AgreementMakerLight Results for OAEI 2013

Daniel Faria[1], Catia Pesquita[1], Emanuel Santos[1],
Isabel F. Cruz[2], and Francisco M. Couto[1]

[1] LASIGE, Dept Informatics, Faculty of Sciences of the University of Lisbon, Portugal
[2] ADVIS Lab, Dept Computer Science, University of Illinois at Chicago, USA

Abstract. AgreementMakerLight (AML) is an automated ontology matching framework based on element-level matching and the use of external resources as background knowledge. This paper describes the configuration of AML for the OAEI 2013 competition and discusses its results.
Being a newly developed and still incomplete system, our focus in this year's OAEI were the anatomy and large biomedical ontologies tracks, wherein background knowledge plays a critical role. Nevertheless, AML was fairly successful in other tracks as well, showing that in many ontology matching tasks, a lightweight approach based solely on element-level matching can compete with more complex approaches.

1 Presentation of the system

1.1 State, purpose, general statement

AgreementMakerLight (AML) is an automated ontology matching framework derived from the AgreementMaker system [2, 4]. It was developed with the main goal of tackling very large ontology matching problems such as those in the life science domain, which AgreementMaker cannot handle efficiently.

The key design principles of AML were efficiency and simplicity, although flexibility and extensibility—which are key features of AgreementMaker—were also high on the list [5]. Additionally, AML drew upon the knowledge accumulated in AgreementMaker by reusing, adapting, and building upon many of its components. Finally, one of the main paradigms of AML is the use of external resources as background knowledge in ontology matching.

AML is primarily focused on lexically rich ontologies in general and on life sciences ontologies in particular, although it can be adapted to many other ontology matching tasks, thanks to its flexible and extensible framework. However, due to its short development time (eight months), it does not include components for instance matching or translation yet, and thus cannot handle all ontology matching tasks.

1.2 Specific techniques used

The AML workflow for the OAEI 2013 can be divided into six steps, as shown in Fig. 1: ontology loading, baseline matching and profiling, background knowledge matching (optional), extension matching and selection, property matching (conditional), and repair (optional).

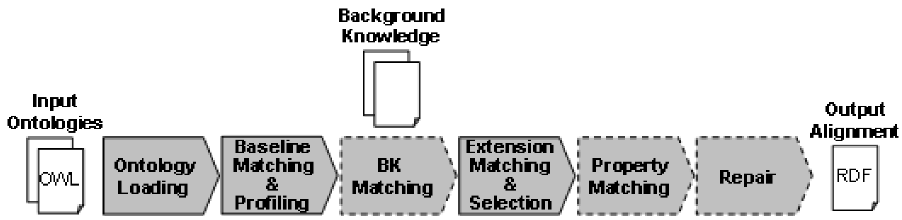

Fig. 1. The AgreementMakerLight Workflow for the OAEI 2013.

Ontology Loading In the ontology loading step, AML reads and processes each of the input ontologies and stores the information necessary for the subsequent steps in its own data structures.
First, AML reads the localName, labels and synonym properties of all classes, normalizes them, and enters them into the *Lexicon* [5] of that ontology. Then, it derives new synonyms for each name in the *Lexicon* by removing leading and trailing stop words [8], and by removing name sections within parenthesis. After class names, AML reads the class-subclass relationships and the disjoint clauses and stores them in the *RelationshipMap* [5]. Finally, AML reads the name, type, domain, and range of each property and stores them in the *PropertyList*.
Note that AML currently does not store or use comments, definitions, or instances.

Baseline Matching and Profiling In the baseline matching and profiling step, AML employs an efficient weighted string-equivalence algorithm, the *Lexical Matcher* [5], to obtain a baseline class alignment between the input ontologies. Then, AML profiles the matching problem by assessing the size (i.e., number of classes) of the input ontologies, the cardinality of the baseline alignment, and the property/class ratio.
Regarding size, AML divides matching problems into three size categories (small, medium or large), which will affect decisions and thresholds during the background knowledge matching and the extension matching and selection steps.
Regarding cardinality, AML also considers three categories (near-one, medium and high), which will determine how selection is performed during the extension matching and selection step.
As for the property/class ratio, it determines whether AML will match properties during the property matching step.

