SSC* in most of the alignment test cases we considered.

References

1. L. M. Cowardin, V. Carter, F. C. Golet, and E. T. LaRoe. Classification of Wetlands and Deepwater Habitats of the United States. U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C. Jamestown, ND: Northern Prairie Wildlife Research Center Online. `http://www.npwrc.usgs.gov/resource/wetlands/classwet/index.htm` (Version 04DEC1998), 1979.
2. I. F. Cruz, W. Sunna, N. Makar, and S. Bathala. A Visual Tool for Ontology Alignment to Enable Geospatial Interoperability. *Journal of Visual Languages and Computing*, 18(3):230–254, 2007.
3. J. Dini, G. Gowan, and P. Goodman. South African National Wetland Inventory, Proposed Wetland Classification System for South Africa. `http://www.ngo.grida.no/soesa/nsoer/resource/wetland/inventory`, 1998.
4. W. Sunna and I. F. Cruz. Structure-based Methods to Enhance Geospatial Ontology Alignment. In *Second International Conference on Geospatial Semantics (GeoS)*, volume 4853 of *Lecture Notes in Computer Science*. Springer, 2007. To appear.