ALIN Results for OAEI 2020

Jomar da Silva[1], Carla Delgado[1], Kate Revoredo[2], and Fernanda Araujo Baião[3]

[1] Graduate Program in Informatics, Federal University of Rio de Janeiro (UFRJ), Brazil
[2] Vienna University of Economics and Business, Vienna, Austria
[3] Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil
jomar.silva@uniriotec.br,carla@ppgi.ufrj.br,
kate.revoredo@wu.ac.at,fbaiao@puc-rio.br

Abstract. [1] ALIN is an ontology matching system specialized in the interactive ontology matching. At OAEI 2020, the new version of ALIN uses natural language processing techniques to improve the standardization of the terms of the ontologies to be aligned. The purpose of using these techniques is to improve the quality of the generated alignment. This paper describes the ALIN participation in the OAEI 2020 and discusses its results.

Keywords: ontology matching, Wordnet, interactive ontology matching, ontology alignment, interactive ontology alignment, natural language processing

1 Presentation of the System

Due to the advances in information and communication technologies, a large amount of data repositories became available. Those repositories, however, are highly semantically heterogeneous, which hinders their integration. Ontology matching has been successfully applied to solve this problem, by discovering mappings between two distinct ontologies which, in turn, conceptually define the data stored in each repository. Ontology matching process seeks to discover correspondences (mappings) between entities of different ontologies [1]. The ontology matching can be done in manual, semi-automatic or automatic way [1]. Among the semi-automatic approaches, the ones that follow an interactive strategy stand out, considering the knowledge of domain experts through their participation [2]. The use of a domain expert is not always possible, as it is an expensive, scarce and time-consuming resource. But when it is possible to use it, this strategy has achieved results superior to automatic (non-interactive) strategies, but there is still room for better results [2], as can be seen in the evaluation of interactive tools in the OAEI[4] (Ontology Alignment Evaluation Initiative). ALIN[3] is a system for interactive ontology matching which has been participating in all OAEI editions since 2016, with improving results.

1.1 State, Purpose and General Statement

The interactive ontology matching systems has generally relied on the domain expert to achieve better results than those of the non-interactive ones. In this version, ALIN will use another professional, one with knowledge of some NLP techniques, to improve the generated alignment. In its ontology matching process, ALIN will use this professional to construct regular (in reality its equivalent regular expressions) and context free grammars and their respective lexical analyzers (scanners) and syntax analyzers (parsers) for the terms of the involved ontologies before the execution of ALIN. The aim is to standardize the terms of the ontologies. The standardization of terms has shown good results in the quality of the generated alignment [4].

1.2 Specific Techniques Used

ALIN handles three sets of mappings: (i) Accepted, which is a set of mappings definitely to be retained in the alignment; (ii) Selected, which is a set of mappings where each is yet to be decided if it will be included in the alignment; and (iii) Suspended, which is a set of mappings that have been previously selected, but (temporarily or permanently) filtered out of the alignment.

Given the previous definitions, ALIN procedure follows 5 Steps, described as follows:

1. Select mappings: select the first mappings and automatically accepts some of them. We explain the selection and acceptance process below;
2. Filter mappings: suspend some selected mappings, using lexical criteria for that;
3. Ask expert: accepts or rejects selected mappings, according to expert feedback
4. Propagate: select new mappings, reject some selected mappings or unsuspend some suspended mappings (depending on newly accepted mappings)
5. Go back to 3 as long as there are undecided selected mappings

All versions of ALIN (since its very first OAEI participation) follow this general procedure. In this 2020 version, ALIN uses a new step where a professional who knows construct scanners and parsers develops grammars to the terms of the ontologies. ALIN uses these grammars to standardize the terms of ontologies and thus improve the ontology matching process. The new step can lead to, for example, correction of spelling errors, unifying different spellings for the same grammatical situation, etc. You can see more detailed examples of possible

[4] Available at http://oaei.ontologymatching.org/2020/results/interactive/index.html, last accessed on Oct, 23, 2020.

standardization of terms in [4]. Once the professional develops the scanner and the parser ALIN uses them in step 1 of the program.

ALIN applies the following techniques:

- Line 1. ALIN selects mappings using linguistic similarities between entity names, after running the scanner and the parser for each term of the ontologies. ALIN uses synonyms and variations in entity name words to automatically accept mappings. At this time, ALIN automatically selects and accepts only concept mappings. ALIN uses the Wordnet and domain-specific ontologies (the FMA Ontology in the Anatomy track) to find synonyms between entities.
- Line 2. ALIN suspends the selected mappings whose entities have low lexical similarity. We use the Jaccard, Jaro-Wrinkler, and n-gram lexical metrics to calculate the lexical similarity of the selected mappings. We based the process of choosing the similarity metrics used by ALIN on the result of these metrics in assessments [5]. It is important to know that these suspended mappings can be unsuspended later, by structural analysis, as proposed in [6].
- Line 3. At this point, the expert interaction begins. ALIN sorts the selected mappings in a descending order according to the sum of similarity metric values. The sorted selected mappings are submitted to the expert.
- Line 4. Initially, the set of selected mappings contains only concept mappings. At each interaction with the expert, if the expert accepts the mapping, ALIN (i) removes from the set of selected mappings all the mappings that compose the mapping anti-pattern [7][8] (we explain mapping anti-pattern below) with the accepted mappings; (ii) selects data property (like [9]) and object property mappings related to the accepted concept mappings; (iii) unsuspends all concept mappings whose both entities are subconcepts of the concept of an accepted mapping, following a similar technique proposed in our previous work [6].
- Line 5. The interaction phase continues until there are no selected mappings.

