ALIN Results for OAEI 2018

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

Graduated Program in Informatics, Department of Applied Informatics
Federal University of the State of Rio de Janeiro (UNIRIO), Brazil
{jomar.silva,katerevoredo,fernanda.baiao}@uniriotec.br

Abstract. ALIN is an ontology matching system specialized in the interactive ontology matching. The main characteristic of ALIN is the use of expert feedback to improve the set of mapping suggestions. ALIN uses semantic and structural techniques to improve it. ALIN has obtained the alignment with the highest quality in the interactive tracking for Conference data set. This paper describes its configuration for the OAEI 2018 competition and discusses its results.

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

1 Presentation of the system

A large amount of data repositories became available due to the advances in information and communication technologies. 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 participated in the OAEI 2016 and OAEI 2017 evaluations.

1.1 State, purpose, general statement

ALIN has the following steps to perform the interactive ontology matching process: First, ALIN generates an initial set of mappings. This set is called the set of mapping suggestions, that are the mappings to receive expert feedback. After, the interactive phase begins, where, at each interaction, the expert gives his feedback for some mapping suggestions. After each expert feedback, ALIN modifies the set of mapping suggestions according to the expert feedback. The modification of the set of mapping suggestions is by the use of the structural analysis of ontologies and the use of alignment anti-patterns. The interactions continue until there are no more mapping suggestions left.

1.2 Specific techniques used

The ALIN algorithm is shown in algorithm 1.

Algorithm 1 ALIN algorithm

Input: Two ontologies to be aligned
Output: Alignment between the two ontologies

1: Loading of ontologies
2: Generation of the initial set of mapping suggestions
3: Move of mappings by automatic classification from the set of mappings suggestions to the alignment
4: Move of mappings by the low value of semantic similarity from the set of mapping suggestions to a backup set
5: **while** Set of mapping suggestions is not empty **do**
6:     Choose mapping from the the set of mapping suggestions to submit to the expert
7:     Receive expert feedback to chosen mapping and remove it from the set of mapping suggestions
8:     **if** Mapping is accepted **then**
9:         Remove mappings in an alignment anti-pattern with accepted mapping from the set of mapping suggestions
10:         Insert some data property and object property mappings related to the accepted mapping into set of mapping suggestions
11:         Move some mappings related to the accepted mapping from the backup set to the set of mapping suggestions
12:     **end if**
13: **end while**

The steps of ALIN algorithm are the following:

- Line 1. ALIN loads the ontology classes, object properties, and data properties through the Alignment API [2]. For each entity, some data are stored such as name and label. ALIN saves the class superclasses and disjunctions. ALIN also saves information about the object properties, like their hypernyms and their associated classes, and information about the data properties, like their associated class. ALIN does not use instances. The ALIN can only work with ontologies whose entity names are in English.
- Line 2. For each similarity metric (SimM) a set of mapping is found using a simple matching algorithm. This algorithm treats the matching problem as a stable marriage problem with size list limited to 1 [3], i.e., the algorithm only selects one mapping if similarity value between the two entities of the mapping is the highest considering all the mappings with at least one of these entities. ALIN uses six metrics and runs six times, once for each one, giving rise, each execution, to a set of mappings. The union of the sets

gives origin to the initial set of mapping suggestions. ALIN uses the linguist metrics Jaccard, Jaro-Winkler, and n-Gram, Resnick, Jiang-Conrath, and Lin. Simmetrics API [4] provides the metrics Jaccard, Jaro-Winkler, and n-Gram and HESML API [5] the metrics Resnick, Jiang-Conrath, and Lin. HESML API uses Wordnet. To use Wordnet, the canonical form of the entity names is needed, therefore Stanford CoreNLP API [6] was used. The most frequent synsets of words are used to calculate semantic similarities.

- Line 3. The value of the similarity metrics (Resnick, Jiang-Conrath, Lin, Jaccard, Jaro-Winkler, and n-Gram) varies from 0 to 1 (1 is the maximum value). When one mapping in the set of mapping suggestions has all the six metrics with the maximum value, ALIN moves the mapping from the set of mapping suggestions to the final alignment.
- Line 4. ALIN moves the mappings whose entities has one of its linguistic metrics less than a given threshold from the set of mapping suggestions to a backup set. These mappings can return later, by structural analysis, to the set of mapping suggestions. [7] shows this technique, but with a little difference, it didn't use a threshold. It moves the class mappings that are not in the same Wordnet synset.
- Lines from 5 to 13. At this point, the interactions with the expert begin. ALIN sorts the mappings in the set of mapping suggestions by the sum of similarity metric values, greater sum first. ALIN submits the mappings to the expert. The set of mapping suggestions has, at first, only class mappings. After each expert feedback, if the expert accept the mapping, ALIN moves it from the set of mapping suggestions to the alignment, else ALIN removes it from the set of mapping suggestions. ALIN can remove mappings (besides the mappings that received feedback) from the set of mapping suggestions and can include other mappings into it, depending on the expert feedback.

