ALIN Results for OAEI 2021

Jomar da Silva[1], Josefino Cabral Melo Lima[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,cabrallima@ufrj.br,
kate.revoredo@wu.ac.at,fbaiao@puc-rio.br

Abstract. [4]
ALIN is a system for interactive ontology matching. The ALIN version participating in OAEI 2021 applies natural language processing techniques (NLP) to standardize the concept names of the ontologies that participate in the matching process, like last year's version. In this version, we modified the grammars related to the ontologies to improve the quality of the alignments. This article describes the participation of ALIN at OAEI 2021 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. The Ontology Matching process seeks to discover correspondences (mappings) between entities of different ontologies, and this may be performed manually, semi-automatically or automatically [1]. Among all semi-automatic approaches, the ones that follow an interactive strategy stand out, considering the knowledge of domain experts through their participation during the matching process [2]. The use of a domain expert is not always possible since it is an expensive, scarce and time-consuming resource; when available, however, this strategy has achieved results that are superior to automatic (non-interactive) strategies. Nevertheless, there is still room for improvements [2], as

evidenced by the most recent results from the evaluation of interactive tools in the OAEI[5] (Ontology Alignment Evaluation Initiative). ALIN [3] is a system for interactive ontology matching which has been participating in all OAEI editions since 2016, with increasingly improved results.

1.1 State, Purpose and General statement

Interactive ontology matching systems select mappings for domain expert evaluates. ALIN selects many of these mappings through semantic and lexical metrics. As the concept names of the ontologies are not standardized, these metrics may return lower values than would be the case if they were standardized. This smaller metric may cause ALIN not to select these mappings for evaluation by the domain expert. In its 2020 version, ALIN proposed Natural Language Processing (NLP) techniques such as the development of regular grammars (in reality its equivalent regular expressions) and context free grammars along with their respective lexical analyzers (scanners) and syntax analyzers (parsers), for the concept names of the ontologies to be matched. The use of these NLP resources (scanners and parsers) makes it possible to translate different patterns used in the two ontologies into a unique one. This standardization allows ALIN to select better mappings for the domain expert to evaluate.

To do the standardization, ALIN will have a new phase before the execution of the program. In this phase, an NLP expert develops, manually, grammars to the concept names of the ontologies and their respective scanners and parsers. ALIN uses these scanners and parsers during the execution of the program. This new phase is possible in an interactive ontology matching system because:

1. We know before the program runs which ontologies it will match, as we need to look for experts in the domain of ontologies to interact with the program;
2. The process of searching, meeting, and scheduling a day available for the expert to participate in the process can take a long time, probably a few days.

We can use this time of a few days until the execution of the program to develop the necessary grammars, scanners, and parsers for the ontologies. In this version of ALIN, the authors of this paper played the role of the NLP expert.

1.2 Specific techniques used

During its matching process, 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.

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

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 domain expert: accepts or rejects selected mappings, according to domain 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 included a new step where an NLP expert develops grammars, and their respective scanners, and parsers to the concept names of the ontologies. ALIN uses these scanners and parsers to standardize the concept names of the ontologies and thus improve the generated alignment. The new step can lead to, for example, correcting spelling errors and unifying different spellings for the same concept name. More detailed examples of possible standardization of concept names are presented in [4]. ALIN uses the developed scanners and parsers in step 1 of the program.

ALIN applies the following techniques:

- Step 1. ALIN runs the scanners and the parsers for each concept name of the ontologies, modifying it and standardizing it. ALIN uses a blocking strategy where it discards all data properties and object properties of the ontologies. So, in this step, ALIN selects only concept mappings, using linguistic similarities between the concept names. ALIN automatically accepts concept mappings whose names are synonyms. ALIN uses the Wordnet and domain-specific ontologies (the FMA Ontology in the Anatomy track) to find synonyms between entities.
- Step 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 relevant to know that these suspended mappings can be further unsuspended later, as proposed in [6].
- Step 3. At this point, the domain 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 domain expert.
- Step 4. Initially, the set of selected mappings contains only concept mappings. At each interaction with the domain expert, if s/he accepts the mapping, ALIN (i) removes from the set of selected mappings all the mappings that compose an instantiation of a mapping anti-pattern [7][8] (we explain mapping anti-patterns 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].
- Step 5. The interaction phase continues until there are no selected mappings.

There are logical constraints which should apply to several ontologies. For example, 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.

In its 2021 version, we modified the grammars used in ALIN to improve the quality of the generated alignments.

1.3 Link to the system and parameters file

ALIN is available [6] as a package to be run through the SEALS client.

2 Results

Interactive ontology matching is the focus of the ALIN system. If you compare the participation of ALIN in 2021 and 2020 (Tables 4 and 5), you will see an improvement in the quality of the generated alignment, showing the effectiveness of the changes in the grammars.

2.1 Comments on the participation of ALIN in non-interactive tracks

The modified grammars generated an increase in the F-Measure of non-interactively generated alignments, both in the Anatomy track and in the Conference track (Table 1).

2.2 Comments on the participation of 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 2). In the Conference track, ALIN was first in quality and third in total requests (Table 3).

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

Table 1. Participation of ALIN in Anatomy Non-Interactive Track - 2020[10]/2021 and Conference Non-Interactive Track - 2020[10]/2021

	Year	Precision	Recall	F-measure
Anatomy track	2020	0.986	0.72	0.832
	2021	0.983	0.726	0.835
	Year	Precision	Recall	F-measure
Conference track	2020	0.82	0.43	0.56
	2021	0.87	0.46	0.59

Table 2. Participation of ALIN in Anatomy Interactive Track - OAEI 2021 - Error Rate 0.0

Tool	Precision	Recall	F-measure	Total Requests
ALIN(est 2021)	0.986	0.887	0.934	372
AML(2020)	0.972	0.933	0.952	189
LogMap(2020)	0.988	0.846	0.912	388

Table 3. Participation of ALIN in Conference Interactive Track - OAEI 2021 - Error Rate 0.0

Tool	Precision	Recall	F-measure	Total Requests
ALIN(est 2021)	0.916	0.718	0.799	253
AML(2020)	0.91	0.698	0.79	221
LogMap(2020)	0.886	0.61	0.723	82

Interactive Anatomy Track In this track, ALIN increased the number of interactions with the domain expert and the quality of the generated alignment. So modifications in the grammars improved the quality of this alignment in this track. (Table 4).

Interactive Conference Track In this track, ALIN also increased the quality of the generated alignment and the number of domain expert interactions (Table 5).

3 General comments

Evaluating the OAEI 2021 results, ALIN has improved the quality of the generated alignment in the interactive track but with a increased number of requests to the user.

Table 4. Participation of ALIN in Anatomy Interactive Track - OAEI 2016[11]/2017[12]/2018[13]/2019[14]/2020[10]/2021 - 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
2021(est)	0.986	0.887	0.934	372

Table 5. Participation of ALIN in Conference Interactive Track - OAEI 2016[11]/2017[12]/2018[13]/2019[14]/2020[10]/2021 - 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
2021(est)	0.916	0.718	0.799	253

3.1 Conclusions

In this 2021 version, we modified ALIN by modifying the grammars used by the ontologies. This modification have been effective in increasing the quality of the generated alignment but with a increased number of requests to the user.

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 2020. In: Proceedings of the 15th International Workshop on Ontology Matching. OM'20 (2020) 139–146
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. 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
14. 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