

Using partial reference alignments to align ontologies

Patrick Lambrix and Qiang Liu

Department of Computer and Information Science
Linköpings universitet
581 83 Linköping, Sweden

Abstract. In different areas ontologies have been developed and many of these ontologies contain overlapping information. Often we would therefore want to be able to use multiple ontologies. To obtain good results, we need to find the relationships between terms in the different ontologies, i.e. we need to align them. Currently, there already exist a number of ontology alignment systems. In these systems an alignment is computed from scratch. However, recently, some situations have occurred where a partial reference alignment is available, i.e. some of the correct mappings between terms are given or have been obtained. In this paper we investigate whether and how a partial reference alignment can be used in ontology alignment. We use partial reference alignments to partition ontologies, to compute similarities between terms and to filter mapping suggestions. We test the approaches on previously developed gold standards and discuss the results.

This is a preprint of the paper published by Springer-Verlag:

Lambrix P, Liu Q, Using partial reference alignments to align ontologies, Proceedings of the 6th European Semantic Web Conference - ESWC09, LNCS 5554, 188-202, Heraklion, Greece, 2009.

The original publication is available at www.springerlink.com.

doi: 10.1007/978-3-642-02121-3_17

1 Introduction

In recent years many ontologies have been developed. The benefits of using ontologies include reuse, sharing and portability of knowledge across platforms, and improved documentation, maintenance, and reliability (e.g. [5]). Ontologies lead to a better understanding of a field and to more effective and efficient handling of information in that field. Many of the currently developed ontologies contain overlapping information. For instance, Open Biomedical Ontologies (OBO, <http://www.obofoundry.org/>) lists 26 different anatomy ontologies (October 2008). Often we would want to be able to use multiple ontologies. For instance, companies may want to use community standard ontologies and use them together with company-specific ontologies. Applications may need to use ontologies from different areas or from different views on one area. Ontology builders may want to use already existing ontologies as the basis for the creation of new ontologies by extending the existing ontologies or by combining knowledge from different smaller ontologies. In each of these cases it is important to know the relationships between the terms in the different ontologies. Further, the data in different

data sources in the same domain may have been annotated with different but similar ontologies. Knowledge of the inter-ontology relationships would in this case lead to improvements in search, integration and analysis of data. It has been realized that this is a major issue and much research has recently been done on ontology alignment, i.e. finding mappings between terms in different ontologies (e.g. [4]).

In the current alignment systems an alignment is computed from scratch. However, recently, some situations have occurred where a partial reference alignment is available, i.e. some of the correct mappings between terms are given or have been obtained. One example is the development of Bioportal (bioportal.bioontology.org, [13]) where mappings between different ontologies in the biomedical domain have been collected. Bioportal also supports collaborative ontology alignment (one of the challenges for ontology alignment described in [15]) where experts can focus on their piece of expertise. In this case for some parts of the ontologies mappings will be available while they are still lacking for other parts of the ontologies. Another situation is an iterative ontology alignment methodology where people and ontology alignment systems interact to iteratively align and improve the quality of the mappings. In systems such as SAMBO [7], users can input their own mappings as well as accept or reject mapping suggestions generated by the system. Both the mappings given by the users and the rejection and acceptance of system-generated mappings by users influence future iterations of ontology alignment. Finally, the Ontology Alignment Evaluation Initiative (OAEI, <http://oaei.ontologymatching.org/>) introduced in 2008 a new task in the Anatomy track [10] where a partial reference alignment was given and participating systems could use the partial reference alignment to improve the quality of their mapping suggestions.

In this paper we investigate whether and how a partial reference alignment (PRA) can be used in ontology alignment. We use PRAs in the different steps of ontology alignment. We use PRAs in a preprocessing step to partition the ontologies into mappable parts that are likely to contain correct mappings and therefore not every term in the first ontology needs to be compared to every term in the second ontology. We also use PRAs in the computation of similarities between terms. Further, we use PRAs to filter mapping suggestions. We test the approaches on previously developed gold standards (from [7] and the Anatomy track of OAEI) and discuss the results.

The remainder of the paper is organized as follows. In section 2 we give background information. We describe a framework for ontology alignment as well as the systems, the gold standards and the evaluation measures that we have used in our experiments. The experiments are described in section 3. We investigate the use of PRAs in the different components of the ontology alignment framework. We also investigate the influence of the size of the PRA on the quality of the alignments. The results of the experiments are shown and discussed in section 4. The paper concludes in section 5.

2 Background

2.1 Ontology alignment

A large number of ontology alignment systems have been developed (see, for instance, review papers [7, 14, 12, 6], the book [4] on ontology matching, and the ontology matching web site at <http://www.ontologymatching.org/>). Many ontology alignment systems

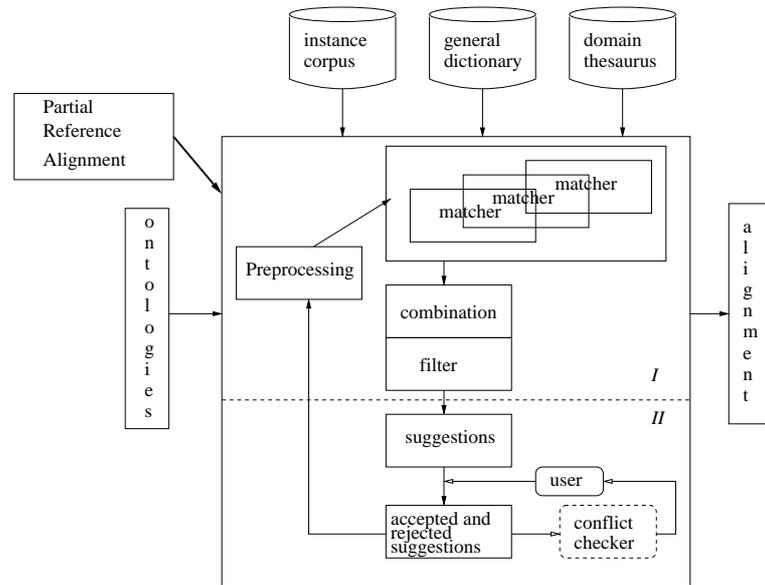


Fig. 1. Alignment framework (modified from [7]).

