

# Structure-based Methods to Enhance Geospatial Ontology Alignment<sup>\*</sup>

William Sunna and Isabel F. Cruz

Department of Computer Science  
University of Illinois at Chicago  
851 S. Morgan St. (M/C 152), Chicago, IL 60607, USA  
[wsunna,ifc]@cs.uic.edu

**Abstract.** In geospatial 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. To enable scalability both in the size and the number of the ontologies involved, the alignment method should be automatic. In this paper, we propose two fully automatic alignment methods that use the structure of the ontology graphs for contextual information, thus providing the matching process with more semantics. We have tested our methods on a set of geospatial ontologies pertaining to the domain of wetlands and on four sets that belong to an ontology repository that is becoming the standard for testing ontology alignment techniques. We have compared the effectiveness and efficiency of the proposed methods against two previous approaches. The effectiveness results that we have obtained with at least one of the new methods are as good or better than the results obtained with the previously proposed methods.

## 1 Introduction

Geospatial data and metadata are highly dependent on the regions for which they have been defined. Such heterogeneity can, for example, be caused by the autonomic and often uncoordinated development of classification schemes by diverse local government organizations or even by different countries. Other causes include the adaptation of those schemes to particular characteristics of the regions that they describe. Therefore, geospatial data sharing and interoperability will require the matching of metadata concepts across a variety of classification schemes.

In our work, classification schemes are represented by ontologies and the matching of concepts is achieved by aligning those ontologies. Ontology alignment encompasses a wide variety of techniques, which include the matching of single concepts [1, 2, 17], the matching of several concepts at a time taking into

---

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

account the structure of the ontologies [14, 16, 18], or even the data associated with the ontological concepts [10, 12, 13]. In this paper, we concentrate on the structure of the ontologies that we want to align.

Two types of architecture can be considered: a *centralized architecture* and a *peer-to-peer architecture*. In the former case, each of the ontologies associated with the heterogeneous data sources is mapped to the global ontology. In the latter case, mappings are established between pairs of ontologies, as needed. In both cases, the ontology from which the mapping is defined is called the *source* and the other ontology is called the *target*. Once a pair of ontologies is mapped, queries posed in terms of one of the ontologies can be automatically translated to the other ontology. A full discussion of these architectures and associated query mechanisms has been discussed elsewhere [5, 8].

As ontologies grow in size or the number of ontologies grows, their alignment should ideally be automatic or require minimum user intervention. Much attention has been recently placed on the automatic alignment of ontologies. For example, the Ontology Alignment Evaluation Initiative (OAEI) [19] promotes the comparison of automatic alignment methods by publishing every year sets of ontologies so as to compare the effectiveness (in terms of recall and precision) of the methods proposed by the contestants. Each set contains a source ontology, a target ontology, and the expected alignment results between them.

In our previous work, we have explored ontology alignment for geospatial applications [4–6], leading to a multi-layered approach, consisting currently of four layers [7]. Two of these layers use automatic methods, one uses a semi-automatic method, and the other one uses only a manual method. The overall process is supervised by a domain expert.

In the first layer, an *automatic mapping by definition* process is undertaken that compares each concept in the first ontology to each concept in the second ontology according to their definition, as provided by a dictionary. A similarity measure from 0% (no match) to 100% (exact match) between the concepts being compared is returned. If a dictionary is not consulted, the procedure will be performed by comparing only the concept names and any associated descriptions or properties of the concepts.

In this paper, we propose an enhancement to our first layer of mapping by introducing two (fully) automatic structure-based methods: the *Descendants' Similarity Inheritance (DSI)* method, which uses the relationships between ancestor concepts, and the *Sibling's Similarity Contribution (SSC)* method, which uses the relationships between sibling concepts.

