ALIN Results for OAEI 2023

Jomar da Silva[1,*], Kate Revoredo[2], Fernanda Araujo Baião[3] and Cabral Lima[1]

[1]*Graduate Program in Informatics*
Federal University of Rio de Janeiro (UFRJ), Brazil
[2]*Humboldt-Universität zu Berlin, Berlin, Germany*
[3]*Department of Industrial Engineering*
Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil

Abstract
Alin is a system for interactive ontology matching that has been participating in all OAEI editions since 2016. Since 2020, Alin applies natural language processing (NLP) techniques to standardize the concept names of the ontologies that participate in the matching process.

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 (ICT) in general, a large amount of data repositories became available as valuable assets for enabling integrated data exchange platforms across organizations. However, those repositories 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 OAEI[1] (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 in the Anatomy interactive track.

OM 2023: The 18th International Workshop on Ontology Matching collocated with the 22nd International Semantic Web Conference ISWC-2023 November 7th, 2023, Athens, Greece
*Corresponding author.
jomar.silva@ufrj.br (J. d. Silva); kate.revoredo@hu-berlin.de (K. Revoredo); fbaiao@puc-rio.br (F. A. Baião); cabrallima@ufrj.br (C. Lima)

[1]Available at https://oaei.ontologymatching.org/2023/results/interactive/index.htm, last accessed on Nov, 11, 2023.

Alin uses Seals[4] to perform interactive ontology matching. Seals is a tool that allows, among other things, simulating the participation of an expert in the interactive ontology matching process. In addition to simulating the expert's response, Seals also keeps track of interactions with the expert. The counting process for interactions with the expert in Seals was modified this year, leading to a decreased number of interactions with the expert during the execution of Alin compared to the execution of Alin in the year 2022.

1.1. State, Purpose and General statement

Interactive ontology matching systems select mappings for domain expert evaluation. Alin selects many of these mappings through semantic and lexical metrics. Concept name standardization could lead to a more correct value in the metrics used. No standardized names may cause Alin not to select correct mappings for evaluation by the domain expert.

Since its 2020 version, Alin uses 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).

These NLP resources make it possible to translate different patterns used in the two ontologies into unique one. This standardization allows Alin to select better mappings for the domain expert to evaluate.

To perform the standardization step, since 2020 Alin has a new phase before the program runs. In this phase, an NLP expert provides grammars, and their respective scanners and parsers, to the ontologies. Alin uses these scanners and parsers during the execution of the program. This standarization step is possible in an interactive ontology matching system because:

1. We know before the program runs which ontologies Alin 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. Detailed in the 'Description of the used techniques' paragraph below;
2. Filter mappings: suspend some selected mappings, using lexical and semantic 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 first OAEI participation) follow this general procedure. Since its 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 [5]. Alin uses the developed scanners and parsers in step 1 of the procedure.

1.2.1. 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 and semantic 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 [6]. These suspended mappings can be further unsuspended later, returning to the status of selected mappings, as proposed in [7].
- 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. Alin can present up to three mappings together to the domain expert if a full entity name in a candidate mapping is the same as another entity name in another candidate mapping.
- 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 [8][9] (we explain mapping anti-patterns below in the 'Mapping anti-patterns' paragraph) with the accepted mappings; (ii) selects data property (as proposed in [10]) 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 [7]).
- Step 5. The interaction phase continues until there are no selected mappings.

1.2.2. 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 constraints. 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.

1.3. Link to the system and parameters file

ALIN is available [2] as a SEALS package (It can be run with MELT).

2. Results

The comparison between the participation of ALIN in 2022 and 2023 (Tables 1 and 2) shows a decrease in the number of interactions with the expert due to the modification of Seals. There were no modifications to the F-measure, Precision, and Recall metrics. Thus, there were no modifications in the results of ALIN's participation in the non-interactive tracks: Anatomy track and Conference track. Please refer to http://oaei.ontologymatching.org/2023/results/ for the results of the ALIN in the OAEI 2023 Anatomy and Conference tracks.

Table 1
Participation of ALIN in Anatomy Interactive Track - OAEI 2016[11]/2017[12]/2018[13]/2019[14]/2020[15]/2021[16]/2022[17]/2023[18] - 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
2022	0.987	0.92	0.952	579
2023	0.987	0.92	0.952	514

2.1. Comments on the participation of ALIN in interactive tracks

In the Anatomy interactive track, ALIN 2023 was better than LogMap in quality (F-Measure) but worse in total requests (Table 3). In the Conference track, ALIN 2023 was, again, better than LogMap in quality (F-Measure) but worse in total requests (Table 4).