are based on the computation of similarity values between terms in different ontologies and can be described as instantiations of the general framework in figure 1. The framework consists of two parts. The first part (*I* in figure 1) computes mapping suggestions. The second part (*II*) interacts with the user to decide on the final alignment.

An alignment algorithm receives as input two source ontologies. A preprocessing step can be used to modify the original ontologies or to partition the ontologies into mappable parts. The algorithm can include several matchers that calculate similarities between the terms from the different source ontologies or mappable parts of the ontologies. They can implement strategies based on linguistic matching, structure-based strategies, constraint-based approaches, instance-based strategies, strategies that use auxiliary information or a combination of these. Each matcher utilizes knowledge from one or multiple sources. Mapping suggestions are then determined by combining and filtering the results generated by one or more matchers. By using different matchers and combining and filtering the results in different ways we obtain different alignment strategies. The suggestions are then presented to the user who accepts or rejects them. The acceptance and rejection of suggestions may influence further suggestions. Further, a conflict checker is used to avoid conflicts introduced by the mapping suggestions. The output of the alignment algorithm is a set of mappings between terms from the source ontologies. In this paper we add a partial reference alignment as input and study how this can be used in the different components of this framework.

2.2 SAMBO and SAMBOdtf

SAMBO and SAMBOdtf are ontology alignment systems based on the framework described in the previous section. They both currently contain five basic matchers [7]: two terminological matchers (a basic matcher and an extension using WordNet; extension described below), a structure-based matcher (which uses the is-a and part-of hierarchies of the source ontologies), a matcher based on domain knowledge (described below), and a learning matcher (which uses literature that is related to the concepts in the ontologies to define a similarity value between the concepts). In addition to these techniques we have also experimented with other matchers [9, 17].

In our evaluations we use the versions of SAMBO and SAMBOdtf as used in OAEI 2008 [8]. These systems performed best and second best, respectively, in the Anatomy track of OAEI 2008. These versions contain the matchers *TermWN* and *UMLSKSearch*. The matcher *TermWN* contains matching algorithms based on the textual descriptions (names and synonyms) of concepts and relations. In the current implementation, the matcher includes two approximate string matching algorithms (n-gram and edit distance), and a linguistic algorithm that also uses WordNet (<http://wordnet.princeton.edu/>) to find synonyms and is-a relations. Our matcher *UMLSKSearch* uses the Metathesaurus in the Unified Medical Language System (UMLS, <http://www.nlm.nih.gov/research/umls/>). The similarity of two terms in the source ontologies is determined by their relationship in UMLS. In our experiments we used the UMLS Knowledge Source Server to query the UMLS Metathesaurus with source ontology terms. The querying is based on searching the normalized string index and normalized word index provided by the UMLS Knowledge Source Server. We used version 2008AA of UMLS. As a result we obtain concepts that have the source ontology term as their synonym. We assign a similarity value of 0.99 if the source ontology terms are synonyms of the same concept and 0 otherwise.

The combination algorithm used for the OAEI versions of SAMBO and SAMBOdtf is a maximum-based algorithm. The similarity value for a pair of terms is the maximum of the values obtained from *TermWN* and *UMLSKSearch* for this pair of terms.

The filtering method in SAMBO is single threshold filtering. Pairs of concepts with a similarity value higher than or equal to a given threshold value are returned as mapping suggestions to the user. SAMBOdtf implements the double threshold filtering method developed in [1]. The double threshold filtering approach uses the structure of the ontologies. It is based on the observation that (for the different approaches in the evaluation in [7]) for single threshold filtering the precision of the results is decreasing and the recall is increasing when the threshold is decreasing. Therefore, we propose to use two thresholds. Pairs with similarity value equal or higher than the upper threshold are retained as suggestions. The intuition is that this gives suggestions with a high precision. Further, pairs with similarity values between the lower and the upper threshold are filtered using structural information and the rest is discarded. We require that the pairs with similarity values between the two thresholds are 'reasonable' from a structural point of view.¹ The intuition here is that the recall is augmented by adding new suggestions, while at the same time the precision stays high because only structurally

¹ In our implementation we have focused on the is-a relation.

reasonable suggestions are added. The double threshold filtering approach contains the following three steps. (i) Find a consistent suggestion group from the pairs with similarity value higher or equal than the upper threshold. We say that a set of suggestions is a consistent suggestion group if each concept occurs at most once as first argument in a pair, at most once as second argument in a pair and for each pair of suggestions $\langle A, A' \rangle$ and $\langle B, B' \rangle$ where A and B are concepts in the first ontology and A' and B' are concepts in the second ontology: $A \subset B$ iff $A' \subset B'$. (ii) Use the consistent suggestion group to partition the original ontologies. (iii) Filter the pairs with similarity values between the lower and upper thresholds using the partitions. Only pairs of which the elements belong to corresponding pieces in the partitions are retained as suggestions. For details we refer to [1].