At each interaction with the expert:

- ALIN removes from the set of mapping suggestions all the mappings that are in alignment anti-pattern [8][9] with the accepted mapping;
- ALIN inserts into the set of mapping suggestions, data property and object property mappings related to the accepted class mappings. [10] shows how ALIN inserts data properties into the set of mapping suggestions.
- ALIN moves from the backup set to the set of mapping suggestions all mappings whose both entities are subclasses of the classes of an accepted mapping. [7] shows a similar technique.

This step continues until the set of mapping suggestion is empty.

1.3 Link to the system and parameters file

ALIN is available through Google drive (
https://drive.google.com/open?id=1myVtcRoKKdUDHQTKNKsomna8AFbukanf) as a package for running through the SEALS client.

2 Results

Interactive ontology matching is the focus of the ALIN system. The system performs better when the number of data and object properties is proportionately large. In the interactive phase the system select as mappings to be submitted to the expert property mappings related with accepted class mappings, thus allowing increase the recall. When the number of properties in the ontologies is small, the system still generates an alignment with good precision, but its recall tends to be not so good.

Another characteristic of ALIN is its reliance on the interactive phase to generate an alignment with good quality. The non-interactive phase of the system is quite simple, based on maximum string similarity. Its goal is to achieve a high precision without worrying about the recall. The recall increases in the interactive phase, thus improving the F-measure. So, ALIN is not robust to expert mistakes. The system uses techniques that take advantage of the expert feedback to include new mappings into the set of mapping suggestions. When the expert gives a wrong answer, it is propagated generating other errors, thereby decreasing the F-measure.

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

As expected the participation of ALIN in non-interactive matching tracks showed the following results: high precision and not so high recall when compared to the other tools, as can be seen in Anatomy track[1] shown in Table 1.

2.2 Comments on the participation of the ALIN in interactive tracks

Interactive Anatomy Track In this track, the program ALIN showed the highest precision among the four evaluated tools when the error rate is zero, as can be seen in Table 2. When the error rate increases, both the precision as the recall falls, so falling the F-measure, as we can see in Table 3. Dependence on expert feedback to ensure precision and to increase recall explains this decline in quality when the expert makes mistakes.

As ontologies of the Anatomy Track contains almost no properties, ALIN cannot utilize some interactive techniques like the selection of property mappings related to accepted class mappings. Not using these techniques has limited the increase in recall, which influenced the F-measure.

[1] Results for OAEI 2018 - Anatomy track. Available at http://oaei.ontologymatching.org/2018/results/anatomy/ Last accessed on Oct, 02, 2018.

Table 1. Participation of ALIN in Anatomy non-interactive track

Tool	Precision	Recall	F-Measure
AML	0.95	0.936	0.943
LogMapBio	0.888	0.908	0.898
POMAP++	0.919	0.877	0.897
XMap	0.929	0.865	0.896
LogMap	0.918	0.846	0.88
SANOM	0.888	0.844	0.865
FCAMapX	0.941	0.791	0.859
KEPLER	0.958	0.741	0.836
Lily	0.872	0.795	0.832
LogMapLite	0.962	0.728	0.828
ALOD2Vec	0.996	0.648	0.785
StringEquiv	0.997	0.622	0.766
DOME	0.997	0.615	0.761
ALIN	0.998	0.611	0.758
Holontology	0.976	0.294	0.451

Table 2. Participation of ALIN in Anatomy interactive track - Error rate 0.0

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.994	0.826	0.902	602
AML	0.964	0.948	0.956	240
LogMap	0.982	0.846	0.909	388
XMap	0.929	0.867	0.897	35

Table 3. Participation of ALIN in Anatomy interactive track - Error rate 0.1

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.914	0.802	0.854	578
AML	0.952	0.946	0.948	268
LogMap	0.961	0.832	0.892	388
XMap	0.929	0.867	0.897	35

Interactive Conference Track In this track, ALIN stood out, showing the greatest F-measure among the four tools when the error rate is zero, as can be

Table 4. Participation of ALIN in Conference interactive track - Error rate 0.0

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.921	0.721	0.809	276
AML	0.912	0.711	0.799	270
LogMap	0.886	0.61	0.723	82
XMap	0.719	0.62	0.666	16

Table 5. Participation of ALIN in Conference interactive track - Error rate 0.1

Tool	Precision	Recall	F-measure	Total Requests
ALIN	0.725	0.686	0.705	264
AML	0.838	0.698	0.762	277
LogMap	0.85	0.596	0.7	82
XMap	0.719	0.62	0.666	16

seen in 4, as with a loss of F-measure when the error rate increases, as can be seen in Table 5.

