ABSTRACT

ONTOMETRY ALIGNMENT USING SEMANTIC SIMILARITY WITH REFERENCE ONTOLOGIES

by Pramit Silwal

Measuring the similarity between concepts in two different ontologies can be done in many different ways as evidenced by the numerous kinds of matchers found in ontology alignment (OA) systems. Some produce a mapping between two concepts in different ontologies by finding an identical bridge concept in a reference ontology to which both concepts are mapped. A new matcher incorporating semantic similarity measurement within one or more reference ontologies has been integrated into the AgreementMaker OA system. Experiments using the Ontology Alignment Evaluation Initiative (OAEI) anatomy track are performed with both the Uberon and the FMA ontologies as reference ontologies. The results of these experiments are compared to the OAEI 2011 and the 2012 results for the anatomy track and show this approach performs better than most other OA systems in the OAEI competition. This thesis demonstrates that using semantic similarity measures within a reference ontology can improve the ontology alignment process.
ONTOMETRY ALIGNMENT USING SEMANTIC SIMILARITY WITH REFERENCE ONTOLOGIES

A Thesis

Submitted to the
Faculty of Miami University
in partial fulfillment of
the requirements for the degree of
Master of Science

Department of Computer Science and Software Engineering

by
Pramit Silwal

Miami University

Oxford, Ohio

2012

Advisor________________________________
Valerie Cross, PhD.

Reader________________________________
Michael Zmuda, PhD.

Reader________________________________
James Kiper, PhD.
Contents

1. Introduction ................................................................................................................................................. 1
2. Ontology Alignment ....................................................................................................................................... 3
   2.1 AgreementMaker : A Competitive OA System .......................................................................................... 3
   2.2 AgreementMaker : A Competitive OA System .......................................................................................... 5
   2.3 Evaluation of Ontology Alignment Systems ............................................................................................ 6
3. Related Research ........................................................................................................................................... 8
   3.1 OA Systems Using Semantic Similarity ................................................................................................ 8
   3.2 Two Current OA Systems Using Reference Ontologies .......................................................................... 9
4. Mediating Matcher with Semantic Similarity (MMSS) .................................................................................. 11
5. Initial Experimental Investigations ............................................................................................................ 12
6. Experiments with Variations of LWC Matchers ...................................................................................... 18
7. Overriding PSM and Increasing the MMSS Semantic Similarity Threshold ............................................. 20
8. Using Another Intermediate Ontology ...................................................................................................... 24
9. Conclusions and Future Research ........................................................................................................... 29
10. References ................................................................................................................................................ 32
List of Tables

Table 1: OA Systems using Semantic Similarities ................................................................................................................................. 9
Table 2: Experimental Results with the OAEI Anatomy Track using Uberon ontology ................................................................. 15
Table 3: New Mappings, OAEI MMSS but not OAEI MM .................................................................................................................. 16
Table 4: Lost Mappings, OAEI MM but not OAEI MMSS .................................................................................................................... 17
Table 5: Various Matcher Combinations for LWC Experimental Results ............................................................................................................. 19
Table 6: LWC2 Preserving PSM 1.0 confidence mapping ....................................................................................................................... 21
Table 7: 0.95 Threshold and Combined with Preserving PSM 1.0 confidence mappings ........................................................................ 23
Table 8: OAEI 2012 Anatomy Track for OA systems with F-measure > 0.800 ....................................................................................... 23
Table 9: Comparing bridge mappings for FMA ontology .................................................................................................................. 26
Table 10: Comparison of FMA and Uberon results ............................................................................................................................. 27
Table 11: Combining two mediating ontologies .............................................................................................................................. 29
List of Figures

Figure 1: Using a mediating ontology in Ontology Alignment adapted from (Cruz et al. 2011)............................2
Figure 2: Architecture of Ontology Alignment (Ehrig 2006).................................................................................4
Figure 3: AgreementMaker Architecture (Cruz et al., 2009).................................................................................6
Figure 4: MMSS algorithm........................................................................................................................................12
Figure 5: OAEI 2011 Anatomy track Matcher configuration .....................................................................................13
Figure 6: OAEI 2011 Anatomy track Matcher configuration with MMSS.................................................................14
ACKNOWLEDGEMENTS

I would like to give my sincere thanks and appreciation to Prof. Valerie Cross for her advice, support and constant encouragement. Without her guidance, I could not have finished this work. I also want to thank my thesis committee members, Prof. Michael Zmuda and Prof. James Kiper for their time in reviewing this work. I would also like to thank Cosmin Stroe from UIC for his support with the Agreementmaker software. Lastly I would like to thank my family and friends for their constant encouragement and support.
1. Introduction

An ontology is a method of formal knowledge representation related to any subject domain. It defines various concepts pertaining to entities from that subject domain and the relationships between them and thus describes the semantics of a given domain formally. Ontologies facilitate automatic discovery of knowledge by search agents across a distributed environment. Ontologies have not only been used extensively in the development of the Semantic Web but also in research in a variety of fields such as information retrieval, artificial intelligence, natural language processing, biomedical informatics, etc.

A single massive ontology modeling the whole world would not only be challenging but since any ontology is based on expert knowledge, it would be impractical to create one. Thus, ontologies are usually created by experts or organizations in order to model their subject domain of interest. As a consequence there are numerous ontologies in existence. Even for the same domain, two ontologies can be defined differently by different experts depending on their own perspectives of the domain. This heterogeneity can create problems for Semantic Web applications that build on top of ontologies and rely on integration of semantic information from multiple ontologies. Such integration is sometimes necessary for performing sophisticated tasks such as semantic web searching.

Ontology alignment (OA) supports interoperability between ontologies on the Semantic Web. The objective of OA is to determine correspondences between the entities (concepts and relationships) in two separate ontologies. It typically produces a set of mapping pairs \((s_i, t_i)\) between the source ontology \(O_S\) and target ontology \(O_T\) and each mapping has a similarity degree, also referred to as a confidence degree of correspondence in the range \([0, 1]\). Ontology alignment techniques differ in the features of the ontologies that are used in determining the correspondences, for example, the descriptions of the entities such as labels, definitions, and comments and the specific instances of the concepts (Ehrig 2005).
This thesis research contributes to the improvement of the ontology alignment process by investigating the use of semantic similarity measurement within one or more external knowledge resources, typically referred to as intermediate, reference, or mediating ontologies to produce indirect mappings between the source and target ontologies. These ontologies are in the same domain as the source and target ontologies and provide additional information related to the matching task of the alignment process (Cruz et al. 2011).

Figure 1 illustrates the use of an intermediate ontology. The source ontology $O_S$ is aligned to the intermediate ontology $O_I$ (1) using an efficient but simpler matcher process to produce a set of mappings $M_{SI}$. Similarly the target ontology $O_T$ is aligned to the intermediate ontology $O_I$ (2) to produce a set of mappings $M_{TI}$. The recent approaches (Cruz et al. 2011) (Gross et al. 2011) create a set of mappings $M_{ST}$ based on an exact agreement between the bridge concepts, (3). These indirect mappings in $M_{ST}$ can be used when the OA process has not found direct mappings between the two ontologies or to support or improve the confidence of existing direct mappings already discovered in the OA process.

Figure 1: Using a mediating ontology in Ontology Alignment adapted from (Cruz et al. 2011).

