

# AML Results for OAEI 2015

Daniel Faria<sup>1</sup>, Catarina Martins<sup>2</sup>, Amruta Nanavaty<sup>3</sup>  
, Daniela Oliveira<sup>2</sup>, Booma S. Balasubramani<sup>3</sup>, Aynaz Taheri<sup>3</sup>  
, Catia Pesquita<sup>2</sup>, Francisco M. Couto<sup>2</sup>, and Isabel F. Cruz<sup>3</sup>

<sup>1</sup> Instituto Gulbenkian de Ciência, Portugal

<sup>2</sup> LaSIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

<sup>3</sup> ADVIS Lab, Department of Computer Science, University of Illinois at Chicago, USA

**Abstract.** AgreementMakerLight (AML) is an automated ontology matching system based primarily on element-level matching and on the use of external resources as background knowledge. This paper describes its configuration for the OAEI 2015 competition and discusses its results.

For this OAEI edition, we focused mainly on the Interactive Matching track due to its expansion, as handling user interactions on large-scale tasks is a critical challenge in ontology matching.

AML's participation in the OAEI 2015 was successful, as it obtained the highest F-measure in 6 of the 7 ontology matching tracks. Notably, it obtained the highest F-measure in all tasks of the Interactive Matching track while posing less queries to the user than comparable participating systems.

## 1 Presentation of the system

### 1.1 State, purpose, general statement

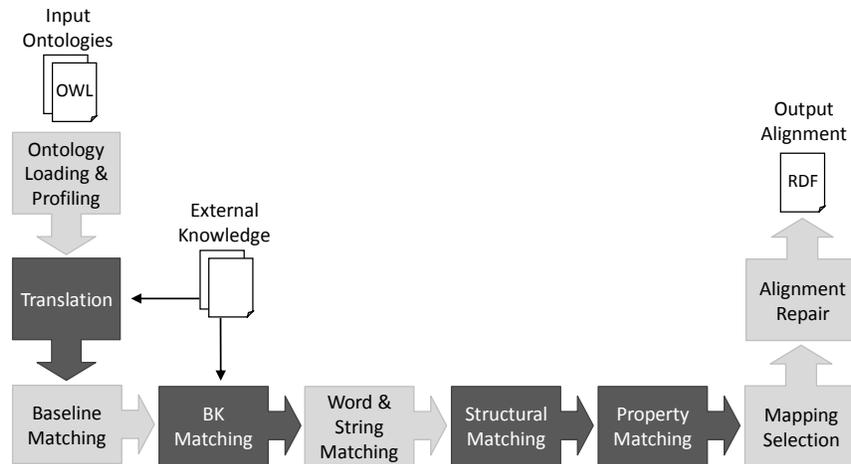
AgreementMakerLight (AML) is an automated ontology matching system based primarily on lexical matching techniques, with an emphasis on the use of external resources as background knowledge and on alignment coherence. While originally focused on the biomedical domain, AML's scope has been expanded, and it can now be considered a general-purpose ontology matching system, as evidenced by its results in last year's OAEI.

AML was derived from AgreementMaker [1, 2] and combines its design principles (flexibility and extensibility) with a strong focus on efficiency and scalability [5]. It draws on the knowledge accumulated in AgreementMaker by reusing and adapting some of its components, but also includes a number of novel components such as an alignment repair module [11] and an automatic background knowledge source selection algorithm [4].

This year, our development of AML for the OAEI competition focused primarily on the Interactive Matching track, due to its expansion to include the Anatomy and Large Biomedical Ontologies datasets. Handling user feedback on large-scale tasks is a critical challenge in ontology matching, and was an aspect in which AML still had room for improvement.

## 1.2 Specific techniques used

The AML workflow for the OAEI 2015 is the same as last year, comprising the nine steps shown in Figure 1: ontology loading and profiling, translation, baseline matching, background knowledge matching, word and string matching, structural matching, property matching, selection, and repair.



**Fig. 1.** The AgreementMakerLight matching workflow for the OAEI 2015.  
Steps in dark gray are conditional.

**Ontology Loading & Profiling** AML employs the OWL API [6] to read the input ontologies and retrieve the necessary information to populate its own data structures [5]:

- Local names, labels and synonym annotations of Classes, Object Properties and Data Properties are normalized and stored into the *Lexicon* of the corresponding ontology. AML automatically derives new synonyms for each name by removing leading and trailing stop words [10], and by removing name sections within parenthesis.
- Domains and ranges of Object and Data Properties are stored in the *Ontology* in *Property* objects.
- Relations between classes (including disjointness) and between properties are stored in a global *RelationshipMap*.
- Cases of implicit disjointness between classes that have incompatible property restrictions in their definition (e.g., different values of a Functional Data Property such as *has mass*) are inferred and made explicit in the *RelationshipMap* as well.