In contrast to the original versions of SAMBO and SAMBOdtf where a term in one ontology can be suggested to be mapped to different terms in the other ontology, in the OAEI versions this is not the case. We retain only suggestions where the similarity between the terms in the suggestion is higher than or equal to the similarity of these terms to any other term according to the suggestion list. In the case there are different possibilities, one is randomly chosen. (In the implementation the first in the list is chosen.)

2.3 Gold standard alignments

In our tests we use alignments that have been used in previous evaluations ([7] and OAEI). Essentially, for our purpose all used ontologies can be seen as taxonomies (concepts and is-a and part-of relations). In [7] a number of smaller test cases were introduced to evaluate different alignment strategies. For the first two cases we use a part of a Gene Ontology ontology (GO) [2] together with a part of Signal Ontology (SigO) [16]. The first case, *B* (behavior), contains 57 terms from GO and 10 terms from SigO. Its reference alignment (RA) contains 4 mappings. The second case, *ID* (immune defense), contains 73 terms from GO and 17 terms from SigO. Its RA contains 8 mappings. The other cases are taken from the anatomy category of Medical Subject Headings (MeSH, [11]) and the Adult Mouse Anatomy (MA, available from OBO): *nose* (15 terms from MeSH, 18 terms from MA, and 7 mappings in the RA), *ear* (39 terms from MeSH, 77 terms from MA, and 27 mappings in the RA), and *eye* (45 terms from MeSH, 112 terms from MA, and 27 mappings in the RA).

A larger test case that we use is the case of the Anatomy track of OAEI. OAEI is a yearly initiative that was started in 2004. The goals are, among others, to assess the strengths and weaknesses of alignment systems, to compare different techniques and to improve evaluation techniques. This is to be achieved through controlled experimental evaluation. For this purpose OAEI publishes different cases of ontology alignment problems, some of which are open (RA is known beforehand), but most are blind (RA is not known - participants send their alignment suggestions to organizers who evaluate the performance). In the *Anatomy* case (version 2008) participants were required to align the Adult Mouse Anatomy (2744 concepts) and the NCI Thesaurus - anatomy (3304 concepts). The anatomy case is a blind case. The RA contains 1523 equivalence mappings of which 934 are deemed trivial (i.e. they can be found by a relatively basic string-based matcher).

2.4 Evaluation measures

The results of our experiments are given in terms of the quality of the mapping suggestions. We use precision, recall, $recall_{PRA}$ and f-measure. *Precision* measures how many of the mapping suggestions were correct. It is defined as the number of correct suggestions divided by the number of suggestions. *Recall* measures how many of the correct mappings are found by the alignment algorithm. It is defined as the number of correct suggestions divided by the number of correct mappings. We also introduce the measure $recall_{PRA}$ which measures how many of the correct mappings that are not in a PRA are found by the alignment algorithm. It is defined as the number of correct mapping suggestions not in the PRA divided by the number of correct mappings not in the PRA. *F-measure* is the weighted harmonic mean of precision and recall. In our tests precision and recall are weighted evenly.

	preprocessing	matchers	combination	filter
SAMBO	none	TermWN + UMLSKSearch	maximum	single threshold
SAMBOdtf	none	TermWN + UMLSKSearch	maximum	double threshold
mgPRA	partitioning	TermWN + UMLSKSearch	maximum	single threshold filter with PRA
mgfPRA	fixing and partitioning	TermWN + UMLSKSearch	maximum	single threshold filter with PRA
pmPRA	none	TermWN + UMLSKSearch pattern-based augmentation	maximum	single threshold filter with PRA
fPRA	none	TermWN + UMLSKSearch	maximum	single threshold filter with PRA
dtfPRA	none	TermWN + UMLSKSearch	maximum	double threshold with PRA filter with PRA
pfPRA	none	TermWN + UMLSKSearch	maximum	filter based on EM and PRA filter with PRA

Table 1. Alignment strategies.

3 Experiments

In this section we define a number of experiments to test the usefulness of using a PRA for ontology alignment. We consider the use of PRAs in the different components of the alignment framework in figure 1. As base systems we use SAMBO and SAMBOdtf as described in section 2.2 and modify their components. We describe the experiments and provide an overview of the resulting strategies in table 1.

An immediate observation regarding a PRA is that the mappings in the PRA are deemed to be correct and therefore should be included in the final result. For the same reason, if there are different suggestions for mapping a term, then a suggestion contained in the PRA is preferred. Therefore, in all alignment strategies using a PRA, we

add the mappings in the PRA to the list of suggestions with a special status. These mappings cannot be removed in any filtering step. This technique we call *filter with PRA*.

3.1 Use of PRA in the preprocessing step

A first question we want to investigate is whether we can use a PRA in the preprocessing phase. Most systems compute a similarity value between all terms from the first ontology and all terms from the second ontology. Some work has been done on partitioning the ontologies to find mappable parts of the ontologies (e.g. [3]). The motivation for that work is scalability. When the size of the ontologies grows, some techniques may become infeasible. In our work we investigate whether we can use a PRA to partition the ontologies into mappable parts and test whether, in addition to the fact that we do not have to compute similarity values between all terms from the first ontology and all terms from the second ontology, this also leads to a better quality of the mapping suggestions.

In the first approach we partition the ontologies into mappable parts using the partitioning step of the double threshold filtering described in section 2.2 and [1]. A part of the PRA satisfying the consistent group property is used as a consistent group. The resulting alignment strategy (this preprocessing, the SAMBO matchers, combination and filters, and filter with PRA) we call *mgPRA* (mappable groups with PRA).