Our chosen application domain of wetlands demonstrates the importance of ontology alignment in the geospatial domain. Organizations monitoring the wetlands data inventory have an interest in sharing data. The lack of standard classification has long been identified as an obstacle to the development, implementation, and monitoring of wetland conservation strategies both at the national and regional levels [9]. In defining wetlands, the United States adopts the “Cowardin” Wetland Classification System [3]. In contrast, European nations use the International Ramsar Convention Definition ([www.ramsar.org](http://www.ramsar.org)) and

South Africa uses the National Wetland Classification Inventory [9]. Most classifications recognize the need for regionalization because of the variations in climate, geology, soils, and vegetation. Regionalization is designed to facilitate three activities: (i) planning, where it is necessary to study management problems and potential solutions on a regional basis, (ii) organization and retrieval of data gathered in a resource inventory, and (iii) interpretation of inventory data, including differences in indicator plants and animals among the regions. It can thus be concluded that it is extremely difficult to have a standardized classification system between nations and also between regions of a country with a large geographic area [3].

We implemented our proposed methods and tested them against our previous technique [7], which provides us with a “base case”. In addition, we tested our methods against the implementation of a structure-based algorithm, the Similarity Flooding algorithm by Melnik *et al.* [14]. Our experiments involve aligning five pairs of ontologies. In particular, we have covered in detail the alignment of ontologies describing the classification schemes of wetlands, so as to illustrate the main principles that underly our structure-based methods. Our experiments show that at least one of our structure-based methods is as effective or better than both our base case method and the Similarity Flooding algorithm.

The rest of this paper is organized as follows. In Section 2, we give an overview of related work in the area. We present a brief description of our multi-layered approach to ontology alignment and an overview of our alignment tool in Section 3. In Section 4, we present our automatic structure-based methods that support the first layer of mapping in our multi-layered approach along with the experimental results of applying these methods on five ontology sets. Finally, in Section 5, we draw conclusions and outline future work.

## 2 Related Work

In their survey paper, Shvaiko and Euzenat [19] provide a comparative review of recent schema and ontology matching techniques in the context of a new classification system they propose, where the techniques are classified as element level or structure level. In the element level category, the techniques can be based on strings, language, linguistic considerations, constraints, or alignment reuse. In the structure level category, the techniques are further classified as graph-based, taxonomy-based, or model-based. In order to derive mappings between concepts during the alignment process, the element level techniques consider the labels of concepts, their definitions, the language they are expressed in, and any possibility to reuse previous mappings to derive new ones. The structure level techniques consider the location of the concept in the ontology structure (e.g., tree, graph) and how the mappings of concepts can contribute to the mappings of adjacent concepts. According to their classification system, our alignment techniques fall into their element level category because of our definition mapping layer (base technique), and structure level category because of our mapping by context layer [7] and of the new methods proposed in this paper.

OLA is an alignment tool, whose main purpose is to align ontologies expressed in OWL [11]. OLA offers parsing and visualization of OWL-Lite and OWL-DL ontologies. In addition it offers similarity computations between concepts of the ontologies being aligned. OLA employs linguistic element level and structure level techniques and supports both manual mappings and automated mappings. The available knowledge about the concepts in the aligned ontologies is taken into consideration prior to the alignment process by allowing appropriate alignment methods to be chosen. OLA tries to achieve the highest level of automation, by letting users provide a minimal set of parameters at the initial steps of the alignment process and then leaving it to the tool to end the alignment. Unlike in our approach, similarities between concepts do not contribute to the similarities of their neighbors.

RiMOM (Risk Minimization based Ontology Mapping) is a system that intends to combine different strategies to achieve optimal alignment from a source ontology to a target ontology [21]. There are two types of defined strategies in the system: linguistic-based techniques (includes edit-distance and statistical-learning), and structure-based techniques (includes similarity-propagation, property-to-property propagation, and concept-to-property propagation). RiMOM first examines the structural similarity of the ontologies and the label similarity of the concepts in the ontologies to determine which strategies to use in the alignment process. For example, if there is high similarity in the labels, RiMOM will rely more on linguistic-based strategies to find the matchings between concepts. RiMOM then applies the selected alignment strategies; each strategy outputs its own independent results and the results are then combined using a linear-interpolation method. Finally, RiMOM applies a refinement procedure to prune alignments that are not considered good. Compared to our approach, we are also using multiple matching techniques and allowing for the determination of which techniques will play a more important role for each matching. However, we offer structure level matching, whereas RiMOM does not.

