Using a Reference Ontology with Semantic Similarity in Ontology Alignment

Valerie Cross*, Pramit Silwal, and David Morell
Computer Science and Software Engineering, Miami University Oxford OH USA

ABSTRACT

The current use of semantic similarity with a reference ontology in ontology alignment (OA) systems is reviewed. An extended matcher is described that incorporates semantic similarity with the use of a reference ontology. This matcher has been implemented using as a basis AgreementMaker's mediating matcher. Specific experiments using the OAEI anatomy track are performed using the Uberon ontology as the reference ontology. The results of these experiments are compared to the OAEI 2011 results for the anatomy track. These show that semantic similarity measures can be useful for discovering mappings missed by the original mediating matcher. The use of semantic similarity with a reference ontology should be further investigated in the effort to improve the OA process.

1 INTRODUCTION

Ontology alignment (OA) systems typically produce a set M_{ST} of mapping pairs (s_i, t_i) between a source ontology O_S and a target ontology O_T with each pair having a similarity degree d_{sim} in (0, 1]. The mapping indicates that the concept s_i in O_S is similar to the concept t_i in O_T with d_{sim}. Most matchers in OA systems rely on only the internal information available within the ontologies to be aligned. External knowledge sources are increasingly being used to improve the alignment process (Shvaiko & Euzenat, 2012). A standard approach has been to create a matcher that uses a reference ontology or creates a lexicon using a thesaurus. The main operation typically is some function of the overlap between the synonym sets found in the reference ontology or the lexicon for the source and target concepts. The problem occurs when no overlap between the two sets exists. Semantic similarity measures can be used to find a possible mapping from a source concept to a target concept based on the similarity between the source's identified concept and the target's identified concept in the reference ontology.

Measuring similarity between a source concept *s* and a target concept *t* in the two different ontologies can then be translated into finding corresponding bridge concepts b_S and b_T in the reference ontology and then measuring the degree of similarity between b_S and b_T. Several important issues to using background knowledge sources have been identified (Shvaiko & Euzenat, 2012). For example the selection of the reference ontology should ensure that it has suitable coverage of the ontologies being aligned. Another important consideration is the means of finding the corresponding entities b_S and b_T in the reference ontology.

The contribution of this research is the use of a reference ontology and semantic similarity measurement within the reference ontology to improve the OA process. Section 2 overviews semantic similarity and its use with background knowledge in existing OA systems. Section 3 first describes a recent experiment to use different biomedical ontologies as reference ontologies without using semantic similarity to improve alignment results for the OAEI anatomy track. Section 4 presents the proposed method that extends the previous approaches with semantic similarity measurement. The experiments results using this method on the OAEI anatomy track are described and compared with those of one the experiments described in section 3. Finally, conclusions and a summary of the research efforts as well as future research plans are presented in section 5.

2 REFERENCE ONTOLOGY WITH SEMANTIC SIMILARITY

Much research is being undertaken to use background knowledge sources to aid the ontology alignment process. Many forms of background knowledge have been used such as partial alignments, existing alignments, domain specific corpora, web pages, linked data, upper ontologies and domain specific ontologies (Shavaiko & Euzenat, 2012). However, the use of simple background knowledge sources such as thesauri, for example, WordNet, has been widespread for some time. More recently research has examined the use of domain specific ontologies especially in the medical domain or a collection of ontologies selected from the Semantic Web. These ontologies have been referred to as reference (Sabou *et al*., 2008), intermediate (Gross *et al.*, 2011) or mediating ontologies (Cruz *et al.*, 2011).

The outcome of several OAEI competitions has not been consistent when it comes to OA systems using background knowledge (Shvaiko & Euzenat, 2012). For example, in the 2007 and 2008 OAEI competitions, the OA systems utilizing background knowledge were undoubtedly the best performing. The best performing OA system in 2009, however, did not use any background knowledge. In 2011 the best performing systems in the anatomy track made use of domain specific ontologies (Euzenat *et al.*, 2011). For the OA systems actually competing in the OAEI competition, the background knowledge sources are manually selected.