This thesis research develops a new method that when $b_s \neq b_t$, uses semantic similarity between $b_s$ and $b_t$ within the intermediate ontology to determine if a possible indirect mapping can be
added into $M_{ST}$ between the corresponding source and target entities for $b_s$ and $b_t$. This new method is integrated into an existing OA system. As part of this research, experiments are performed to evaluate the effectiveness of this new method and compare it to an existing approach. These experiments use two different ontologies as intermediate ontologies and apply the mentioned new method using semantic similarity on an evaluation test case which is a standard track of the Ontology Alignment Evaluation Initiative (OAEI) (Euzenat 2011).

2. Ontology Alignment

First the general process of ontology alignment and its steps are presented. Then a current OA system AgreementMaker (Cruz et al. 2011) is described in more detail. This OA system has been used in previous thesis research and work (Cross et al. 2011), (Cross et al. 2012) and was selected as the software to be extended to perform this current thesis research. Finally, the standard performance measures used to evaluate the results of an ontology alignment system are discussed as well as the Ontology Alignment Evaluation Initiative (OAEI) (Euzenat 2011), a competition which numerous OA systems participate in every year since its official inception in 2005.

2.1 AgreementMaker : A Competitive OA System

The general idea behind ontology alignment is to find the corresponding entities between two ontologies with regards to a certain relationship that can be established between those entities. More formally, ontology alignment can be defined as a function in the following manner (Ehrig 2006):

A general ontology alignment function, genalign, based on the vocabulary, $E$, of all terms $e \in E$ based on the set of possible ontologies, $O$, and based on possible alignment relations, $M$, is a partial function

$$\text{genalign} : E \times O \times O \rightarrow E \times M,$$

with $\forall e \in O1, f \in O2, \exists m \in M : \text{genalign}(e, O1, O2) = (f, m)$. 
The relationship type between two entities denoted by M is not restricted to the equivalence type, but could also be of the generalization or specialization type. Furthermore, mapping cardinalities between entities from O₁ and O₂ include one-to-one, one-to-many and many-to-many mappings. Regardless of the specific details used by an ontology alignment system, the architecture of an ontology alignment process can generally be modeled by the following steps shown in Figure 2.

**Figure 2 : Architecture of Ontology Alignment (Ehrig 2006)**

In figure 2, **input** refers to two or more ontologies which are to be aligned, whereas **output** is the final alignment which is typically represented with an alignment table or in a standard alignment API format. The individual steps are:

1. **Feature Engineering**: Selection of features describing the entities to be compared out of the whole ontology definition, for example, string labels, definitions, comments.
2. **Search Step Selection**: Selection of entity pairs to be compared. While making the selection, it is common to compare only entities of the same type (concepts, instances, properties, etc.) or to compare all entities of first ontology with the ones from the second ontology.
3. **Similarity Computation**: For the chosen pairs of entities, compute some measure of similarity.
4. **Similarity Aggregation**: This step is used to combine multiple similarity values for the same pair of entities if more than one type of similarity computation was employed.
5. **Interpretation**: Initial candidate pairs are determined to have alignment or not based on some screening criteria.
6. Iteration: Alignments created from the previous step can influence the similarity values of neighboring pairs. For example, instances of concepts that are aligned with each other can increase the similarity values of those concepts. Several iterations may produce new alignments and the process stops after certain specified conditions are met, for example, no new alignments can be found.

A variety of OA systems have been developed over the past decade and typically 10 to 15 different systems participate every year in the OAEI competition. For the most part, these systems typically have improved on the previous systems by adding new types of matchers in step 3 of the overall process which calculated the similarity between the concepts in the input ontologies. Examples of some of these matchers are presented in the following section based on those used in AgreementMaker. Other variations in these OA systems occur based on the methods used to aggregate similarity measures and how the mappings are interpreted.

2.2 AgreementMaker: A Competitive OA System

AgreementMaker (Cruz et al. 2011) is an extensible framework for ontology alignment that has been in development for over a decade. AgreementMaker has been performing well in the OAEI competitions from 2009 to 2011. The OAEI competition in 2011 saw AgreementMaker perform better than all of its competitors in the anatomy matching track with respect to standardly used performance measures which are discussed in the following section.

A key feature is its flexibility to combine different matchers in order to handle different matching criteria. This feature allows focusing on development of new matching techniques and their easy integration into the AgreementMaker architecture. The two main categories for AgreementMaker’s matchers are concept-based, which employ multiple string similarity measures, and structural, which search for shared patterns in the hierarchical structure of the ontologies. The concept-based similarities include the Base Similarity Matcher (BSM), the Advanced Similarity Matcher (ASM), the Parametric String-based Matcher (PSM), and the Vector-based Multi-Word Matcher (VMM). The BSM calculates the similarity between two concepts by comparing all the strings associated with those two concepts, those strings include
the concept name, label, and comments. PSM is also a string-based matcher but more complicated since it uses a substring measure and an edit distance measure. VMM compiles a virtual document for every concept of an ontology by concatenating the strings of related concepts and annotations, transforms the resulting strings into TF-IDF vectors, and computes the similarity between those vectors using the cosine similarity measure. The structural matcher is the Descendent’s Similarity Inheritance matcher (DSI). The DSI matcher considers the ancestors of the two concepts \((s_i, t_i)\) in a mapping in order to increase the similarity of the mapping. The DSI matcher uses the heuristic that if two nodes are matched with high similarity, then the similarity between the descendants of those nodes should increase. Figure 3 shows the general architecture of AgreementMaker.

![Figure 3: AgreementMaker Architecture](Cruz et al., 2009).

AgreementMaker version 0.22 extended the string-based matchers by integrating two lexicons: (1) the Ontology Lexicon, built from synonym and definition annotations existing in the ontologies themselves, and (2) the WordNet Lexicon, created by starting with the ontology lexicon and adding any non-duplicated synonyms/definitions found in WordNet. The matchers using the lexicons in their algorithms are annotated with a lex superscript, as in \(\text{BSM}^{\text{leCx}}, \text{PSM}^{\text{lex}},\) and \(\text{VMM}^{\text{lex}}\). The Linear Weighted Combination (LWC) matcher produces a single combined alignment by using mapping quality measures to choose the best mappings from each matcher.

### 2.3 Evaluation of Ontology Alignment Systems


The evaluation of the OA systems is determined using three performance measures adapted from information retrieval evaluation methods. These are precision, recall, and f-measure (Ehrig 2005). A reference alignment is required to determine these performance measures. It is considered as correct alignment result and is also known as the gold standard alignment. The following defines these three performance measures given the reference alignment R and the alignment A produced by the OA system.

Precision is defined as the fraction of output mappings produced by the OA system that are actually correct, i.e., how many of its mappings are found in the reference alignment.

\[ P(A, R) = \frac{|R \cap A|}{|A|} \]

Recall is defined as the fraction of mappings in the reference alignment that are also produced by the OA system, i.e., how many of its mappings are actually found by the OA system.

\[ P(A, R) = \frac{|R \cap A|}{|R|} \]

F-measure: Given a reference alignment R, precision, and recall, the f-measure of some alignment A is given by

\[ F(A, R) = \frac{(b^2 + 1) \cdot P(A, R) \cdot R(A, R)}{b^2 \cdot P(A, R) + R(A, R)} \]

where \( b \) is a parameter that can be set to emphasize either precision or recall or an equal balance.