Silva *et al.* discuss the situation when different mapping agents establish different semantic bridges between the concepts in the source and target ontologies [20]. Due to the inherent and subjective nature of ontologies, different agents establish different semantic bridges for the same set of ontologies. This may cause conflicts. To address this issue, they propose an approach to ontology mapping negotiation where various agents are able to achieve consensus among them. In our approach, multiple alignment layers are supported, such that each layer proposes a set of mappings between the source ontology and the target ontology. In our case, a consolidation mapping layer is applied where it is up to a mapping expert to specify the priority scheme across the different layers.

Melnik *et al.* propose a simple structural model-based level technique, the *Similarity Flooding* algorithm, that can be used in matching a variety of data structures (referred to as models) [14]. Models can be data schemas, data instances, or a mixture of both. In their approach, models are converted to directed labeled graphs. For their algorithm to work, they rely on the fact that concepts from the two graphs are similar when their adjacent concepts on the graphs are

similar. The algorithm starts by obtaining initial mappings between concepts in the two input graphs using a string matching function that returns initial similarities between matched concepts. Having established the initial mappings, the algorithm proceeds iteratively to establish more mappings between other concepts based on the assumption that whenever any two concepts in the input models match with some similarity measure, the similarity of their adjacent concepts increases. The iterations continue “flooding” the similarities across the concepts in the graphs until a fixed point is reached where similarity measures for all concepts have been stabilized. Of the matching techniques that we surveyed, this one is the closest to our vision of what a structure level approach should be, hence we have implemented their algorithm so as to compare its results with those of the methods that we propose in this paper.

### 3 The AgreementMaker Framework

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 mapping layer [7]. As mentioned in Section 1, 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.

Our mapping layers use element-based alignment techniques (first layer) and structure-based alignment techniques (first and third layers). In addition, domain experts can use their knowledge and contribute to the alignment process (second and third layers).

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. 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 [7].

### 4 Automatic Similarity Methods

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 three matching methods: (1)

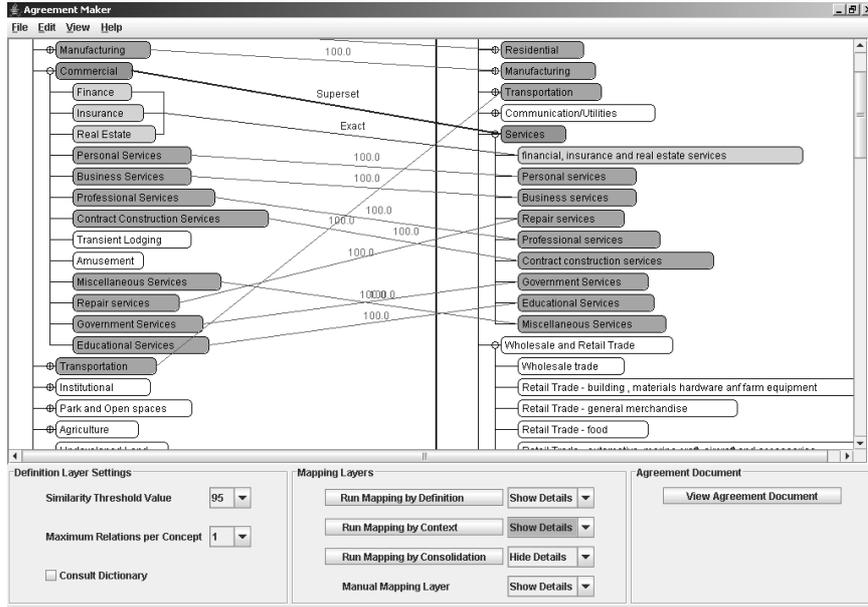


Fig. 1. Results of running three of the mapping layers.

applying the base similarity calculations only, (2) applying the base similarity calculations followed by the Descendant’s Similarity Inheritance (*DSI*) method, or (3) applying the base similarity calculations followed by the Sibling’s Similarity Contribution (*SSC*) method. Both the *DSI* and the *SSC* methods have been introduced to enhance the alignment results that were obtained from using the base similarity method previously proposed [7]. We apply one of these methods in our first mapping layer.

#### 4.1 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 both the *DSI* and *SSC* methods. 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 definitions as provided by a dictionary [7]. 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.,  $base\_sim(C, C') \geq TH$ ).

