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 match-
ing system developed to tackle large ontology matching problems, par-
ticularly 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). In this demo, we will be demonstrating
AML both from the developer and the end-user perspective, using ontol-
ogy matching tasks from the Ontology Alignment Evaluation Initiative
and ontologies collected from BioPortal as examples.

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
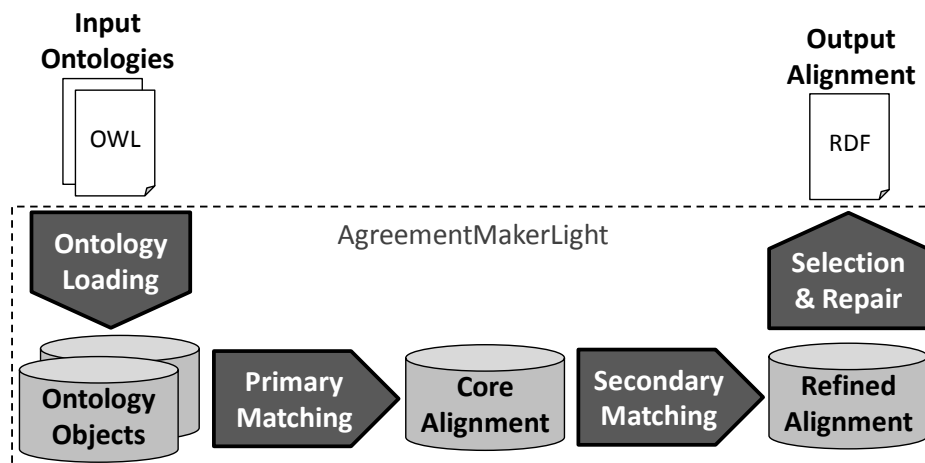complexity).

---

[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 [3]. 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 [5].

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 a linear time complexity matchers and therefore
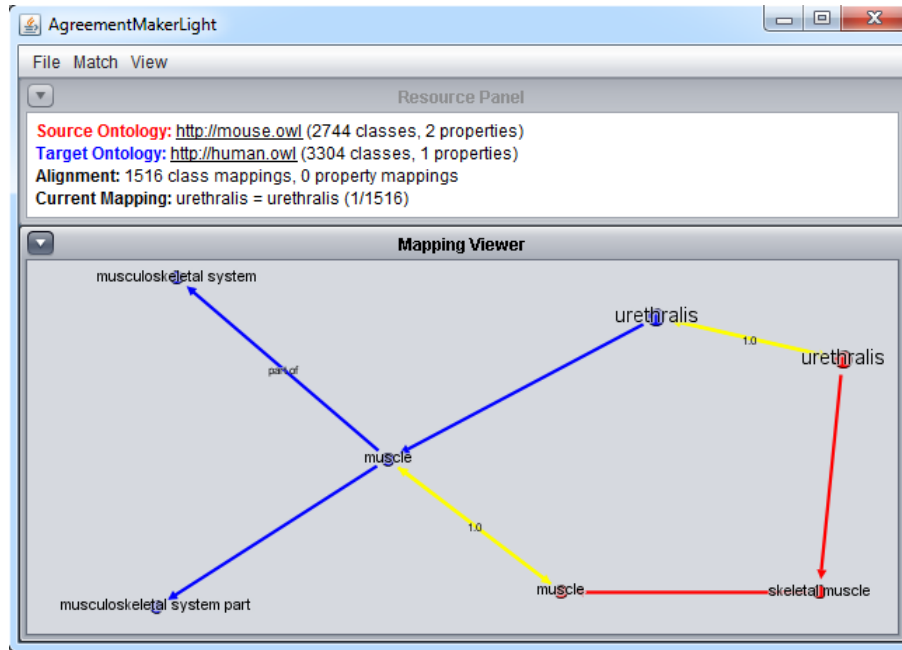
Fig. 2. AgreementMakerLight graphical user interface.

can be employed globally in all matching problems, whereas secondary matchers have polynomial time complexity and thus can only be applied locally on 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.

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 [6].

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 [4]. The result is a simple and easy to use GUI which is shown in Figure 2.

3 Demonstration

We will demonstrate AML both from the developer and the end-user perspective, focusing on one or the other according to the interest of the audience. From the developer perspective, we will navigate through the AML project, and make and run custom matching configurations. From the end-user perspective, we will showcase the AML graphical user interface, and employ it both to visualize precomputed alignments and to perform live ontology matching. We will use medium-sized ontology matching tasks (i.e., with ontologies up to 10,000 classes) from the Ontology Alignment Evaluation Initiative and/or from BioPortal as examples, also depending on the interest of the audience. We will not tackle larger ontology matching tasks due to time constraints, as these take several minutes to run.

Acknowledgments

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, 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.
4. 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.
5. 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.
6. E. Santos, D. Faria, C. Pesquita, and F. M. Couto. Ontology alignment repair through modularization and confidence-based heuristics. *CoRR*, arXiv:1307.5322, 2013.