These performance measures are used in the annual OAEI competition whose purpose is to evaluate, compare, and improve the ontology alignment process. The annual competition has several tracks including the benchmark, the anatomy and the conference tracks. The benchmark track is used to identify the strength and weakness of specific ontology alignment algorithms and is meant to simply determine how well alignment basics are handled by an OA system. The anatomy track uses two real-world ontologies the Adult Mouse Anatomy and NCI human anatomy to see how OA systems perform in more sophisticated ontologies. Conference track which contains 16 ontologies from the domain of conference organization is used to determine how well OA systems can find correct alignments in a set of heterogeneous ontologies from the
same domain.

3. Related Research

First the development of the use of semantic similarity measures with background knowledge sources in several OA systems is reviewed. Then two more recent OA systems that use domain specific ontology as an intermediate or mediating ontologies without the use of semantic similarity are described in detail.

3.1 OA Systems Using Semantic Similarity

A few OA systems have been using semantic similarity with background knowledge resources in the ontology alignment process. Table 1 summarizes the results of an investigation into the use of semantic similarity in the OA process.

<table>
<thead>
<tr>
<th>System</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLA (Euzenat and Valtchev 2003)</td>
<td>A pair of identifiers are first converted into two sets of atomic terms. Then the Wu-Palmer (Wu and Palmer 1994) similarity measure computes semantic similarity between pairs of terms from the set using WordNet. The degree of proximity between the two sets is calculated to find the similarity between two identifiers.</td>
</tr>
<tr>
<td>Imapper (Su 2004)</td>
<td>Use of semantic similarity to increase the similarity value of produced mapping. Descriptive labels of concepts are looked up in WordNet for which simple path based measures are calculated. If there is no path between the terms, similarity value is left unmodified.</td>
</tr>
<tr>
<td>CIDER (Gracia and Mena 2008)</td>
<td>The alignment process incorporates a sense similarity measure that calculates semantic similarity based on the different senses and synonyms of keywords using WordNet.</td>
</tr>
<tr>
<td>ASMOV (Jean-Mary et al. 2009)</td>
<td>The Lin information-content (IC) based measure may be used when external resources such as Wordnet or UMLS are available. The similarity between two concept labels are assigned as follows: identical concepts (1.0), concepts found in identical</td>
</tr>
</tbody>
</table>
WordNet synsets (0.99), concept antonyms (0.0), otherwise compute the Lin semantic similarity.

| UFOme (Pirro and Talia 2010) | Similar to ASMOV, its WordNet similarity matcher uses the Lin similarity between synsets of concept terms when the concepts do not map to the identical lexical concept as represented by its synset. |

Table 1: OA Systems using Semantic Similarities

In most of these systems the primary background knowledge source has been WordNet which has a standard semantic similarity measure library that can be used to produce semantic similarity between its lexical concepts. ASMOV, a commercial OA system has created its own interface for the UMLS ontology to compute semantic similarity between its concepts.

3.2 Two Current OA Systems Using Reference Ontologies

The two primary background knowledge sources used for ontology alignment are the UMLS and WordNet. The simplest use of these two ontologies is to find synonyms for the source concept and synonyms for the target concept in the background knowledge sources. If a common synonym exists between the two, then the similarity or confidence of the source and target concept mapping is set higher.

Two recent research efforts (Gross et al. 2011) (Cruz et al. 2011) have employed the Uberon ontology as a background resource to improve alignment results for the OAEI anatomy track. They differ in the methods used to match source and target ontology concepts to the reference or intermediate ontology concepts (steps 1 and 2 in Figure 1) and also in the aggregation they use to produce the final mapping. Neither uses any kind of semantic similarities within the reference ontology while performing the alignment.

In (Gross et al. 2011) the composition-based OA system first performs intermediate alignment between source $O_S$ and intermediate ontologies $O_I$ and similarly between target $O_t$ and intermediate ontologies $O_I$ to produce sets of mappings $M_{SI}$ and $M_{IT}$. 

9
Any alignments between source and target concepts is found using

\[ M_{ST} = \{(c_S, c_T, \text{aggSim}(\text{mapSim}_{SI}, \text{mapSim}_{IT})) \mid c_S \in O_S, c_I \in O_I, c_T \in O_T: \exists (c_S, c_I, \text{mapSim}_{SI}) \in M_{SI} \land \exists (c_I, c_T, \text{mapSim}_{IT}) \in M_{IT}\} \]

The aggregation operator \(\text{aggSim}\) combines mapping similarities for \(M_{SI}\) and \(M_{IT}\). Different operators could be used. Average was used in their experiments. \(M_{SI}\) and \(M_{IT}\) were determined using linguistic trigram similarity between concept names and synonyms with a 0.8 threshold. In effect, two simple ontology alignments are first performed to create the mappings \(M_{SI}\) and \(M_{IT}\) before composition-based mapping is done. The researchers also suggest that existing mappings found in BioPortal (Noy et al. 2009) could replace these intermediate source-intermediate and intermediate-target ontology mappings.

Four different reference ontologies (FMA, Uberon, RadLex and UMLS) were used in their experiments. The system was able to produce a F-measure of 88.2% for the OAEI anatomy track which exceeded any F-measure achieved in the OAEI 2010 competition for the anatomy track.

AgreementMaker (Cruz et al. 2011) introduced a mediating matcher (MM) for the OAEI2011 competition. Similar in concept to the previous system from (Gross et al. 2011), their mediating matcher uses a lexical similarity matcher to align the mouse anatomy (MA) and the human anatomy portion of NCI (HA) ontologies to a reference ontology. Uberon was selected based on its performance in (Gross et al. 2011). The specific matcher (BSM\textsuperscript{lex}) uses a lexicon to compare all the strings associated with two concepts (namely concept name, label and comments.) Therefore, AgreementMaker’s approach in aligning the source and target ontologies to a reference ontology is more sophisticated than the trigram similarity methods used in (Gross et al. 2011). Both approaches, however, require an exact match on the bridge concept, i.e. \(b_S = b_T\).

In the reported OAEI 2011 results (Euzenat et al. 2011), AgreementMaker had the best performance with respect to the F-measure (91.7%). These results are better than those in (Gross et al., 2011). AgreementMaker used only one reference ontology Uberon while the best results in
(Gross et al. 2011) were based on merging results using all four reference ontologies. Another difference is that AgreementMaker’s final mappings are determined by a hierarchical arrangement of its Linear Weighted Combination (LWC) matchers. A single combined alignment is produced using mapping quality measures to choose the best mappings from each matcher, of which its mediating matcher is only one of these.

4. Mediating Matcher with Semantic Similarity (MMSS)

The motivation for using semantic similarity is to find those source and target concepts that were mapped to different mediating concepts but that have a high semantic similarity, for example, very close related sibling concepts or parent/child concepts. The mediating matcher (MM) from AgreementMaker is extended to incorporate semantic similarity calculations. Figure 4 describes the MMSS algorithm. The following steps provide more details of this process.

**Step 1:** Uses the original mediating matcher to find all source-target alignments where \( b_s = b_T \), i.e., source bridge concept is the same as target bridge concept. This creates a mapping set \( M_{ST} \).

**Step 2:** Determine the set of all the source concepts, \( U_S \) and set of all target concepts, \( U_T \) that did not get directly mapped in step 1.

