

## SAMBO

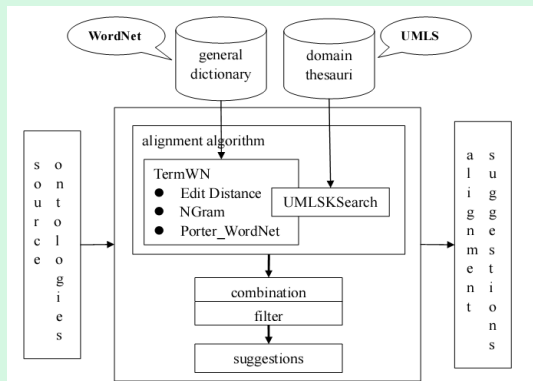
### System for Aligning and Merging Biomedical Ontologies

JAVA web application built on our ontology alignment framework [1]

The screenshot shows the SAMBO web application interface. On the left, there are 'matchers' for 'nose\_MA' and 'nose\_MeSH' with checkboxes for TermBasic, TermWN, UMLSKSearch, Hierarchy, and BayesLearning. A 'threshold' is set to 0.6. A 'weighted sum combination and threshold filtering' note is present. On the right, 'Alignment Candidate Details' are shown for 'nasal cavity epithelium' (nose\_MA) and 'nasal mucosa' (nose\_MeSH), including their definitions and synonyms. A 'comment on the alignment' field and a 'new name for the alignment' field are also visible.

### SAMBO Adaptation for the Ontology Alignment Evaluation Initiative

1. Only non-interactive part of the system.
2. A concept can appear in at most one alignment suggestion.
3. Maximum-based combination algorithm.



## SAMBOdtf

### SAMBO with Double Threshold Filtering

Intuition [2]:

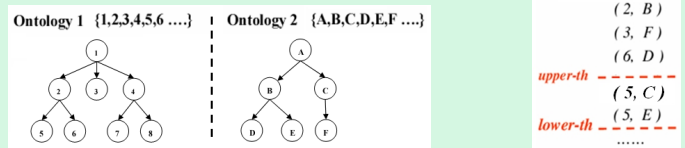
The idea is based on the observation that for single threshold filtering often the precision of the results is decreasing and the recall is increasing when the threshold is decreasing. Therefore, we propose to use two thresholds during the filtering phase, in which:

- 1) Pairs with similarity value equal or higher than the upper threshold are retained as suggestions, and
- 2) Pairs with similarity value between the lower and upper threshold are filtered using structural information (such as is-a relationships).

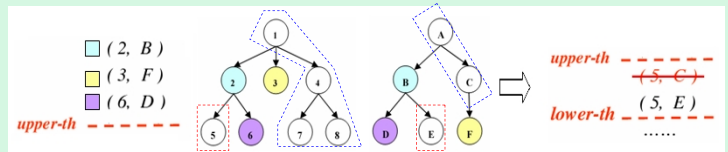
The intuition is that the recall is augmented by adding new suggestions, while at the same time the precision stays high because only structurally reasonable suggestions are added.

### A Simplified Example:

1. Suppose we have two ontologies and their is-a relationships.
2. Calculate similarity values between their concepts.



3. Use suggestions above upper threshold to partition the ontologies into match groups, which are used to filter the suggestions between the two thresholds.



4. Final suggestions consist of 1) pairs with similarity value above upper threshold and 2) pairs of concepts with similarity value between the two thresholds for which the concepts belong to related match groups.

## Results for SAMBO and SAMBOdtf for the Anatomy Case

### Anatomy Case:

Align two biological ontologies:

- Adult Mouse Anatomy (2744 concepts)
- NCI Thesaurus – human anatomy (3304 concepts)

We participate in two subtasks:

- Task 1 : Align the ontologies.
- Task 4 : Align the ontologies using a partial reference alignment.

### System Parameter Setting:

#### Matchers

- TermWN (approximate string matching + dictionary)
- UMLSKSearch (domain knowledge)

#### Combination

- Maximum of TermWN and UMLSKSearch results

#### Filtering

- SAMBO: single threshold filtering (threshold 0.6)
- SAMBOdtf: double threshold filtering (upper threshold 0.8, lower threshold 0.4)

### Results:

The table below shows the number of suggestions in the final results. For SAMBOdtf, the number after '/' indicates the number of suggestions with similarity value between the two thresholds that are filtered out.

System	Anatomy Task 1	Anatomy Task 4
SAMBO	1465 suggestions	1494 suggestions
SAMBOdtf	1527 suggestions / 44	1547 suggestions / 49

### Evaluation Result for Anatomy Task 1:

SAMBO and SAMBOdtf achieved the best and second best F-value, respectively, among all participating systems in 2008.

System	Runtime	BK	Precision	Recall	Recall+	F-value
SAMBO	≈ 12h	yes	0.869 <sub>0.845</sub>	0.836 <sub>0.797</sub>	0.586 <sub>0.601</sub>	0.852 <sub>0.821</sub>
SAMBOdtf	≈ 17h	yes	0.831	0.833	0.579	0.832

### Evaluation Result for Anatomy Task 4:

Use of partial reference alignment:

- SAMBO : The alignments in the partial reference alignment were included in the suggestions and cannot be removed by the filtering algorithm.
- SAMBOdtf : The alignments in the partial reference alignment were used to generate "match groups".

Among all participating systems in 2008

- SAMBO and SAMBOdtf achieved the best and second best F-value, respectively.
- SAMBOdtf obtained the highest increase in F-value.
- Only SAMBOdtf improved both precision and recall with respect to the unknown part of the reference alignment.

System	Δ-Precision	Δ-Recall	Δ-F-value
SAMBO	+0.024	-0.002	+0.011
SAMBOdtf	+0.040	+0.008	+0.025

The above table shows the overall quality increase/decrease of the generated suggestions with respect to the unknown part of the reference alignment.

## References

- [1] Lambrich P, Tan H, SAMBO - A System for Aligning and Merging Biomedical Ontologies, *Journal of Web Semantics*, 4(3):196-206, 2006.
- [2] Chen B, Tan H, Lambrich P, Structure-based filtering for ontology alignment, *IEEE WETICE Workshop on Semantic Technologies in Collaborative Applications*, 364-369, 2006.
- [3] Lambrich P, Tan H, Liu Q, SAMBO and SAMBOdtf Results for the Ontology Alignment Evaluation Initiative 2008, *International Workshop on Ontology Matching*, 2008.

Contact: P. Lambrich  
Department of Computer and Information Science  
Linköping universitet, Sweden

sambo@ida.liu.se  
<http://www.ida.liu.se/~iislab/projects/SAMBO/>