According to our experience in aligning ontologies we know that the structure of the source ontologies is not always perfect. For instance, given the two ontologies and the PRA in the Anatomy task of OAEI 2008, it can be deduced that many is-a relations are missing in at least one of the source ontologies. Based on this observation we experiment with a second approach where we add to the source ontologies the missing is-a relationships that can be deduced from the source ontologies and the PRA.² After this 'fixing' of the source ontologies the PRA will satisfy the consistent group property. The resulting strategy is called *mgfPRA* (mappable groups and fixing with PRA).

3.2 Use of PRA in a matcher

One way to create a matcher based on a PRA, is to use underlying properties of the mappings in the PRA. We have previously observed that sometimes for two given source ontologies, common patterns can be found between the correct mappings. For instance, in the PRA of the OAEI 2008 Anatomy we find the mappings <lumbar vertebra 5, l5 vertebra> and <thoracic vertebra 11, t11 vertebra> which share a similar linguistic pattern. Also the mappings <forebrain, fore brain>, <gallbladder, gall bladder>, and the mappings <stomach body, body stomach> and <stomach fundus, fundus stomach> share similar linguistic patterns, respectively. When using different linguistic matchers the similarity values according to these matchers for the mappings sharing similar patterns are therefore very similar as well.

² This is in line with the challenge for ontology alignment on discovering missing background knowledge as described in [15].

Based on this observation we developed a matcher that augments previously generated similarity values for term pairs when these term pairs display a similar (linguistic) pattern as mappings in the PRA. For the experiment we used TermWN and UMLSKeywordSearch to compute the original similarity values. Further, for each term pair we compute a vector with, in this case, three similarity values, based on the different components of TermWN (n-gram, edit distance and a linguistic algorithm that uses WordNet; as described in section 2.2). Based on this similarity vector we compute the Euclidean distance of a term pair to the mappings in the PRA and count how many of the PRA mappings are within a predefined radius. (In the experiment we used 0.1 for the radius.) Based on the value of this count, the matcher may augment the original similarity value for the term pair. (In the experiment we used $(\text{count} * 0.06)$ as augmentation value and augmented only pairs with a similarity value lower than 0.9. We also set the limit for the augmented similarity value to 0.9.) The resulting alignment strategy is called *pmPRA* (pattern matcher with PRA).

3.3 Use of PRA in the filter step

Another question that we want to investigate is whether we can use a PRA for filtering the list of mapping suggestions.

As mentioned before, all strategies using PRAs implement the filter with PRA approach. In a first experiment we only add this filter approach to SAMBO. The resulting alignment strategy is called *fpRA*.

In addition to filter with PRA, the second strategy also uses a variant of the double threshold filtering approach of SAMBOdtf. While in the original double threshold filtering approach, a consistent suggestion group is computed based on mapping suggestions with a high similarity value, in this approach a part of the PRA satisfying the consistent group property is used as a consistent group. The resulting alignment strategy is called *dtfPRA*.

The last strategy we use is called *pfPRA* (pattern filter with PRA). This strategy is based on the observation described above that some correct mappings share similar patterns. Similar to *pmPRA*, we assign a vector with similarity values based on the different components of TermWN to each term pair in the suggestion list and the PRA. We compute clusters based on these similarity vectors of the term pairs in the suggestion list using the expectation-maximization algorithm. (For the small cases we set the number of clusters. For *Anatomy* we did not use a predefined number of clusters.) Then, each mapping in the PRA is assigned to the cluster for which the distance between the cluster center and the PRA mapping is the smallest. Further, we compute for each cluster the average distance between a PRA mapping in the cluster and its cluster center. The filter strategy retains the suggestions in a cluster with a distance to the cluster center that is smaller or equal to the computed average for that cluster and discards the others.

3.4 Influence of the size of the PRA

We also want to investigate the influence of the size of the PRA. For this purpose we compare the results of the approaches for the *Anatomy* case with two PRAs: PRA-F and its subset PRA-H. For PRA-F we take the PRA as provided by the OAEI 2008

Anatomy task. PRA-H contains half of the trivial and half of the non-trivial mappings from PRA-F.

Case	Th	SAMBO	SAMBOdtf	mgPRA	mgfPRA	pmPRA	fPRA	dtfPRA	pfPRA
<i>B</i>	0.4	6/2	6/2	6	2	6	6	4	3
	0.6	5/2	5/2	5	2	5	5	4	3
	0.8	4/2	-	4	2	4	4	-	3
<i>ID</i>	0.4	12/4	11/3	12	12	12	12	11	12
	0.6	8/4	7/3	5	5	8	8	8	8
	0.8	7/3	-	4	5	8	8	-	8
<i>nose</i>	0.4	7/4	7/4	7	4	7	7	7	6
	0.6	7/4	7/4	7	4	7	7	7	6
	0.8	7/4	-	7	4	7	7	-	6
<i>ear</i>	0.4	30/14	29/14	28	18	4	30	30	25
	0.6	29/14	29/14	27	18	3	29	29	25
	0.8	26/14	-	24	18	1	26	-	24
<i>eye</i>	0.4	31/13	30/13	30	13	6	31	31	23
	0.6	36/12	26/12	26	13	3	27	27	22
	0.8	24/11	-	25	13	2	26	-	22
<i>Anatomy</i>	0.4	1575/943	1527/940	1690	1663	1625	1601	1552	1251
	0.6	1466/942	1438/940	1488	1444	1598	1498	1474	1221
	0.8	1297/933	-	1308	1271	1528	1342	-	1139

Table 2. Number of mapping suggestions. For SAMBO and SAMBOdtf also the number of PRA mappings found.

