XMap : Results for OAEI 2017

Warith Eddine DJEDDI[a,b], Mohamed Tarek KHADIR[a] and Sadok BEN YAHIA[b]

[a]LabGED, Computer Science Department, University Badji Mokhtar, Annaba, Algeria
[b]Faculty of Sciences of Tunis, University of Tunis El-Manar, LIPAH-LR 11ES14, 2092, Tunisia
{djeddi,khadir}@labged.net
sadok.benyahia@fst.rnu.tn

Abstract. We describe in this paper the XMap system and the results achieved during the 2017 edition of the Ontology Alignment Evaluation Initiative. XMap aims to tackle the issue of matching large scale ontologies by involving particular parallel matching on multiple cores or machines.

1 Presentation of the system

XMap, as for eXtended Mapping, is one of the leading ontology matching systems for large-scale ontology matching relying on the notion of context in order to deal with lexical ambiguity as well as a divide-and-conquer approach to tackle the issue of matching large ontologies.

In XMap, the measurement of lexical similarity in ontology matching is performed using a synset, defined in WordNet [1] and UMLS [2]. In our approach, the similarity between two entities of different ontologies is evaluated not only by investigating the semantics of the entities names, but also taking into account the context, through which the effective meaning is described. The translation into many languages is based on the Microsoft ®Translator. Our system stores locally all translation results from Microsoft ®Translator in dictionary files. The translator will also be queried only when no stored translation are found in order to gain time and avoid overloading the server.

2 State, purpose, general statement

XMap using an oracle by modifying the validation process of the candidate mappings according to the quality of the interactive matching in terms of F-measure and number of required interactions. This process is performed after each round of candidate retrieving. Our approach is based on semantic techniques and on a parallel execution strategy adapted from [3], to address the challenge of scalability and efficiency of matching techniques. One of the main trusts of the introduced approach is the increasing scalability and speed of ontology alignment by matching linguistic and structural features.

At a glance, the mapping process of XMap is depicted in Figure 1. XMap uses various similarity measures of different categories such as string, linguistic, and structural based similarity measures, each contributing to some extent to the alignment results. Afterwards, the alignments from all matchers can be aggregated to obtain a final alignment through the use of sequential composition [4]. Finally, a fast repair method is applied
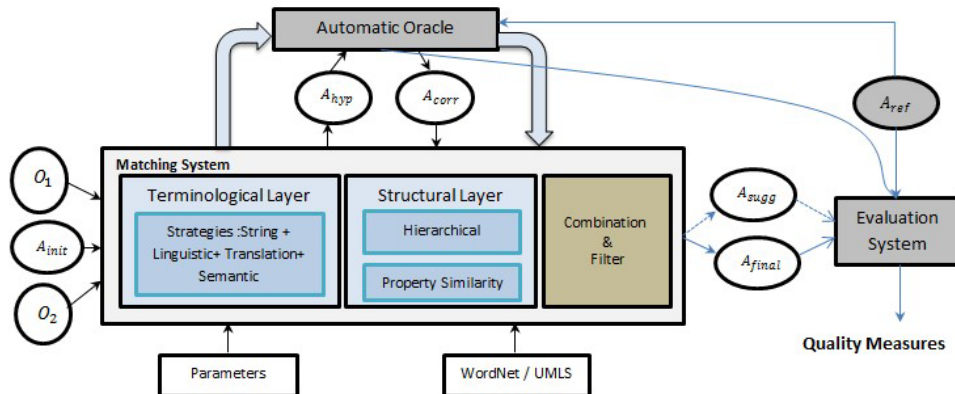
Fig. 1. The different steps for scoring a multiple network alignment.

so as to detect and remove the inconsistent classes by "'Applying Logical Constraints on Matching Ontologies'" (ALCOMO) [5]. The main goal is to try to remove less unsatisfiable classes (discovering disjointness relationships) without having an impact on the F-measure score.

3   Results

In this section, we present the evaluation results obtained by running XMap under the SEALS client with *Anatomy*, *Conference*, *Multifarm*, *Interactive matching evaluation*, *Large Biomedical Ontologies* and *Disease and Phenotype* tracks.

Anatomy   The Anatomy track consists of finding an alignment between the Adult Mouse Anatomy (2744 classes) and a part of the NCI Thesaurus (3304 classes) describing the human anatomy. XMap achieves a good F-Measure value of ≈89% in a reasonable amount of time (37 sec.) (see Table 1). In terms of F-Measure/runtime, XMap is ranked 2nd among the tools participated in this track.

Table 1. Results for Anatomy track.

System	Precision	F-Measure	Recall	Time(s)
XMap	0.926	0.893	0.863	37

Conference   The Conference track uses a collection of 16 ontologies from the domain of academic conferences. Most ontologies were equipped with OWL-DL axioms of various types; this opens a useful way to test our semantic matchers. For each reference

alignment, three evaluation modalities are applied : a) crisp reference alignments, b) the uncertain version of the reference alignment, c) logical reasoning.

Table 2. Results based on the crisp reference alignments.

	Precision	F-Measure 1	Recall
Original reference alignment (ra1)			
ra1-M1	0.84	0.73	0.64
ra1-M2	0.75	0.32	0.2
ra1-M3	0.84	0.68	0.57
Entailed reference alignment (ra2)			
ra2-M1	0.79	0.67	0.58
ra2-M2	0.83	0.35	0.22
ra2-M3	0.79	0.63	0.52
Violation reference alignment (rar2)			
rar2-M1	0.78	0.68	0.6
rar2-M2	0.83	0.35	0.22
rar2-M3	0.78	0.65	0.55
Uncertain reference alignments (Sharp)			
-	0.84	0.57	0.6

Table 3. Results based on the uncertain version of the reference alignment.