**Step 3:** For each pair \((s, t)\) in \( U_S \times U_T \), all bridge concepts for \( s \) and \( t \) are considered. Semantic similarity between each pair of the bridge concepts is determined. The maximum is then used.

\[
bridgeSim = \max_{b_s \in O_s \, \text{and} \, b_T \in O_T} \text{semSim}(b_s, b_T).
\]

**Step 4:** The mapping is then added to an enhanced mapping set based on aggregation the confidences of the mappings between the source and target to the intermediate ontology and the bridge similarity as follows. To filter out weak mappings from \( E_{ST} \), a threshold value can be set \( E_{ST} = \{(s, t, \text{agg}(\text{mapSim}_{SI}, \text{mapSim}_{TI}, \text{bridgeSim})) \mid s \in U_S, b_s \in O_s, b_T \in O_T, t \in U_T: \exists (s, b_s, \text{mapSim}_{SI}) \in M_{SI} \land \exists (t, b_T, \text{mapSim}_{TI}) \in M_{TI} \} \).
\[ \text{bridgeSim} = \max \ b_S, \ b_T \in O_i \left( \text{semSim}(b_S, b_T) \right). \]

**Step 5:** MMSS returns \( M_{ST} \cup E_{ST} \).

Figure 4: MMSS algorithm

### 5. Initial Experimental Investigations

After developing the MMSS algorithm and implementing it within AgreementMaker as a new matcher, the first objective was to compare the performance of the MMSS to the AgreementMaker’s MM. For this comparison, an ontology alignment between the mouse anatomy (MA) and NCI human anatomy (MA) ontologies was performed as done in the anatomy track of the OAEI competition.

Different measures can be used for the semantic similarity function \( \text{semSim} \). For the experiments reported below, the standard Lin semantic similarity measure (Lin 1998) is used with information content as defined in (Seco et al., 2004) since it has frequently been used in bioinformatics research and has performed well. An additional similarity threshold value may be set in the
MMSS to eliminate mappings in $E_{ST}$ whose aggregated similarity falls below the specified threshold.

As presented in the overview of AgreementMaker in section 2.2, several matchers are configured using a hierarchical aggregation. Figure 5 illustrates the configuration used by AgreementMaker in the OAEI 2011 anatomy track competition.

![Diagram of Matcher Configuration](image)

Figure 5: OAEI 2011 Anatomy track Matcher configuration

The combination stack of matchers demonstrates that mappings selected by the MM influences the results of the LWC1 and the final mapping results produced by the LWC3; however, LWC2 results do not change based on the use of the MM. AgreementMaker created this runtime hierarchical matcher configuration for the OAEI anatomy track using its ontology-profiling algorithm on the ontology pair to be aligned. The explanation in (Cruz et al 2011) simply states that this configuration first combined more similar matching algorithms and then the final alignment results were produced by combining those previously combined results. For this initial experiment, the aggregation configuration of matchers was left exactly the same and only the MM was replaced with the MMSS as shown in Figure 6.
To be consistent with previous work described in section 3.2, the 2011 OAEI anatomy track was used. The reference alignment $R$ for this track contains 1516 mappings between the Mouse Anatomy (MA) and the Human Anatomy (HA). Before experimenting with the complete OAEI 2011 LWC configuration, a direct comparison of only the MM to the MMSS is performed. Only the mappings produced by the MM are compared to only those produced by the MMSS with varying thresholds used for the aggregated similarity so that AgreementMaker’s LWC configuration of matchers does not affect these results. The results of this MM to MMSS comparison are listed in the five rows before the row labeled OAEI 2011 in Table 2. The thresholds used are listed after the MMSS label in each row.

<table>
<thead>
<tr>
<th></th>
<th>Mapped</th>
<th>Correct</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MM$</td>
<td>1200</td>
<td>1143</td>
<td>95.2</td>
<td>75.4</td>
<td>84.2</td>
</tr>
<tr>
<td>$MMSS, 0.0$</td>
<td>1322</td>
<td>1152</td>
<td>87.1</td>
<td>76</td>
<td>81.2</td>
</tr>
<tr>
<td>$MMSS, 0.65$</td>
<td>1301</td>
<td>1151</td>
<td>88.5</td>
<td>75.9</td>
<td>81.7</td>
</tr>
<tr>
<td>$MMSS, 0.85$</td>
<td>1240</td>
<td>1150</td>
<td>92.7</td>
<td>75.9</td>
<td>83.5</td>
</tr>
</tbody>
</table>

Figure 6: OAEI 2011 Anatomy track Matcher configuration with MMSS
<table>
<thead>
<tr>
<th>Method</th>
<th>Matches</th>
<th>Correct</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSS, 0.90</td>
<td>1229</td>
<td>1148</td>
<td>93.4</td>
<td>75.7</td>
<td>83.6</td>
</tr>
<tr>
<td>MMSS, 0.95</td>
<td>1209</td>
<td>1147</td>
<td>94.9</td>
<td>75.7</td>
<td>84.2</td>
</tr>
<tr>
<td>OAEI 2011</td>
<td>1443</td>
<td>1350</td>
<td>93.6</td>
<td>89.1</td>
<td>91.2</td>
</tr>
<tr>
<td>MM</td>
<td>1447</td>
<td>1348</td>
<td>93.2</td>
<td>88.9</td>
<td>91.0</td>
</tr>
<tr>
<td>MMSS, 0.85</td>
<td>1447</td>
<td>1350</td>
<td>93.3</td>
<td>89.1</td>
<td>91.1</td>
</tr>
<tr>
<td>MMSS, 0.90</td>
<td>1447</td>
<td>1350</td>
<td>93.4</td>
<td>75.7</td>
<td>83.6</td>
</tr>
</tbody>
</table>

Table 2: Experimental Results with the OAEI Anatomy Track using Uberon ontology

The first row after the heading of table 2 is the performance of the unmodified mediating matcher, whereas the next four rows show the performance of MMSS using various threshold values. As can be clearly seen, the MMSS finds more correct mappings in the anatomy track at any threshold value and, therefore, produces a higher recall value than the MM. However, the precision and consequently the F-measures suffer due to the MMSS producing more incorrect mappings. As the threshold increases, the MMSS is still able to find more correct mappings than the MM and also improve its precision.

The three rows after the row labeled OAEI 2011 compare the two different mediating matchers within the configuration of matchers used in the OAEI 2011 anatomy track competition. This comparison investigates the interaction between the mappings of the MMSS at different threshold values and those produced by the other OAEI 2011 matchers as well as the effects of the LWC matchers combining the various mappings results.

First examining the nine more correct mappings (1152-1143) found by only the MMSS at the 0 threshold as compared to only the MM revealed that four were also found by the OAEI 2011 configuration of matchers with the MM. The reason is the MA concept string name is an exact match or a substring of the HA concept and so the PSM matcher of the OAEI 2011 configuration was able to identify these mappings. Using only the MMSS these four mappings were found using semantic similarity measurement within Uberon.

At first glance, the substitution of MMSS for MM does not seem to have improved the performance, both produced the same number of correct mappings at 1350 but the MM produced
slightly less number of mappings (1443 vs 1447). What the numbers do not tell is that the OAEI 2011 configuration with MMSS produces 3 “new” correct mappings that the OAEI 2011 configuration with MM is not able to produce. The OAEI 2011 configuration with MMSS, however, at the same time is losing 3 correct mappings which the OAEI 2011 configuration with the MM finds.