Background Knowledge Matching For the OAEI 2013, AML employs three sources of background knowledge: Uberon [6], UMLS [1] and WordNet [10]. When using background knowledge, AML tests how well each source fits the matching problem by comparing the coverage of its alignment with the coverage of the baseline alignment.
The *Uberon Matcher* uses the Uberon ontology (in OWL) and a table of pre-processed Uberon cross-references (in a text file). Each input ontology is matched both against the Uberon ontology using the *Lexical Matcher* and directly against the cross-reference

table, and AML determines which form of matching is best (giving priority to the cross-references, since they are more reliable). When Uberon is a good fit for the matching problem, it is selected as the only source of background knowledge and is used to extend the *Lexicons* of the input ontologies [8]. When it is a reasonable fit, its alignment is merged with the baseline alignment.
The *UMLS Matcher* uses a pre-processed version of the MRCONSO table from the UMLS Metathesaurus (in a text file). Each input ontology is matched against the whole UMLS table, then AML decides whether to use a single UMLS source (by comparing the coverage of all sources) or the whole table. When UMLS is a good fit for the matching problem, its alignment is used exclusively, and the extension matching and selection step is skipped. Otherwise, if it is a reasonable fit, its alignment is merged with the baseline alignment.
The *WordNet Matcher* queries the WordNet database for synonyms of each name in the *Lexicons* of the input ontologies, using the Jaws API CITATION. These synonyms are used to create temporary extended *Lexicons*, which are matched with the *Lexical Matcher*. Because WordNet is prone to induce errors, AML uses it only to extend the baseline alignment, meaning that it matches only previously unmatched classes.

Extension Matching and Selection The extension matching and selection step comprises two matching sub-steps that alternate with two selection sub-steps. First, AML employs a word-based similarity algorithm, the *Word Matcher* [5], to extend the current alignment globally, followed by a selection algorithm to reduce the alignment to the desired cardinality. Then AML employs the *Parametric String Matcher* [5], which implements the *Isub* string similarity metric [11], to extend the resulting alignment locally (i.e., by matching the children, parents and siblings of already matched class pairs). This is followed by a final selection sub-step.
When the matching problem is profiled as 'large', the *Word Matcher* is skipped because it is too memory intensive to be used globally, and its local use is subsumed by that of the *Parametric String Matcher* [3].
In the interactive matching track, AML employs an interactive selection algorithm, which asks the user for feedback about mappings in case of conflict or below a given similarity threshold, until a given number of negative answers is reached.

Property Matching In the property matching step, AML matches the ontology properties. AML compares the properties' types, domains and ranges, looking for mappings in the class alignment when the domains/ranges are classes. Then, if the properties have attributes in common, AML measures the word-based similarity between their names (as per the *Word Matcher* [5]), employing also WordNet when background knowledge is turned on.

Repair In the repair step, AML employs a heuristic repair algorithm [9] to ensure that the final alignment is coherent with regard to disjoint clauses. The repair algorithm was used by default in all OAEI tracks, except for the Large Biomedical Ontologies track where we ran AML both with and without repair.

1.3 Link to the system and parameters file

The AML system and the alignments it produced for the OAEI 2013 are available at the SOMER project page (http://somer.fc.ul.pt/).

2 Results

2.1 Benchmark

AML had a very high precision (100%) but a fairly low recall (40%) in the Benchmark track, returning empty alignments in several of the tests. This is a consequence of AML's simple framework, which is exclusively based on element-level matching and does not handle instances. Nevertheless, it was interesting to note that AML had the highest F-measure/time ratio, which attests to its efficiency.

2.2 Anatomy

The AML Anatomy results are shown in Table 1. AML ran in this track both with and without background knowledge (AML-BK and AML respectively). In the case of this track, AML-BK selects Uberon exclusively as the source of background knowledge, and uses it for *Lexicon* extension. Thus, the only difference between AML and AML-BK is that the latter has *Lexicons* enriched with Uberon synonyms.

Table 1. AgreementMakerLight results in the Anatomy track.

Configuration	Precision	Recall	F-Measure	Recall+
AML	95.4%	82.7%	88.6%	54.5%
AML-BK	95.4%	92.9%	94.2%	81.7%

The results of AML-BK were very good, with a fairly high precision, and the highest recall, F-measure and recall+ in this year's evaluation. However, the AML results without background knowledge were also good, ranking fourth overall in F-measure, and second if we exclude the systems using Uberon. In fact, we believe that the AML results are near-optimal for a strategy based solely on element-level matching, and that background knowledge is required to obtain substantial improvements. The impact and quality of the Uberon cross-references is clear when we note that AML-BK gained 10% recall over AML without any loss in precision. Finally, it is also noteworthy that AML was one of only two systems to produce coherent alignments.