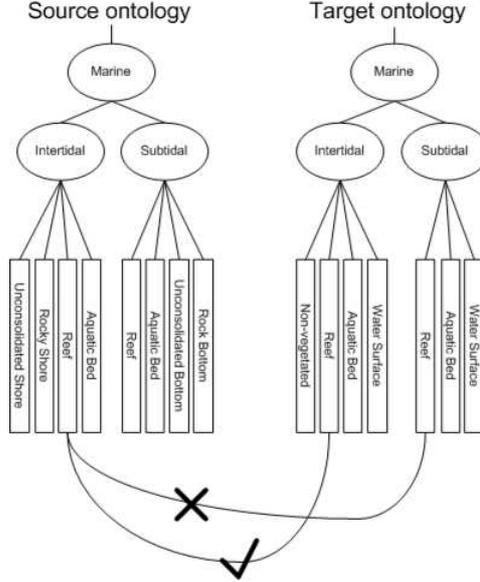
Establishing base similarities between concepts of the source ontology and concepts of the target ontology may not be sufficient to achieve a high degree of precision in relating concepts in the two ontologies. To exemplify this point, we give an example in the domain of wetland classification. The first ontology describes the “Cowardin” wetland classification system and the second ontology describes the South African wetland classification system. 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.

## 4.2 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  $i$ th 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  $parent_1(C') = B'$ .



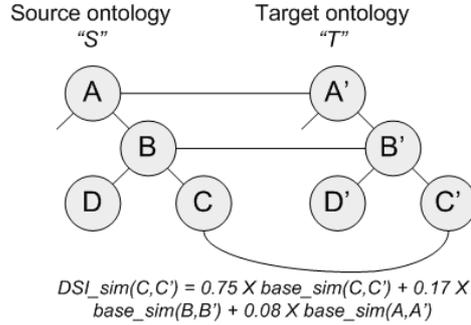
**Fig. 2.** An example of a case where misleading mappings may occur when two concepts have the same label.

- 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  $DSLsim(C, C')$ .
- We compute  $DSLsim(C, C')$  as follows:

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

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

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



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

that 75% for the value of the *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.

### 4.3 Sibling's Similarity Contribution (SSC) method

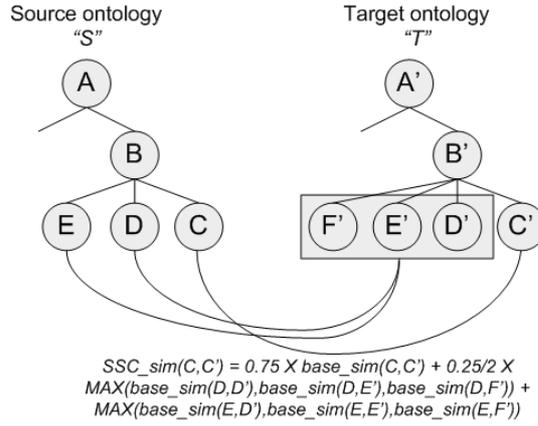
In this method, siblings of a concept contribute to the identification of the concept. This may further enhance the quality of the automatic alignment process. Similarly to the *DSI* method, the *SSC* method reconfigures the base similarities between concepts. We define the *SSC* similarity between a concept  $C$  in  $S$  and a concept  $C'$  in  $T$  as  $SSC\_sim(C, C')$ . In what follows, we present the details on how to determine this similarity.

- Let  $sibling\_count(C)$  be the number of sibling concepts of concept  $C$  in  $S$ . For example, in Figure 4,  $sibling\_count(C) = 2$ .
- Let  $sibling\_count(C')$  be the number of sibling concepts of concept  $C'$  in  $T$ . For example, in Figure 4,  $sibling\_count(C') = 3$ .

- Let  $SS(C)$  be the set of all the concepts that are siblings of  $C$  in  $S$  and  $SS(C')$  be the set of all the concepts that are siblings of  $C'$  in  $T$ .
- Let  $S_i$  be the  $i$ th sibling of concept  $C$  where  $S_i \in SS(C)$ , and  $1 \leq i \leq sibling\_count(C)$ .
- Let  $S'_j$  be the  $j$ th sibling of concept  $C'$  where  $S'_j \in SS(C')$ , and  $1 \leq j \leq sibling\_count(C')$ .
- 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  $SSC\_sim(C, C')$ .
- If both  $SS(C)$  and  $SS(C')$  are not empty, we define  $SSC\_sim(C, C')$  as follows:

$$MCP \cdot base\_sim(C, C') + \frac{1 - MCP}{n} \sum_{i=1}^n \max(base\_sim(S_i, S'_1), \dots, base\_sim(S_i, S'_m))$$

where  $n = sibling\_count(C)$  and  $m = sibling\_count(C')$ .



