Using AgreementMaker to Align Ontologies for OAEI 2011*

Isabel F. Cruz, Cosmin Stroe, Federico Caimi, Alessio Fabiani, Catia Pesquita, Francisco M. Couto, Matteo Palmonari

ADVIS Lab, Department of Computer Science, University of Illinois at Chicago
{ifc|cstroe1|fcaimi|fabiani|pmatteo}@cs.uic.edu

Facultade de Ciencias da Universitade de Lisboa, Portugal
cpesquita@xldb.di.fc.ul.pt, fcouto@di.fc.ul.pt

University of Milan-Bicocca, Italy
matteo.palmonari@disco.unimib.it

Abstract. The AgreementMaker system is unique in that it features a powerful user interface, a flexible and extensible architecture, an integrated evaluation engine that relies on inherent quality measures, and semi-automatic and automatic methods. This paper describes the participation of AgreementMaker in the 2011 OAEI competition in four tracks: benchmarks, anatomy, conference, and instance matching. After its successful participation in 2009 and 2010, the goal in this year's participation is to explore previously unused features of the ontologies in order to improve the matching results. Furthermore, the system should be able to automatically adapt to the matching task, choosing the best configuration for the given pair of ontologies. We believe that this year we have made considerable progress in both of these areas.

1 Presentation of the system

We have been developing the AgreementMaker system since 2001, with a focus on real-world applications [5,9] and in particular on geospatial applications [4,6,8,10,11,12,13,15]. However, the current version of AgreementMaker, whose development started in 2008, represents a whole new effort. The code base has more than doubled since then, with the AgreementMaker framework being expanded to accomodate many types of ontology matching techniques.

* Partially supported by NSF Awards IIS–0513553 and IIS-0812258 and by the Intelligence Advanced Research Projects Activity (IARPA) via Air Force Research Laboratory (AFRL) contract number FA8650-10-C-7061. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, AFRL, or the U.S. Government.

1.1 Purpose and state of the system

The AgreementMaker system [1,2,3,7] is an extensible ontology matching framework that has been expanded to include many types of matching algorithms in order to handle many different matching scenarios. At the heart of the system is its ability to efficiently combine the results from several matching algorithms into one single and better result [2]. This capability allows us to focus on developing new matching algorithms and later combine them with our existing approach in order to improve our results.

2 Schema Matching Techniques Introduced in OAEI 2011

As compared to previous years, we have introduced several matching techniques in order to improve our matching algorithms.

2.1 Automatic Configuration Selection via Ontology Profiling Metrics

The AgreementMaker system can be run with different configurations that optimize the system accuracy and coverage depending on the specific ontologies to be aligned. Changing the composition of the matcher stack (e.g. an instance matcher is used only when instances are available) has a high impact on the system performance. We developed an approach to adaptively optimize the configuration of AgreementMaker depending on the ontologies to be aligned.

The approach we adopted can be sketched as follows: the ontologies to be aligned are profiled using several metrics proposed in the literature (e.g. relationship richness, inheritance richness, WorldNet coverage and so on [16]). The metric-based profiles are used to automatically classify the matching task into a configuration class with specific settings. The classification is based on a supervised machine learning framework trained with a subset of the OEAI dataset for which a reference alignment is available.

Our learning framework is very flexible: we can use many combinations of matchers and parameters, various types of classifiers (KStar, Naive Bayes, Multilayer Perceptron etc.) and new metrics. The experimental results show that the use of the automatic configuration methods improved the overall performance of AgreementMaker in the competition. In particular, in this paper we show the importance of this method for the significant improvements we achieved in the Benchmark and Conference tracks. The new AgreementMaker 's matching process follows the steps represented in Figure 1: the ontology pair to be matched is classified by the ontology-profiling algorithm; based on the classification, a run configuration is created, and an ontology matching algorithm is instantiated to create an alignment.

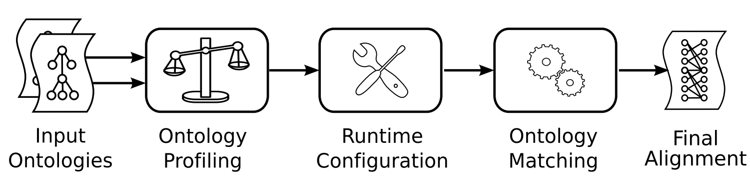

Fig. 1. AgreementMaker OAEI2011 Automatic configuration selection.

2.2 Lexicon Expansion via a Mediating Ontology

One approach to matching two domain specific ontologies is to use a third ontology from the same domain as a mediating ontology, with the mediating ontology to provide missing information relevant to the matching task. Shown in Figure 2, the source and target ontologies, O_S and O_T respectively, are first matched with the mediating ontology O_M. Mappings between the source and target ontologies are then created based on the distance between the concepts in the mediating ontology to which they have been mapped previously.

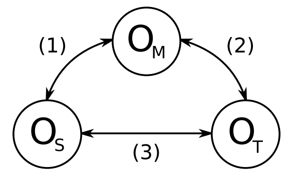

Fig. 2. Using a mediating ontology.

For the specific problem of matching the Mouse Anatomy ontology to the Human Anatomy ontology a successful approach has been to use the UBERON cross species anatomy ontology as a mediating ontology [14]. We have adapted this approach to our lexicon framework, using the BSM_{lex} to match the MA and HA ontologies with UBERON thereby making use of the extra synonyms defined in UBERON.

