AgreementMakerLight 2.0: Towards Efficient Large-Scale Ontology Matching

Daniel Faria[1], Catia Pesquita[1,2], Emanuel Santos[2], Isabel F. Cruz[3], and
Francisco M. Couto[1,2]

[1] LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal
[2] Dept. Informática, Faculdade de Ciências, Universidade de Lisboa, Portugal
[3] ADVIS Lab, Dept. of Computer Science, University of Illinois at Chicago, USA

Abstract. Ontology matching is a critical task to realize the Semantic Web vision, by enabling interoperability between ontologies. However, handling large ontologies efficiently is a challenge, given that ontology matching is a problem of quadratic complexity.

AgreementMakerLight (AML) is a scalable automated ontology matching system developed to tackle large ontology matching problems, particularly for the life sciences domain. Its new 2.0 release includes several novel features, including an innovative algorithm for automatic selection of background knowledge sources, and an updated repair algorithm that is both more complete and more efficient.

AML is an open source system, and is available through GitHub [1] both for developers (as an Eclipse project) and end-users (as a runnable Jar with a graphical user interface).

1 Background

Ontology matching is the task of finding correspondences (or mappings) between semantically related concepts of two ontologies, so as to generate an alignment that enables integration and interoperability between those ontologies [2]. It is a critical task to realize the vision of the Semantic Web, and is particularly relevant in the life sciences, given the abundance of biomedical ontologies with partially overlapping domains.

At its base, ontology matching is a problem of quadratic complexity as it entails comparing all concepts of one ontology with all concepts of the other. Early ontology matching systems were not overly concerned with scalability, as the matching problems they tackled were relatively small. But with the increasing interest in matching large (biomedical) ontologies, scalability became a critical aspect, and as a result, traditional all-versus-all ontology matching strategies are giving way to more efficient anchor-based strategies (which have linear time complexity).

---

[1] https://github.com/AgreementMakerLight

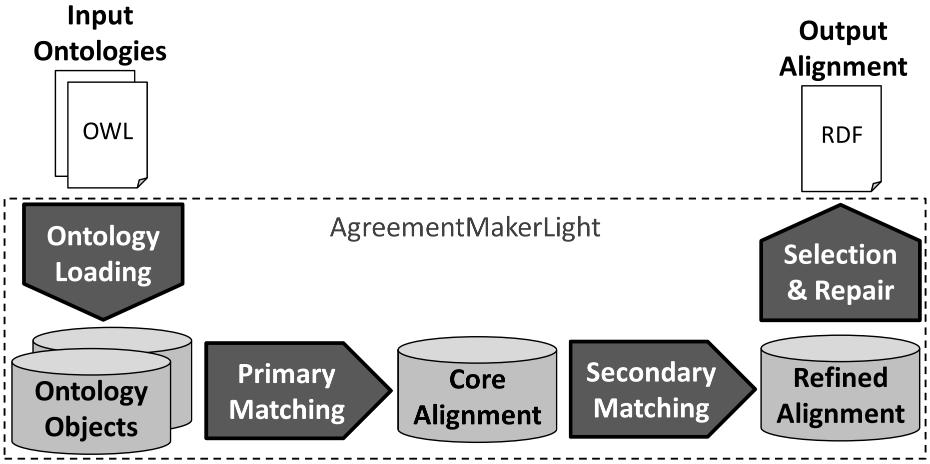


Fig. 1. AgreementMakerLight ontology matching framework.

2 The AgreementMakerLight System

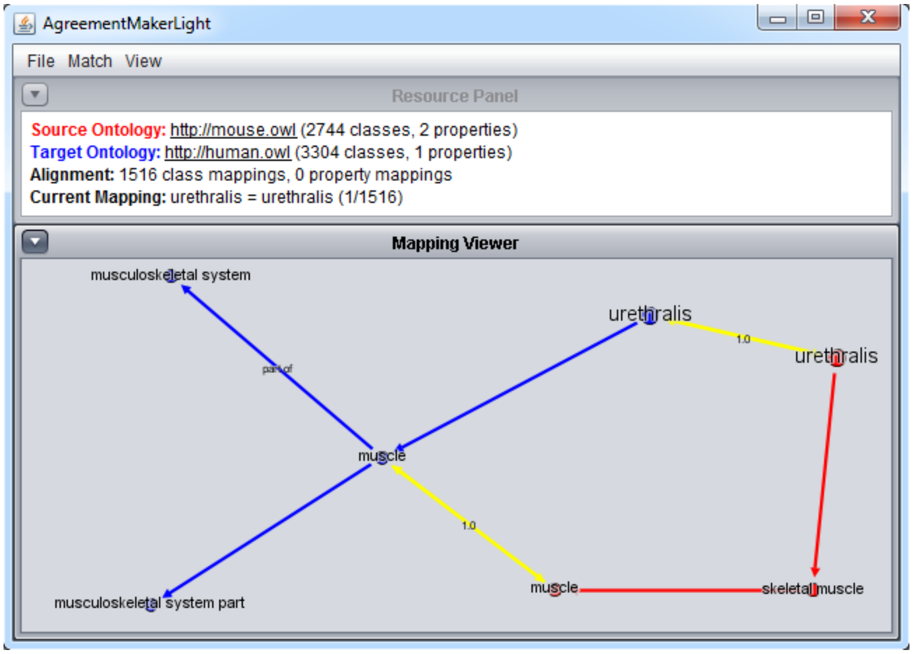
AgreementMakerLight (AML) is a scalable automated ontology matching system developed to tackle large ontology matching problems, and focused in particular on the biomedical domain. It is derived from AgreementMaker, one of the leading first generation ontology matching systems [1], and adds scalability and efficiency to the design principles of flexibility and extensibility which characterized its namesake.