**Fig. 4.** Applying the SSC method to calculate the similarity between  $C$  and  $C'$ .

The main characteristic of the  $SSC$  method is that it allows for the siblings of a given concept to play a role in the identification process of the concept. In Figure 4 we show how the  $SSC$  similarity is determined between the concept  $C$  in the source ontology  $S$  (shown on the left) and the concept  $C'$  in the target ontology  $T$  (shown on the right) when applying the  $SSC$  method with an  $MCP$  value of 75%. The  $SSC$  similarity is determined by adding 75% of the base similarity between  $C$  and  $C'$  to (1) 12.5% of the maximum base similarity between  $D$  and  $D'$ ,  $D$  and  $E'$ , and  $D$  and  $F'$  and to (2) 12.5% of the maximum base similarity between  $E$  and  $D'$ ,  $E$  and  $E'$ , and  $E$  and  $F'$ . As for the  $DSI$  method, the value of 75% for the  $MCP$  factor was found to work well in practice.

Ontology set	Depth	Number of concepts in the source ontology	Number of concepts in the target ontology
<i>Wetlands</i>	5	29	29
<i>Weapons</i>	6	153	213
<i>People and pets</i>	4	65	93
<i>Computer networks</i>	5	90	89
<i>Russia</i>	5	86	87

Table 1. Depth and number of concepts in the ontology sets.

Algorithm	Ontology1	Ontology2	Ontology3	Ontology4	Ontology5
Base Similarity	125	1516	236	391	484
DSI	156	3656	562	579	844
SSC	172	4344	938	719	1891
Similarity Flooding	187	2266	703	906	1796

Table 2. Performance results for the base similarity, DSI, SCC, and Similarity Flooding algorithms in milliseconds.

#### 4.4 Evaluation

To validate our approach from the point of view of efficiency and of effectiveness, we have aligned the two geospatial wetland ontologies mentioned in Section 4.2 using our own base similarity method, the *DSI* method, and the *SSC* method. We have also used our implementation of the Similarity Flooding algorithm in the alignment of the set of wetland ontologies. In addition, to further evaluate our methods, we run experiments on the alignment of four sets of ontologies provided by the Ontology Alignment Evaluation Initiative (OAEI) [19]. Of these, the first set contains two ontologies describing classifications of various weapon types, the second set contains two ontologies describing attributes of people and pets, the third set contains two ontologies describing classifications of computer networks and equipments, and, finally, the fourth set contains general information about Russia. Each set contains a source ontology, a target ontology, and the expected alignment results between them. Table 1 displays the depth and number of concepts in the five ontology sets we consider.

Similarly to the Similarity Flooding algorithm [14], both our *DSI* and *SSC* methods depend on establishing initial similarities between concepts before they can be executed. However, unlike the Similarity Flooding algorithm, our *DSI* and *SSC* methods do not run in multiple iterations that keep reconfiguring the similarities between concepts until the similarities become stable.

We conducted experiments to determine the running time of all the four methods (base similarity, *DSI*, *SSC*, and Similarity Flooding) for the previously mentioned five ontology sets. We have implemented all the methods using Java and have run them on an 1.6 GHz Intel Centrino Duo with 1GB of RAM, running Windows XP. The results are shown in Table 2.