[2]https://osf.io/pu7fv/?view_only=736f83561cfb421eac34db819939bc31

Table 2
Participation of Alin in Conference Interactive Track - OAEI 2016[11]/2017[12]/2018[13]/2019[14]/2020[15]/2021[16]/2022[17]/2023[18] - 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
2022	0.919	0.744	0.822	309
2023	0.919	0.744	0.822	274

Table 3
Participation of Alin in Anatomy Interactive Track - OAEI 2023[18] - Error Rate 0.0

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.987	0.92	0.952	514
LogMap	0.988	0.846	0.912	388

Table 4
Participation of Alin in Conference Interactive Track - OAEI 2023[18] - Error Rate 0.0

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.919	0.744	0.822	274
LogMap	0.886	0.61	0.723	82

3. General comments

Alin uses Seals[4] to perform interactive ontology matching. Seals is a tool that allows, among other things, simulating the participation of an expert in the interactive ontology matching process. The counting process for interactions with the expert in Seals was modified this year, resulting in a decreased number of interactions with the expert during the execution of Alin.

Acknowledgments

Kate Revoredo is funded by the Berliner Chancengleichheitsprogramm (BCP) as part of the DiGiTal Graduate Program.

Fernanda Baiao is partially funded by FAPERJ (grants 200.514/2023 and 211.308/2019) and CNPq (grants 312059/2022-1 and 422810/2021-5).

References

[1] J. Euzenat, P. Shvaiko, Ontology Matching - Second Edition, Springer-Verlag, 2013.

[2] H. Li, Z. Dragisic, D. Faria, V. Ivanova, E. Jimenez-Ruiz, P. Lambrix, C. Pesquita, User validation in ontology alignment: functional assessment and impact, The Knowledge Engineering Review (2019). doi:10.1017/S0269888919000080.

[3] J. Da Silva, K. Revoredo, F. Baião, J. Euzenat, Alin: improving interactive ontology matching by interactively revising mapping suggestions, The Knowledge Engineering Review 35 (2020). doi:10.1017/S0269888919000249.

[4] S. N. Wrigley, R. García-Castro, L. Nixon, Semantic evaluation at large scale (seals), in: Proceedings of the 21st International Conference on World Wide Web, WWW '12 Companion, Association for Computing Machinery, New York, NY, USA, 2012, p. 299–302. URL: https://doi.org/10.1145/2187980.2188033. doi:10.1145/2187980.2188033.

[5] F. J. Q. Real, G. Bella, F. McNeill, A. Bundy, Using domain lexicon and grammar for ontology matching, 2020.

[6] M. Cheatham, P. Hitzler, String similarity metrics for ontology alignment, in: Proceedings of the 12th International Semantic Web Conference - Part II, ISWC '13, Springer-Verlag New York, Inc., New York, NY, USA, 2013, pp. 294–309.

[7] J. Silva, F. Baião, K. Revoredo, J. Euzenat, 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, pp. 13–24.

[8] A. Guedes, F. Baião, R. Shivaprabhu, Revoredo, On the Identification and Representation of Ontology Correspondence Antipatterns, in: Proc. 5th Int. Conf. Ontol. Semant. Web Patterns (WOP'14), CEUR Work. Proc., 2014.

[9] A. Guedes, F. Baião, K. Revoredo, Digging Ontology Correspondence Antipatterns, in: Proceeding WOP'14 Proc. 5th Int. Conf. Ontol. Semant. Web Patterns, volume 1032, 2014, pp. 38––48.

[10] J. Silva, K. Revoredo, F. A. Baião, J. Euzenat, Interactive Ontology Matching: Using Expert Feedback to Select Attribute Mappings, in: CEUR Workshop Proceedings, volume 2288, 2018, pp. 25–36.

[11] J. Silva, F. Baião, K. Revoredo, Alin results for oaei 2016, in: OM-2016: Proceedings of the Eleventh International Workshop on Ontology Matching, OM'16, 2016, pp. 130–137.

[12] J. Silva, F. Baião, K. Revoredo, Alin results for oaei 2017, in: OM-2017: Proceedings of the Twelfth International Workshop on Ontology Matching, OM'17, 2017, pp. 114–121.

[13] J. Silva, F. Baião, K. Revoredo, Alin results for oaei 2018, in: Ontology Matching: OM-2018: Proceedings of the ISWC Workshop, OM'18, 2018, pp. 117–124.

[14] J. Silva, C. Delgado, K. Revoredo, F. Baião, Alin results for oaei 2019, in: Proceedings of the 14th International Workshop on Ontology Matching, OM'19, 2019, pp. 94–100.

[15] J. Silva, C. Delgado, K. Revoredo, F. Baião, Alin results for oaei 2020, in: Proceedings of the 15th International Workshop on Ontology Matching, OM'20, 2020, pp. 139–146.

[16] J. Silva, , K. Revoredo, F. Baião, C. Lima, Alin results for oaei 2021, in: Proceedings of the 16th International Workshop on Ontology Matching, OM'21, 2021, pp. 109–116.

[17] J. Silva, , K. Revoredo, F. Baião, C. Lima, Alin results for oaei 2022, in: Proceedings of the 17th International Workshop on Ontology Matching, OM'22, 2022, pp. 129–136.

[18] Results for oaei 2023 - interactive track, 2023. URL: https://oaei.ontologymatching.org/2023/results/interactive/index.htm, accessed: 2023-11-11.