

Results of AML in OAEI 2017*

Daniel Faria¹, Booma Sowkarthiga Balasubramani², Vivek R. Shivaprabhu²,
Isabela Mott³, Catia Pesquita³, Francisco M. Couto³, and Isabel F. Cruz²

¹ Instituto Gulbenkian de Ciência, Portugal

² ADVIS Lab, Department of Computer Science, University of Illinois at Chicago, USA

³ LaSIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

Abstract. AgreementMakerLight (AML) is an automated ontology matching system that was developed with both extensibility and efficiency in mind. This paper describes its configuration for the OAEI 2017 competition and discusses its results. For this OAEI edition, we built upon the instance matching foundations we laid last year, and tackled the new Hobbit track and its new evaluation platform. AML was the only system to participate in all OAEI tracks this year, and was the top performing system or among the top performing ones in nearly all tracks, including the new Hobbit track. It was awarded the IBM Research prize for the best performing system in all instance matching related tracks.

1 Presentation of the System

1.1 State, Purpose, General Statement

AgreementMakerLight (AML) is an ontology matching system inspired by AgreementMaker [2, 3] but more concerned with efficiency, in order to tackle large-scale matching problems [7]. While it originally focused primarily on the biomedical domain, it has since been expanded to address a broad range of ontology and instance matching problems. AML relies heavily on lexical matching techniques [10], with an emphasis on the use of background knowledge [6], but also includes structural components for both matching and filtering—namely it features a logical repair algorithm [11].

This year, our development of AML centered on the instance matching tasks from the new Hobbit track, and to a lesser degree on the new tasks in the Process Model Matching and Instance Matching tracks.

We maintained the solution of using configuration files we adopted last year, but only for the instance matching tasks, as only for these is the goal of the matching tasks not always inferable from the datasets (e.g., it is generally not possible to infer when the goal is to match only instances of a given type).

1.2 Specific Techniques Used

For the sake of brevity, this section focuses mainly on the features of AML that are new for this edition of the OAEI. For a complete description of AML’s matching strategy,

* Copyright

please refer to the results papers of the last two OAEI editions [4, 5].

1.2.1 AML-Hobbit

The Hobbit track datasets required profound adaptations to AML. First, although the ontology files were included in the training sets, in the Hobbit client only the instances were provided to the matching systems. This meant that the datasets could not be correctly parsed using OWL API [8], and required us to create an N-Triples parser tailored to these datasets (i.e., with the contextual information from the ontology files hard-coded into the parser). Second, the unusual characteristics of the matching tasks, which involve matching traces based on their geographical points, required that we implement dedicated data structures and matching algorithms.

Linking

The Linking task focused on finding equivalent traces by matching their geographical points. The information available for points could include geographical coordinates, address, timestamp, and velocity. The target dataset resulted from a transformation of the source dataset, where some information was omitted and other were altered. Of particular note was the conversion of the geographical coordinates to different coordinate systems. This required us to do the reverse conversion to the decimal system, which we performed during parsing.

The main difficulty of the task was its size, as each trace included on average ≈ 2000 points, and the full task consisted in matching 10000 traces. An efficient matching strategy was therefore paramount.

To enable such a strategy, we adopted a *HashMap*-based data structure with inverted indexes, analogous to AML's other matching structures, but where geographical points were used as keys. To this end, we defined a hash code for points based on the combination of their coordinates. This made it possible to find matching points in $O(1)$ time and therefore match the trace datasets in $O(n)$ time, with n being the total number of points in the ontology with the least points. We used the address and timestamp of the points to filter the matches, and found the velocity to be unnecessary.

Spatial

The Spatial tasks focused on determining whether traces were related according to a number of different topological relations (e.g., contains, crosses, disjoint). In this case, the traces were given as a list of coordinate pairs corresponding to their points, and no transformation of the data was necessary.

To tackle these tasks, we adopted the ESRI Geometry API, which can be used for constructing geometries and performing spatial operations and topological relationship tests on them.

1.2.2 AML-SEALS

Only a few changes were made to AML's matching strategy for the SEALS tracks since the OAEI 2016 edition [5].

Ontology Parser

We made a few changes to AML's ontology parser to cope with typical omissions in instance matching datasets, such as undeclared properties. By default, the OWL API interprets undeclared properties to be annotation properties, which leads to erroneous parsing of the dataset, and hinders AML's performance.

Additionally, we also modified the ontology parser to process OBO logical definitions directly from OWL, as the new versions of the Disease and Phenotype track datasets already included these definitions (last year they did not, and that required us to use external files with the definitions).

Translator

We improved AML's Translator by adding a translation to English of the input ontologies in addition to the reciprocal translation we were already performing. This not only increases the likelihood that a direct match can be found between ontology entities, but also enables the use of WordNet [9].