Table 3 shows the three correct mappings produced with the OAEI 2011 configuration with MMSS that are not produced with the MM.

<table>
<thead>
<tr>
<th>MA Source</th>
<th>HA MMSS Target</th>
<th>Uberon Bₜ</th>
<th>Uberon Bₛ</th>
</tr>
</thead>
<tbody>
<tr>
<td>gastrointestinal system mesentery</td>
<td>Mesentery</td>
<td>gastrointestinal system mesentery</td>
<td>Mesentery</td>
</tr>
<tr>
<td>Limb long bone</td>
<td>Long bone</td>
<td>Limb long bone</td>
<td>Long bone</td>
</tr>
<tr>
<td>Brain ependyma</td>
<td>Ependyma</td>
<td>Brain ependyma</td>
<td>Ependyma</td>
</tr>
</tbody>
</table>

Table 3. New Mappings, OAEI MMSS but not OAEI MM

For the three new correct mappings found by MMSS, none of the AgreementMaker matchers (PSM, VMM, LSM, and MM) found the third mapping. The PSM found the second mapping but the VSM incorrectly mapped the “forelimb long bone” to “long bone” instead with a higher confidence than the PSM had. LWC2 which combines the VSM and PSM produced the VSM mapping. Only the VSM produced the first mapping. Since the PSM did not, the LWC2 did not produce this correct mapping. LWC1 could not produce any of three mappings since it combines the LSM and MM, neither of which produced any of these mappings.

Table 4 shows the three correct mappings produced by the OAEI 2011 with MM that are not produced with the MMSS. The second column shows the correct mapping made by the MM in
the target HA ontology. The MMSS incorrectly mapped the MA sources to the HA concepts shown in column 4 due to both the high semantic similarity between the Uberon $B_S$ and the Uberon $B_T$ concepts in columns 3 and 4 of Table 4 and the interaction between the configuration of the LWC matchers.

<table>
<thead>
<tr>
<th>MA Source</th>
<th>HA MM Target</th>
<th>Uberon $B_S$</th>
<th>Uberon $B_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain arachnoid matter</td>
<td>Cerebral Arachnoid Membrane</td>
<td>Brain arachnoid mater</td>
<td>Leptomeninges</td>
</tr>
<tr>
<td>Iliac circumflex artery</td>
<td>circumflex iliac artery</td>
<td>Iliac circumflex artery</td>
<td>Deep circumflex iliac artery</td>
</tr>
<tr>
<td>Vagina squamous epithelium</td>
<td>Vagina squamous epithelium</td>
<td>Vagina squamous epithelium</td>
<td>Stratified squamous epithelium</td>
</tr>
</tbody>
</table>

Table 4. Lost Mappings. OAEI MM but not OAEI MMSS

For the three correct mappings lost with the MMSS, the PSM did produce all three, and the VSM did produce the first two. The MMSS, however, mapped the MA sources to incorrect targets for all three. The LWC2 did produce the three correct mappings but the LWC1 using the MMSS and LSM produced the three incorrect mappings. When LWC3 combines the LWC1 and LWC2 results, the LWC1 results had higher confidence values so the second and third MMSS incorrect mappings were selected. The first incorrect MMSS mapping overrides the correct mapping in LWC3 probably because its quality evaluation does not satisfy the cutoff threshold.

From these initial experiments, it can be seen that the MMSS is successful at discovering more correct mappings than AgreementMaker’s original MM. The drawback, however, is it produces more mappings overall. The pattern of interactions of the MMSS with the other matchers in the OAEI 2011 configuration is unclear. The objective is to incorporate the additional correct
mappings the MMSS without overriding the correct mappings of the other matchers with the incorrect ones of the MMSS. To better understand this interaction, next experiments using other possible LWC hierarchical aggregation schemes and their results are described.

6. Experiments with Variations of LWC Matchers

The results of the initial experiments were examined in more detail to better understand which mappings are produced by which individual matchers within AgreementMaker and how the final mappings in the alignment are determined when conflicts occur. Several aspects made this analysis challenging: the interaction among its matchers, AgreementMaker’s local quality measures (LQM) and their use as weighting factors in the LWC matchers, and the hierarchical organization of the LWC matchers. These considerations have subtle effects on the mappings which are eventually selected for the final alignment result. Because of these challenges, the approach taken is to perform experiments that vary how the LWC matchers are created. The goal of these experiments is to investigate other compositions of the LWCs which include the OAEI-MMSS to determine if it is possible to produce the three new mappings as well as retain the three lost correct mappings found by OAEI-MM. To this end several experiments were performed combining different matchers and the order of the matcher combinations.

Examining how the three correct MM mappings were being lost from the MMSS results revealed that because the MMSS produces incorrect source to target mappings with high confidence values, these incorrect mappings of the MMSS replace correct mappings produced by other matchers such as PSM and the VMM. Figure 6 shows the LWC configuration using the original OAEI 2011 anatomy track version with the MMSS. It combines matchers LSM and MMSS into a linear weighted combination LWC1, and similarly PSM and VMM into another linear weighted combination LWC2. LWC1 and LWC2 are then combined using yet another linear weighted combination matcher LWC3. This hierarchical organization can be represented as LWC3(LWC1(LSM+MMSS) + LWC2(PSM + VMM)).
In these experiments the four individual matchers are combined in different orders as specified below and listed in Table 5:

1) LWC(LSM+VMM+PSM+MMSS),
2) LWC3(LWC1(LSM+PSM) + LWC2(VMM +MMSS)),
3) LWC3(LWC1(LSM+VMM)+ LWC2(PSM +MMSS)),
4) LWC3(LWC1(MMSS) +LWC2(LSM+PSM+VMM));

The results produced by these various LWC combinations are shown in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>Produced</th>
<th>Correct</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAEI-MM</td>
<td>1439</td>
<td>1348</td>
<td>93.7</td>
<td>88.9</td>
<td>91.2</td>
</tr>
<tr>
<td>OAEI-MMSS</td>
<td>1441</td>
<td>1348</td>
<td>93.5</td>
<td>88.9</td>
<td>91.2</td>
</tr>
<tr>
<td>MM – 1</td>
<td>1413</td>
<td>1326</td>
<td>93.8</td>
<td>87.5</td>
<td>90.5</td>
</tr>
<tr>
<td>MMSS – 1</td>
<td>1416</td>
<td>1330</td>
<td>93.9</td>
<td>87.7</td>
<td>90.7</td>
</tr>
<tr>
<td>MM – 2</td>
<td>1402</td>
<td>1319</td>
<td>94.0</td>
<td>87.0</td>
<td>90.4</td>
</tr>
<tr>
<td>MMSS – 2</td>
<td>1403</td>
<td>1321</td>
<td>94.2</td>
<td>87.1</td>
<td>90.5</td>
</tr>
<tr>
<td>MM – 3</td>
<td>1396</td>
<td>1314</td>
<td>94.1</td>
<td>86.7</td>
<td>90.2</td>
</tr>
<tr>
<td>MMSS – 3</td>
<td>1400</td>
<td>1318</td>
<td>94.1</td>
<td>86.9</td>
<td>90.4</td>
</tr>
<tr>
<td>MM – 4</td>
<td>1433</td>
<td>1340</td>
<td>93.4</td>
<td>88.4</td>
<td>90.8</td>
</tr>
<tr>
<td>MMSS – 4</td>
<td>1440</td>
<td>1346</td>
<td>93.5</td>
<td>88.8</td>
<td>91.1</td>
</tr>
</tbody>
</table>

Table 5. Various Matcher Combinations for LWC Experimental Results

As can be seen, the different combination approaches did not improve the recall rate; that is, none of the LWC combinations with the MMSS produced more correct mappings than that of the OAEI original anatomy track configuration with the MMSS. Combinations 1 through 3 did, however, slightly improve the precision by reducing the number of incorrect mappings. It is interesting to note that Combination 4 with the MMSS results combined with the aggregated results of the other three matchers produced the next highest number of correct mappings while
producing close to the same number of total mappings.

