ALIN Results for OAEI 2019

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

[1] Graduate Program in Informatics, Federal University of Rio de Janeiro (UFRJ), Brazil
[2] Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil
jomar.silva@uniriotec.br,carla@ppgi.ufrj.br,katerevoredo@ppgi.ufrj.br, fbaiao@puc-rio.br

Abstract. [1] ALIN is an ontology matching system specialized in the interactive ontology matching, and its main characteristic is the use of expert feedback to improve the set of selected mappings, using semantic and structural techniques to make this improvement. This paper describes its configuration for the OAEI 2019 competition and discusses its results.

Keywords: ontology matching, Wordnet, interactive ontology matching, ontology alignment, interactive ontology alignment

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. Among the various ontology matching approaches that exist in the literature, interactive ontology matching includes the participation of domain experts to improve the quality of the final alignment [1]. ALIN is an interactive ontology matching system and has been participating in all OAEI editions since 2016, with improving results.

1.1 State, Purpose and General Statement

ALIN is a system for interactive ontology matching that consists of two steps: one non-interactive and one interactive. In the first step, ALIN chooses the first mappings, among which some are directly placed in the alignment and others are presented to the expert. In the 2019 version, ALIN uses new techniques to improve the first step, thus placing more mappings directly in the alignment without having to present them to the expert.

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 2019 version, however, we introduced modifications in Step 1. In previous versions, ALIN automatically accepted only the entities with the same name. In this version, ALIN also automatically accepts the entities whose names are synonyms or with variations in name words. ALIN searches synonyms in the Wordnet. In the Anatomy track, ALIN uses the FMA ontology too.

ALIN applies the following techniques:

- Line 1. ALIN selects mappings using linguistic similarities between entity names. ALIN uses synonyms and variations in entity name words to automatically accept mappings. At this time, ALIN automatically selects and accepts only concept mappings. To do that, ALIN uses linguistic metrics. 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 [2]. It is important to know that these suspended mappings can be unsuspended later, by structural analysis, as proposed in [3].
- 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 [4][5] (we explain mapping anti-pattern below) with the accepted mappings; (ii) selects data property (like [6]) 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 [3].
– 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

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

2 Results

Interactive ontology matching is the focus of the ALIN system. Comparing its results in the 2019 campaign to its previous participations (Table 5), ALIN improvements include an expressive reduction on the number of interactions with the expert and the increase of the quality of the generated alignment.

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

ALIN used new techniques to automatically accept mappings. These techniques led to an increase in the F-Measure of non-interactively generated alignment, which shows the effectiveness of the techniques. (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 - OAEI 2018[7]/2019[8]

Year	Precision	Recall	F-measure
2018	0.998	0.611	0.758
2019	0.974	0.698	0.813

[2] $https://drive.google.com/file/d/1SxJL6fLRVqI84epm8DbA_MlcscEoGbgZ/view?usp=sharing$

Table 2. Participation of ALIN in Conference Non-Interactive Track - OAEI 2018/2019[9]

Year	Precision	Recall	F-measure
2018	0.81	0.42	0.55
2019	0.82	0.43	0.56

2.2 Comments on the Participation of the ALIN in Interactive Tracks

In the Anatomy track, ALIN was tied for second in quality (F-Measure) with slightly lower total requests (Table 3). In the Conference track, ALIN was tied for first in quality with a slightly higher total request (Table 4).

Table 3. Participation of ALIN in Anatomy Interactive Track - Error Rate 0.0[10]

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.979	0.85	0.91	365
AML	0.968	0.948	0.958	236
LogMap	0.982	0.846	0.909	388

Table 4. Participation of ALIN in Conference Interactive Track - Error Rate 0.0[10]

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.914	0.695	0.79	228
AML	0.91	0.698	0.79	221
LogMap	0.886	0.61	0.723	82

Interactive Anatomy Track In this track, ALIN has had a decrease in the number of expert interactions and an increase in the quality of the generated alignment, showing that the new techniques used to automatically accept correct mappings are effective (Table 5).

ALIN used the FMA ontology to help find synonyms between the two ontologies of the Anatomy track. The Foundational Model of Anatomy Ontology (FMA) is a reference ontology for the domain of Human anatomy [3].