Looking at the performance results of Table 2, the running time for the *DSI*, *SSC*, and Similarity Flooding algorithms include the running time for the base similarity method because they rely on it to run. Therefore, the base similarity algorithm takes the least amount of time. Examining the results with the exclusion of the base similarity method, the *DSI* method has the best run time performance for four of the test cases, while the Similarity Flooding algorithm has the best running time performance for one test case only. The *SSC* method has the worst performance in three test cases while it performs better than the Similarity Flooding algorithm in two test cases. As compared to the Similarity Flooding algorithm, the *DSI* method only runs once to complete, whereas the Similarity Flooding algorithm will need several iterations to complete. The *SSC* method depends on the number of siblings for a given concept, therefore the larger the number of siblings the worse it performs. In other words, if the ontology trees are wide, then the performance of *SSC* will suffer. Similarly, the running time of the *DSI* method degrades for deep ontology trees. In our future work we are planning to examine ways to improve the running time of the *DSI* and *SSC* methods.

To compare the effectiveness of the four methods, we started by aligning the set of ontologies for the wetlands as described in Section 4.2 and did the same for the other four sets of ontologies. In the wetlands example, we have captured the number of discovered relations between the concepts of the source ontology (“Cowardin”) and the concepts of the target ontology (“South African”) for each method. Each relationship represents a mapping from a concept  $C$  in the source ontology  $S$  to a matching target ontology concept  $C' \in MS(C)$  with the highest similarity measure. We note that there may be concepts in  $S$  that are not mapped to any concepts in the target ontology (corresponding to an empty mapping set). After capturing the discovered relations, we count how many of these relations are valid when compared with the expected alignment results. Having figured the number of correct relations, we calculate both the precision and the recall values. The precision is calculated by dividing the number of discovered valid relations to the total number of discovered relations, the recall is calculated by dividing the number of discovered valid relations to the total number of valid relations as provided by the expected alignment results.

In the alignment of the wetland ontologies, the *DSI* method yielded slightly higher precision and recall values than the Similarity Flooding algorithm which in turn yielded higher values than the *SSC* method. Overall, these three methods significantly enhanced the precision and recall values obtained by applying the base similarity method only. Table 3 shows the complete results for this test case. The following tests pertain to the four sets of ontologies of the OAEI initiative. In the alignment of the ontologies in the first OAEI set (Weapons), the *DSI* method yielded slightly higher precision and recall values than both the *SSC* and the Similarity Flooding methods as shown in Table 4.

All four methods yielded the same results for recall and precision in the alignment of the second OAEI set (People and pets) as shown in Table 5. This is

Algorithm	Total correct relations	Discovered relations	Correct relations	Precision	Recall
Base similarity	54	39	24	61.54%	44.44%
DSI	54	39	37	94.87%	68.52%
SSC	54	39	29	74.36%	53.70%
Similarity flooding	54	39	36	92.31%	66.67%

**Table 3.** Applying the base similarity, *DSI*, *SSC*, and Similarity Flooding algorithms to align the geospatial wetland ontologies.

Algorithm	Total correct relations	Discovered relations	Correct relations	Precision	Recall
Base similarity	73	78	64	82.05%	87.67%
DSI	73	78	66	84.62%	90.41%
SSC	73	78	65	83.33%	89.04%
Similarity flooding	73	78	65	83.33%	89.04%

**Table 4.** Applying the base similarity, *DSI*, *SSC*, and Similarity Flooding algorithms on the ontology set describing weapons.

Algorithm	Total correct relations	Discovered relations	Correct relations	Precision	Recall
Base similarity	74	81	49	60.49%	66.22%
DSI	74	81	49	60.49%	66.22%
SSC	74	81	49	60.49%	66.22%
Similarity flooding	74	81	49	60.49%	66.22%

**Table 5.** Applying the base similarity, *DSI*, *SSC*, and Similarity Flooding algorithms on the ontology set describing people and pets.

Algorithm	Total correct relations	Discovered relations	Correct relations	Precision	Recall
Base similarity	29	23	17	73.91%	58.62%
DSI	29	23	14	60.87%	48.28%
SSC	29	23	19	82.61%	65.52%
Similarity flooding	29	23	18	78.26%	62.07%

**Table 6.** Applying the base similarity, *DSI*, *SSC*, and Similarity Flooding algorithms on the ontology set describing computer networks.

an indication that the locality of all the concepts in the ontologies of the second set are irrelevant in distinguishing their identity.

The *SSC* method yielded better recall and precision results than the Similarity Flooding algorithm, which in turn yielded better results than the *DSI* method when aligning the third OAEI set (Computer networks) as shown in Table 6. Finally, as shown in Table 7, in the alignment of the fourth OAEI set (Russia), the *DSI* method yielded the highest results for precision and recall than either the *SSC* method or the Similarity Flooding algorithm.

