

SAMBO and SAMBOdtf Results for the Ontology Alignment Evaluation Initiative 2008

Patrick Lambrix, He Tan, and Qiang Liu

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

Abstract. This article describes a base system for ontology alignment, SAMBO, and an extension, SAMBOdtf. We present their results for the benchmark, anatomy and FAO tasks in the 2008 Ontology Alignment Evaluation Initiative. For the benchmark and FAO tasks SAMBO uses a strategy based on string matching as well as the use of a thesaurus. It obtains good results in many cases. For the anatomy task SAMBO uses a combination of string matching and the use of domain knowledge. This combination performed well in former evaluations using other anatomy ontologies. SAMBOdtf uses the same strategies but, in addition, uses an advanced filtering technique that augments recall while maintaining a high precision.

1 Presentation of the system

In this section we present the purpose of SAMBO and SAMBOdtf, the framework on which they are built, the specific techniques that are used and the adaptations made for the evaluation.¹

1.1 State, purpose, general statement

Although several of our methods and techniques are general and applicable to different areas, when developing SAMBO, we have focused on biomedical ontologies. We chose this field because ontologies are recognized as important in some of the grand challenges in the biomedical domain, and many biomedical ontologies have been developed and are publicly available and have overlapping information. This has, however, had an influence on the approaches on which we focused. In general, ontologies may contain concepts, relations, instances and axioms. Most biomedical ontologies are controlled vocabularies, taxonomies, or thesauri. This means that they may contain concepts, is-a and part-of relations, and sometimes a limited number of other relationships. Therefore, we have mainly developed methods that are based on these ontology components. For some approaches we have also used documents about a concept as instances for that concept. We have not dealt with axioms. SAMBOdtf is an extension of SAMBO that uses an advanced filtering method.

¹ Some parts of the description of the system are the same as last year's description in [11].

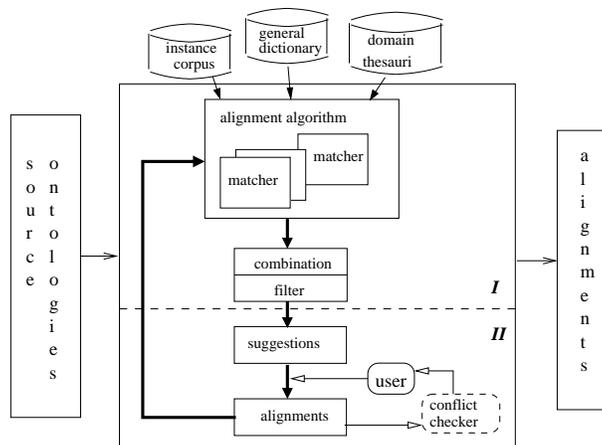


Fig. 1. Alignment framework [4].

1.2 Framework

SAMBO and SAMBOdtf are based on the framework shown in figure 1 [4]. The framework consists of two parts. The first part (*I* in figure 1) computes alignment suggestions. The second part (*II*) interacts with the user to decide on the final alignments. An alignment algorithm receives as input two source ontologies. The algorithm includes one or several matchers, which calculate similarity values between the terms from the different source ontologies. The matchers may use knowledge from different sources. Alignment 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 a suggestion may influence further suggestions. Further, a conflict checker is used to avoid conflicts introduced by the alignment relationships. The output of the alignment algorithm is a set of alignment relationships between terms from the source ontologies.

1.3 Specific techniques used

In this section we describe the matchers, and combination and filtering techniques that are available in SAMBO and SAMBOdtf. These matchers and techniques were previously evaluated using test cases for aligning Gene Ontology and Signal Ontology, and for aligning Medical Subject Headings (MeSH) and the Anatomical Dictionary for the Adult Mouse (MA) [4] using the KitAMO evaluation environment [5].² In addition to these techniques we have also experimented with other matchers [7, 9, 12]. We are also working on methods for recommendation of alignment strategies [10] which we intend to integrate into SAMBO in the future.

² An introduction to SAMBO and KitAMO can be found in [6].

Matchers SAMBO and SAMBOdtf contain currently five basic matchers: two terminological matchers, a structure-based matcher, a matcher based on domain knowledge, and a learning matcher. We describe the matchers used in OAEI-2008, and mention the others briefly.