An ontology may have construction constraints, such as a concept cannot be equivalent to its superconcept. An alignment may have other constraints like, for example, an entity of ontology O cannot be equivalent to two entities of the ontology O'. A mapping anti-pattern is a combination of mappings that generates a problematic alignment, i.e., a logical inconsistency or a violated constraint.

1.3 Link to the System and Parameters File

In this version, I developed only scanners and parsers for the conference track and anatomy track ontologies. ALIN will work only with these ontologies.

ALIN is available [2] as a package to be run through the SEALS client.

[2] $https : //drive.google.com/file/d/1ZM3g_0aOgUha - 9Vp_tUbqk9nmnkFCl7L/view?usp = sharing$

2 Results

Interactive ontology matching is the focus of the ALIN system. If you compare the participation of ALIN in 2020 and 2019 (Table 5), you will see an improvement in the quality of the generated alignment, showing the effectiveness of the techniques used.

2.1 Comments on the Participation of the ALIN in Non-Interactive Tracks

ALIN used NLP techniques to improve the standardization of ontology terms. The use of NLP techniques led to an increase in the F-Measure of non-interactively generated alignments (Table 1 and Table 2). Conference track, unlike the Anatomy track, has relationship mappings and attribute mappings that ALIN does not automatically accept, thus making the F-Measure on the Conference track, although higher than last year, still low.

Table 1. Participation of ALIN in Anatomy Non-Interactive Track - /2019[10]/2020[11]

Year	Precision	Recall	F-measure
2019	0.974	0.698	0.813
2020	0.986	0.72	0.832

Table 2. Participation of ALIN in Conference Non-Interactive Track - 2019[10]/2020[12]

Year	Precision	Recall	F-measure
2019	0.87	0.44	0.58
2020	0.87	0.46	0.60

2.2 Comments on the Participation of the ALIN in Interactive Tracks

In the Anatomy track, ALIN was better than LogMap in both quality (F-Measure) and total requests, but worse in both aspects than AML (Table 3). In the Conference track, ALIN was first in quality and third in total requests (Table 4).

Table 3. Participation of ALIN in Anatomy Interactive Track - Error Rate 0.0[13]

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.988	0.856	0.917	360
AML	0.972	0.933	0.952	189
LogMap	0.988	0.846	0.912	388

Table 4. Participation of ALIN in Conference Interactive Track - Error Rate 0.0[13]

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.915	0.705	0.796	233
AML	0.91	0.698	0.79	221
LogMap	0.886	0.61	0.723	82

Interactive Anatomy Track In this track, Alin had a decrease in the number of expert interactions and an increase in the quality of the generated alignment, showing that the use of the NLP techniques are effective for this track (Table 5).

Interactive Conference Track In this track, Alin had an increase in the quality of the generated alignment but an increase in the number of expert interactions. (Table 7).

2.3 Comparison of the Participation of ALIN in OAEI 2020 with his Participation in OAEI 2019

In this version, Alin uses NLP techniques to improve the standardization of ontology terms. This use proved to be effective to increase the quality of the generated alignments.

The quality of the alignment generated by Alin is dependent on the correct expert feedback, as Alin uses expert responses to select new mappings. When Alin selects wrong mappings, the quality of the generated alignment tends to decrease. If we compare this year's quality decline with last year's, we see that this fall is more sharp (Table 6 and Table 8).

The run time of Alin this year was shorter than last year (Table 9 and Table 10). In my machine, Alin has run 20% faster this year than last year. The execution in OAEI had a bigger reduction in the run time, but other systems also had a reduction, so I don't know if all the difference in the run time in OAEI is due to the modifications made in Alin or to changes in the configuration of the machine they use.

Table 5. Participation of ALIN in Anatomy Interactive Track - OAEI 2016[14]/2017[15]/2018[16]/2019[10]/2020[13] - Error Rate 0.0

Year	Precision	Recall	F-measure	Total Requests
2016	0.993	0.749	0.854	803
2017	0.993	0.794	0.882	939
2018	0.994	0.826	0.902	602
2019	0.979	0.85	0.91	365
2020	0.988	0.856	0.917	360

Table 6. F-Measure of ALIN in Anatomy Interactive Track - OAEI /2019[10]/2020[13]

Year	Error rate 0.0	Error rate 0.1
2019	0.91	0.889
2020	0.917	0.887

Table 7. Participation of ALIN in Conference Interactive Track - OAEI 2016[14]/2017[15]/2018[16]/2019[10]/2020[13] - Error Rate 0.0