Precision	F-Measure 1	Recall
Uncertain reference alignments (Sharp)		
0.84	0.68	0.57
Uncertain reference alignments (Discrete)		
0.79	0.72	0.67
Uncertain reference alignments (Continuous)		
0.81	0.73	0.67

As depicted in Table 2 and 3, XMap produces fairly consistent alignments when matching the conference ontologies. Finally, XMap generated only one incoherent alignment for the evaluation based on logical reasoning.

Multifarm This track is based on the translation of the OntoFarm collection of ontologies into 9 different languages. XMap have low performance due to many internal exceptions. The results are showed in Table 4.

Interactive matching evaluation The goal of this evaluation is to imitate interactive alignment [6, 7], where a oracle user is involved to validate the correspondences found

Table 4. Results for Multifarm track.

System	Different ontologies			Same ontologies		
	P	F	R	P	F	R
XMap	0.24	0.06	0.04	0.66	0.10	0.06

by the alignment approach by checking the reference alignment, and changing error values in order to assess their influence on the performance of alignment systems. For the 2017 edition, participating systems are evaluated on the Conference, Anatomy, Large biomedical and Phenotype datasets using an oracle based on the reference alignment.

XMap uses various similarity measures to generate candidate mappings. It applies two thresholds to filter the candidate mappings: one for the mappings that are directly added to the final alignment and another for those that are presented to the user for validation. The latter threshold is selected to be high in order to minimize the number of requests and the rejected candidate mappings from the oracle; the requests are mainly about incorrect mappings. The mappings accepted by the user are moved to the final alignment. For the two years 2016 and 2017, XMap preserved roughly the same F-Measure value, and it benefits the least from the interaction with the oracle. All XMap's measures differ with less than 0.2% from the non-interactive runs, and performance does not change at all with the increasing error rates.

Large biomedical ontologies This track consists of finding alignments between the Foundational Model of Anatomy (FMA), SNOMED CT, and the National Cancer Institute Thesaurus (NCI). The results obtained by XMap are depicted by Table 5.

Table 5. Results for the Large BioMedical track.

Test set	Precision	Recall	F-Measure	Time(s)
Small FMA-NCI	0.977	0.901	0.937	20
Whole FMA-NCI	0.884	0.847	0.865	130
Small FMA-SNOMED	0.974	0.847	0.906	62
Whole FMA- Large SNOMED	0.774	0.843	0.807	625
Small SNOMED-NCI	0.894	0.566	0.693	106
Whole SNOMED-NCI	0.819	0.553	0.660	563

In general, we can conclude that XMap achieved a good precision/recall values. The high recall value can be explained by the fact that UMLS thesaurus contains definitions of highly technical medical terms.

Disease and Phenotype This track based on a real use case where it is required to find alignments between disease and phenotype ontologies. Specifically, the selected ontologies are the Human Phenotype Ontology (HPO), the Mammalian Phenotype Ontology (MP), the Human Disease Ontology (DOID), and the Orphanet and Rare Diseases Ontology (ORDO).

XMap achieved fair results according to the three evaluation (Silver standard, Manually generated mappings and Manual assessment of unique mappings).

4 General comments

4.1 Comments on the results

This is the 5th time that we participate in the OAEI campaign. The official results of OAEI 2017 show that XMap is competitive with other well-known ontology matching systems in all OAEI tracks.

4.2 Comments on the OAEI 2017 procedure

As a fifth participation, we found the OAEI procedure very convenient and the organizers very supportive. The OAEI test cases are various, and this leads to a comparison on different levels of difficulty, which is very interesting. We found that SEALS platform is a precious tool to compare the performance of our system with the others.

5 Conclusion

In this paper, we presented the results achieved during the 2017 edition of the OAEI campaign. The used benchmark helped greatly identify the power and weaknesses of the algorithm. In addition, XMap showed the feasibility of our approach especially on large-scale biomedical ontologies which was a thriving challenge in ontology matching domain.

References

1. Christiane D. Fellbaum. *WordNet – An Electronic Lexical Database*. MIT Press, 1998.
2. Olivier Bodenreider. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32(Database-Issue):267–270, 2004.
3. Anika Gross, Michael Hartung, Toralf Kirsten, and Erhard Rahm. On matching large life science ontologies in parallel. In *Data Integration in the Life Sciences*, pages 35–49. Springer, 2010.
4. Jérôme Euzenat and Pavel Shvaiko. *Ontology Matching*. Springer-Verlag New York, Inc., Berlin, Heidelberg, 2007.
5. Christian Meilicke. *Alignment incoherence in ontology matching*. PhD thesis, University of Mannheim, 2011.
6. Heiko Paulheim, Sven Hertling, and Dominique Ritzei. Towards evaluating interactive ontology matching tools. In *The Semantic Web: Semantics and Big Data, 10th International Conference, ESWC 2013, Montpellier, France, May 26-30, 2013. Proceedings*, pages 31–45, 2013.
7. Zlatan Dragisic, Valentina Ivanova, Patrick Lambrix, Daniel Faria, Ernesto Jimenez-Ruiz, and Catia Pesquita. User validation in ontology alignment. In *Proceedings of the International Semantic Web Conference*, volume 9981 of *LNCS*, October 2016.