Terminological matchers. The basic terminological matcher, *Term* 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. An n-gram is a set of n consecutive characters extracted from a string. Similar strings will have a high proportion of n-grams in common. Edit distance is defined as the number of deletions, insertions, or substitutions required to transform one string into the other. The greater the edit distance, the more different the strings are. The linguistic algorithm computes the similarity of the terms by comparing the lists of words of which the terms are composed. Similar terms have a high proportion of words in common in the lists. A Porter stemming algorithm is employed to each word. These algorithms were evaluated in [3] using MeSH anatomy (ca 1400 terms) and MA (ca 2350 terms). *Term* computes similarity values by combining the results from these three algorithms using a weighted sum. The combination we use in our experiments (weights 0.37, 0.37 and 0.26 for the linguistic algorithm, edit distance and n-gram, respectively) outperformed the individual algorithms in our former evaluations [3]. Further, the matcher *TermWN* is based on *Term*, but uses a general thesaurus, WordNet (<http://wordnet.princeton.edu/>), to enhance the similarity measure by looking up the hypernym relationships of the pairs of words in WordNet.

Structural matcher. The structural matcher is an iterative algorithm based on the is-a and part-of hierarchies of the ontologies. The algorithm requires as input a list of alignment relationships and similarity values and can therefore not be used in isolation. The intuition behind the algorithm is that if two concepts lie in similar positions with respect to is-a or part-of hierarchies relative to already aligned concepts in the two ontologies, then they are likely to be similar as well.

Use of domain knowledge. Another strategy is to use domain knowledge. Our matcher *UMLSKeywordSearch* 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 1³ if the source ontology terms are synonyms of the same concept and 0 otherwise.⁴

³ For the anatomy task we assign a value of 0.99 in order to introduce a preference of exact string matching over *UMLSKeywordSearch*. Although this is not useful for SAMBO, it is used in the adaptations made especially for OAEI.

⁴ Observe that this is slightly different from the version reported in [4] where we used version 2005AA of UMLS and we assigned a similarity value of 1 for two terms with the exact same names, 0.6 if the source ontology terms are synonyms of the same concept, and 0 otherwise.

Learning matcher. The matcher makes use of life science literature that is related to the concepts in the ontologies. It is based on the intuition that a similarity measure between concepts in different ontologies can be defined based on the probability that documents about one concept are also about the other concept and vice versa.

Combinations The user is given the choice to employ one or several matchers during the alignment process. The similarity values for pairs of concepts are then determined based on the similarity values computed by one matcher, or as a weighted sum of the similarity values computed by different matchers.

Filtering 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 alignment suggestions to the user.

SAMBOf 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 [4]) for single threshold filtering the precision of the results is decreasing and the recall is increasing when the thresholds are 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.⁵ 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 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 (A,A') and (B,B') 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].

1.4 Adaptations made for the evaluation

SAMBO and SAMBOf are interactive alignment systems. The alignment suggestions calculated by SAMBO and SAMBOf are normally presented to the user who accepts or rejects them. Alignment suggestions with the same concept as first item in the pair are shown together to the user. Therefore, the systems show the user the different alternatives for aligning a concept. This is a useful feature, in particular when the system

⁵ In our implementation we have focused on the is-a relation.

computes similarity values which are close to each other and there is no or only a small preference for one of the suggestions. Further, the acceptance and rejection of a suggestion may influence which suggestions are further shown to the user.

The computation of the alignment suggestions in SAMBO and SAMBOdtf is based on the computation of a similarity value between the concepts. The computation of the similarity values does not take into account what the relationship of the alignment should be. However, when an alignment is accepted, the user can choose whether the alignment relationship should be an equivalence relation or an is-a relation.

As the OAEI evaluation only considers the non-interactive part of the system and the computation of the similarity values does not take the relationship into account, we had to modify the computation of the suggestions. It would not make sense to have alignment suggestions where a concept appears more than once as the user would not be able to make a choice. Therefore, we decided to filter our systems' alignment suggestion lists such that only suggestions are retained where the similarity between the concepts in the alignment suggestion is higher than or equal to the similarity of these concepts to any other concept according to the alignment suggestion list. (In the case there are different possibilities, one is randomly chosen. In the implementation the first in the list is chosen.)

1.5 Link to the system and parameters file

The SAMBO (and SAMBOdtf) project page is at <http://www.ida.liu.se/~iislab/projects/SAMBO/>.

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

The suggested alignments are available at <http://www.ida.liu.se/~iislab/projects/SAMBO/OAEI/2008/>.

2 Results

We have provided alignment suggestions for the tasks 'benchmark', 'anatomy' and 'FAO'. Tests were performed on a IBM R61i Laptop, WinXP Intel(R) Pentium(R) Dual T2370 @ 1.73GHz, 1.73GHz, 1.99G RAM.

2.1 Benchmark

For the benchmark task the results for SAMBO were obtained by using TermWN with threshold 0.6. We introduced a preprocessing step where we used two strategies to generate names and for each case used the one that gave the best result for TermWN. The first strategy splits names based on capital letters occurring within a name. For instance, 'InCollection' was split into 'In Collection'. In the second strategy we remove stop words such that, for instance, 'is part of' is converted into 'part'. We did not use the comment field. The results may be improved using also this field.

