

EVOCROS: Results for OAEI 2019

Juliana Medeiros Destro¹, Javier A. Vargas¹, Julio Cesar dos Reis¹, and
Ricardo da S. Torres²

¹Institute of Computing, University of Campinas, Campinas-SP, Brazil

²Norwegian University of Science and Technology (NTNU), Ålesund, Norway
{juliana.destro,jreis}@ic.unicamp.br
ricardo.torres@ntnu.no, jalvarm.acm@gmail.com

Abstract. This paper describes the updates in EVOCROS, a cross-lingual ontology alignment system suited to create mappings between ontologies described in different natural language. Our tool combines syntactic and semantic similarity measures with information retrieval techniques. The semantic similarity is computed via NASARI vectors used together with *BabelNet*, which is a domain-neutral semantic network. In particular, we investigate the use of rank aggregation techniques in the cross-lingual ontology alignment task. The tool employs automatic translation to a pivot language to consider the similarity. EVOCROS was tested and obtained high quality alignment in the Multifarm dataset. We discuss the experimented configurations and the achieved results in OAEI 2019. This is our second participation in OAEI.

Keywords: cross-lingual matching · semantic matching · background knowledge · ranking aggregation

1 Presentation of the system

There is a growing number of ontologies described in different natural languages. The mappings among different ontologies are relevant for the integration of heterogeneous data sources to facilitate the exchange of information between systems. EVOCROS is our approach to automatic cross-lingual ontology matching. In our previous participation, in OAEI 2018, EVOCROS employed a weighted combination of similarity and semantic measures. The new version, submitted in OAEI 2019, combines syntactic and semantic similarity measures with information retrieval techniques. In this section, we describe the modifications to the system and the implemented techniques.

1.1 State, purpose, general statement

EVOCROS is a cross-lingual ontology alignment tool. The newest version of the tool leverages supervised methods of ranking aggregation techniques exploiting

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

labeled information (*i.e.*, training data) and ground-truth relevance to boost the effectiveness of a new ranker. Our goal is to leverage rank aggregation in cross-lingual mapping, by generating ranked lists based on distinct similarity measurements between the concepts of source and target ontologies.

1.2 Specific techniques used

The tool is developed in Python 3 and uses learning to rank techniques implemented in the well-known library *RankLib*. We model the mapping problem as an information retrieval query. Figure 1 depicts the workflow of the proposed technique. The inputs are source and target ontologies written in Web Ontology Language (OWL). These ontologies are converted to objects. The first step is the pre-processing of the source and target input ontologies, converting them into owlready2 objects. Each concept of the source ontology is compared to all concepts of the target ontology.

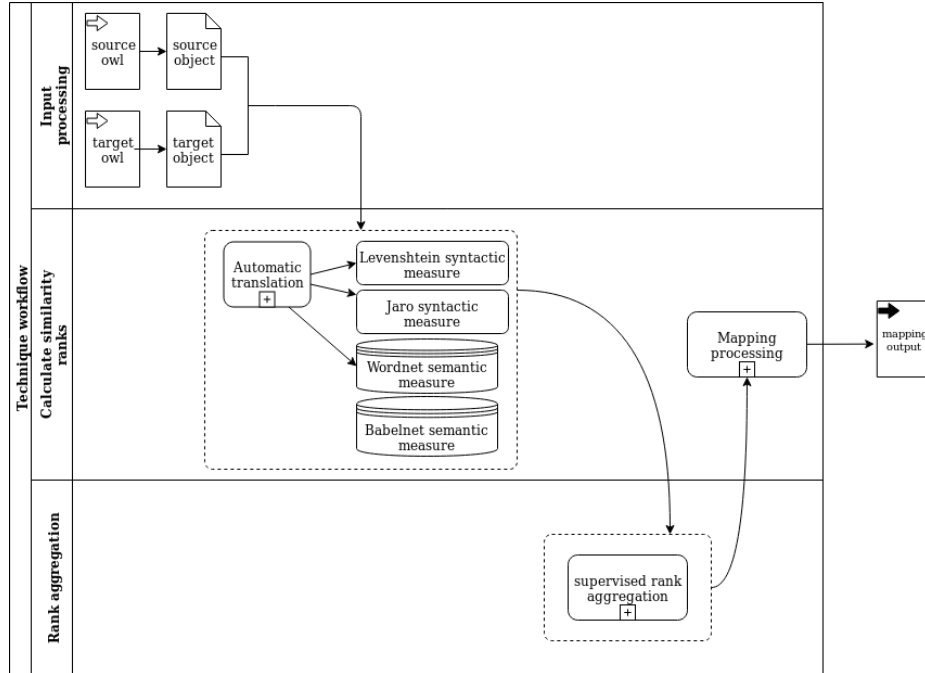


Fig. 1. General description of the technique. The mapping processing stage is where the top-1 entity of the final ranking is mapped to the input concept e_1 .

RankLib: <https://sourceforge.net/p/lemur/wiki/RankLib/> (As of November 16, 2019).

Python 3 library to manipulate ontologies as objects.

