AML and AMLC results for OAEI 2021

Daniel Faria[1], Beatriz Lima[1], Marta Contreiras Silva[1],
Francisco M. Couto[1], and Catia Pesquita[1]

LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

Abstract. AgreementMakerLight (AML) is an ontology matching system with a scalable and extensible framework that enables it to tackle a variety of ontology matching tasks. For the OAEI 2021, AML's development focused exclusively on expanding its range of complex matching algorithms, which feature in its complex matching version, AMLC. AML remains one of the systems with the broadest coverage of OAEI tracks and with the best overall performance.

1 Presentation of the System

1.1 State, Purpose, General Statement

AgreementMakerLight (AML) is an ontology matching system with a scalable and extensible framework, inspired by AgreementMaker [1, 2, 11], but developed to tackle the efficient matching of very large ontologies [7]. Despite an emphasis on the biomedical domain, AML is a broad purpose ontology matching system capable of tackling problems across the whole spectrum of ontology matching, and of any domain.
AML relies primarily on lexical matching algorithms [9], but also features structural algorithms for both matching and filtering, as well as a logical repair algorithm [10]. It makes extensive use of external sources of background knowledge, featuring an automatic selection algorithm to identify the background knowledge sources that should be used for any given matching task [6].
AMLC is a version of AML developed for complex ontology matching, which remains separate from the main AML codebase and OAEI submission, as tackling complex ontology matching required refactoring most of the AML framework. This year, our work was exclusively focused on furthering the development to AMLC.
As has been the case in recent years, we also participated in the SPIMBENCH and Link Discovery tracks via the HOBBIT platform. In the case of SPIMBENCH, we participated with the HOBBIT adaptation of the main AML codebase. In the case of Link Discovery, we participated with a specialized version of AML, AML-Spatial, due to the unique characteristics of the matching tasks in this track and to the unavailability of the TBox assertions in the HOBBIT datasets.

1.2 Specific Techniques Used

This section describes only the features of AMLC that are new for OAEI 2021. For further information on AML's simple matching strategy, please consult AML's original paper [7] as well as the AML OAEI results publications of 2016-2018 [4, 3, 5].

Our main development this year was the extension of our modular association rule mining (ARM) framework for ontology matching, inspired by Zhou et al. [12], but where the fact that the complex matching patterns are known *a priori* is exploited to steer the ARM process, rather than just to filter the final results. Our framework features a central ARM implementation that selects patterns (i.e., mappings) based on their confidence and support, and a suite of algorithms devoted to finding individual types of patterns and computing their confidence and support from among the set of shared instances in a matching task. While for last year, we had only implemented algorithms for detecting simple class and property mappings, as of the OAEI 2021, AMLC includes algorithms for most types of complex patterns.

1.3 Adaptations Made for the Evaluation

As has been the case in recent OAEI editions, the Link Discovery submission of AML is adapted to these particular tasks and datasets, as their specificities (namely the absence of a Tbox) demand a dedicated submission. The same is also true to some extent of AML's Complex Matching submission.
As usual, our submission included precomputed dictionaries with 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

AML's and AMLC's OAEI 2021 results are summarized in Table 1. Results that were different from OAEI 2020 [8] are discussed in the subsections below.

2.1 Biodiversity and Ecology

AML's results were the same as last year for the ANAEETHES-GEMET task, but on the ENVO-SWEET and AGROVOC-NALT tasks it had a lower precision but higher recall than last year, which we can only ascribe to differences in the datasets, since AML was unchanged. AML ranked first in all three tasks.

Table 1: Summary of AML's and AMLC's OAEI 2021 results.

Task	Precision	Recall/ Coverage	F-measure	Run time (s)	Rank[1]
—— Anatomy ——					
Mouse-Human	0.956	0.927	0.941	29	1
—— Biodiversity & Ecology ——					
ENVO-SWEET	0.745	0.895	0.813	47	1
ANAEETHES-GEMET	0.976	0.764	0.839	21	1
AGROVOC-NALT	0.853	0.904	0.877	139.5	1[a]
—— Common Knowledge Graphs ——					
Mouse-Human	0.0	0.0	0.0	319	8
—— Complex ——					
Conference	0.31	0.37	0.34	-	1[a]
Hydrography	0.49	0.04	0.08	-	3
Geolink	0.49	0.22	0.30	-	2
Populated Geolink	0.49	0.22	0.30	-	2
Populated Enslaved	0.46	0.12	0.18	-	3
—— Conference ——					
OntoFarm (rar2-M3)	0.78	0.62	0.69	-	1
OntoFarm (Discrete)	0.79	0.77	0.78	-	1
OntoFarm (Continuous)	0.80	0.74	0.77	-	1
DBpedia-OntoFarm	0.50	0.73	0.59	-	2
—— Disease & Phenotype ——					
HP-MP	0.911	0.72	0.804	117	3
DOID-ORDO	0.691	0.833	0.755	231	1
—— Interactive Matching ——					
Anatomy (error 0.0)	0.972	0.933	0.952	16	1
Anatomy (error 0.1)	0.961	0.931	0.946	16	1
Anatomy (error 0.2)	0.952	0.927	0.939	16	1
Anatomy (error 0.3)	0.942	0.925	0.933	16	1
Conference (error 0.0)	0.91	0.698	0.79	21	2
Conference (error 0.1)	0.845	0.687	0.758	21	1
Conference (error 0.2)	0.777	0.665	0.717	21	1
Conference (error 0.3)	0.724	0.65	0.685	21	1
—— Knowledge Graph ——					
Aggregate (class)	0.98	0.81	0.89	–	1
Aggregate (property)	0.92	0.57	0.70	–	6
Aggregate (instance)	0.90	0.80	0.85	–	3[b]
Aggregate (all)	0.90	0.80	0.85	3026	3[b]
—— Large Biomedical Ontologies ——					
FMA-NCI small	0.958	0.91	0.933	44	1
FMA-NCI whole	0.806	0.881	0.842	92	1
FMA-SNOMED small	0.923	0.762	0.835	124	1
FMA-SNOMED whole	0.685	0.710	0.697	183	3