We assume that ontology builders use a reasonable naming scheme and thus we did not tackle the cases where labels were replaced by a random one. Therefore, the recall for tests 201-202, 248-254, 257-262, 265-266. For these cases we may use other kind of information in the ontology such as the comment field or the structure. For the tests that were new for this year [*-2,4,6,8] where the labels are scrambled, the precision is high. In general, the recall is high when few of the labels are scrambled and drops when more labels are scrambled. We also did not focus on different natural languages (206-207, 210) or subsumption relationships (302).

Regarding the other cases we received high precision and recall except for cases 205 and 209. For 205 and 209 we had expected that using WordNet would be an advantage. Therefore, we compared the results with a run using Term (without WordNet). The differences between the results for Term and TermWN were small for all cases, including cases 205 and 209.

For SAMBOdtf we used the same matcher with upper threshold 0.8 and lower threshold 0.4. In the cases where there is no is-a hierarchy, SAMBOdtf with upper threshold 0.8 gives the same results as SAMBO with threshold 0.8. This is also the case when there are no suggestions with similarity value above 0.8, or no suggestions with similarity value between 0.4 and 0.8. Most of the test cases for benchmark belonged to one of these categories. For other test sets, we got the same result as SAMBO for [252-2,4,6,8], [259-2,4,6,8], [261-2,4,6,8], and 301. We obtained a little better recall for 204-210 and 304, since the lower threshold introduced some new alignments, most of which were correct.

2.2 Anatomy

Task 1 The results for the anatomy task for SAMBO were obtained by first running exact string matching and retaining the pairs with similarity values 1. On the remainder we run UMLSKSearch and retain the pairs with similarity value at least 0.99. Finally, we run TermWN⁶ with threshold 0.6 on the remainder of the pairs. With respect to the computation of the suggestions, this would be similar to having a matcher that returns as similarity value for a pair the maximum of the similarity value for the pair according to UMLSKSearch and the similarity value for the pair according to TermWN, and then using 0.6 as threshold. SAMBO generated 1465 alignment suggestions. SAMBO reached a precision of 0.869, a recall of 0.836 and an f-value of 0.852. Further, it reached a recall+ of 0.586. This was the best result for all 9 participating systems in OAEI 2008.⁷ In 2007 we used a version of SAMBO that used Term instead of TermWN and a previous version of UMLS. The 2007 version reached a better recall for non-trivial alignments, but at the cost of an overall decrease of precision and recall. A possible explanation for this is our strategy for choosing maximum one alignment suggestion per concept. In 2008 exact matching strings were preferred, while in 2007 there was no preference between pairs that had exact matching strings or pairs that were proposed based on domain knowledge.

⁶ Last year we used Term instead of TermWN.

⁷ The system with best f-measure in 2007 obtained 0.928 precision, 0.815 recall, 0.523 recall+ and 0.868 f-measure.

For SAMBOdtf, the same strategy is used, but with upper threshold 0.8 and lower threshold 0.4. SAMBOdtf generates 1527 alignment suggestions. Of these suggestions, 1440 have a similarity value between 0.6 and 0.8. This means that SAMBOdtf filtered out 25 of the suggestions obtained by SAMBO with threshold 0.6. (A manual check seems to suggest that most of these are correctly filtered out, but some are wrongly filtered out.) Further, SAMBOdtf also filtered out 19 suggestions with similarity values between 0.4 and 0.6. (A manual check seems to suggest that these were correctly filtered out.) SAMBO reached a precision of 0.831, a recall of 0.833, an f-value of 832 and a recall+ of 0.579. This was the second best result for all 9 participating systems in OAEI 2008.

The running time for SAMBO was ca 12 hours and for SAMBOdtf ca 17 hours.

Task 4 For task 4, we augmented SAMBO and SAMBOdtf in the following ways.

For SAMBO we added the alignments in the partial reference list to the list of alignment suggestions, but with a special status. These alignments could not be removed in the special filtering step that was introduced for OAEI (see section 1.4). SAMBO generated 1494 suggestions of which 988 are also in the partial reference list. SAMBO obtained the best results of the participating systems. With respect to the unknown part of the reference alignment, its precision increased with 0.024, its recall decreased with 0.002 and its f-value increased with 0.011

For SAMBOdtf we also added the alignments in the partial reference list to the list of alignment suggestions with the special status. In addition, we used the partial reference list in the double threshold filtering step. We used a consistent part⁸ of the partial reference list as a consistent suggestion group. For upper threshold 0.8 and lower threshold 0.4 we obtained 1547 alignment suggestions. SAMBOdtf obtained the second best results of the participating systems. With respect to the unknown part of the reference alignment, its precision increased with 0.040, its recall with 0.008 and its f-value with 0.025. SAMBOdtf was the system with the highest increase in f-value and was the only system that used the partial reference alignment to increase both precision and recall. This result is most likely due to the fact that, in contrast to task 1 where the consistent suggestion group consists of suggestions, in this task the consistent suggestion group consists of true alignments. Therefore, the suggestions with similarity value between the two thresholds that are retained are structurally reasonable with respect to true alignments and not just (although with high confidence) suggestions.