* To whom correspondence should be addressed: crossv@muohio.edu

2.1 Semantic Similarity in Ontologies

In ontology alignment, numerous similarity measures are used to determine the similarity between concepts in two different ontologies. The purpose is to create a list of concept mappings between the two ontologies. Semantic similarity, however, unlike similarity measurement typically used within OA, measures the similarity between two concepts within a single ontology. Due to space limitations, only a historical review of such measures is presented. These measures or slight variations represent those used in the OA systems described in the next section. A detailed overview of current semantic similarity measures and research can be found in (Yu, 2010) and (Cross and Yu, 2012).

The earliest semantic distance measures were developed for use in semantic networks and were simple path distance measures, i.e., the count of the number of edges or nodes, between two concepts (Rada *et al*., 1989). This simple path-based distance has been used in ontologies viewed as graphs. Wu and Palmer (Wu & Palmer, 1994) improved upon the early path-based semantic distance measures by proposing a semantic similarity measure between two concepts that is the ratio of twice the distance of their lowest common subsumer to the root concept and the sum of the distance of each concept from the root concept.

Another approach to semantic similarity is based on using a measure of information content (IC) for a concept. IC measures how specific a concept is within a given ontology. The more specific a concept is the higher its information content, the more general the lower its IC. IC has been determined by either a corpus-based (Resnik, 1995) or an ontology-based method (Seco *et al*., 2004). The corpus-based IC uses an external resource such as an associated corpus for the problem domain and is determined using the negative log of the probability of the concept with respect to the corpus. The ontology-based IC method simply uses the structure of ontology itself to determine a concept's IC value. It is a function of the number of descendents of a concept and the total number of concepts in the ontology.

The first IC-based semantic similarity measure is defined as the maximum information content two concepts share (Resnik, 1995). The common ancestor of the two concepts having the maximum IC value must be found and its IC value is taken as the semantic similarity between the two. An improvement to Resnik's measure was proposed by Lin (1998). It is formulated as the ratio of twice the maximum shared information content between the two concepts and the sum of each concept's individual information content.

2.2 OA Systems Using Semantic Similarity

Here a brief survey of only OA systems using a background knowledge source, WordNet, UMLS, or both as a reference ontology with semantic similarity is presented. They apply standard semantic similarity measures or their variations between the concepts within the reference ontology and not between the concepts from the two ontologies being aligned. The systems are presented in chronological order of their references. A complete overview of the state of the art for OA systems can be found in (Euzenat *et al.*, 2011).

2.3.1 OLA **(**Euzenat and Valtchev, 2003**)**. A modified version of the Wu-Palmer semantic similarity measure (Wu and Palmer, 1989) is used in determining lexical similarity between a pair of identifiers which are each first converted into a set of atomic terms. Next pairs of terms, one from each set, are compared using WordNet. The pair's similarity is calculated as the ratio between the depth of the most specific common hypernym (ancestor in the WordNet hierarchy) and the sum of depth of each term. Then a degree of proximity between the sets of terms is calculated.

2.3.2 Imapper **(**Su, 2004). The similarity value determined for the mapping between two concepts may be increased using the distance of the two concepts in WordNet. The concepts are found in WordNet using their descriptive labels. A simple path based semantic distance between two terms *x* and *y* found in WordNet is used. If they belong to the same synset in WordNet, then the path distance is 1. Otherwise, the path length is determined by the number of nodes rather than the links in the path so that the length between sibling nodes is 3. If no path can be found between them (they exist in unconnected WordNet subontologies), then they are unrelated. Their similarity value is, therefore, not strengthened.

2.3.3 ASMOV (Jean-Mary et al., 2009). Semantic similarity measures may be used in determining the lexical similarity between concept labels. If the string labels for the source and target concepts are identical, the lexical similarity is 1.0. If they are not identical and an external ontology such as WordNet or UMLS is available, then various thesaurus relationships are used. If the source label string is in the synonym set of the target label, then their lexical similarity is set to 0.99. If one is an antonym of the other, then their lexical similarity is set to 0.0. If neither of those relationships hold and if both string labels exist in the external ontology, their lexical similarity is set to the Lin (1998) semantic similarity measure between the two. Otherwise, the minimum inclusion measure between the two sets of tokens is used.

