

ALIN Results for OAEI 2021

Jomar da Silva¹,
Kate Revoredo², Fernanda Araujo Baião³, and Cabral Lima¹

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, kate.revoredo@wu.ac.at,

fbaiao@puc-rio.br, cabrallima@ufrj.br

Abstract. ⁴

ALIN is a system for interactive ontology matching. In last year's version, ALIN applied natural language processing techniques (NLP) to standardize the concept names of the ontologies that participate in the matching process. In the current version, we modified the grammars used during the process in order 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]. The interactive approach, which considers the knowledge of domain experts through their participation during the matching process, has stood out among semi-automatic ones [2]. A domain expert is an expensive, scarce, and time-consuming resource; when available, however, this resource has improved the achieved results. Nevertheless, there is still room for improvements [2], as evidenced by the most recent results from the evaluation of interactive tools in the OAEI⁵ (Ontology

⁴ Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

⁵ Available at <http://oei.ontologymatching.org/2021/results/interactive/index.htm>, last accessed on Oct, 20, 2021.

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 in the Anatomy interactive track.

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 the use of Natural Language Processing (NLP) resources such as regular grammars (in fact, their equivalent regular expressions) and context-free grammars along with their respective lexical analyzers (scanners) and syntax analyzers (parsers).

The use of these NLP resources 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 has a new phase before the program runs. In this phase, an NLP expert manually develops grammars, and their respective scanners and parsers, to the ontologies. 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. To test the new techniques, the authors of this paper played the role of the NLP experts.

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 selected mappings.

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 techniques in the 'Description of the used techniques' paragraph 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 to step 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 procedure.

Description of the used 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 standardized concept names. ALIN automatically accepts concept mappings whose standardized 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]. These suspended mappings can be further unsuspended later, returning to the status of selected mappings, 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 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 in the 'Mapping anti-patterns' paragraph) with the accepted mappings; (ii) selects data property (as proposed in [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 (as proposed in [6]).
- Step 5. The interaction phase continues until there are no selected mappings.

Mapping anti-patterns An anti-pattern mapping can be a logical inconsistency, a construction constraint on the ontology, or an alignment constraint. An ontology may have construction constraints, such as a concept cannot be equivalent to its superconcept. The alignment between two ontologies can have a constraint like, for example, an entity of ontology O cannot be equivalent to two entities of the ontology O' . Anti-pattern mapping 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 by 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.

The comparison between the participation of ALIN in 2021 and 2020 (Tables 4 and 5) shows an improvement in the quality of the generated alignment, attesting 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 in the Anatomy track. The modified grammars generated a minor increase in the F-Measure of non-interactively generated alignments in the Anatomy track. There was a decrease in precision and an increase in recall. The current version of ALIN did not participate in the Conference track. (Table 1).

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

Table 1. Participation of ALIN in Anatomy Non-Interactive Track in OAEI 2020[10] and OAEI 2021[11]

	Year	Precision	Recall	F-measure
Anatomy track	2020	0.986	0.72	0.832
	2021	0.983	0.726	0.835

2.2 Comments on the participation of Alin in interactive tracks

In the Anatomy track, ALIN was better than LogMap in quality (F-Measure) but worse in total requests, and worse in both aspects than AML (Table 2). In the Conference track, ALIN was first in quality and third in total requests (Table 3).

Table 2. Participation of ALIN in Anatomy Interactive Track - OAEI 2021[12] - Error Rate 0.0

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.986	0.887	0.934	404
AML	0.972	0.933	0.952	189
LogMap	0.988	0.846	0.912	388

Table 3. Participation of ALIN in Conference Interactive Track - OAEI 2021[12] - Error Rate 0.0

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.916	0.718	0.799	281
AML	0.91	0.698	0.79	220
LogMap	0.886	0.61	0.723	82

2.3 Alin participations in OAEI 2021 and OAEI 2020

ALIN improved the quality of the alignment generated for both the Anatomy and the Conference tracks, with a higher improvement for the Anatomy scenario. The number of interactions has also increased on both tracks (Tables 4 and 5).

The quality of the alignment generated by ALIN depends on the correct feedback from the domain expert, as ALIN uses this feedback to select new mappings. When ALIN selects wrong mappings, the quality of the generated

alignment tends to decrease. The comparison between the two runs of the table 6 shows a quality decline sharper in the current version.

The run time of ALIN 2021 was shorter than last year’s (Table 7), but other systems also had a similar reduction. So there is a chance that this reduction may be due to changes in the computational environment.

Table 4. Participation of ALIN in Anatomy Interactive Track - OAEI 2016[13]/2017[14]/2018[15]/2019[16]/2020[10]/2021[12] - 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	0.986	0.887	0.934	404

Table 5. Participation of ALIN in Conference Interactive Track - OAEI 2016[13]/2017[14]/2018[15]/2019[16]/2020[10]/2021[12] - 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	0.916	0.718	0.799	281

3 General comments

The analysis of OAEI 2021 results showed that ALIN 2021 has improved the quality of the generated alignment in the interactive track. Moreover, an increase in the expert error rate caused a sharper fall in the alignment quality. The number of interactions with the expert was increased (from 233 to 281 requests in the Conference track and from 360 to 404 requests in the Anatomy track). Finally, ALIN had a decrease in run time, like all other participants (Table 7), so the drop seems to be caused by changes in the computing environment rather than changes made to the program.

Table 6. F-Measure of ALIN in Anatomy Interactive Track - OAEI /2020[10]/2021[12] and in Conference Interactive Track - OAEI /2020[10]/2021[12] - with Different Error Rates

	Year	Error rate 0.0	Error rate 0.1
Anatomy	2020	0.917	0.887
	2021	0.934	0.899
	Year	Error rate 0.0	Error rate 0.1
Conference	2020	0.796	0.713
	2021	0.799	0.698

Table 7. Run Time (sec) in Anatomy Interactive Track - OAEI /2020[10]/2021[12] and in Conference interactive track - OAEI /2020[10]/2021[12]

	Tool	2020	2021
Anatomy	ALIN	1152	880.66
	AML	37,3	16.101
	LogMap	7,6	4.209
	Tool	2019	2020
Conference	ALIN	136,9	82.381
	AML	30.1	20.727
	LogMap	37.96	23.999

3.1 Conclusions

This paper presented the newest version of ALIN (ALIN 2021), an interactive matching tool that makes use of natural language processing (NLP) resources, and its performance results obtained in OAEI 2021. When compared to its previous version, ALIN 2021 used modified grammars during the NLP process to allow for a higher level of standardization among the terms used by the ontologies being aligned. These grammar modifications were tested on the Anatomy and Conference tracks and proved to be efficient in increasing the quality of the generated alignment, obtaining the highest F-measure value in the Conference interactive track, with an expert error rate of zero.

In future versions, we intend to implement new grammar changes to further improve the quality of the generated alignment.

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. : Results for oaei 2021 - anatomy track. <http://oaei.ontologymatching.org/2021/results/anatomy/index.html> Accessed: 2021-10-19.
12. : Results for oaei 2021 - interactive track. <http://oaei.ontologymatching.org/2021/results/interactive/index.htm> Accessed: 2021-10-19.
13. 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
14. 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
15. 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
16. 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