2.3 Conference

The AML Conference results with reference alignment 1 are shown in Table 2 (the results with reference alignment 2 are slightly worst for all systems, but do not affect their ranking). AML ran in this track with and without background knowledge, with

Table 2. AgreementMakerLight results in the Conference track with reference alignment 1.

Configuration	Precision	Recall	F-Measure
AML	87%	56%	68%
AML-BK	87%	58%	70%

AML-BK using WordNet as the only source of background knowledge (to match both classes and properties).

The results of both AML-BK and AML were good, having the highest precision of this year's evaluation and ranking second and tied for third in terms of F-measure, respectively. An important part of the success of AML in this task was the property matching algorithm, which found 9 and 11 property mappings with 100% precision, with and without background knowledge respectively.

2.4 Multifarm

As we expected, the performance of AML in the Multifarm track was poor, with F-measures of only 4% and 3% when comparing different ontologies and the same ontologies respectively. Participation in this track was beyond our scope, as AML does not handle translations or employ structural-level matching, which are essential for success in this track.

2.5 Library

The results of AML in the Library track were reasonable, as it ranked 4th in terms of F-measure (with 73%) and had the second highest recall of this year's OAEI (87.7%). Nevertheless, there is clearly room for improvement regarding precision, which was significantly lower than that of other top systems (62.5%) likely due to the fact that AML does not take the language of labels into account. Indeed, the results of AML were very similar to the MatcherAllLabels benchmark.

2.6 Interactive Matching

The AML Interactive Matching results are shown in Table 3. AML ran with the same configurations used in the Conference track, except that in this track the selection algorithm employed is interactive, rather than automatic.

Table 3. AgreementMakerLight results in the Interactive Matching track.

Configuration	Precision	Recall	F-Measure	Interactions
AML	91%	60.7%	71.5%	138
AML-BK	91.2%	62.7%	73%	140

The results show that AML's interactive selection algorithm was effective, gaining both precision and recall in comparison with the conference results. Nevertheless, this algorithm is far from optimized, and it should be possible to reduce the number of user interactions without sacrificing F-measure.

2.7 Large Biomedical Ontologies

The AML Large Biomedical Ontologies results are shown in Table 4. AML ran in this track with six different configurations: without background knowledge (AML); with background knowledge (AML-BK); with specialized background knowledge (AML-SBK); and in all three cases with (-R) and without repair. AML-BK selects Uberon in all six tasks of this track (although never for *Lexicon* extension) and selects WordNet only in the SNOMED-NCI small task. AML-SBK is given access to UMLS, and selects it exclusively for all six tasks.
Our goals in testing all these configurations were: to assess the impact of using domain background knowledge both unrelated (Uberon) and directly related (UMLS) to the reference alignments; to assess the effect of using repair on the quality of the results; and to contribute to improve the quality of the reference alignments.

Table 4. Summary AgreementMakerLight results in the Large Biomedical Ontologies track.

Configuration	Precision	Recall	F-Measure	Incoherence
AML	92.6%	68.3%	78.3%	43.1%
AML-R	93.9%	66.6%	77.6%	0.028%
AML-BK	90.8%	70.9%	79.2%	44.2%
AML-BK-R	92.1%	69.2%	78.5%	0.027%
AML-SBK	96.2%	96.1%	96.2%	55%
AML-SBK-R	97.6%	92.5%	95%	0.015%

The results of AML-SBK were very good, with a marked advantage over all other systems in this year's evaluation. This is unsurprising given that AML-SBK derived its alignments from UMLS using an automatic strategy that is likely analogous to that used to build the reference alignments in the first place. This evidently gives AML-SBK an advantage over systems that do not use UMLS. Note, however, that the strategy employed by AML is a general-purpose strategy for reusing preexisting mappings and cross-references, which is used for both UMLS and Uberon. The only issue is that the reference alignments were also automatically derived from UMLS, which makes the evaluation of AML-SBK positively biased.
The results of AML-BK were also good, ranking second overall in recall and F-measure if we exclude the systems that used UMLS. However, in this case the evaluation of AML-BK is negatively biased by the reference alignments. The reason for this is that AML-BK uses Uberon, and many of the mappings derived from Uberon are not present in UMLS despite being correct. This is particularly evident in the FMA-NCI matching problem with whole ontologies, where the contribution of Uberon (based on cross-references which are manually curated) was approximately neutral, decreasing the precision as substantially as it increased the recall (in relation to AML). Perhaps extending the reference alignments by compiling mappings from multiple reliable data sources