AML does not store or use comments, definitions, or instances.

After loading, the matching problem is profiled taking into account the size of the ontologies, their language(s), and their property/class ratio.

**Translation** AML features an automatic translation module based on Microsoft® Translator, which is called when there is no significant overlap between the language(s) of the input ontologies. AML employs this module to translate the names of all classes and properties from the language(s) of the first ontology to the language(s) of the second and vice-versa. The translation is done by querying Microsoft® Translator for the full name (rather than word-by-word) in order to help provide context. To improve performance, AML employs a cache strategy, by storing locally all translation results in dictionary files, and queries the Translator only when no stored translation is found.

**Baseline Matching** AML employs an efficient, and generally precise, weighted string-equivalence algorithm, the *Lexical Matcher* [5], to obtain a baseline class alignment between the input ontologies.

**Background Knowledge Matching** AML has available four sources of background knowledge which can be used as mediators between the input ontologies: the Uber Anatomy Ontology (Uberon) [8], the Human Disease Ontology (DOID) [12], the Medical Subject Headings (MeSH) [9], and the WordNet [7].

The WordNet is only used for small English language ontologies, as it is prone to produce erroneous mappings in other settings (particularly in domains with specialized vocabularies, such as the Life Science domain). It is used through the JAWS API <sup>1</sup> and with the *Lexical Matcher*. The remaining three background knowledge sources are all specific to the biomedical domain, and thus are tested for all non-small English language ontologies, given that biomedical ontologies are seldom small. They are tested by measuring their mapping gain over the baseline alignment [4]. When the mapping gain is high ( $\geq 20\%$ ), the source is used to extend the *Lexicons* of the input ontologies [10]; otherwise, when it is above the minimum threshold (2%) they are used merely as mediators and their alignment is added to the baseline alignment.

Uberon and DOID are both used in OWL format, and each has an additional table of pre-processed cross-references (in a text file). They can be used directly through the cross-references or with the *Lexical Matcher*. MeSH is used as a stored *Lexicon* file, which was produced by parsing its XML file, and is used only with the *Lexical Matcher*.

**Word & String Matching** To further extend the alignment, AML employs a word-based similarity algorithm (the *Word Matcher*) and a string similarity algorithm (the *Parametric String Matcher*) [5]. The former is not used for very large ontologies, because it is error prone. The latter is used globally for small ontologies, but only locally for larger ones as it is time-consuming.

For small ontologies, AML also employs the *Multi-Word Matcher*, which matches closely related multi-word names that have matching words and/or words with common WordNet synonyms or close hypernyms, and the new *Acronym Matcher*, which attempts to match acronyms to the corresponding full name.

---

<sup>1</sup> <http://lyle.smu.edu/tspell/jaws/>

**Structural Matching** For small and medium-sized ontologies, AML also employs a structural matching algorithm, called *Neighbor Similarity Matcher*, that is analogous to AgreementMaker’s Descendants Similarity Inheritance algorithm [3]. This algorithm computes similarity between two classes by propagating the similarity of their matched ancestors and descendants, using a weighting factor to account for distance.

**Property Matching** When the input ontologies have a high property/class ratio, AML also employs the *PropertyMatcher*. This algorithm first ensures that properties have the same type and corresponding/matching domains and ranges. If they do, it compares the properties’ names by doing a full-name match and computing word similarity, string similarity, and WordNet similarity.

**Selection** AML employs a greedy selection algorithm, the *Ranked Selector* [5], to reduce the cardinality of the alignment. Depending on the size of the input ontologies, one of three selection strategies is used: strict, permissive, or hybrid. In strict selection, no concurrent mappings (i.e., different mappings for the same class/property) are allowed and a strict 1-to-1 alignment is produced; in permissive selection, concurrent mappings are allowed if their similarity score is exactly the same; in hybrid selection, up to two mappings per class are allowed above 75% similarity, and permissive selection is applied below this threshold. For very large ontologies, AML employs a selection variant that consists on combining the (lexical) similarity between the classes with their structural similarity, prior to performing ranked selection. This strategy enables AML to select mappings that “fit in” structurally over those that are outliers but have a high lexical similarity.