SNOMED-NCI small	0.906	0.746	0.818	1026	1
SNOMED-NCI whole	0.862	0.687	0.765	375	1
—— Link Discovery ——					
Spatial (mainbox)	1.0	1.0	1.0	–	1[b]
—— Material Sciences and Engineering ——					
—— Multifarm ——					
MultiFarm	0.72	0.35	0.47	10200	1
—— SPIMBENCH ——					
SPIMBENCH (mainbox)	0.839	0.884	0.860	46517	2

[1]according to F-measure; [a] only system with results; [b] tied with other systems

2.2 Common Knowledge Graphs

In this new track, AML only generated instance mappings as its automatic configuration detected this as an instance matching problem. However, the track is in fact a class matching problem, so the instance mappings generated by AML were not contemplated in the evaluation, and as a result its alignment had no valid mappings and an F-measure of 0.

2.3 Complex Matching

AMLC's results were either identical or slightly worse than those of last year, indicating that, despite our efforts, more work is needed on our complex matching algorithms. Nevertheless, it is worth mentioning that AMLC remains the only system capable of producing complex mappings in some of the complex tasks.

2.4 Conference

AML had the exact same results as in recent years on the OntoFarm suite of tasks, ranking first in all evaluation modalities. It had a higher precision and recall than last year on the DBpedia-OntoFarm suite (ranking second) which we can only ascribe to changes in the dataset.

2.5 Disease and Phenotype

AML had similar results to those of recent years, with minor changes being attributed to the fact that the evaluation is based on a silver standard, as there were no changes on the side of AML.

2.6 Interactive Matching

AML had the same performance as last year in tasks with 0 error rate, as expected. In the tasks with error rate greater than 0, small differences can be observed between the results of this year and last year, due to the stochastic nature of the Oracle's errors that makes each evaluation run unique. The differences between the two years are small because the evaluation is the average of 10 runs.

2.7 Multifarm

AML's results were slightly better than last years', with a 2% increase in F-measure in the different ontologies modality and a 1% increase in the same ontologies modality. These differences are due to correcting a minor configuration problem when using AML's word-matching algorithm in a multilingual setting.

3 Conclusions

In 2021, AML was once again one of the systems that successfully tackled most OAEI tracks and datasets, as well as one of the best performing systems overall.
The fact that AML remains the best performing system in some of the oldest recurring tracks, despite no new developments on our side, suggests that either there has been no interest from the ontology matching community in tackling these challenges, or AML's results are sufficiently close to the best that can be achieved with an automated ontology matching algorithm to make it difficult to surpass them.

With regard to complex matching, AMLC featured an expand suite of algorithms, but nevertheless failed to improve upon last year's results. Clearly more work is needed, in what is the most challenging sub-field of ontology matching, to attain a performance that would enable practical use of AMLC.

Acknowledgments

The authors dedicate this article to the memory of Isabel F. Cruz, whose guidance and encouragement were instrumental to the development of AML and its continued success throughout the years.

This work was supported by FCT through the LASIGE Research Unit (UIDB/00408/2020 and UIDP/00408/2020). It was also partially supported by the KATY project which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101017453.

References

1. 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.
2. 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. CEUR-WS.org, 2011.
3. D. Faria, B. S. Balasubramani, V. R. Shivaprabhu, I. Mott, C. Pesquita, F. M. Couto, and I. F. Cruz. Results of AML in OAEI 2017. In *ISWC International Workshop on Ontology Matching (OM)*, volume 2032 of *CEUR Workshop Proceedings*, pages 122–128. CEUR-WS.org, 2017.

4. 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 *ISWC International Workshop on Ontology Matching (OM)*, volume 1766, pages 138–145. CEUR-WS.org, 2016.
5. D. Faria, C. Pesquita, B. S. Balasubramani, T. Tervo, D. Carriço, R. Garrilha, F. M. Couto, and I. F. Cruz. Results of AML Participation in OAEI 2018. In *ISWC International Workshop on Ontology Matching (OM)*, volume 2288 of *CEUR Workshop Proceedings*, pages 125–131. CEUR-WS.org, 2018.
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 AgreementMakerLight Ontology Matching System. In *OTM Conferences - ODBASE*, pages 527–541, 2013.
8. B. Lima, D. Faria, F. M. Couto, I. F. Cruz, and C. Pesquita. Oaei 2020 results for aml and amlc. In *ISWC International Workshop on Ontology Matching (OM)*, volume 2788 of *CEUR Workshop Proceedings*, 2020.
9. 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.
10. 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.
11. W. Sunna and I. F. Cruz. Structure-Based Methods to Enhance Geospatial Ontology Alignment. In *International Conference on GeoSpatial Semantics (GeoS)*, volume 4853 of *Lecture Notes in Computer Science (LNCS)*, pages 82–97. Springer, 2007.
12. L. Zhou, M. Cheatham, and P. Hitzler. Towards Association Rule-Based Complex Ontology Alignment. In X. Wang, F. A. Lisi, G. Xiao, and E. Botoeva, editors, *Semantic Technology*, pages 287–303, Cham, 2020. Springer International Publishing.