4 Results and discussion

Table 2 shows the number of suggestions generated by the algorithms in the experiments. For SAMBO and SAMBOdtf we also added the number of mappings in the PRA that were found by those algorithms. We note that for all thresholds for *B*, *nose* and *ear*, for thresholds 0.4 and 0.6 for *ID* and for threshold 0.4 for *eye* SAMBO actually finds all mappings in the PRA. This means that in these cases adding the PRA to the solutions in itself does not improve recall. For *Anatomy* SAMBO does not find between 45 (threshold 0.4) and 55 (threshold 0.8) of the 988 mappings in the PRA.

The results in terms of precision, recall, f-measure and $recall_{PRA}$ are given in tables³ 3 to 7. The results in the tables are truncated values. Note also that as SAMBO and SAMBOdtf do not use a PRA, we have set $recall = recall_{PRA}$ for SAMBO and SAMBOdtf. Th is the threshold for the filtering for the single threshold approaches. For

³ As *Anatomy* is a blind case at OAEI we do not have the RA available. Therefore, we have sent the mapping suggestions for *Anatomy* to the organizers of the OAEI 2008 Anatomy track who have returned the values for the different evaluation measures.

SAMBOdtf and dtfPRA the upper threshold is always 0.8 while Th is the lower threshold. There are no results for SAMBOdtf and dtfPRA for upper threshold 0.8 as this would be the same as using SAMBO and fPRA with single threshold 0.8, respectively.

4.1 PRA in preprocessing

Case	RA	PRA	Th	SAMBO	mgPRA	mgfPRA
<i>B</i>	4	2	0.4	0.66/1.00/0.80/1.00	0.66/1.00/0.80/1.00	1.00/0.50/0.66/0.00
			0.6	0.80/1.00/0.88/1.00	0.80/1.00/0.88/1.00	1.00/0.50/0.66/0.00
			0.8	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00	1.00/0.50/0.66/0.00
<i>ID</i>	8	4	0.4	0.50/0.75/0.60/0.75	0.41/0.62/0.50/0.25	0.41/0.62/0.50/0.25
			0.6	0.75/0.75/0.75/0.75	1.00/0.62/0.76/0.25	1.00/0.62/0.76/0.25
			0.8	0.71/0.62/0.66/0.62	1.00/0.62/0.76/0.25	1.00/0.62/0.76/0.25
<i>nose</i>	7	4	0.4	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00	1.00/0.57/0.72/0.00
			0.6	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00	1.00/0.57/0.72/0.00
			0.8	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00	1.00/0.57/0.72/0.00
<i>ear</i>	27	14	0.4	0.86/0.96/0.91/0.96	0.85/0.88/0.87/0.76	1.00/0.66/0.80/0.30
			0.6	0.89/0.96/0.92/0.96	0.88/0.88/0.88/0.76	1.00/0.66/0.80/0.30
			0.8	0.96/0.92/0.94/0.92	1.00/0.88/0.94/0.76	1.00/0.66/0.80/0.30
<i>eye</i>	27	13	0.4	0.80/0.92/0.86/0.92	0.80/0.88/0.84/0.78	1.00/0.48/0.65/0.00
			0.6	0.92/0.88/0.90/0.88	0.92/0.88/0.90/0.78	1.00/0.48/0.65/0.00
			0.8	0.91/0.81/0.86/0.81	0.92/0.85/0.88/0.71	1.00/0.48/0.65/0.00
<i>Anatomy</i>	1523	988	0.4	0.82/0.85/0.83/0.85	0.78/0.87/0.82/0.64	0.78/0.85/0.81/0.58
			0.6	0.88/0.84/0.86/0.84	0.88/0.86/0.87/0.61	0.88/0.84/0.86/0.55
			0.8	0.94/0.80/0.87/0.80	0.96/0.82/0.89/0.50	0.96/0.80/0.88/0.45

Table 3. Using the PRA in the preprocessing phase (precision/recall/f-measure/recall_{PRA}).

The results of the experiments for using PRAs in the preprocessing step are given in table 3. The intuition behind mgPRA and mgfPRA is to partition the ontologies into mappable parts. Therefore, we can only generate mapping suggestions that are reasonable from a structural point of view. This suggests that, comparing to the base systems, the precision may become higher as suggestions that do not conform to the structure of the source ontologies cannot be made. As we add the PRA to the result, the recall may be increased as some of the PRA mappings may not be found by the base systems. However, the similarity values between the terms do not change and it is therefore not likely that new mappings are found. (The only way to find new mappings compared to the base system is when a mapping suggestion with high similarity in the base system cannot be suggested by using mgPRA and mgfPRA because the terms were in incompatible parts of the ontologies. In that case other suggestions involving these terms may be generated.) The results give some support to these intuitions. For threshold 0.8 the precision of mgPRA and mgfPRA is always equal to or higher than the precision for SAMBO. This is also almost always the case for threshold 0.6. For

threshold 0.4 there is no conclusive result. We also notice that, except for threshold 0.4 in *ID* and *Anatomy*, mgfPRA always gives better precision than SAMBO.

As expected, the recall for mgPRA and mgfPRA is equal to or less than the recall for SAMBO in most cases. For the large test case the recall is always higher for mgPRA and equal for mgfPRA. In the cases there is a loss of recall, this is due to a different modeling in the source ontologies. When the PRA satisfies the consistent group property, mgPRA and mgfPRA give the same results. When this is not the case, mgfPRA 'fixes' the source ontologies by adding missing is-a relationships. In most cases this has led to an improvement in precision, in the other cases the precision stayed the same. However, contrary to the intuition in many cases fixing the source ontologies has led to a decrease in recall. This is due to the use of is-a in the source ontologies. For instance, the hierarchical relation in MeSH covers both is-a and part-of. Therefore, not all of these relations should be treated as is-a. However, as mgfPRA cannot distinguish between these, it may fix the source ontology in the wrong way, by adding a hierarchical link (which mgfPRA interprets as is-a, but which should have been interpreted as part-of). For instance, in the *nose* case, having $\langle \text{nose}, \text{nose} \rangle$ in the PRA would lead to introducing is-a relations in MA between nose and its parts. Therefore, fixing the ontologies may lead to worse results. As for all alignment strategies using structural information, the quality of the underlying ontologies, the completeness of the structure and the correct use of the structural relations, has an important influence on the quality of the results.