2.3.4 CIDER (Gracia & Mena, 2008). The alignment process uses a modified version of a sense semantic similarity measure to evaluate similarity between the possible senses of a keyword and their synonyms to perform disambiguation. The techniques used in CIDER are adapted from the PowerMap WordNet based algorithm (Lopez *et al*., 2006).

2.3.5 UFOme (Pirro and Talia, 2010). A set of matchers, many of which have already been developed previously for numerous OA systems, are integrated into UFOme. Its WordNet matcher also uses the Lin semantic similarity measure between WordNet synsets when the concepts do not map to the same synset in WordNet.

3 RECENT EXPERIMENTS WITH REFERENCE ONTOLOGIES

Two very recent experiments using reference ontologies to improve the alignment mapping process are presented. In (Gross *et al.,* 2011), the reference ontology is called an intermediate ontology and in (Cruz *et al.,* 2011) it is called a mediating ontology. Both follow a very similar approach. The differences exist in the alignment methods used to produce the mappings from the source and target ontologies to the reference ontology and what aggregation method of similarity values are used to produce the final mapping from a source concept to a target concept through a reference concept. Neither incorporates semantic similarity measurement between concepts within the reference ontology

3.1 Composition-Based Matching

In (Gross et al., 2011) the OA system uses intermediate ontologies O_I to composes mappings M_{SI} from the source O_S to O_I with mappings M_{IT} from O_I to the target O_T to produce a set of mappings M_{ST} from the O_S to the O_T. More formally, the final alignment result is defined as

$$M_{ST} = \{(c_S, c_T, aggSim\,(mapSim_{SI}, mapSim_{IT})) \mid c_S \in O_S, c_I \in O_I, c_T \in O_T : \exists (c_S, c_I, mapSim_{SI},) \in M_{SI} \wedge \exists (c_I, c_T, mapSim_{IT},) \in M_{IT}\} \quad (1)$$

The aggregation operator aggSim combines the mapping similarities for M_{SI} and M_{IT}. Different operators could be used. They state average was used. They suggest that M_{SI} and M_{IT} could be existing mappings such as those in BioPortal. M_{SI} and M_{IT} in their experiments were determined using linguistic trigram similarity between concept names and synonyms with a threshold of 0.8. In effect, two simplified ontology alignments were first performed to create the mappings M_{SI} and M_{IT} before the composition-based mapping is done. One point not clear is the method if multiple c_I exist, i.e., if 1-1 mapping is not enforced. The method to produce intermediate mappings may enforce 1-1 mappings. An optional step tries to find direct mappings from the set of unmapped concepts in O_S to the set of unmapped concepts in O_T. These two sets are matched against each other using a string similarity match algorithm.

They evaluate the proposed composition approach using the Adult Mouse Anatomy ontology (MA) and the anatomical part (human anatomy HA) of the NCI Thesaurus, the OAEI anatomy track. The four reference ontologies are FMA, Uberon, RadLex, and UMLS, all late 2010 versions. Separate experiments were done for each of the ontologies. Only F-measures are reported. Uberon produced the best results (F-measure of 88.2%) with the two step process 1) produce mappings first using Uberon as the intermediate ontology and 2) add direct mappings between the MA and HA. Their paper points out that none of the previous approaches participating in OAEI 2010 anatomy track exceeded an 87% F-measure. .

3.2 AgreementMaker Mediating Matcher

For OAEI 2011, AgreementMaker (Cruz *et al.*, 2011) added a new matcher, the mediating matcher (MM). The mediating matcher inputs two ontologies to be aligned and a reference ontology and then uses AgreementMaker's BSM^{lex} (base similarity matcher with lexicon) to match the MA and the HA ontologies with the reference ontology. The BSM^{lex} matcher is calculates the similarity between two concepts by comparing all the strings associated with those two concepts, that is, the concept name, label, and comments.

