

LYAM++ Results for OAEI 2016

Abdel Nasser Tigrine, Zohra Bellahsene, Konstantin Todorov

{lastname}@lirmm.fr
LIRMM / University of Montpellier, France

Abstract. LYAM++ is a fully automatic ontology matching system based on the use of external sources. Our approach applies a novel orchestration of the components of the matching workflow. We present our results on anatomy, conference large biomedical and Multifarm tracks of OAEI2016.

1 Presentation of the System

In spite of the considerable advance that has been made in the field of ontology matching recently, many questions remain open [1]. The current work addresses the challenge of using background knowledge with a focus on aligning cross-lingual ontologies, i.e., ontologies defined in different natural languages [2].

Indeed, considering multilingual and cross-lingual information is becoming more and more important, in view particularly of the growing number of web content-creating non-English users and the clear demand of cross-language interoperability. In the context of the web of data, it is important to propose procedures for linking vocabularies across natural languages, in order to foster the creation of a veritable global information network.

The use of different natural languages in the concepts and relations labeling process is becoming an important source of ontology heterogeneity. The methods that have been proposed to deal with it most commonly rely on automatic translation of labels to a single target language [3,4] or apply machine learning techniques [2]. However, machine translation tolerates low precision levels and machine learning methods require large training corpus that is rarely available in an ontology matching scenario. An inherent problem of translation is that there is often a lack of exact one-to-one correspondence between the terms in different natural languages.

1.1 State, Purpose, General Statement

We present LYAM++ (Yet Another Matcher - Light)[5], a fully automatic ontology matching system based on the use of external sources. LYAM++ does not rely on machine translation for cross-lingual ontology matching. Instead, we make use of the openly available general-purpose multilingual semantic network BabelNet¹ in order to recreate the missing semantic context in the matching

¹ <http://babelnet.org/>

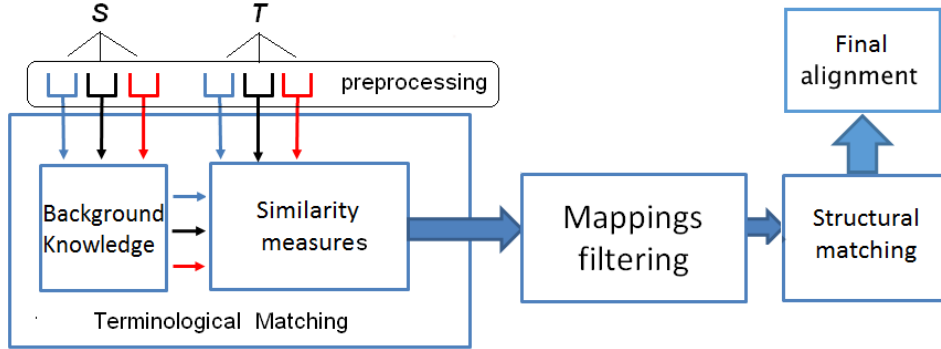


Fig. 1: The processing pipeline of LYAM++.

process. Another original feature of our approach is the choice of orchestration of the matching workflow. The novel workflow orchestration provides better results compared to the classical one. We refer the reader to the results reported in [5].

1.2 Specific Techniques Used

The workflow of LYAM++ is given in Fig 1. The overall process consists of four main components: a terminological matcher, a mapping selection module and, finally, a structural matcher. One of the original contributions of this work is the choice of orchestration of these components. Indeed, the places of the mapping selection module and the structural matcher are reversed in the existing OM tools [6]. However, we wanted to ensure that we feed only good quality mappings to the structural matcher, therefore we decided to filter the discovered correspondences right after producing the initial alignment. This decision is supported experimentally in [5].

The *terminological matching* module, the second contribution described in this paper, acts on the one hand as a preprocessing component and, on the other hand – as a light-weight terminological matcher between labels. We start by splitting the elements of each ontology in three groups: labels of classes, labels of object properties and labels of data object properties (in colors blue, black and red in the figure), since these groups of elements are to be aligned separately. A standard preprocessing procedure is applied on these sets of labels, comprising character normalization, stop-words filtering, tokenization and lemmatization.

For the cross-lingual ontology matching, at first every token of a given label s in the source ontology S is enriched by related terms and synonyms from BabelNet and all of these terms are represented in the language l_T (language of the target ontology), which makes these terms comparable to the tokens of the labels in the target ontology T . A simple similarity evaluation by the help of the Jaccard coefficient selects the term in each set of related terms corresponding to a given token from s that has the highest score with respect to every token in each label of T . This helps to reconstitute the label s in the language l_T . Finally,

the labels in each group of S and T , seen as sets of tokens, are compared by using the Soft TFIDF similarity measure [7], which produces an intermediate terminological alignment. For monolingual ontology matching, the system uses the relations such as "hasSynonyms" present in a given BK to match between two concepts.

The three remaining components are standard OM modules [6], although ordered in a new manner. The *Mapping selection* is a module that transforms the initial 1 to many mapping to a 1:1 mapping based on the principle of iteratively retaining the pairs of concepts with maximal value of similarity. Finally, the *structural matcher* component filters the trustworthy pairs of aligned concepts by looking at the similarity values produced for their parents and their children in the ontology hierarchies.

1.3 Adaptations made for the evaluation

The adaptation made for the evaluation is in the preprocessing step. LYAM++ uses (1) Uberon [8] for anatomy and BioMed tracks, (2) BabelNet [9] for conference and multifarm tracks.