[3] "Foundational Model of Anatomy Ontology". Available at http://sig.biostr.washington.edu/projects/fm/AboutFM.html Last accessed on Oct, 11, 2019.

Interactive Conference Track In this track, ALIN has had a decrease in the number of expert interactions keeping a good quality of the generated alignment (Table 7).

2.3 Comparison of the Participation of ALIN in OAEI 2019 with his Participation in OAEI 2018

In this version, ALIN uses new techniques to automatically accept mappings. These techniques use synonyms and word variations to find equal entities between the two ontologies. ALIN also started to use FMA ontology as an external resource.

The use of the new techniques proved to be effective as it reduced the number of interactions while keeping a good level of quality. The new techniques also increased the quality of the alignment generated in Anatomy interactive tracking, where ALIN used the FMA ontology.

It is not always possible to use an external resource to find synonyms between entities of two ontologies, but when this is possible, the results showed that it is worth it.

The quality of the alignment generated by ALIN is dependent on the correct expert feedback, as expert responses are used to select new mappings. When ALIN selects wrong mappings, the quality of the generated alignment tends to decrease. But if we compare this year's quality decline with last year's, we see that this fall is less sharp (Table 6 and Table 8). The less sharp decline in quality is because we need less user interaction as we are automatically accepting more mappings.

The organization of FMA ontology in memory and the search for synonyms and word variations led to longer run time (Table 9 and Table 10)

Table 5. Participation of ALIN in Anatomy Interactive Track - OAEI 2016[11]/2017[12]/2018[7]/2019[10] - 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

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;

Table 6. F-Measure of ALIN in Anatomy Interactive Track - OAEI /2018[7]/2019[10] - with Different Error Rates

Year	Error rate 0.0	Error rate 0.1
2018	0.902	0.854
2019	0.91	0.889

Table 7. Participation of ALIN in Conference Interactive Track - OAEI 2016[11]/2017[12]/2018[7]/2019[10] - 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

Table 8. F-Measure of ALIN in Conference Interactive Track - OAEI /2018[7]/2019[10] - with Different Error Rates

Year	Error rate 0.0	Error rate 0.1
2018	0.809	0.705
2019	0.79	0.725

Table 9. Run Time (sec) in Anatomy Interactive Track - OAEI /2018[13]//2019[10]

Tool	2018	2019
ALIN	317	2132
AML	48	82
LogMap	23	29

Table 10. Run Time (sec) in Conference interactive track - OAEI /2018[13]/2019[10]

Tool	2018	2019
ALIN	106	397
AML	35	34
LogMap	37	37

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

And there was a worsening run time, where we could improve too.

3.1 Conclusions

ALIN used new techniques to automatically accept new mappings. They have been effective in reducing the number of interactions, while also keeping good quality in the generated alignment. In the case of the Anatomy track, these new techniques both decreased the number of interactions and increased the quality of the generated alignment. We can explain this quality improvement in this track by the use of the FMA ontology as a new external resource. With the use of the new techniques in both Anatomy and Conference tracks, there has been a less sharp drop in quality as the expert makes mistakes. Nevertheless, ALIN had an increase in run time due to the use of the new techniques, which may be addressed in future work.

References

1. Paulheim, H., Hertling, S., Ritze, D.: Towards Evaluating Interactive Ontology Matching Tools. Lecture Notes in Computer Science **7882** (2013) 31–45
2. 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
3. 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
4. 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)
5. 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
6. Silva, J., Revoredo, K., Baião, F.A., Euzenat, J.: Interactive Ontology Matching: Using Expert Feedback to Select Attribute Mappings. (2018)
7. 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
8. : Results for oaei 2019 - anatomy track. http://oaei.ontologymatching.org/2019/results/anatomy/ Accessed: 2019-10-11.
9. : Results of evaluation for the conference track within oaei 2019. http://oaei.ontologymatching.org/2019/results/conference/index.html Accessed: 2019-10-11.
10. : Results for oaei 2019 - interactive track. http://oaei.ontologymatching.org/2019/results/interactive/ Accessed: 2019-10-11.
11. 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
12. 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
13. : Results for oaei 2018 - interactive track. http://oaei.ontologymatching.org/2018/results/interactive/index.html Accessed: 2019-10-11.