AgreementMaker's approach is similar to that in (Gross et al., 2011). Both require an exact match on the bridge concept, i.e., $b_S = b_T$. It differs in the sophistication of the matcher used to find the bridge concepts for the source and target ontologies in the reference ontology, i.e., BSM^{lex} algorithm versus linguistic trigram similarity. Based on the success of the Uberon ontology as a reference ontology in (Gross *et al*., 2011), AgreementMaker also chose to use it as the mediating ontology for the OAEI 2011 anatomy track. The BSM^{lex} also used Uberon to develop its lexicon in matching the MA and HA ontologies to Uberon to take advantage of the extra synonyms defined in Uberon.

In the reported OAEI 2011 results (Euzenat *et al.*, 2011), AgreementMaker had the best performance with respect to F-measure (91.7%). These results are better than those in (Gross et al., 2011). AgreementMaker used only the one reference ontology Uberon while the best results in (Gross et al., 2011) were based on merging results using four different reference ontologies. Another difference is that AgreementMaker's final mappings are determined by a hierarchically 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 MM is only one.

Each matcher produces a similarity matrix between the source concepts and the target concepts. A LWC takes as input two or more matchers' similarity matrix and produces a weighted aggregation of them. The output is another matrix mapping the source and target concepts.

AgreementMaker's OAEI 2011 final matcher used three different LWCs. LWC1 produces a weighted average of the similarity matrices for the LSM (Lexical Similarity Matcher) and the MM. LWC2 produces a weighted average for the PSM (Parametric String-based Matcher) and the VMM (Vector-based Multi-word Matcher). LWC3 determines the final confidence factor for each alignment as a weighted average of the LWC1 and LWC2 similarity matrices.

4 MEDIATING MATCHER + SEMANTIC SIMILARITY

This proposed method of combining a reference ontology with semantic similarity builds on the work of early OA systems as described section 2.2. The recent uses of composition-based mapping and a mediating matcher described in section 3.1 and 3.2, respectively, also motivate this work. Neither OA system presented in those two sections, however, makes use of semantic similarity measures with a reference ontology. Our research extends AgreementMaker's mediating matcher and has produced a new mediating matcher that incorporates semantic similarity measurement (MMSS) between the corresponding bridge concepts in the mediating ontology. First the extension is described and then the experimental results are presented.

First AgreementMaker's MM is used in a first pass to produce the mappings between the source and target concepts where there is an exact match on the bridge concepts in the mediating ontology, i.e., $b_S = b_T$. When an exact match occurs, MM produces a mapping between *s* and *t* as

$$M_{ST} = \{(s, t, mapSim_{SI} * mapSim_{TI}) \mid s \in O_S, b_S, b_T \in O_I, t \in O_T : \exists(s, b_S, mapSim_{SI}) \in M_{SI} \wedge \exists(t, b_T, mapSim_{TI}) \in M_{TI} \wedge b_S = b_T\} \quad (2)$$

Here M_{SI} is the mapping from the source O_S to the intermediate O_I using BSM^{lex}. Similarly, M_{TI} is the mapping from the target O_T to the intermediate O_I using BSM^{lex}. The next step is to determine U_S and U_T, all the source concepts *s* in the mapping set from source to mediating ontology and all the target concepts *t* in the mapping set from target to mediating ontology, respectively, which did not get selected by the original mediating matcher. These two sets are given as

$$U_S = \{s \mid s \in O_S : \exists(s, b_S, mapSim_{SI}) \in M_{SI} \wedge \nexists\, t \in O_T : (s, t, sim_{ST}) \in M_{ST}\}$$
$$U_T = \{t \mid t \in O_T : \exists(t, b_T, mapSim_{TI}) \in M_{TI} \wedge \nexists\, s \in O_S : (s, t, sim_{ST}) \in M_{ST}\}. \quad (3)$$