Case	RA	PRA	Th	SAMBO	pmPRA
<i>B</i>	4	2	0.4	0.66/1.00/0.80/1.00	0.66/1.00/0.80/1.00
			0.6	0.80/1.00/0.88/1.00	0.80/1.00/0.88/1.00
			0.8	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00
<i>ID</i>	8	4	0.4	0.50/0.75/0.60/0.75	0.50/0.75/0.60/0.50
			0.6	0.75/0.75/0.75/0.75	0.75/0.75/0.75/0.50
			0.8	0.71/0.62/0.66/0.62	0.75/0.75/0.75/0.50
<i>nose</i>	7	4	0.4	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00
			0.6	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00
			0.8	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00
<i>ear</i>	27	14	0.4	0.86/0.96/0.91/0.96	0.86/0.96/0.91/0.92
			0.6	0.89/0.96/0.92/0.96	0.89/0.96/0.92/0.92
			0.8	0.96/0.92/0.94/0.92	0.96/0.92/0.94/0.84
<i>eye</i>	27	13	0.4	0.80/0.92/0.86/0.92	0.80/0.92/0.86/0.85
			0.6	0.92/0.88/0.90/0.88	0.89/0.92/0.90/0.85
			0.8	0.91/0.81/0.86/0.81	0.92/0.88/0.90/0.78
<i>Anatomy</i>	1523	988	0.4	0.82/0.85/0.83/0.85	0.78/0.83/0.81/0.54
			0.6	0.88/0.84/0.86/0.84	0.79/0.83/0.81/0.54
			0.8	0.94/0.80/0.87/0.80	0.83/0.83/0.83/0.52

Table 4. Using the PRA in a matcher (precision/recall/f-measure/recall_{PRA}).

4.2 PRA in a matcher

The results of the experiment using PRAs in a matcher are found in table 4. The intuition of the matcher was to augment the similarity values of suggestions that had a similar linguistic pattern as mappings in the PRA. For *B*, *nose* and *ear* the results for SAMBO and pmPRA are the same. The augmentation did not have any influence. For *B* and *nose* the recall was already 1, so that no influence was expected. For *ear* the found correct mappings had already a high similarity value and the missed correct mappings (two for threshold 0.8 and one for thresholds 0.4 and 0.6) did not have a similar linguistic pattern as the mappings in the PRA. For *ID* the recall and the precision were equal or became higher (because of the addition of the PRA to the results). For *eye* the recall improved or was the same. For *Anatomy* the precision and, for low thresholds, also the recall decreased. The recall increased for high thresholds. To investigate the increase of recall for *Anatomy*, we compared the results of using PRA-F with PRA-H. In this case augmenting allowed to find <lateral cuneiform, external cuneiform bone foot> and <brain arachnoid matter, cerebral arachnoid membrane>. For each of these there were 7 mappings in PRA-H with similar linguistic patterns.

Case	RA	PRA	Th	SAMBO	fPRA	pfPRA
<i>B</i>	4	2	0.4	0.66/1.00/0.80/1.00	0.66/1.00/0.80/1.00	1.00/0.75/0.85/0.50
			0.6	0.80/1.00/0.88/1.00	0.80/1.00/0.88/1.00	1.00/0.75/0.85/0.50
			0.8	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00	1.00/0.75/0.85/0.50
<i>ID</i>	8	4	0.4	0.50/0.75/0.60/0.75	0.50/0.75/0.60/0.50	0.50/0.75/0.60/0.50
			0.6	0.75/0.75/0.75/0.75	0.75/0.75/0.75/0.50	0.75/0.75/0.75/0.50
			0.8	0.71/0.62/0.66/0.62	0.75/0.75/0.75/0.50	0.75/0.75/0.75/0.50
<i>nose</i>	7	4	0.4	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00	1.00/0.85/0.92/0.66
			0.6	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00	1.00/0.85/0.92/0.66
			0.8	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00	1.00/0.85/0.92/0.66
<i>ear</i>	27	14	0.4	0.86/0.96/0.91/0.96	0.86/0.96/0.91/0.92	1.00/0.92/0.96/0.84
			0.6	0.89/0.96/0.92/0.96	0.89/0.96/0.92/0.92	1.00/0.92/0.96/0.84
			0.8	0.96/0.92/0.94/0.92	0.96/0.92/0.94/0.84	1.00/0.88/0.94/0.76
<i>eye</i>	27	13	0.4	0.80/0.92/0.86/0.92	0.80/0.92/0.86/0.85	0.95/0.81/0.88/0.64
			0.6	0.92/0.88/0.90/0.88	0.92/0.92/0.92/0.85	1.00/0.81/0.89/0.64
			0.8	0.91/0.81/0.86/0.81	0.92/0.88/0.90/0.78	1.00/0.81/0.89/0.64
<i>Anatomy</i>	1523	988	0.4	0.82/0.85/0.83/0.85	0.83/0.88/0.86/0.66	0.91/0.74/0.82/0.28
			0.6	0.88/0.84/0.86/0.84	0.89/0.87/0.88/0.64	0.93/0.74/0.82/0.27
			0.8	0.94/0.80/0.87/0.80	0.95/0.84/0.89/0.54	0.97/0.72/0.83/0.22

Table 5. Using the PRA during the filter phase - 1 (precision/recall/f-measure/recall_{PRA}).

