ALIN Results for OAEI 2024

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. In this new version, we modified the lexical analyzers used in Alin. Additionally, we used ChatGPT to enhance the selected mappings. Both modifications have reduced the number of interactions with the expert.

Keywords
ontology matching, Wordnet, interactive ontology matching, ontology alignment, interactive ontology alignment, lexical analyzer, ChatGPT

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.

OM-2024: The 19th International Workshop on Ontology Matching collocated with the 23rd International Semantic Web Conference (ISWC 2024), November 11th, Baltimore, USA
*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/2024/results/interactive/, last accessed on Oct, 08, 2024.

In an interactive process, besides the F-Measure, which assesses the quality of the generated alignment, the number of interactions with the expert is also important—the fewer the interactions, the better. In the interactive process, a key step is selecting mappings for the expert. An improvement in the alignment process would occur if we could find a way to reduce the number of mappings in this set without lowering the F-Measure, or at least reduce the number of mappings at a greater rate than the decrease in the F-Measure.

In this year's version of Alin, we made two modifications that decreased the number of interactions; however, this also led to a decrease in quality, albeit to a lesser degree. The first modification was effective for both the Anatomy and Conference tracks, as we incorporated ChatGPT to filter the selected mappings.

Alin uses lexical analyzers to standardize entity names before evaluating them with similarity metrics. In this new version, we developed programs that helped us improve the lexical analyzers for the mouse and human ontologies in the Anatomy track. The second modification involved using these new lexical analyzers in conjunction with modifications to the suspension of selected mappings.

1.1. State, Purpose and General statement

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 'Specific techniques used' subsection below;
2. Filter mappings: suspend some selected mappings, using lexical and semantic criteria for that. In Alin 2024, we introduced a new criterion, and before suspending mappings, we excluded some selected mappings using ChatGPT.
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.

1.2. Specific techniques used

- Step 1. Alin employs a blocking strategy where it does not consider data and object properties from the ontologies at this step. It selects only concept mappings based on linguistic similarities between previously standardized concept names. Alin automatically accepts mappings with standardized names that are synonyms, using WordNet and domain-specific ontologies, such as the FMA Ontology in the Anatomy track, to identify these synonyms.

- Step 2. ALIN excludes all mappings rejected by ChatGPT from the set of selected mappings. It then suspends some selected mappings that exhibit low lexical and semantic similarity in their entity names, removing them from the set of selected mappings. We use the Jaccard, Jaro-Winkler, and n-gram lexical metrics to calculate the lexical similarity of the selected mappings. We also used a semantic metric called the ALIN metric. These suspended mappings can be further unsuspended later, returning to the set of selected mappings, as proposed in [4]. We employ a threshold for suspension, where we suspend a mapping if all its similarity values are below this threshold. Until last year, we used a threshold of 0.9 for both the Anatomy and Conference tracks. In 2024, we adjusted the threshold to 0.96 for the Anatomy track.
- 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 [5][6] (we explain mapping anti-patterns below in the 'Mapping anti-patterns' paragraph) with the accepted mappings; (ii) selects data property (as proposed in [7]) 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 [4]).
- Step 5. Go to step 3 until there are no selected mappings.

1.2.1. 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. Modifications made in the 2024 version of ALIN

Since 2020, we have employed lexical analyzers to standardize the names before evaluating them with the similarity metrics. In 2024, we improved the lexical analyzers used in the Anatomy track by using programs to automate the search for entity names that should have their spelling unified. This improvement increased the F-Measure but also significantly raised the number of interactions.

To address this issue, we decided to raise the threshold to suspend selected mappings, which led to a slight decrease in quality but a substantial reduction in the number of interactions.

Until last year, this value was set to 0.9. This year, we maintained this value for the Conference track but adjusted it to 0.96 for the Anatomy track.

Additionally, we employed ChatGPT to exclude mappings from the set of selected mappings, which resulted in a decrease in the number of interactions, with a slight decline in the quality of the generated alignment in both the Anatomy and Conference tracks.

1.4. 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 2023 and 2024 (Tables 1 and 2) shows a decrease in interactions with the expert. There was also a decrease in the quality of the generated alignment, although to a much lesser extent compared to the reduction in interactions.

Table 1
Participation of ALIN in Anatomy Interactive Track - OAEI 2016[8]/2017[9]/2018[10]/2019[11]/2020[12]/2021[13]/2022[14]/2023[15]/2024[16] - 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
2024	0.986	0.878	0.929	262

2.1. Comments on the participation of ALIN in interactive tracks

In the Anatomy interactive track, ALIN 2024 was better than LogMap in quality (F-Measure) and in total requests (Table 3). In the Conference track, ALIN 2024 was better than LogMap in quality (F-Measure) but worse in total requests (Table 4). Please refer to https://oaei.ontology-matching.org/2024/results/interactive/ for the results of the ALIN in the OAEI 2024 Interactive track.

3. General comments

This new version of ALIN uses ChatGPT to filter out mappings before expert feedback in interactive ontology matching. In the Anatomy track, this version also includes improvements

[2] https://osf.io/x7qhf

Table 2
Participation of ALIN in Conference Interactive Track - OAEI 2016[8]/2017[9]/2018[10]/2019[11]/2020[12]/2021[13]/2022[14]/2023[15]/2024[16] - 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
2024	0.915	0.702	0.794	187

Table 3
Participation of ALIN in Anatomy Interactive Track - OAEI 2024[16] - Error Rate 0.0

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.986	0.878	0.929	262
LogMap	0.988	0.846	0.912	388

Table 4
Participation of ALIN in Conference Interactive Track - OAEI 2024[16] - Error Rate 0.0

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.915	0.702	0.794	187
LogMap	0.886	0.61	0.723	82

to the lexical analyzers for standardizing entity names and an increase in the threshold for suspending selected mappings. The results indicate that these changes led to fewer interactions with the expert and a slight decrease in the quality of the generated alignment.

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] 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.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] 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.
[12] 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.
[13] 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.
[14] 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.
[15] J. Silva, , K. Revoredo, F. Baião, C. Lima, Alin results for oaei 2023, in: Proceedings of the 18th International Workshop on Ontology Matching, OM'23, 2023, pp. 140–145.
[16] Results for oaei 2024 - interactive track, 2024. URL: https://oaei.ontologymatching.org/2024/results/interactive/, accessed: 2024-10-8.