In interactive matching mode, AML employs an interactive selection algorithm instead. This algorithm uses patterns in the similarity values produced by AML’s various matching algorithms to detect suspicious mappings. Above the high similarity threshold of 70%, AML queries the user for suspicious mappings, and accepts all other mappings as true. Below this threshold, AML automatically rejects suspicious mappings, and queries the user for all other mappings, until the minimum threshold of 45% is reached, the limit of consecutive negative answers is reached, or the query limit is reached, whichever happens first. The query limit is 45% of the alignment for small ontologies, and 15% of the alignment for all other ontologies (with a further 5% of the alignment reserved for interactive repair). It ensures that the workload for the user is kept within reasonable boundaries.

**Repair** AML employs a heuristic repair algorithm to ensure that the final alignment is coherent [11].

For the interactive matching track, AML employs an interactive variant of this algorithm, wherein the user is asked for feedback about the mappings selected for removal. This variant is not used on the Large Biomedical Ontologies dataset due to its particular evaluation, wherein mappings repaired from the reference alignment are ignored but considered true by the Oracle.

### 1.3 Adaptations made for the evaluation

The only adaptations made for the evaluation were the preprocessing of cross-references from Uberon and DOID for use in the Anatomy and Large Biomedical Ontologies tracks (due to namespace differences), and the precomputing of translations for the Multifarm track (due to 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>) as an Eclipse project, as a stand-alone Jar application, and as a package for running through the SEALS client.

## 2 Results

### 2.1 Anatomy

AML had almost identical results to last year, with an F-measure of 94% and a recall++ of 82%, making it the best performing system in this track this year as well. The only difference from last year’s alignment was one missing mapping due to a change in the structural matching algorithm.

### 2.2 Benchmark

AML had a small improvement in the Biblio Benchmark over last year, from 55% to 57%, likely due to the few refinements made in the processing of properties. However, its performance on the new Energy Benchmark was poor, with a recall of only 2%, and consequently a low F-measure as well (18%). This remains the only OAEI track where AML’s performance is sub-par, mainly due to the fact that involves instances, which AML currently does not read or process in any way.

### 2.3 Conference

AML had the best performance overall in the Conference track, with the highest F-measure on the full reference alignments ra1 and ra2 (74% and 70% respectively). It also had the highest F-measure in the class-only alignments, and the second-highest in the property-only alignments (notably with 100% precision). In comparison with last year, AML improved its F-measure by 3% with regard to ra2, thanks to the addition of the *Acronym Matcher* and to a few refinements in the processing of properties. Concerning the logical reasoning evaluation, AML was one of the five systems that produced alignments without consistency principle violations, and it had an average number of conservativity principle violations of 1.86 which is the sixth lowest overall, and a reasonable figure considering that some of these violations are false positives.

## 2.4 Interactive Matching

AML obtained the highest F-measure in all interactive tasks, with 96.2% in Anatomy (with no error), 81.8% in Conference and an average of 84.5% in LargeBio. AML also had the lowest number of queries among comparable systems in all datasets (i.e., LogMap and ServOMBI, as JarvisOM called upon the Oracle in an active learning approach rather than to filter mapping candidates, which enabled it to make a minimal number of queries in Anatomy, but resulted in it having the worst F-measure as well). It should be noted, however, that AML had the highest non-interactive F-measure on all tracks, so it is unsurprising that it could remain ahead of the other systems while making less queries. Thus, it is important to add that AML also had the highest F-measure-gain-per-query ratio among comparable systems in all datasets (again, excluding JarvisOM), meaning it was more efficient in exploring the user feedback.

With regard to the introduction of Oracle errors, AML was the only system where their impact was linear, with all other systems being impacted superlinearly. The evidence lies in the fact that AML's F-measure was approximately constant when evaluated by the Oracle (i.e., when considering the errors made by the Oracle to be correct) whereas the other systems' F-measures decreased as the error increased. This implies that other systems are drawing inferences from the Oracle's replies, and deciding on the outcome of multiple mappings based on a single query, whereas AML is treating each mapping more or less independently, and thus is less sensitive to the impact of Oracle errors.

## 2.5 Large Biomedical Ontologies

AML's performance in this track was exactly the same as last year, with an average F-measure of 81.9%, as none of the developments made affect this track. As last year, AML had the highest F-measure in each individual task (among valid participants), and thus the highest average F-measure as well. Furthermore, it also had the lowest average degree of unsatisfiabilities, though it was closely followed by LogMap.