4.3 PRA in a filter

The results for the experiments regarding the use of PRAs in the filter step are given in tables 5 and 6. In the filter phase mapping suggestions are removed. Both correct and

Case	RA	PRA	Th	SAMBOdtf	dtfPRA
<i>B</i>	4	2	0.4	0.66/1.00/0.80/1.00	1.00/1.00/1.00/1.00
			0.6	0.80/1.00/0.88/1.00	1.00/1.00/1.00/1.00
<i>ID</i>	8	4	0.4	0.45/0.62/0.52/0.62	0.54/0.75/0.63/0.50
			0.6	0.71/0.62/0.66/0.62	0.75/0.75/0.75/0.50
<i>nose</i>	7	4	0.4	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00
			0.6	1.00/1.00/1.00/1.00	1.00/1.00/1.00/1.00
<i>ear</i>	27	14	0.4	0.89/0.96/0.92/0.96	0.86/0.96/0.91/0.92
			0.6	0.89/0.96/0.92/0.96	0.89/0.96/0.92/0.92
<i>eye</i>	27	13	0.4	0.83/0.92/0.87/0.92	0.80/0.92/0.86/0.85
			0.6	0.92/0.88/0.90/0.88	0.92/0.92/0.92/0.85
<i>Anatomy</i>	1523	988	0.4	0.84/0.84/0.84/0.84	0.86/0.87/0.87/0.65
			0.6	0.89/0.84/0.86/0.84	0.90/0.87/0.88/0.64

Table 6. Using the PRA during the filter phase - 2 (precision/recall/f-measure/recall_{PRA}).

wrong suggestions can be removed and therefore both precision and recall can change compared to the base system. All strategies implement filter with PRA. Therefore, as we add the PRA to the result, the recall may be increased as some of the PRA mappings may not be found by the base systems. The precision and recall may increase when wrong suggestions are removed because there are mappings in the PRA involving the same terms. This intuition is supported by the test cases. For all cases fPRA has a higher or equal precision and recall than SAMBO.

In dtfPRA we use, in addition to filter with PRA, also the structure of the source ontologies to filter the suggestion list by filtering out the suggestions which are not reasonable with respect to the structure of the ontologies and the given PRA. The intuition is that wrong suggestions may be removed and that it is better to use a PRA than a computed suggestion group. For *B*, *ID* and *Anatomy*⁴ dtfPRA has a higher precision and recall than SAMBOdtf. For *ear* and *eye* the recall is equal or higher for dtfPRA than for SAMBOdtf. For lower threshold 0.4 the precision is lower for dtfPRA than for SAMBOdtf, while it is higher or equal for lower threshold 0.6. The lower precision for *ear* at threshold 0.4 comes from a suggestion <inner ear epithelium, inner hair cell> which was filtered out by SAMBOdtf, but not by dtfPRA. One reason could be that in the SAMBOdtf case the consistent group consisted of 17 mapping suggestions while the consistent part of the PRA only consisted of 9 mappings (as the whole PRA did not satisfy the consistent group property). The partitioning for SAMBOdtf could therefore result in smaller mappable parts. We also compared dtfPRA to SAMBOdtf with just adding the PRA to the results. In most cases we have the same precision and recall. For *ID* threshold 0.4 precision for dtfPRA is slightly better, while for *ear* threshold 0.4 it is

⁴ According to the results of OAEI 2008 Anatomy task [10], with respect to the unknown part of the RA, dtfPRA's precision increased with 0.040, its recall with 0.008 and its f-value with 0.025. dtfPRA was the system with the highest increase in f-value and was the only system that used the PRA to increase both precision and recall. In [10] dtfPRA is called 'SAMBOdtf for task 4'.

slightly worse. For *Anatomy* just adding the PRA to SAMBOdtf actually gives a slightly higher recall than dtfPRA (3 more correct mappings are found), but a lower precision.

The third filter strategy removes suggestions that do not have similar linguistic patterns than the mappings in the PRA. We expect therefore that some correct suggestions obtained through UMLS will be removed and therefore the recall may go down. This is indeed the case in our experiments. The precision for pfPRA is, however, always higher or equal to the precision for SAMBO. This is because the suggestions that had a linguistically similar pattern as mappings in the PRA were usually correct.

4.4 Influence of the size of the PRA

Strategy	Th	PRA-F	new-F	PRA-H	new-H	NF
mgPRA	0.4	0.78/0.87/0.82	345	0.80/0.85/0.82	351	44
	0.6	0.88/0.86/0.87	327	0.88/0.83/0.85	337	46
	0.8	0.96/0.82/0.89	281	0.95/0.80/0.86	281	50
mgfPRA	0.4	0.78/0.85/0.81	313	0.79/0.81/0.80	336	85
	0.6	0.88/0.84/0.86	295	0.87/0.80/0.83	321	87
	0.8	0.96/0.80/0.88	243	0.95/0.76/0.84	268	89
pmPRA	0.4	0.78/0.83/0.81	290	0.77/0.83/0.80	313	26
	0.6	0.79/0.83/0.81	290	0.79/0.83/0.81	312	26
	0.8	0.83/0.83/0.83	282	0.84/0.82/0.83	294	28
fPRA	0.4	0.83/0.88/0.86	356	0.83/0.86/0.84	357	25
	0.6	0.89/0.87/0.88	347	0.88/0.86/0.87	348	26
	0.8	0.95/0.84/0.89	293	0.95/0.82/0.88	294	30
pfPRA	0.4	0.91/0.74/0.82	152	0.90/0.74/0.81	179	32
	0.6	0.93/0.74/0.82	148	0.92/0.74/0.82	175	33
	0.8	0.97/0.72/0.83	118	0.96/0.71/0.82	136	34
dtfPRA	0.4	0.86/0.87/0.87	350	0.84/0.86/0.85	355	26
	0.6	0.90/0.87/0.88	344	0.89/0.86/0.87	348	26