such as Uberon could enable a fairer evaluation of the systems competing in this track, and make the tasks less trivial for systems using background knowledge.
The use of repair led to clearly more coherent alignments, as all AML configurations with repair obtained very low degrees of unsatisfiability. However, in terms of quality of the results, the use of repair led to a minor increase in F-measure in some cases, but a substantial decrease in others, and thus had a negative effect overall. This is tied to yet another bias in the reference alignments, caused by the fact that they were automatically repaired [7]. Employing a repair strategy that differs from that used to build the reference alignments can be more penalizing than not doing any repair at all, since for each different decision a repair algorithm makes, it will remove a "correct" mapping and keep an "incorrect" one, whereas without repair we would only have the latter. The problem is that such decisions are essentially arbitrary regarding correctness.

3 General comments

3.1 Comments on the results

On the whole, the results of AML (without background knowledge) were interesting, and show that, for many ontology matching tasks, a lightweight approach based solely on element-level matching can compete with more complex approaches. It is worth highlighting that AML was among the quickest systems in all tracks, and thus had a consistently high F-measure/time ratio in all tracks except for Multifarm. However, the results in the Multifarm track, and to a lesser degree those in the Benchmark track, remind us that AML is still a system in development.
The results of AML-BK (and SBK) show that using suitable background knowledge is critical in specialized domains such as the biomedical, but can be advantageous even for more typical matching problems (such as those in the Conference track).

3.2 Discussions on the way to improve the proposed system

Implementing efficient and effective structural-level matching algorithms will be critical to improve the performance of AML overall. Language handling and translation will also be important to expand the scope of AML, and allow it to tackle tasks such as those in the Multifarm track. Finally, the inclusion of more sources of background knowledge will undoubtedly contribute to improve the performance of AML in tasks beyond the biomedical domain.

4 Conclusion

The participation of AML in the OAEI 2013 was a success overall, with very good results in the Anatomy, Conference, Interactive Matching and Biomedical Ontologies tracks, and reasonable results in the Library track. These results validate the background knowledge paradigm of AML, and demonstrate the effectiveness of a lightweight ontology matching strategy based solely on element-level matching. Nevertheless, it is also

clear from the results that AML is not a complete ontology matching system yet, and that it can benefit from the addition of new tools to its base strategy.
Regarding its namesake, AML was able to build upon the success AgreementMaker had in the Anatomy track in previous OAEI competitions, and was able to transpose this success to the Large Biomedical Ontologies track.

Acknowledgments

DF, CP, ES and FMC were funded by the Portuguese FCT through the SOMER project (PTDC/EIA-EIA/119119/2010) and the multi-annual funding program to LASIGE. CP was also funded by the FLAD-NSF 2013 PORTUGAL-U.S. Research Networks Program through the project "Turning Big Data into Smart Data". The research of IFC was partially supported by NSF Awards IIS-0812258, IIS-1143926, IIS-1213013, and CCF-1331800, by a UIC Area of Excellence Award, and by a IPCE Civic Engagement Research Fund Award.

References

1. O. Bodenreider. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res*, 32(Database issue):267–270, 2004.
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, F. Palandri Antonelli, C. Stroe, U. Keles, and A. Maduko. Using AgreementMaker to Align Ontologies for OAEI 2009: Overview, Results, and Outlook. In *ISWC International Workshop on Ontology Matching (OM)*, volume 551 of *CEUR Workshop Proceedings*, pages 135–146, 2009.
4. 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.
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. 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.
7. C. Pesquita, D. Faria, E. Santos, and F. M. Couto. Using AgreementMaker to Align Ontologies for OAEI 2011. In *ISWC International Workshop on Ontology Matching (OM)*, CEUR Workshop Proceedings, page To appear, 2013.
8. C. Pesquita, C. Stroe, D. Faria, 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)*, page To appear, 2013.
9. E. Santos, D. Faria, C. Pesquita, and F. M. Couto. Ontology alignment repair through modularization and confidence-based heuristics. arXiv:1307.5322, 2013.
10. B. Spell. Java API for WordNet Searching (JAWS). http://lyle.smu.edu/∼tspell/jaws/, 2009.
11. G. Stoilos, G. Stamou, and S. Kollias. A string metric for ontology alignment. In *International Semantic Web Conference (ISWC)*, pages 624–637, 2005.