Year	Precision	Recall	F-measure	Total Requests
2016	0.957	0.735	0.831	326
2017	0.957	0.731	0.829	329
2018	0.921	0.721	0.809	276
2019	0.914	0.695	0.79	228
2020	0.915	0.705	0.796	233

Table 8. F-Measure of ALIN in Conference Interactive Track - OAEI /2019[10]/2020[13] - with Different Error Rates

Year	Error rate 0.0	Error rate 0.1
2019	0.79	0.725
2020	0.796	0.713

Table 9. Run Time (sec) in Anatomy Interactive Track - OAEI /2019[10]/2020[13]

Tool	2019	2020
ALIN	2132	1152
AML	82	37,3
LogMap	29	7,6

Table 10. Run Time (sec) in Conference interactive track - OAEI /2019[10]/2020[13]

Tool	2019	2029
ALIN	397	136,9
AML	34	30.1
LogMap	37	37.96

3 General Comments

Evaluating the results, we can see that the system has improved, although it can improve even further, towards:

- handling user error rate;
- generating a higher quality initial alignment in its non-interactive phase;
- reducing the number of interactions with the expert;

Another consideration is that this version of ALIN generates the need for a new professional involved in the process, an expert in writing grammars. This professional may not always be available, but if he is, the results have shown that his work can improve the quality of the generated alignment.

3.1 Conclusions

ALIN used NLP techniques to improve the standardization of the ontology terms. They have been effective in increase the quality in the generated alignment with a small change in the total requests. ALIN had a decrease in run time but a more sharp fall in the alignment quality when the domain expert makes mistakes. A problem that arose with the inclusion of NLP techniques is the need to generate the scanner and the parser for the ontologies involved in the matching and a professional to develop them may not always be available.

References

1. Euzenat, J., Shvaiko, P.: Ontology Matching - Second Edition. Springer-Verlag (2013)
2. Li, H., Dragisic, Z., Faria, D., Ivanova, V., Jimenez-Ruiz, E., Lambrix, P., Pesquita, C.: User validation in ontology alignment: functional assessment and impact. The Knowledge Engineering Review (2019)
3. Da Silva, J., Revoredo, K., Baião, F., Euzenat, J.: Alin: improving interactive ontology matching by interactively revising mapping suggestions. The Knowledge Engineering Review **35** (2020)
4. Real, F.J.Q., Bella, G., McNeill, F., Bundy, A.: Using domain lexicon and grammar for ontology matching. (2020)
5. Cheatham, M., Hitzler, P.: String similarity metrics for ontology alignment. In: Proceedings of the 12th International Semantic Web Conference - Part II. ISWC '13, New York, NY, USA, Springer-Verlag New York, Inc. (2013) 294–309

6. Silva, J., Baião, F., Revoredo, K., Euzenat, J.: Semantic interactive ontology matching: Synergistic combination of techniques to improve the set of candidate correspondences. In: OM-2017: Proceedings of the Twelfth International Workshop on Ontology Matching. Volume 2032. (2017) 13–24
7. Guedes, A., Baião, F., Shivaprabhu, Revoredo, R.: On the Identification and Representation of Ontology Correspondence Antipatterns. In: Proc. 5th Int. Conf. Ontol. Semant. Web Patterns (WOP'14), CEUR Work. Proc. (2014)
8. Guedes, A., Baião, F., Revoredo, K.: Digging Ontology Correspondence Antipatterns. In: Proceeding WOP'14 Proc. 5th Int. Conf. Ontol. Semant. Web Patterns. Volume 1032. (2014) 38—-48
9. Silva, J., Revoredo, K., Baião, F.A., Euzenat, J.: Interactive Ontology Matching: Using Expert Feedback to Select Attribute Mappings. In: CEUR Workshop Proceedings. Volume 2288. (2018) 25–36
10. Silva, J., Delgado, C., Revoredo, K., Baião, F.: Alin results for oaei 2019. In: Proceedings of the 14th International Workshop on Ontology Matching. OM'19 (2019) 94–100
11. : Results for oaei 2020 - anatomy track. http://oaei.ontologymatching.org/2020/results/anatomy/ Accessed: 2020-10-23.
12. : Results of evaluation for the conference track within oaei 2020. http://oaei.ontologymatching.org/2020/results/conference/index.html Accessed: 2020-10-23.
13. : Results for oaei 2020 - interactive track. http://oaei.ontologymatching.org/2020/results/interactive/index.html Accessed: 2020-10-23.
14. Silva, J., Baião, F., Revoredo, K.: Alin results for oaei 2016. In: OM-2016: Proceedings of the Eleventh International Workshop on Ontology Matching. OM'16 (2016) 130–137
15. Silva, J., Baião, F., Revoredo, K.: Alin results for oaei 2017. In: OM-2017: Proceedings of the Twelfth International Workshop on Ontology Matching. OM'17 (2017) 114–121
16. Silva, J., Baião, F., Revoredo, K.: Alin results for oaei 2018. In: Ontology Matching: OM-2018: Proceedings of the ISWC Workshop. OM'18 (2018) 117–124