For each pair *(s, t)* in U_S x U_T, the semantic similarity between all bridge concepts for *s* and all bridge concepts for *t* are calculated, and the maximum is used in determining the enhanced mapping set as

$$E_{ST} = \{(s, t, agg(mapSim_{SI}, mapSim_{TI}, bridgeSim)) \mid s \in U_S, b_S, b_T \in O_I, t \in U_T : \exists(s, b_S, mapSim_{SI}) \in M_{SI} \wedge \exists(t, b_T, mapSim_{TI}) \in M_{TI} : bridgeSim = \max_{b_S, b_T \in O_I}(semSim(b_S, b_T))\}. \quad (4)$$

$M_{ST} \cup E_{ST}$ is returned as the result of the MMSS and is input to the LWC1 in place of simply M_{ST}. Different *agg* operators may be used. For the experiments reported below, the minimum is used since this aggregator looks for the weakest similarity between the three pairs of concepts. The final mapping between *s* and *t* is not considered any stronger than the weakest similarity of the three being aggregated. Different measures can be used for *semSim*. For the experiments reported below, the standard Lin semantic similarity measure is used with IC as defined in (Seco et al., 2004) since it has frequently been used in current OA systems. An additional threshold value may be set to eliminate mappings in E_{ST} whose aggregated similarity falls below the threshold.

To be consistent with previous work in section 3, the OAEI anatomy track was used. Its reference alignment contains 1516 mappings. Table 1 shows the results of the experiments which are divided into two groups. First, only the mappings from the MM are compared to only those from the MMSS with varying thresholds as listed. The results of the first group are listed in the rows before the row labeled OAEI 2011. AgreementMaker's LWC matchers are not affecting these results. The second group compares the two different mediating matchers with the full OAEI 2011 AgreementMaker LWC matchers as described at the end of section 3.2. The second group investigates the interaction between the mappings of the MMSS and those produced by the other OAEI 2011 matchers as well as the effects of its LWC matchers combining the various mappings results.

For the first group, the MMSS with no threshold had the best recall but the worst precision. As the threshold increases the MMSS is still able to find more correct mappings than the MM and improve its precision. Of the nine more correct ones (1152-1143) found by the MMSS, four were also found by the OAEI 2011 matcher with the MM. The reason is the MA concept string name is an exact match or a substring of the HA concept. The MMSS found these four through using semantic similarity within Uberon.

The OAEI 2011 results using MMSS always produced more mappings than that using the MM. An interesting observation though is the 1350 correct for the MM and the MMSS with 0.90 threshold are not the same ones. Each found 3 different correct ones from each other. The goal is to study the interaction among the other OAEI 2011 matchers with the MMSS and the MM to try to keep both sets of 3 correct matches instead of replacing them with each other.

	Mapped	Correct	Precision	Recall	F-measure
MM	1200	1143	95.2	75.4	84.2
MMSS, 0.0	1322	1152	87.1	76	81.2
MMSS, 0.65	1301	1151	88.5	75.9	81.7
MMSS, 0.85	1240	1150	92.7	75.9	83.5
MMSS, 0.90	1229	1148	93.4	75.7	83.6
OAEI 2011					
MM	1443	1350	93.6	89.1	91.2
MMSS, 0.85	1447	1348	93.2	88.9	91.0
MMSS, 0.90	1447	1350	93.3	89.1	91.1

Table 1. Experimental Results on the OAEI Anatomy Track

Table 2 shows thethree correct mappings produced with the OAEI 2011 matcher and MMSS and not produced with MM. Table 3 shows the three correct mappings produced by the OAEI 2011 with MM and not produced with MMSS. The MMSS incorrectly mapped the MA sources to the HA concepts matching the Uberon B_T column of Table 3 since each of these concepts exists in the HA ontology and were mapped from the HA to the corresponding Uberon concept.

MA Source	HA MMSS Target	Uberon B_S	Uberon B_T
gastrointestinal system mesentery	Mesentery	gastrointestinal system mesentery	Mesentery
Limb long bone	Long bone	Limb long bone	Long bone
Brain ependyma	Ependyma	Brain ependyma	Ependyma

Table 2. New Mappings, OAEI MMSS but not OAEI MM

MA Source	HA MM Target	Uberon B_S	Uberon B_T
Brain arach-noid matter	Cerebral Arachnoid Membrane	Brain arach-noid mater	leptomeninges
Iliac circum-flex artery	circumflex iliac artery	Iliac circum-flex artery	Deep circumflex iliac artery
Vagina squamous epithelium	Vagina squamous epithelium	Vagina squamous epithelium	Stratified squamous epithelium

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

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 "fore-limb 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.