Other results, including results with different error rates, can be seen on the OAEI 2018[2] page.

2.3 Comparison of the participation to ALIN in OAEI 2017 with his participation in OAEI 2016

- One modification made in ALIN was the withdrawal of additional criteria for the automatic classification of mappings. At the beginning of its execution, ALIN, automatically, select mappings with the entities with the same name to put into the alignment. In the OAEI 2017, ALIN used additional criteria for that, that is, if a mapping had the two entities with the same name, but had met one of those criteria, ALIN didn't put it into the alignment. In the conference data set, the use of these criteria increased the precision of the alignment, and thus its quality, but also the number of interactions. In the Anatomy data set, the use of these criteria increased only the number of interactions. For OAEI 2018, ALIN focused on reducing its number of interactions. So, ALIN doesn't use the additional criteria for the automatic classification of mappings anymore. This modification reduced the number of interactions (Total Requests) in both the conference track and the anatomy track, without decreasing the quality (F-measure) on the anatomy track, as can be seen in Tables 7 and 6.
- Another modification was the selection of new mappings to the set of mapping suggestions. For OAEI, one interactive matching system can place up

[2] Results for OAEI 2018 - Interactive Track . Available at http://oaei.ontologymatching.org/2018/results/interactive/ Last accessed on Oct, 2, 2018.

to three related mappings in an interaction. To take advantage of this rule, in 2018, ALIN selects new mappings, with at least one entity equal to other already selected, to put into the set of mapping suggestions. This selection increases the likelihood of raising the recall. This modification increased the recall on the anatomy track but not increased enough on the conference track to compensate for the first modification, as can be seen in Tables 7 and 6.
- ALIN has stopped using the WS4J API[3]. ALIN had already stopped using WS4J to calculate similarity in OAEI 2017, starting to use HESML. ALIN was only using WS4J to find the most common synset to an entity name, but now it is directly accessing the wordnet files.

Table 6. Participation of ALIN in Anatomy interactive track - OAEI 2016[11]/2017[12]/2018- 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

Table 7. Participation of ALIN in Conference interactive track - OAEI 2016[11]/2017[12]/2018- 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

3 General Comments

Evaluating the results it can be seen that the system can be improved towards:

- handling user error rate;
- generating a higher quality (especially w.r.t. recall) initial alignment in its non-interactive phase;
- reducing the number of interactions with the expert.

[3] 'WS4J'. Available at https://github.com/Sciss/ws4j Last accessed on Jan, 16, 2018.

3.1 Conclusions

The ALIN system stands out in the interactive ontology matching process when ontologies have some characteristics, such as many documented properties, and the expert does not make mistakes.

The second author was partially funding by project PQ-UNIRIO N01/2017 (" Aprendendo, adaptando e alinhando ontologias:metodologias e algoritmos.") and CAPES/PROAP.

References

1. Paulheim, H., Hertling, S., Ritze, D.: Towards Evaluating Interactive Ontology Matching Tools. Lecture Notes in Computer Science **7882** (2013) 31–45
2. David, J., Euzenat, J., Scharffe, F., Trojahn dos Santos, C.: The alignment api 4.0. Semant. web **2**(1) (January 2011) 3–10
3. Irving, R.W., Manlove, D.F., O'Malley, G.: Stable marriage with ties and bounded length preference lists. Journal of Discrete Algorithms **7**(2) (2009) 213 – 219 Selected papers from the 2nd Algorithms and Complexity in Durham Workshop ACiD 2006.
4. Surhone, L.M., Timpledon, M.T., Marseken, S.F.: SimMetrics. VDM Publishing (2010)
5. Lastra-Daz, J.J., Garca-Serrano, A., Batet, M., Fernndez, M., Chirigati, F.: Hesml. Inf. Syst. **66**(C) (June 2017) 97–118
6. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: Association for Computational Linguistics (ACL) System Demonstrations. (2014) 55–60
7. 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
8. 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 (WOP14), CEUR Work. Proc. (2014)
9. Guedes, A., Baião, F., Revoredo, K.: Digging Ontology Correspondence Antipatterns. In: Proceeding WOP14 Proc. 5th Int. Conf. Ontol. Semant. Web Patterns. Volume 1032. (2014) 38–48
10. Silva, J., Revoredo, K., Baião, F.A., Euzenat, J.: Interactive Ontology Matching: Using Expert Feedback to Select Attribute Mappings. (2018)
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