1.4 Links to the System and to the Set of Provided Alignments

Last year, the system was not available online because it depends heavily on the use of BabelNet 3.0 version, which is under a non-free licence. In this year, we used old version of BabelNet 2.0 which is under free license.

The alignments produced by LYAM++ for this year's can be found under the following link: <http://www.lirmm.fr/benellefi/Alignements.rar>. LYAM++ can be found under the following link: <http://www.lirmm.fr/benellefi/Lyam++.rar>

2 Results

We have evaluated our approach on data coming from the ontology alignment evaluation initiative (OAEI)² and particularly anatomy, conference, large biomedical and multifarm.

Anatomy This track aims to discovering alignments between a human anatomy ontology, part of the NCI Thesaurus³ and a mouse anatomy ontology. This track is considered as a large-scale matching task because the input ontologies are of a large size and very rich semantically. Table 1 presents the results obtained by LYAM++ on this year's

Conference This track contains 16 ontologies from the scientific publication field. Table 2 presents the results obtained by LYAM++ on this year's

² <http://oaei.ontologymatching.org/>

³ <https://ncit.nci.nih.gov/ncitbrowser/>

Table 1: Results of LYAM++ for anatomy .

	F-M	Recall	Precision
LYAM++	0.87	0.88	0.86

Table 2: Results of LYAM++ for conference .

	F-M	Recall	Precision
ra1-M1	0.36	0.18	0.48
ra1-M2	0.34	0.57	0.13
ra1-M3	0.29	0.15	0.38
ra2-M1	0.36	0.19	0.52
ra2-M2	0.35	0.59	0.13
ra2-M3	0.31	0.16	0.41

Large biomedical ontologies This track aims at aligning three large biomedical ontologies, namely FMA, SNOMED and the NCI Thesaurus. Table 3 presents the results obtained by LYAM++ on this year's

Table 3: Results of LYAM++ for BioMed.

	F-M	Recall	Precision
Small FMA-NCI	0.79	0.88	0.72

MultiFarm is a benchmark designed for evaluating cross-lingual ontology matching systems. Multifarm data consist of a set of 7 ontologies originally coming from the *Conference* benchmark of OAEI, translated into 8 languages. Two evaluation tasks are defined: *task 1* consists in matching two different ontologies given in different languages, while *task 2* aims to align different language versions of one single ontology.

Table 4 presents the results obtained by LYAM++ on this year's Multi-farm evaluation campaign. What we see is the average F-measure value for all language-pairs without any threshold on the confidence measure. The value in the parenthesis corresponds to the average F-measure value for the generated alignments only (the pairs of languages that the system handles).

3 Conclusion

In this paper, we present the over view of the LYAM++ system and our results on the OAEI2016 tracks . In this year, our goal was to participate on monolingual ontology matching scenarios. We used Babelnet 2.0 version instead of Babelnet 3.0 version due to the licenses problems. Subjects of ongoing and

Table 4: Results of LYAM++ for Multifarm.

	Task1	Task2
LYAM++	0.01	0.02

future work are (1) testing and evaluating different sources of external knowledge, (2) applying semantic mappings selection methods to improve the results, (3) adaptation of the approach to the large scale ontology matching scenarios.

References

1. P. Shvaiko and J. Euzenat, “Ontology matching: state of the art and future challenges,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, no. 1, pp. 158–176, 2013.
2. D. Spohr, L. Hollink, and P. Cimiano, “A machine learning approach to multilingual and cross-lingual ontology matching,” in *The Semantic Web–ISWC 2011*, pp. 665–680, Springer, 2011.
3. D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, and F. M. Couto, “The agreementmakerlight ontology matching system,” in *On the Move to Meaningful Internet Systems: OTM 2013 Conferences*, pp. 527–541, Springer, 2013.
4. D. Ngo and Z. Bellahsene, “YAM++ : A multi-strategy based approach for ontology matching task,” in *Knowledge Engineering and Knowledge Management - 18th International Conference, EKAW 2012, Galway City, Ireland, October 8-12, 2012. Proceedings*, pp. 421–425, 2012.
5. A. N. Tigrine, Z. Bellahsene, and K. Todorov, “Light-weight cross-lingual ontology matching with LYAM++,” in *On the Move to Meaningful Internet Systems: OTM 2015 Conferences - Confederated International Conferences: CoopIS, ODBASE, and C&TC 2015, Rhodes, Greece, October 26-30, 2015, Proceedings*, pp. 527–544, 2015.
6. D. Ngo, Z. Bellahsene, and K. Todorov, “Opening the black box of ontology matching,” in *The Semantic Web: Semantics and Big Data*, pp. 16–30, Springer, 2013.
7. W. W. Cohen, P. D. Ravikumar, and S. E. Fienberg, “A comparison of string distance metrics for name-matching tasks,” in *IJWeb*, pp. 73–78, 2003.
8. M. Haendel, J. P. Balhoff, F. B. Bastian, D. C. Blackburn, J. A. Blake, Y. Bradford, A. Comte, W. M. Dahdul, T. Dececchi, R. E. Druzinsky, T. F. Hayamizu, N. Ibrahim, S. E. Lewis, P. M. Mabee, A. Niknejad, M. Robinson-Rechavi, P. C. Sereno, and C. J. Mungall, “Unification of multi-species vertebrate anatomy ontologies for comparative biology in uberon,” *J. Biomedical Semantics*, vol. 5, p. 21, 2014.
9. R. Navigli and S. P. Ponzetto, “Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network,” *Artif. Intell.*, vol. 193, pp. 217–250, 2012.