Additional experiments were run to compare the MMSS using the various LWC combinations to that of the MM with the same LWC combinations. These results are shown in Table 5 in italics and a different font. The results show that in each case the MMSS performed better than the MM with the exact same LWC combination.

None of the MMSS combinations were able to retain the three lost mappings. However, Combination 1 through Combination 3, while reducing the number of incorrect mappings, produced several mappings not found by either the OAEI-MM or OAEI-MMSS. In total, 9 new correct mappings were found that were not found by the original OAEI MM combination. Although none of the combinations produced more correct mappings than the OAEI 2011 anatomy track configuration, experimenting with these various LWC combinations shows that finding additional correct mappings is possible but that more research and study are needed to determine how to recover these nine correct mappings in one ontology alignment result.

7. **Overriding PSM and Increasing the MMSS Semantic Similarity Threshold**

Since the various LWC combinations did not improve the overall final results, the results produced by the MM and the MMSS with original OAEI 2011 anatomy track configuration, were studied in more detail. Examining the three lost mappings showed that one is a source-target mapping produced by the PSM where the string labels are almost identical. This correct mapping is being replaced by an incorrect MMSS mapping in the LWC3 aggregation. Specifically, the PSM mapped source concept MA_0001731 (vagina_squamous_epithelium) to NCI_C49314 (Vaginal_ Squamous_Epithelium) with confidence value 1.0. However, the MMSS mapped the source to target NCI_C13180 (Stratified_Squamous_Epithelium). On investigating the LWC2 matcher, it became clear that when the above-mentioned result of PSM was combined with result of VMM, the same mapping lost confidence because VMM did not
produce this mapping. Therefore, to try to preserve this mapping, the LWC2 matcher was modified to directly output all the PSM mappings with a 1.0 similarity as their confidence factor no matter what was produced by VMM. This approach proved useful as shown in Table 6 since it increased the number of correct mappings by 4 to 1352 of the 1440 the MMSS produced and eliminated one incorrect mapping. However, still none of the lost 3 mappings are recovered. Instead the modification in general improved the LWC2 matcher algorithm.

<table>
<thead>
<tr>
<th></th>
<th>Mapped</th>
<th>Correct</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAEI2011-MM</td>
<td>1439</td>
<td>1348</td>
<td>93.7</td>
<td>88.9</td>
<td>91.2</td>
</tr>
<tr>
<td>OAEI2011-MMSS</td>
<td>1441</td>
<td>1348</td>
<td>93.6</td>
<td>88.9</td>
<td>91.2</td>
</tr>
<tr>
<td>OAEI-MMSS-PSM-identical mappings-preserved</td>
<td>1440</td>
<td>1352</td>
<td>93.9</td>
<td>89.2</td>
<td>91.5</td>
</tr>
</tbody>
</table>

Table 6. LWC2 Preserving PSM 1.0 confidence mapping

In a further effort to recover the lost three mappings, more analysis of the mapping results was performed. In particular the confidence levels for the mappings produced by the various matchers were compared. The MMSS confidence for the three lost mappings are lower than those for three new mappings discovered by the MMSS. Furthermore, the similarity matrix produced by the MMSS between the source-target pairs plays a crucial role in the calculation of the LQM used in linear combinations of the matchers. On observing the similarity matrix for select source concepts under the MMSS, while for a particular source, only one non-zero similarity exists. That is, the source is only mapped to one target. On the other hand, the similarity matrices of PSM as well as VMM are drastically different with very few 0.0 in each
matrix row. These matchers assign some similarity source-target pairs with even very different strings labeling. That, in turn influences the LQM which is calculated as average similarity of selected target mappings minus the average similarity of unselected target mappings. The result is LQMs for MMSS are usually higher than for PSM or VMM; therefore, the MMSS seems to dominant in determining the final LWC3 results.

To try to regulate this dominance of the MMSS, an experiment was run with the threshold of semantic similarity in the MMSS raised so that only strongly semantically similar source-target mappings are allowed in the MMSS mappings. Raising the threshold from 0.9 to 0.95 eliminated the three incorrect mappings produced by the MMSS for the lost three source concepts. Eliminating these incorrect mappings allowed the three mappings that were produced by MM to be kept in the final alignment. This new threshold, however, caused one of the three new mappings found by MMSS to be removed. Table 7 in the third row table illustrates how by simply raising the threshold for the MMSS to 0.95 improved both precision and recall over the 0.90 threshold. The fourth row of Table 7 is the result of both preserving identical source-target mappings, those with 1.0 confidence produced by the PSM, and increasing the MMSS threshold to 0.95. All performance measures are higher than with the OAEI 2011 MM. This result is achieved while only losing one of the correct mappings produced using the MM.

<table>
<thead>
<tr>
<th></th>
<th>Mapped</th>
<th>Correct</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAEI2011-MM</td>
<td>1439</td>
<td>1348</td>
<td>93.7</td>
<td>88.9</td>
<td>91.2</td>
</tr>
<tr>
<td>OAEI2011-MMSS</td>
<td>1441</td>
<td>1348</td>
<td>93.6</td>
<td>88.9</td>
<td>91.2</td>
</tr>
<tr>
<td>OAEI-MMSS 0.95threshold</td>
<td>1441</td>
<td>1350</td>
<td>93.7</td>
<td>89.1</td>
<td>91.3</td>
</tr>
<tr>
<td>OAEI-MMSS-0.95 with identical PSM mappings preserved</td>
<td>1443</td>
<td>1353</td>
<td>93.8</td>
<td>89.2</td>
<td>91.4</td>
</tr>
</tbody>
</table>
The results of the above experiments in the section and the previous section demonstrate that the use of semantic similarity with a reference ontology does find more correct mappings. Although the improvements in the performance measures are small, they can increase the performance of Agreement in comparison to the results of other OA systems in the OAEI 2012 anatomy track presented in Table 8.