Each entity of the source ontology is compared with all entities of the same type found in the target ontology (*i.e.*, classes are matched to classes and properties are matched to properties). In this sense, for each entity e_i in the source ontology O_X , we calculate the similarity value with each entity e_j in the target ontology O_Y (Figure 2), thus generating a ranked list $\{rank1, rank2, rank3, rank4\}$ for each similarity measure used (*cf.* Figure 3).

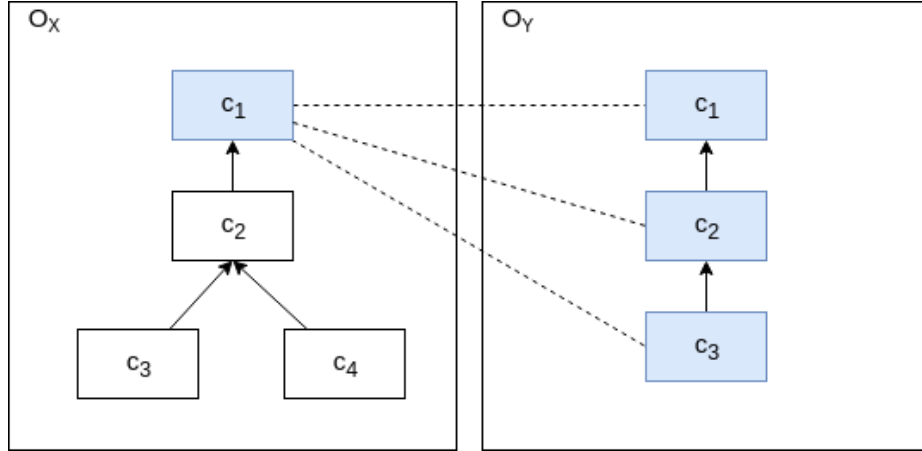


Fig. 2. Concept $c_1 \in O_X$ is compared against all concepts $c_n \in O_Y$.

For similarity measures that rely on monolingual comparison (*i.e.*, syntactic and WordNet), the automatic translation of labels of entities $e_i \in O_X$ and $e_j \in O_Y$ to a pivot language is used by leveraging Google Translate API during runtime. These similarity comparisons generate k ranks, each one based on a different similarity measure. We use the measures to generate the ranks, thus adding the flexibility to the use or the addition of different similarity measures without disrupting the technique.

The ranks are then aggregated using LambdaMART [7] because this technique has the best score among the majority of languages during the execution phase of OAEI 2019. Figure 4 presents that the set of multiple ranks are aggregated in a final rank. The Top-1 result of the aggregated rank $c_2 \in \mathcal{C}_{O_Y}$, is mapped to the source ontology entity $c_1 \in \mathcal{C}_{O_X}$, thus generating the candidate mapping $m(c_1, c_2)$ (*cf.* Figure 5). The mapping output follows the standard used by the Alignment API [?].

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

Alignment results are available at <https://github.com/jmdestro/evocros-results> (As of November 16, 2019).

Query	Rank1	Rank2	Rank3	Rank4
O_{x-c_1}	O_{Y-c_2}	O_{Y-c_2}	O_{Y-c_1}	O_{Y-c_2}
	O_{Y-c_1}	O_{Y-c_3}	O_{Y-c_2}	O_{Y-c_1}
	O_{Y-c_3}	O_{Y-c_1}	O_{Y-c_3}	O_{Y-c_3}

Fig. 3. Ranked lists generated by each similarity measure used.

2 Results

In this section, we describe the results obtained in the experiments conducted in OAEI 2019.

2.1 Multifarm

We consider the *MultiFarm* dataset [5], version released in 2015. Our experiments built cross-language ontology mappings by using English as a pivot language for Levenshtein [4], Jaro [3], and WordNet similarity measures. The semantic similarity relying on the *Babelnet* does not require a translation as it can retrieve the synsets used in NASARI vectors [1], by using the concepts original language. The application of each similarity measure in our technique generated a rank.

A subset of all languages was used for training and validation. The subsets are 10% of queries for training set, 15% queries for validation set, and 75% queries for testing. These subsets were generated per language and then combined, so the algorithms were trained, validated and tested using all languages at once. The comparable gold standard (*i.e.*, MultiFarm manually curated mappings) were adjusted to contain only the queries related to the testing subset. In this sense, a lower number of entities was considered in the tests, because we removed the set of queries used in training and validation from the reference mappings to ensure consistency.