Algorithm	Total correct relations	Discovered relations	Correct relations	Precision	Recall
Base similarity	117	51	45	88.24%	38.46%
DSI	117	51	48	94.12%	41.03%
SSC	117	51	46	90.20%	39.32%
Similarity flooding	117	51	46	90.20%	39.32%

**Table 7.** Applying the base similarity, *DSI*, *SSC*, and Similarity Flooding algorithms on the ontology set about Russia.

The differences found in the recall and precision values for a given method when applied across different test cases are mainly due to the characteristics of the ontologies. For example, in the first OAEI set (Weapons) and the second OAEI set (People and pets), the relations between the concepts, their parents, and their siblings do not contribute to refining the base similarity results. However, the relationships between the concepts and their siblings added value in refining the base similarity results when aligning the third OAEI set (Computer networks). The relationships between the concepts and their parents added value in refining the results when aligning the fourth OAEI set (Russia). Therefore, the selection of an appropriate matching method should be done after a preliminary examination of the concepts in the ontologies and how they relate to each other. Mochol *et al.* [15] present a methodology on how to select an appropriate matching method for a specific alignment case by having a domain expert fill a questionnaire about the nature of the ontologies to be aligned.

## 5 Conclusions

The subject of automatic ontology alignment has been receiving a lot of attention recently. In this paper, we have proposed two methods that will enhance our multi-layer approach to ontology alignment, which is supported by a visual interface. Our methods use the structure of the ontology graph for contextual information thus providing the matching process with more semantics.

The two methods that we propose, the *Descendants' Similarity Inheritance (DSI)* method and the *Sibling's Similarity Contribution (SSC)* method use respectively the information associated with the descendants and with the siblings of each concept. Our main test case is provided by a geospatial domain application for wetlands. Other ontologies were also tested in the spirit of the Ontology Alignment Evaluation Initiative (OAEI) [19], which currently does not include geospatial ontologies in their repository of ontologies, but is now widely regarded as the repository with which to study the effectiveness of ontology alignment methods. The pairs of ontologies in the OAEI repository have associated with them the correct mappings that should be derived by any automatic alignment method, thus enabling an objective effectiveness comparison.

In addition to implementing our own methods, we have also implemented the Similarity Flooding algorithm [14] and tested our new methods against: (1) our

base technique that uses a similarity comparison among individual concepts and (2) the Similarity Flooding algorithm. The experimental results show that from an effectiveness viewpoint at least one of our new methods is as good or better than the results obtained with the previously proposed methods.

Much work remains to be accomplished in the general area of ontology alignment and in the particular area of geospatial ontology alignment. A research subject involves the determination of which methods to use depending on the ontologies involved and on their particular topologies. For example, the fact that the most effective method is not always the same and that sometimes all the four methods have similar results shows that: (1) the best method depends on the topology of the ontology graph and (2) for certain topologies, structure-based methods do not play an important role. Both of these conclusions have been arrived at by others [15] and they further justify our multi-layered approach where several techniques can be used and combined [7].

The knowledge of the best method to apply will directly impact our consolidation layer in which priority weights are given to the different matching layers. If such priority weights can be automatically determined, then our overall approach will further attain automation. Another subject of research would be the “fusion” in the same method of different techniques (e.g., *DSI*, *SCC*, and Similarity Flooding), where such fusion could be guided again by the characteristics of the topologies at hand. A comparison of these two alternatives can then be undertaken.

Many more test cases and studies are needed: the introduction of geospatial ontologies in the OAEI repository will allow for a wide variety of researchers to explore their methods in the geospatial domain; also, there is the need for many more geospatial ontologies to become available. In particular, initiatives such as the Open Geospatial Consortium (<http://www.opengeospatial.org/>) will likely bring about a plethora of standardized and much larger ontologies that must be semantically aligned to promote data sharing and interoperability.

## Acknowledgments

We would like to thank Sarang Kapadia for his help with the implementation of the Similarity Flooding algorithm.