<table>
<thead>
<tr>
<th>Matcher</th>
<th>Runtime</th>
<th>Size</th>
<th>Precision</th>
<th>F-Measure</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOMMA-bk</td>
<td>15</td>
<td>1534</td>
<td>0.917</td>
<td>0.923</td>
<td>0.928</td>
</tr>
<tr>
<td>YAM++</td>
<td>69</td>
<td>1378</td>
<td>0.943</td>
<td>0.898</td>
<td>0.858</td>
</tr>
<tr>
<td>CODI</td>
<td>880</td>
<td>1297</td>
<td>0.966</td>
<td>0.891</td>
<td>0.827</td>
</tr>
<tr>
<td>LogMap</td>
<td>20</td>
<td>1392</td>
<td>0.920</td>
<td>0.881</td>
<td>0.845</td>
</tr>
<tr>
<td>GOMMA</td>
<td>17</td>
<td>1264</td>
<td>0.956</td>
<td>0.870</td>
<td>0.797</td>
</tr>
<tr>
<td>MapSSS</td>
<td>453</td>
<td>1212</td>
<td>0.935</td>
<td>0.831</td>
<td>0.747</td>
</tr>
<tr>
<td>LogMapLt</td>
<td>6</td>
<td>1147</td>
<td>0.963</td>
<td>0.829</td>
<td>0.728</td>
</tr>
<tr>
<td>WeSeE</td>
<td>15833</td>
<td>1266</td>
<td>0.911</td>
<td>0.829</td>
<td>0.761</td>
</tr>
<tr>
<td>TOAST*</td>
<td>3464</td>
<td>1339</td>
<td>0.854</td>
<td>0.801</td>
<td>0.755</td>
</tr>
</tbody>
</table>

Table 8. OAEI 2012 Anatomy Track for OA systems with F-measure > 0.800

http://oaei.ontologymatching.org/2012/anatomy/results.html

In the OAEI 2012 anatomy track 17 different OA systems participated. Only two of these are from U.S. research groups, that is, Optima and MapSSS. AgreementMaker from the University of Illinois, Chicago, did not participate in OAEI 2012. This thesis research using AgreementMaker was completed after the deadline for entering the competition. When comparing the OAEI 2011 MMSS performance measures to the top 2012 performers, only GOMMA-bk has a higher F-measure and recall than the OAEI 2011 MMSS; however, OAEI
2011 MMSS has a higher precision than GOMMA-bk and a higher precision than 4 of the other 8 systems in Table 8.

GOMMA-bk (Generic Ontology Matching and Mapping Management with background knowledge) (Groß et al 2012) is a new and improved version of the composition-based matching OA system described in section 2.2. It uses mappings to three intermediate or reference ontologies (UMLS, Uberon or FMA). But as previously discussed, it like the mediating matcher of AgreementMaker requires an exact match on the source and target bridge concept in the intermediate ontologies.

AgreementMaker with the MMSS is quite competitive with the other OA systems in the 2012 anatomy track. Two issues deserving further investigation for incorporating the MMSS matcher into AgreementMaker is the development of a more sophisticated LQM measure that can better balance the confidence levels of the various matchers so that the number of incorrect mappings can be reduced. Hopefully the reduction in incorrect mappings produced by the various matchers will allow more correct mappings to be produced.

8. Using Another Intermediate Ontology

In order to evaluate the use of semantic similarity within a second mediating ontology in the OAEI anatomy track, an additional ontology – the FMA ontology was selected as a reference ontology. FMA (Foundational Model of Anatomy) is a domain ontology representing declarative knowledge of the human anatomy and extensible to other species. It has a medium level of coverage for both the Mouse Anatomy (MA) at 57% and NCI Thesaurus (NCIT) at 67% (Groß et al 2011). This is due to the presence of relatively few names and synonyms per concept in FMA ontology. This shortage affects linguistic matching when constructing bridges. However, it was selected as a second mediating ontology primarily due to both its use in GOMMA and Agreementmaker’s memory restrictions when trying to work with another larger reference
ontology UMLS.

Bridge mappings between FMA and the MA and NCIT HA ontologies were constructed using AgreementMaker’s lexical synonym matcher, which performs alignment based on matching lexicons or groups of synonyms related to a particular concept. These bridge mappings are labeled as AM Generated. A second set of bridge mappings was obtained from Dr. Anika Groß whose GOMMA system (Groß et al 2012) also uses FMA along with several other mediating ontologies for ontology alignment. This set of bridge mappings is labeled Groß. The Groß bridge mappings, constructed using linguistic trigram similarity between concept names and synonyms, provided more correct mappings in the context of both MM and MMSS matchers for aligning the MA and NCIT HA ontologies.

Table 9 illustrates the precision, recall and F-measure of both MM and MMSS matchers when used with the two sets of bridge mappings. Only the MM and MMSS matchers are used without any of the matchers in the OAEI 2011 anatomy configuration.

<table>
<thead>
<tr>
<th>Bridge mappings</th>
<th>Mappings Found</th>
<th>Correct</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM generated</td>
<td>845</td>
<td>728</td>
<td>86.2</td>
<td>48.0</td>
<td>61.7</td>
</tr>
</tbody>
</table>
Table 9: Comparing bridge mappings for FMA ontology

For the following experimental results reported in this section, the Groß bridge mappings between FMA and both the MA and NCIT HA ontologies are used because they produce many more correct mappings than the AM generated mappings.

Table 10 shows the results of OAEI2011 anatomy track configuration using the FMA and the Uberon ontologies with the MM and the MMSS. The three rows after the Uberon – MM row and the three rows after the Uberon – OAEI2011 list experimental results discussed previously and are repeated here for comparison to the results when using FMA as the reference ontology.
An obvious observation is that MM produces more mappings using Uberon than by using FMA. This result agrees with the observation in (Groß et al 2011) that Uberon covers more concepts from NCI HA than FMA does (80 percent Uberon vs 57 percent FMA). MMSS adds 5 more mappings to MM using Uberon at a threshold of 0.90 and 4 more mappings at 0.95. However, MMSS fails to add any more mappings to MM using FMA at 0.95 and only 1 more mapping at 0.90. Thus, the OAEI2011 MMSS configuration produces only a very small variation from the MM with the FMA as compared to the Uberon as a reference ontology.
<table>
<thead>
<tr>
<th>Uberon - MM</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MM</td>
<td>1200</td>
<td>1143</td>
<td>95.2</td>
<td>75.4</td>
</tr>
<tr>
<td>MMSS 0.90</td>
<td>1229</td>
<td>1148</td>
<td>93.4</td>
<td>75.7</td>
</tr>
<tr>
<td>MMSS 0.95</td>
<td>1209</td>
<td>1147</td>
<td>94.9</td>
<td>75.7</td>
</tr>
<tr>
<td>FMA - MM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>1033</td>
<td>985</td>
<td>95.4</td>
<td>65.0</td>
</tr>
<tr>
<td>MMSS 0.90</td>
<td>1040</td>
<td>986</td>
<td>94.8</td>
<td>65.0</td>
</tr>
<tr>
<td>MMSS 0.95</td>
<td>1033</td>
<td>985</td>
<td>95.4</td>
<td>65.0</td>
</tr>
<tr>
<td>Uberon-OAEI2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>1439</td>
<td>1348</td>
<td>93.7</td>
<td>88.9</td>
</tr>
<tr>
<td>MMSS at 0.90 PSM preserved</td>
<td>1440</td>
<td>1352</td>
<td>93.9</td>
<td>89.2</td>
</tr>
<tr>
<td>MMSS at 0.95 PSM preserved</td>
<td>1443</td>
<td>1353</td>
<td>93.8</td>
<td>89.2</td>
</tr>
<tr>
<td>FMA-OAEI2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>1424</td>
<td>1335</td>
<td>93.8</td>
<td>88.1</td>
</tr>
<tr>
<td>MMSS at 0.90 PSM preserved</td>
<td>1425</td>
<td>1338</td>
<td>93.8</td>
<td>88.3</td>
</tr>
<tr>
<td>MMSS at 0.95 PSM preserved</td>
<td>1423</td>
<td>1339</td>
<td>94.0</td>
<td>88.3</td>
</tr>
</tbody>
</table>

Table 10: Comparison of FMA and Uberon results

It is to be noted that while much lower thresholds of semantic similarity do tend to produce more mappings in MMSS, the production of incorrect mappings contributes negatively to LWC combination stages in the OAEI2011 configuration.