Table 1 presents the obtained values for precision, recall, and f-measure for each language pair tested. The precision, recall, and f-measure scores have the same value due to the nature of the experiments. Our approach generates n : n mappings, where $n = |O_X| = |O_Y|$ because the ontologies are translations of each other to different natural languages, thus every entity in the source ontology presents a correspondence in the target ontology. In this sense, both the gold standard and the generated mappings have the same size because each query (*i.e.*, each entity in the source ontology) generates a mapping between the query (source entity) and the top-1 result of the final aggregated rank. Results

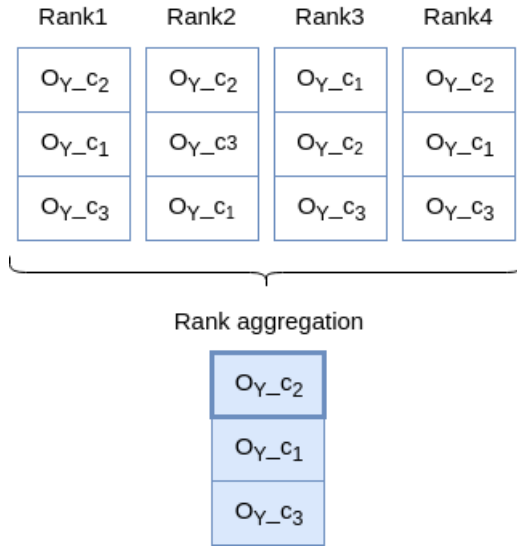


Fig. 4. Rank aggregation of the ranked lists. Each rank aggregation algorithm generates a distinct final rank.



Fig. 5. Mapping generated between source entity $c_1 \in O_X$ and top-1 entity of the final rank generated by the rank aggregation algorithm, $c_2 \in O_Y$.

show competitive results when compared to the other tools participating in the evaluation.

Table 1. Results achieved by different language pair

Language pair	Precision	Recall	F-measure
fr-nl	0.61290	0.61290	0.61290
en-pt	0.59140	0.59140	0.59140
es-nl	0.58065	0.58065	0.58065
cz-nl	0.52688	0.52688	0.52688
cn-pt	0.50538	0.50538	0.50538
es-ru	0.38710	0.38710	0.38710
cn-ru	0.32258	0.32258	0.32258
cz-ru	0.32258	0.32258	0.32258
de-ru	0.32258	0.32258	0.32258

3 General comments

In this section, we discuss our results and the ways to improve the system.

3.1 Comments on the results

The tool had satisfactory results, with competitive f-measure, but the execution time was exceedingly long due even with local caches for *Babelnet* NASARI vectors. This is due to the amount of comparisons required during execution because each concept or attribute in the source ontology is compared against all concepts and attributes of the target ontology.

3.2 Discussions on the way to improve the proposed system

This was the second evaluation of the system and results are encouraging. Our main goals for future work are: **Reduce execution time:** the tool has a long execution time even with local caches. Our future work will explore ontology partitioning during the pre-processing stage of the matching task to reduce the amount of comparisons needed, thus improving the execution time. **Bag of graphs:** ontologies can be represented as graphs, thus allowing for partitioning [2] and comparison of sub-graphs. Bag-of-graphs [6] is a graph matching approach, similar to bag-of-words. It represents graphs as feature vectors, highly simplifying the computation of graph similarity and reducing execution time. We propose as future investigation to use a simple vector-based representation for graphs and investigate it for cross-lingual ontology matching.

3.3 Comments on OAIE

Although we were not participating, our tool was executed on the Knowledge Graph track. There were issues during the evaluation phase, preventing the system to fully participate in both Multifarm and KG tracks.

4 Conclusion

The newest version of EVOCROS proposed an approach considering four similarity measures to build ranks and used a supervised method of rank aggregation. This is the second participation of the system in OAIE. The evaluation with the Multifarm dataset confirmed the quality of mappings generated by our technique. For future work, we plan to improve our cross-lingual alignment proposal by considering different combinations of similarity measures and different ways of computing the syntactic and semantic similarities taking into account additional stages in the pre-processing of the ontology.

Acknowledgements

This work was supported by São Paulo Research Foundation (FAPESP): grant #2017/02325-5.

References

1. Camacho-Collados, J., Pilehvar, M.T., Navigli, R.: Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities. *Artificial Intelligence* **240**, 36–64 (2016)
2. Hamdi, F., Safar, B., Reynaud, C., Zargayouna, H.: Alignment-based partitioning of large-scale ontologies. In: *Advances in knowledge discovery and management*, pp. 251–269. Springer (2010)
3. Jaro, M.A.: Advances in record-linkage methodology as applied to matching the 1985 census of tampa, florida. *Journal of the American Statistical Association* **84**(406), 414–420 (1989)
4. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* **10**, 707–710 (1966)
5. Meilicke, C., García-Castro, R., Freitas, F., Van Hage, W.R., Montiel-Ponsoda, E., De Azevedo, R.R., Stuckenschmidt, H., ŠVáb-Zamazal, O., Svátek, V., Tamin, A., et al.: Multifarm: A benchmark for multilingual ontology matching. *Web Semantics: Science, Services and Agents on the World Wide Web* **15**, 62–68 (2012)
6. Silva, F.B., de O. Werneck, R., Goldenstein, S., Tabbone, S., da S. Torres, R.: Graph-based bag-of-words for classification. *Pattern Recognition* **74**(Supplement C), 266 – 285 (Feb 2018). <https://doi.org/10.1016/j.patcog.2017.09.018>, <http://www.sciencedirect.com/science/article/pii/S0031320317303680>
7. Wu, Q., Burges, C.J., Svore, K.M., Gao, J.: Adapting boosting for information retrieval measures. *Information Retrieval* **13**(3), 254–270 (Jun 2010). <https://doi.org/10.1007/s10791-009-9112-1>