## 2.6 Multifarm

AML had an F-measure of 51% when matching different ontologies and of 64% when matching the same ontologies in different languages, both of which were the highest overall by a considerable margin (the next best system in matching different ontologies was LogMap at 41% F-measure, and at matching the same ontologies was CLONA at 58% F-measure). It also had the highest recall overall in both modes, and the second-highest precision. These results are not directly comparable to last year, due to the introduction of the Arabic language ontologies, but running this year's AML on last year's dataset, we observe a marginal improvement in matching different ontologies (by 0.1% F-measure) but a substantial improvement in matching the same ontologies (by 3.3% F-measure). This improvement is mainly due to the refinements made to structural matching algorithm, which naturally have a higher impact on matching different languages of the same ontology, given that the structure will be the same.

## **2.7 Ontology Alignment for Query Answering**

AML had the best performance in this track this year, with an F-measure of 75.9% using the original reference alignment (ra1) and 74.4% using the repaired reference alignment (rar1). It also had the highest precision (tied with XMap on ra1) and recall (tied with LogMap on both ra1 and rar1). These results reflect the fact that AML was the best performing system in the Conference track, and therefore, is naturally the system best positioned to use its Conference alignments for query answering.

## **3 General comments**

### **3.1 Comments on the results**

In comparison with last year, AML improved its performance in 5 tracks: Benchmark (Biblio dataset), Conference, Interactive Matching, Multifarm, and Ontology Alignment for Query Answering. It's performance in the Anatomy and LargeBio tracks was essentially the same as last year. These improvements are tied to developments made in structural matching, property processing and matching, and interactive selection, which reflect the effort put into AML for this year's OAEI.

### **3.2 Discussions on the way to improve the proposed system**

While AML has established itself as a versatile and effective ontology matching system, there is still an important aspect where it is lacking: handling and matching ontology instances.

### **3.3 Comments on the OAEI test cases**

The expansion of the Interactive Matching track to include more challenging test cases and simulate user error was an important improvement to this track and to the OAEI as a whole. Alas, not all was perfect with this year's evaluation, as the Oracle's behaviour on the LargeBio 'soft' repaired reference alignments severely hindered the performance of any interactive repair algorithm, and led to our decision not to employ ours on the LargeBio datasets. We also believe that a query limit should be enforced to ensure that the usage of the Oracle remains within reasonable boundaries, so that systems cannot employ the Oracle to review all their mapping candidates.

## **4 Conclusion**

For this OAEI edition, our goal was to improving AML's interactive selection algorithm and refine its strategy for matching small ontologies. We decided not to make any developments for the biomedical tracks (Anatomy and Large Biomedical Ontologies) as AML's performance was already very good, and we felt that investing further in these tracks would bring a low return on investment.

The results obtained by AML this year have reflected and rewarded our effort, topping

the tables with regard to F-measure in all ontology matching tasks except for Benchmark, with improvements upon last year's performance in the Interactive Matching track and all tracks based on the Conference dataset, while maintaining the performance in Anatomy and Large Biomedical Ontologies.

Thus the OAEI 2015 results highlight the fact that AML is an effective, efficient, and versatile ontology matching system.

## Acknowledgments

FMC, CM, DO and CP were funded by the Portuguese FCT through the LASIGE Strategic Project (UID/CEC/00408/2013). The research of IFC, AN, BS and AT was partially supported by NSF Awards CCF-1331800, IIS-1213013, and IIS-1143926.

## 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, 2011.
3. I. F. Cruz and W. Sunna. Structural alignment methods with applications to geospatial ontologies. *Transactions in GIS*, 12(6):683–711, 2008.
4. 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.
5. 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.
6. M. Horridge and S. Bechhofer. The owl api: A java api for owl ontologies. *Semantic Web*, 2(1):11–21, 2011.
7. G. A. Miller. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39–41, 1995.
8. C. J. Mungall, C. Torniai, G. V. Gkoutos, S. Lewis, and M. A. Haendel. Uberon, an Integrative Multi-species Anatomy Ontology. *Genome Biology*, 13(1):R5, 2012.
9. S. J. Nelson, W. D. Johnston, and B. L. Humphreys. Relationships in medical subject headings (mesh). In *Relationships in the organization of knowledge*, pages 171–184. Springer, 2001.
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. arXiv:1307.5322, 2013.
12. L. M. Schriml, C. Arze, S. Nadendla, Y.-W. W. Chang, M. Mazaitis, V. Felix, G. Feng, and W. A. Kibbe. Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Research*, 40(D1):D940–D946, 2012.