2.3 FAO

We only show results for the first task in FAO. For SAMBO we used TermWN with threshold 0.6. For SAMBOdtf we used TermWN with upper threshold 0.8 and lower threshold 0.4.

3 General comments

A problem that users face is that often it is not clear how to get the best alignment results given that there are many strategies to choose from. In most systems (including

⁸ The partial reference list is actually not a consistent group.

ours) there usually is no strategy for choosing the matchers, combinations and filters in an optimal way. Therefore, we used our experience from previous evaluations [4, 1] to decide which matchers and thresholds to use for which task. The lack of an optimization strategy is also the reason why we did not provide results for the second and third test for anatomy (optimization with respect to precision and recall, respectively). In the future, however, this may be possible using recommendation methods for alignment strategies such as proposed in [10] that will be able to recommend matchers, combinations and filters based on the alignment task and evaluation methods.

The OAEI deals with the non-interactive part of the alignment systems. This allows for evaluating how good the alignment suggestions are. However, for some systems, such as SAMBO and SAMBOdtf, the list of alignment suggestions is only an initial list and is updated after each acceptance or rejection of a suggestion.

4 Conclusion

We have briefly described our ontology alignment systems SAMBO and SAMBOdtf and some results of them on the alignment tasks of OAEI.

For the benchmark task we have used TermWN and obtained good results in many cases. We expect that the results will still improve when we use more information available in the ontology, such as the comment field and the structure.

Regarding the anatomy task we have used a combination of UMLSKSearch and TermWN, which performed best in former evaluations using other anatomy ontologies. We are currently also evaluating instance-based matchers [7].

A major problem is deciding which algorithms should be used for a given alignment task. This is a problem that users face, and that we have also faced in the evaluation. We expect that recommendation strategies [10, 8, 2] will alleviate this problem.

References

1. B Chen, H Tan, and P Lambrix. Structure-based filtering for ontology alignment. In *Proceedings of the IEEE WETICE Workshop on Semantic Technologies in Collaborative Applications*, pages 364–369, 2006.
2. M Ehrig, S Staab, and Y Sure. Bootstrapping ontology alignment methods with APFEL. In *Proceedings of the International Semantic Web Conference*, pages 186–200, 2005.
3. P Lambrix and H Tan. Merging DAML+OIL ontologies. In J Barzdins and A Caplinskas, editors, *Databases and Information Systems - Selected Papers from the Sixth International Baltic Conference on Databases and Information Systems*, pages 249–258. IOS Press, 2005.
4. P Lambrix and H Tan. SAMBO - a system for aligning and merging biomedical ontologies. *Journal of Web Semantics, Special issue on Semantic Web for the Life Sciences*, 4(3):196–206, 2006.
5. P Lambrix and H Tan. A tool for evaluating ontology alignment strategies. *Journal on Data Semantics, LNCS 4380*, VIII:182–202, 2007.
6. P Lambrix and H Tan. Ontology alignment and merging. In Burger, Davidson, and Baldock, editors, *Anatomy Ontologies for Bioinformatics: Principles and Practice*, pages 133–150. Springer, 2008.

7. P Lambrix, H Tan, and W Xu. Literature-based alignment of ontologies. In *Proceedings of the Third International Workshop on Ontology Matching*, 2008.
8. M Mochol, A Jentzsch, and J Euzenat. Applying an analytic method for matching approach selection. In *Proceedings of the Workshop on Ontology Matching*, 2006.
9. H Tan, V Jakonienė, P Lambrix, J Aberg, and N Shahmehri. Alignment of biomedical ontologies using life science literature. In *Proceedings of the International Workshop on Knowledge Discovery in Life Science Literature, LNBI 3886*, pages 1–17, 2006.
10. H Tan and P Lambrix. A method for recommending ontology alignment strategies. In *Proceedings of the 6th International Semantic Web Conference, LNCS 4825*, pages 494–507, 2007.
11. H Tan and P Lambrix. SAMBO results for the ontology alignment evaluation initiative 2007. In *Proceedings of the Second International Workshop on Ontology Matching*, pages 236–243, 2007.
12. T Wächter, H Tan, A Wobst, P Lambrix, and M Schroeder. A corpus-driven approach for design, evolution and alignment of ontologies. In *Proceedings of the Winter Simulation Conference*, pages 1595–1602, 2006. Invited contribution.