2.1 Ontology Matching Framework

The AML ontology matching framework is represented in Figure 1. It is divided into four main modules: ontology loading, primary matching, secondary matching, and alignment selection and repair

The ontology loading module is responsible for reading ontologies and parsing their information into the AML ontology data structures, which were conceived to enable anchor-based matching [5]. AML 2.0 marks the switch from the Jena2 ontology API to the more efficient and flexible OWL API, and includes several upgrades to the ontology data structures. The most important data structure AML uses for matching is the *Lexicon*, a table of class names and synonyms in an ontology, which uses a ranking system to weight them and score their matches [7].

The primary and secondary matching modules contain AML's ontology matching algorithms, or matchers, with the difference between them being their time complexity. Primary matchers have $O(n)$ time complexity and therefore can be



Source Ontology: http://mouse.owl (2744 classes, 2 properties)
Target Ontology: http://human.owl (3304 classes, 1 properties)
Alignment: 1516 class mappings, 0 property mappings
Current Mapping: urethralis = urethralis (1/1516)

Fig. 2. AgreementMakerLight graphical user interface.

employed globally in all matching problems, whereas secondary matchers have $O(n^2)$ time complexity and thus can only be applied locally in large problems. The use of background knowledge in primary matchers is a key feature in AML, and it includes an innovative automated background knowledge selection algorithm [4].

The alignment selection and repair module ensures that the final alignment has the desired cardinality and that it is coherent (i.e., does not lead to the violation of restrictions of the ontologies) which is important for several applications. AML's approximate alignment repair algorithm features a modularization step which identifies the minimal set of classes that need to be analyzed for coherence, thus greatly reducing the scale of the repair problem [8].

2.2 User Interface

The GUI was a recent addition to AML, as we sought to make our system available to a wider range of users. The main challenge in designing the GUI was finding a way to visualize an alignment between ontologies that was both scalable and useful for the user. Our solution was to visualize only the neighborhood of one mapping at a time, while providing several options for navigating through the alignment [6]. The result is a simple and easy to use GUI which is shown in Figure 2.

3 Performance

AML 1.0 achieved top results in the 2013 edition of the Ontology Alignment Evaluation Initiative (OAEI) [3]. Namely, it ranked first in F-measure in the anatomy track, and second in the large biomedical ontologies, conference and interactive matching tracks. In addition to its effectiveness in matching life sciences ontologies, AML was amongst the fastest systems in all tracks, and more importantly, had consistently a high F-measure/run time ratio.
AML 2.0 is more effective than its predecessor (thanks to the improved handling of background knowledge, the richer data structures and the addition of new matching algorithms) without sacrificing efficiency, so we expect it to perform even better in this year's edition of the OAEI.

Acknowledgments

DF, CP, ES and FMC were funded by the Portuguese FCT through the SOMER project (PTDC/EIA-EIA/119119/2010) and the LASIGE Strategic Project (PEst-OE/EEI/UI0408/2014). The research of IFC was partially supported by NSF Awards CCF–1331800, IIS–1213013, IIS–1143926, and IIS–0812258 and by a UIC-IPCE Civic Engagement Research Fund Award.

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. J. Euzenat and P. Shvaiko. *Ontology Matching.* Springer-Verlag New York Inc, 2007.
3. D. Faria, C. Pesquita, E. Santos, I. F. Cruz, and F. M. Couto. AgreementMakerLight Results for OAEI 2013. In *ISWC International Workshop on Ontology Matching (OM)*, 2013.
4. D. Faria, C. Pesquita, E. Santos, I. F. Cruz, and F. M. Couto. Automatic Background Knowledge Selection for Matching Biomedical Ontologies. *PLoS One*, In Press, 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*, volume 8185 of *LNCS*, pages 527–541, 2013.
6. C. Pesquita, D. Faria, E. Santos, and F. M. Couto. Towards visualizing the alignment of large biomedical ontologies. In *10th International Conference on Data Integration in the Life Sciences*, 2014.
7. 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 *The Semantic Web - ISWC 2013*, volume 8218 of *Lecture Notes in Computer Science*, pages 526–541. Springer Berlin Heidelberg, 2013.
8. E. Santos, D. Faria, C. Pesquita, and F. M. Couto. Ontology alignment repair through modularization and confidence-based heuristics. *CoRR*, arXiv:1307.5322, 2013.