1.3 Adaptations made for the evaluation

The Hobbit submission of AML is, as a whole, an adaptation made for the evaluation, as the specificities of the Hobbit evaluation (namely the absence of a Tbox) and the tasks (which are almost exclusively based on spatial coordinates) demanded a dedicated submission.

In addition, as in previous years, our SEALS submission included precomputed translations, to circumvent Microsoft® Translator's query limit.

1.4 Link to the system and parameters file

AML is an open source ontology matching system and is available through GitHub: <https://github.com/AgreementMakerLight>.

2 Results

2.1 Anatomy

AML's result in the Anatomy track was the same as last year, with 95% precision, 93.6% recall, 94.4% F-measure, and 83.2% recall++. It remains the best performing system in this track.

2.2 Conference

AML's performance in the Conference track was also the same as last year. It remains the best performing system in this track, with the highest F-measure on the full reference alignment 1 (74%), the full reference alignment 2 (70%), and on both evaluation modalities with the uncertain reference alignment (Discrete: 78%; Continuous: 77%).

Concerning the logical reasoning evaluation, AML had no consistency principle violations, but did have conservativity principle violations as this is an aspect AML deliberately doesn't take into account given that many of these violations were empirically found to be false positives.

2.3 Disease and Phenotype

AML generated 2029 mappings in the HP-MP task, 75 of which were unique. It had the highest F-measure according to the 2-vote silver standard, with 87.2%. In the HP-MeSH task, it generated 5638 mappings of which 678 were unique. It also had the highest F-measure according to the 2-vote silver standard, with 87.1%. In the HP-OMIM task, it generated 6681 mappings of which 679 were unique, and was third in F-measure with 87.8%. In the DOID-ORDO task, it generated the most mappings (4779) and the most unique mappings (1520), and as a result had a relatively low F-measure according to the 2-vote silver standard (66.1%).

2.4 Hobbit

AML produced a perfect result (100% F-measure) in Linking and all Spatial tasks, with the sole exception of the Spatial disjoint mainbox task, where it timed out. In Linking, it had the lowest run time in both the sandbox and mainbox modalities (the other participant timed out in the mainbox task). In Spatial, it had generally the highest run time in the sandbox modalities, but had the lowest run time in the mainbox modality of several tasks, which suggests that it is more scalable than the other participants.

2.5 Instance Matching

In the SPIMBENCH sub-track, AML obtained the second highest F-measure in the sandbox modality (91.8%) and the highest F-measure in the mainbox modality (92.2%). In the Doremus sub-track, AML's results were underwhelming, with only 61.3% F-measure in the Heterogeneities task and 58.2% F-measure in the False Positives Trap task. These tasks were considerably more difficult than the homonym tasks of last year.

2.6 Interactive Matching

AML had an equivalent performance to last year, as we were unable to devote time to address the issues we detected on its user interaction module. In the Anatomy dataset, AML had the highest F-measure (95.8% with 0% errors), the second lowest number of oracle requests, and the lowest impact of errors, with a drop in performance under 3% between 0 and 30% errors. In the Conference dataset, it was second in F-measure with 0% errors, but first when errors were introduced (for all error rates). Despite this, it was more impacted by errors than LogMap, due to the fact that it made considerably more user interactions.

2.7 Large Biomedical Ontologies

AML had the same results as last year in this track, except that the alignment it produced for the SNOMED-NCI whole ontologies tasks had more unsatisfiabilities. This is a consequence of the fact that this year we opted to switch off the use of the ELK reasoner when parsing the ontologies, due to the SPIMBENCH ontologies being inconsistent. Although AML's ontology parser captures most of the subclass and equivalence relationships identified by ELK (which is why there are only differences in this task), it doesn't capture all of them. AML obtained either the highest or the second highest F-measure in all tasks, and had the highest average F-measure overall with 82.7% (ignoring the XMAP results, since this system uses the UMLS metathesaurus as background knowledge, which is the basis of the reference alignments).

2.8 Multifarm

AML improved its results in matching different ontologies, and remains the system with the highest F-measure (46%). However, its performance in matching the same ontologies decreased, and it has only the fourth best F-measure (26%). This decrease was reportedly due to some errors in parsing the alignments for which a confidence higher than 1 was generated, an issue which we will investigate and address.

2.9 Process Model

AML obtained the same result as last year in the University Admission dataset, with 70.2% F-measure. This remains the highest F-measure of all OAEI and PMMC [1] participants. In the new Birth Registration dataset, it obtained the highest F-measure among OAEI participants (42.0%), but would rank only fifth among PMMC participants.