As discussed in sections 5, 6 and 7 and indicated in table 10 the best configuration for OAEI2011-MMSS matcher using Uberon was able to produce net five more correct mappings compared to OAEI2011-MM. Two mappings out of the net five were produced solely by MMSS, while the rest of the mappings were produced by other matchers in the OAEI hierarchy. Similarly,
when using FMA, the best configuration for OAEI2011-MMSS produced 4 more mappings; however none of these mappings were produced by MMSS. They were produced by the preserving of the PSM mappings in the LWC2 matcher. Comparing the Uberon to FMA as a reference ontology, the Uberon ontology is slightly better with respect to the performance measures.

Since GOMMA-bk combines multiple mediating ontology mappings, the next experiment combines both Uberon and FMA using a linear weighted combination of their mediating matcher results. This linear weighted combination is also used in place of the mediating matcher in the OAEI2011 matcher hierarchy in the following manner

\[ \text{LWC3(LWC1(LWC0(MM-Uberon+MM-FMA) + LSM)+LWC2(PSM+VMM))} \]

As expected, the linear combination of the two mediating matchers each using a different reference ontology produces more correct mapping(1226) than individual matchers (Uberon at 1143 and FMA at 985). When using the same configuration with MMSS at 0.90, 3 more net mappings are produced as compared to the MM. The OAEI2011 results are the best achieved so far when combining the mappings produced using the two reference ontologies. When using MMSS instead of MM under this combined OAEI2011 hierarchy, a net gain of 1 mapping (1364 vs 1363) was achieved. This extra mapping was produced by the MMSS when used with Uberon.

<table>
<thead>
<tr>
<th>Test</th>
<th>Mappings Found</th>
<th>Correct</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FMA+Uberon Mediating matcher</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>1282</td>
<td>1226</td>
<td>95.6</td>
<td>80.9</td>
<td>87.6</td>
</tr>
<tr>
<td>MMSS at 0.90</td>
<td>1307</td>
<td>1229</td>
<td>94.0</td>
<td>81.1</td>
<td>87.0</td>
</tr>
<tr>
<td>MMSS at 0.95</td>
<td>1287</td>
<td>1227</td>
<td>95.3</td>
<td>80.9</td>
<td>87.5</td>
</tr>
<tr>
<td><strong>FMA+Uberon OAEI2011</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>1449</td>
<td>1363</td>
<td>94.0</td>
<td>89.9</td>
<td>91.9</td>
</tr>
<tr>
<td>MMSS at 0.90- PSM preserved</td>
<td>1453</td>
<td>1362</td>
<td>93.7</td>
<td>89.8</td>
<td>91.7</td>
</tr>
</tbody>
</table>

28
Table 11 Combining two mediating ontologies

| MMSS at 0.95- PSM preserved | 1453 | 1364 | 93.9 | 90.0 | 91.9 |

Although, combining the results of the MMSS using Uberon and the MMSS using FMA did improve the final alignment over that produced using only the Uberon ontology, the f-measure and the recall of the GOMMA-bk OA systems is still slightly higher. This outcome most likely can be attributed to the use by GOMMA-bk of a third reference ontology the UMLS. As part of this thesis research, an experiment was attempted with the UMLS but due to its size, the AgreementMaker system was unable to complete the alignment process with this reference ontology.

9. Conclusions and Future Research

This thesis documents efforts to improve ontology alignment results by using semantic similarity measures within a reference or mediating ontology. This approach is inspired by existing OA systems that use a mediating ontology to map a concept in a source ontology to a concept in a target ontology when both source and target concepts map exactly to the same concept in the mediating ontology. This thesis research has developed a new matcher to use semantically related concepts of the mediating ontology to relate source and target ontologies. The new matcher is integrated as an enhancement to a top performing OA system in the OAEI 2011 competition, the AgreementMaker system developed at the ADVIS Lab at the University of Illinois, Chicago.

The Mediating Matcher with Semantic Similarity (MMSS) extended Agreementmaker’s Mediating Matcher (MM) to allow for connecting source and target concepts through semantically similar reference ontology concepts. Initial experiments comparing only the MMSS to the MM matcher showed that the MMSS produced more correct mappings than the MM, but that it also produce more mappings in general, i.e., it improved recall but not precision. As the
similarity threshold was increased for the MM the number of incorrect mappings was reduced. Then the MMSS was substituted for the MM in the OAEI 2011 AgreementMaker matcher configuration. The MMSS in this configuration produced the same number of correct mappings as the MM but discovered three new mappings and lost three mappings found by the MM. Experiments with various configurations of matchers in AgreementMaker’s LWC were undertaken in hope that all six correct mappings could be retained. This result was not obtained but the various individual configurations of the matchers produced nine new correct mappings. More investigation is needed to better understand the interactions and the linear combination weighting matcher architecture to develop better local quality measures for weighting factors so that more correct mappings can be retained in a single final alignment results.

Based on studying the mappings in the individual alignment results produced, it was determined that the PSM mappings were being overridden by the high similarity values of the MMSS matcher. By creating a new LWC matcher that gives precedence to the PSM mappings and increasing the threshold for similarity degree of the MMSS, the OAEI 2011 MMSS produced final alignment results better than any other systems participating in the evaluation of standard test data in the anatomy track of the OAEI 2011 competition.

The scope of this thesis was chosen as the OAEI anatomy track since the two systems that had already been using mediating ontologies as a medium of ontology alignment had successfully participated in this track. All the experiments reported in this thesis use the Lin semantic similarity measure due to its success in existing bioinformatics research. Much effort was made to study the combination of the MMSS matcher with other matchers in the OAEI2011 matcher configuration in order to create new mappings based on MMSS while preserving mappings already produced. The test framework initially used Uberon as a mediating ontology; however; it was extended to use FMA and a combination of Uberon with FMA to evaluate usefulness of MMSS with several mediating ontology options. The combination of Uberon and FMA produced
better results than all but one of the system participating in the 2012 OAEI anatomy track, the GOMMA-bk system. The GOMMA-bk system is a new improved version of the composition-based systems discussed in section 3.2. The GOMMA-bk system which only participated for the first time in the OAEI 2012 competition has a better f-measure 92.3 compared to 91.9 because of its better recall 91.7 compared to 90.0, but OAEI 2011 MMSS has a better precision of 93.7 compared to 91.7. AgreementMaker did not participate in the 2012 OAEI competition.

Future work related to semantic similarity in mediating ontologies could extend this evaluation framework on other test data pertaining to different knowledge domains. As was observed in the case of Uberon vs FMA, a reference ontology with sufficient coverage of concepts from both ontologies performs better in connecting the two ontologies to be aligned. Furthermore, efforts to optimize the combination of the MMSS matcher with other matchers inside Agreementmaker’s OAEI2011 matcher could incorporate the mappings found by varying combination schemes of these individual matchers.
10. References