## References

1. S. Bergamaschi, S. Castano, and M. Vincini. Semantic Integration of Semistructured and Structured Data Sources. *SIGMOD Record*, 28(1):54–59, 1999.
2. S. Castano, V. D. Antonellis, and S. D. C. di Vimercati. Global Viewing of Heterogeneous Data Sources. *IEEE Transactions on Knowledge and Data Engineering*, 13(2):277–297, 2001.
3. 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.
4. I. F. Cruz, A. Rajendran, W. Sunna, and N. Wiegand. Handling Semantic Heterogeneities using Declarative Agreements. In *ACM GIS 10th International Symposium on Advances in Geographic Information Systems*, pages 168–174, 2002.
  5. I. F. Cruz, W. Sunna, and K. Ayloo. Concept Level Matching of Geospatial Ontologies. In *GISPlanet Second Conference and Exhibition on Geographic Information*, Estoril, Portugal, 2005.
  6. I. F. Cruz, W. Sunna, and A. Chaudhry. Semi-Automatic Ontology Alignment for Geospatial Data Integration. In *3rd Int. Conf. on Geographic Information Science (GIScience)*, volume 3234 of *Lecture Notes in Computer Science*, pages 51–66. Springer, 2004.
  7. 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.
  8. I. F. Cruz and H. Xiao. The Role of Ontologies in Data Integration. *Journal of Engineering Intelligent Systems*, 13(4):245–252, December 2005.
  9. 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\\_classif.htm](http://www.ngo.grida.no/soesa/nsoer/resource/wetland/inventory_classif.htm), 1998.
  10. A. Doan, J. Madhavan, P. Domingos, and A. Y. Halevy. Learning to Map between Ontologies on the Semantic Web. In *11th International World Wide Web Conference (WWW)*, pages 662–673, 2002.
  11. J. Euzenat, P. Guégan, and P. Valtchev. OLA in the OAEI 2005 Alignment Contest. In *Integrating Ontologies '05, K-CAP Workshop on Integrating Ontologies*, volume 156 of *CEUR Workshop Proceedings*, Banff, Canada, 2005.
  12. D. Fossati, G. Ghidoni, B. D. Eugenio, I. F. Cruz, H. Xiao, and R. Subba. The Problem of Ontology Alignment on the Web: a First Report. In *2nd Web as Corpus Workshop (associated with the 11th Conference of the European Chapter of the ACL)*, 2006.
  13. R. Ichise, H. Takeda, and S. Honiden. Rule Induction for Concept Hierarchy Alignment. In *Workshop on Ontologies and Information Sharing at the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, 2001.
  14. S. Melnik, H. Garcia-Molina, and E. Rahm. Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching. In *18th International Conference on Data Engineering (ICDE)*, 2002.
  15. M. Mochol, A. Jentzsch, and J. Euzenat. Applying an Analytic Method for Matching Approach Selection. In *International Workshop on Ontology Matching (OM-2006) collocated with the 5th International Semantic Web Conference (ISWC), Athens, Georgia, USA, 2006*.
  16. N. F. Noy and M. A. Musen. Anchor-PROMPT: Using Non-local Context for Semantic Matching. In *Workshop on Ontologies and Information Sharing at the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, 2001.
  17. L. Palopoli, D. Saccà, and D. Ursino. An Automatic Techniques for Detecting Type Conflicts in Database Schemes. In *7th International Conference on Information and Knowledge Management (CIKM)*, pages 306–313, 1998.

18. M. A. Rodríguez and M. J. Egenhofer. Determining Semantic Similarity among Entity Classes from Different Ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 15(2):442–456, 2003.
19. P. Shvaiko and J. Euzenat. A Survey of Schema-Based Matching Approaches. In *J. Data Semantics IV*, volume 3730 of *Lecture Notes in Computer Science*, pages 146–171. Springer, 2005.
20. N. Silva, P. Maio, and J. Rocha. An Approach to Ontology Mapping Negotiation. In *Integrating Ontologies '05, K-CAP Workshop on Integrating Ontologies*, volume 156 of *CEUR Workshop Proceedings*, Banff, Canada, 2005.
21. J. Tang, J. Li, B. Liang, X. Huang, Y. Li, and K. Wang. Using Bayesian Decision for Ontology Mapping. *Journal of Web Semantics*, 4(4):243–262, 2006.