3 General comments

3.1 Comments on the results

AML was the only system to participate in all tracks this year, and was either the best performing or among the top performing systems in nearly all tasks, including the new Hobbit track and the new datasets in the Process Model Matching and Disease and Phenotype tracks. AML was also consistently among the fastest systems and among those that produced the most coherent alignments. As was the case last year, these results reflect our continued effort to extend and improve AML while ensuring that it remains both effective and efficient.

3.2 Comments on the OAEI test cases

While we welcome the efforts of the OAEI organizers to expand it with new datasets, we must comment on some of the issues we encountered during this year's competition, and suggest some possible improvements for future editions.

In the new Hobbit track, even if it is understandable in a new massive venture such

as the Hobbit evaluation platform, the tardiness of the information on the submission process and evaluation datasets hindered participation. More importantly, the fact that Tbox data was unavailable through the platform meant that participating systems had to be trained specifically to interpret the Hobbit Abox data, which we feel violates the spirit of the OAEI.

We were also not fully satisfied with the evaluation of the Disease and Phenotype track. Generating silver standards from the alignments produced by the participating systems via voting is a reasonable starting point for producing a reference alignment, but they should not be used as-is for evaluating matching systems, as the evaluation will be unreliable and superficial. We hope that future efforts focus on improving the evaluation prior to adding more datasets.

4 Conclusion

In 2017, AML was the only system to participate in all tracks, and was among the best performing systems in nearly all tasks (with the sole exception of the Instance Matching DOREMUS sub-track). However, our efforts to participate in the new Hobbit track left little time for making other improvements to AML, and as a result, its performance in most tracks remained the same as last year. That said, our efforts were fully rewarded, as AML was awarded the IBM Research prize for the best performing system in all instance matching related tracks.

Acknowledgments

DF was funded by the EC H2020 grant 676559 ELIXIR-EXCELERATE. CP and FMC were funded by the Portuguese FCT through the LASIGE Strategic Project (UID/CEC/00408/2013). CP was also funded by FCT (PTDC/EEI-ESS/4633/2014). The research of IFC, BSB and VRS was partially funded by NSF awards CNS-1646395, III-1618126, CCF-1331800, and III-1213013, and by a Bill & Melinda Gates Foundation Grand Challenges Explorations grant.

References

1. G. Antunes, M. Bakhshandeh, J. Borbinha, J. Cardoso, S. Dadashnia, C. Francescomarino, M. Dragoni, P. Fettke, A. Gal, C. Ghidini, et al. The process model matching contest 2015. In *6th EMISA Workshop*, pages 127–155, 2015.
2. I. F. Cruz, F. Palandri Antonelli, and C. Stroe. AgreementMaker: Efficient Matching for Large Real-World Schemas and Ontologies. *PVLDB*, 2(2):1586–1589, 2009.
3. I. F. Cruz, C. Stroe, F. Caimi, A. Fabiani, C. Pesquita, F. M. Couto, and M. Palmonari. Using AgreementMaker to Align Ontologies for OAEI 2011. In *ISWC International Workshop on Ontology Matching (OM)*, volume 814 of *CEUR Workshop Proceedings*, pages 114–121, 2011.
4. D. Faria, C. Martins, A. Nanavaty, D. Oliveira, B. S. Balasubramani, A. Taheri, C. Pesquita, F. M. Couto, and I. F. Cruz. AML results for OAEI 2015. In *Ontology Matching Workshop*. CEUR, 2015.

5. D. Faria, C. Pesquita, B. S. Balasubramani, C. Martins, J. Cardoso, H. Curado, F. M. Couto, and I. F. Cruz. OAEI 2016 results of AML. In *Ontology Matching Workshop*. CEUR, 2016.
6. D. Faria, C. Pesquita, E. Santos, I. F. Cruz, and F. M. Couto. Automatic Background Knowledge Selection for Matching Biomedical Ontologies. *PLoS One*, 9(11):e111226, 2014.
7. D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, and F. M. Couto. The Agreement-MakerLight Ontology Matching System. In *OTM Conferences - ODBASE*, pages 527–541, 2013.
8. M. Horridge and S. Bechhofer. The owl api: A java api for owl ontologies. *Semantic Web*, 2(1):11–21, 2011.
9. G. A. Miller. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39–41, 1995.
10. C. Pesquita, D. Faria, C. Stroe, E. Santos, I. F. Cruz, and F. M. Couto. What’s in a ”nym”? Synonyms in Biomedical Ontology Matching. In *International Semantic Web Conference (ISWC)*, pages 526–541, 2013.
11. E. Santos, D. Faria, C. Pesquita, and F. M. Couto. Ontology alignment repair through modularization and confidence-based heuristics. *PLoS ONE*, 10(12):e0144807, 2015.