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 is lost in LWC3 probably because its quality evaluation does not satisfy the cutoff threshold,

5 CONCLUSIONS AND FUTURE WORK

The MMSS is successful at discovering more correct mappings than AgreementMaker's MM. The drawback, however, is it suggests more mappings. More experimentation is needed to better understand the interaction between the MMSS and the other matchers in the OAEI 2011 configuration so that other possible LWC schemes can be developed to better combine the strengths of the MMSS with the other matchers. In addition, different semantic similarity measures need to be investigated with different reference ontologies Other source and target ontologies with different structures and more varied labeling should also be tested.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Isabel Cruz and Cosmin Stroe for their support in this research effort.

REFERENCES

Cruz, I. F., Stroe, C., Caimi, F., Fabiani, A., Pesquita, C., Couto, F. M., Palmonari, M. (2011) Using AgreementMaker to Align Ontologies for OAEI 2011. Ontology Matching Workshop, International Semantic Web Conference.

Cross, V. and Yu, X. (2011) Investigating ontological similarity theoretically with fuzzy set theory, information content, and Tversky similarity and empirically with the Gene Ontology, Proc. of the 5th International Conference on Scalable Uncertainty Management.

Euzenat, J., Valtchev, P. (2003) An integrative proximity measure for ontology alignment. In: Proc. ISWC-2003 Workshop on semantic information integration, Sanibel Island (FL US), 33—38.

Euzenat, J.et al. (2011) The Results of the Ontology Alignment Evaluation Initiative 2011. Ontology Matching Workshop, International Semantic Web Conference.

Jean-Mary, Y.R. , Shironoshita, E. P., and. Kabuka, M. R (2009) Ontology matching with semantic verification, *Journal of Web Semantics,* vol. 7, no. 3, 235–251.

Gross, A., Hartung, M., Kirsten, T., and Rahm, E. (2011) Mapping Composition for Matching Large Life Science Ontologies. In Proc. of the International Conference on Biomedical Ontology, pages 109–116.

Lin D. (1998) An Information-theoretic Definition of Similarity. In: Proc. of the 15th International Conference on Machine Learning. Morgan Kaufmann. pp. 296-304.

Rada R, Mili H, Bicknell E, Blettner M (1989) Development and Application of a Metric on Semantic Nets. *IEEE Transaction on Systems, Man, and Cybernetics* vol. 19, 17—30.

Pirro, G., Talia D. (2010) UFOme: an ontology mapping system with strategy prediction capabilities. *Data Knowl.Eng.* 69.5, 444-471, 2010,

Resnik, P. (1995) Using Information Content to Evaluate Semantic Similarity in Taxonomy. In: Proc. of the 14th International Joint Conference on Artificial Intelligence, 448—453.

Sabou, M., d'Aquin, M. and Motta, E. (2008) Exploring the semantic web as background knowledge for ontology matching, *Journal on Data Semantics*, vol. XI, 156–190.

Seco N, Veale T, Hayes J. (2004) An Intrinsic Information Content Metric for Semantic Similarity in WordNet. In: ECAI. 1089 -1090.

Shvaiko, P. and Euzenat, J. (2012) Ontology Matching: State of the Art and Future Challenges, *IEEE Transactions on Data and Knowledge Engineering,* to appear in print.

Wu Z, Palmer M. S. (1994) Verb Semantics and Lexical Selection. In: Proc. of the 32nd. Annual Meeting of the Association for Computational Linguistics, 133-138.

Yu, X. (2010). A Mathematical and Experimental Investigation of Ontological Similarity Measures and their Use in Biomedical Domains. Master's Thesis, Computer Science and Software Engineering, Miami University, Oxford OH