Table 7. *Anatomy* (1523 correct mappings in the RA) with PRA-F (988 mappings) and PRA-H (494 mappings) - (precision/recall/f-measure). new-X represents the number of correct mappings not in PRA-F found by using PRA-X. NF is the number of mappings in PRA-F not found by the algorithms using PRA-H.

The results for the experiment regarding the influence of the size of the PRA are shown in table 7. Intuitively, the more correct mappings we have, the higher the recall should be. This is supported by all test cases. As shown in the last column of table 7, several mappings in PRA-F are not found by the algorithms using PRA-H. However, using PRA-H seems to generate more mappings from the unknown part of the RA. On the other hand, it also generates more wrong suggestions. One explanation could be that for most strategies the larger the PRA, the more constraints the suggestions need to satisfy and thus the fewer suggestions (correct and wrong) are generated. For the preprocessing strategies mgPRA and mgfPRA the precision is lower for the larger PRA

when we use a low threshold. However, the precision is better for the larger PRA when the threshold is high. The matcher strategy pmPRA shows a similar tendency. For the three filtering strategies the precision for the larger PRA is always better or equal than the precision of the smaller PRA.

5 Conclusion

In this paper we have investigated whether and how a PRA can be used in ontology alignment by experimenting with using a PRA in the different components of ontology alignment systems. The use of PRA in preprocessing and filtering reduces the number of suggestions and in most cases leads to an improvement in precision. In some cases also the recall improved. Filter with PRA should always be used. For approaches using structural information the quality of the structure in the underlying ontologies has a large impact. The matcher using linguistic patterns in the PRA mappings can be used for finding new suggestions. The differences between the results for the algorithms that use a PRA and the base systems are relatively small. However, considering the nature of the test cases and the fact that SAMBO and SAMBOdtf perform already well on their own, even small improvements are valuable. Also, for the large test case, due to the choice of the PRA all newly found mappings are non-trivial. There are different directions for future work. The algorithms should be tested on other ontologies and with different base algorithms. Combinations and interactions of the methods should be investigated. It would also be interesting to look at other kinds of patterns in alignment data. The approach should also be integrated in an iterative ontology alignment framework.

Acknowledgements We are indebted to Christian Meilicke for running our algorithms on the Anatomy track test case of the OAEI. We also acknowledge the financial support of the Swedish Research Council (Vetenskapsrådet).

References

1. B Chen, H Tan, and P Lambrix. Structure-based filtering for ontology alignment. In *Proc. IEEE WETICE Workshop on Semantic Technologies in Collaborative Applications*, pages 364–369, 2006.
2. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29, 2000.
3. M Ehrig and S Staab. QOM - quick ontology mapping. In *Proc. International Semantic Web Conference*, pages 683–697, 2004.
4. J Euzenat and P Shvaiko. *Ontology Matching*. Springer, 2007.
5. R Jasper and M Uschold. A framework for understanding and classifying ontology applications. In *Proc. IJCAI Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends*, 1999.
6. Y Kalfoglou and M Schorlemmer. Ontology mapping: the state of the art. *The Knowledge Engineering Review*, 18(1):1–31, 2003.
7. P Lambrix and H Tan. SAMBO - a system for aligning and merging biomedical ontologies. *Journal of Web Semantics*, 4(3):196–206, 2006.
8. P Lambrix, H Tan, and Q Liu. SAMBO and SAMBOdtf results for the ontology alignment evaluation initiative 2008. In *Proc. 3rd International Workshop on Ontology Matching*, 2008.

9. P Lambrix, H Tan, and W Xu. Literature-based alignment of ontologies. In *Proc. 3rd International Workshop on Ontology Matching*, 2008.
10. C Meilicke and H Stuckenschmidt. Anatomy track at the Ontology Alignment Evaluation Initiative 2008. Anatomy at <http://oaei.ontologymatching.org/2008/>.
11. MeSH. Medical subject headings. <http://www.nlm.nih.gov/mesh/>.
12. NF Noy. Semantic integration: A survey of ontology-based approaches. *Sigmod Record*, 33(4):65–70, 2004.
13. NF Noy, N Griffith, and M Musen. Collecting community-based mappings in an ontology repository. In *Proc. 7th International Semantic Web Conference*, pages 371–386, 2008.
14. P Shvaiko and J Euzenat. A survey of schema-based matching approaches. *Journal on Data Semantics*, IV:146–171, 2005.
15. P Shvaiko and J Euzenat. Ten challenges for ontology matching. In *Proc. 7th International Conference on Ontologies, Databases, and Applications of Semantics*, 2008.
16. T Takai-Igarashi, Y Nadaoka, and T Kaminuma. A database for cell signaling networks. *Journal of Computational Biology*, 5(4):747–754, 1998.
17. H Tan, V Jakonienė, P Lambrix, J Aberg, and N Shahmehri. Alignment of biomedical ontologies using life science literature. In *Proc. International Workshop on Knowledge Discovery in Life Science Literature, LNBI 3886*, pages 1–17, 2006.