2.3 Extension of Synonyms

This strategy relies on synonyms defined in the OWL ontology itself, currently via the `hasRelatedSynonym` property, to generate a lexicon of *synonym terms* (single or multi-word terms). This is done by finding common terms between ontology synonyms to infer synonyms terms.

For example, in the Human Anatomy ontology, the concept NCI_C12275 ("Maxillary_Sinus") has the synonyms "Antrum, Maxillary" and "Sinus, Maxillary". Our algorithm infers that "sinus" and "antrum" are synonyms as well without any external reference. These synonym terms are then used to create novel synonyms, by interchanging terms in existing synonyms and labels with their synonymous term.

2.4 Alternate Hierarchy Support

In addition to the subclass hierarchy defined as part of the OWL ontologies of the Anatomy track, there is also a "part of" hierarchy defined using the `UNDEFINED_part_of` property. Taking into account this hierarchy in the VMM_{lex} increases the percision and recall of the matching algorithm.

3 Results

In this section, we present the results obtained by AgreementMaker in the OAEI 2011 competition. It participated in four tracks: benchmarks, anatomy, conference, and instance matching. Tests were carried out on a PC running Ubuntu Linux 10.04 with AMD Athlon™ II X4 635 processor running at 2.9 Ghz and 8 GB RAM.

3.1 Link to the system and parameters file

The AgreementMaker system is available at `http://agreementmaker.org/`. The matching algorithm we used is implemented in the "OAEI 2011 Matcher" algorithm, in the "Hybrid" category. The alignment results obtained by AgreementMaker in the OAEI 2010 are available at `http://agreementmaker.org/oaei`.

3.2 Benchmarks Track

Benchmarks Track Results In this track, a source ontology is compared to 111 ontologies that describe the same domain which can be divided into 3 categories: concept tests cases (1xx cases), systematic tests cases (2xx cases), and real ontology test cases (3xx cases). The 2xx benchmarks test cases are subdivided into 3 groups: 1) 201 to 210, 2) 221 to 247 and 3) 248 to 266. As shown in the Table 1, our results are very good in all the tracks, due to the use of a combination of properties (lexical, structural, syntactic, etc.) to match the ontologies. We are able to perform well even if only one of these properties is available for comparison.

	101-104	201-210	221-247	248-266	301-304	H-Mean
Precision	1.00	1.00	0.98	0.96	0.90	0.97
Recall	1.00	0.95	0.99	0.66	0.85	0.87
F-Measure	1.00	0.97	0.99	0.76	0.87	0.91

Table 1. Results of AgreementMaker in the Benchmarks track of the OAEI 2011 competition.

	101-104	201-210	221-247	248-266	301-304	H-Mean
Precision 2010	0.98	0.97	0.95	0.96	0.88	0.95
Precision 2011	1.00	1.00	0.98	0.96	0.90	0.97
Recall 2010	1.00	0.90	0.99	0.74	0.53	0.79
Recall 2011	1.00	0.95	0.99	0.66	0.85	0.87
F-Measure 2010	0.99	0.94	0.97	0.82	0.61	0.84
F-Measure 2011	1.00	0.97	0.99	0.76	0.87	0.91

Table 2. Comparison of the results in the 2010 and 2011 OAEI Benchmarks track.

Benchmarks Track Comments Although we improved our results with respect last year in almost all the tasks, we are particularly satisfied of our improvements on the third group (301-304). This is because the goal of our system and of the matching methods we are using is to improve the performance on real ontology test cases. A detailed comparison between the results achieved in the 2010 and 2011 competitions is shown in Table 2. One of the main reasons for our improvements is the new automatic configuration method we introduced. In fact, the matching tasks of the Benchmark track are very diverse in order to test several aspects of automatic matching methods; therefore, when the user has to manually select only one configuration, she selects the configuration that obtains the best average results on the whole set of matching tasks, but such a configuration cannot be assumed to be the best one for each individual task. Instead, thanks to the new automatic configuration method, the system automatically select the best configuration for each individual matching task.

3.3 Anatomy Track

Anatomy Track Results This track consists of two real world ontologies to be matched, the source ontology describing the Adult Mouse Anatomy (with 2744 classes) and the target ontology is the NCI Thesaurus describing the Human Anatomy (with 3304 classes). This year, the reference alignment is available for this track, which allowed us to have instant evaluation of our improvements, greatly reducing our development time. As shown in Table 3, we have been able to consistently make improvements to our matching algorithms. A large part of these improvements has been increasing the recall by leveraging external sources, including WordNet and other anatomy ontologies. We have also been able to improve precision by managing a finer grained control of our combination algorithms.

Anatomy Track	Runtime	Precision	Recall	F-Measure
2009	≈23 min	86.5%	79.8%	83.1%
2010	≈5 min	90.3%	85.3%	87.7%
2011[1]	≈7 min	95.4%	88.4%	91.8%

Table 3. Comparison of previous results with this year's results.

Anatomy Track Comments This year we have been able to further increase precision and recall by using the UBERON multi-species anatomy ontology as a mediating ontology, an approach demonstrated by others at the International Conference for Biomedical Ontology [14], by extending our lexicon synonyms using *synonym terms*, and by using the part-of hierarchy in our matching algorithms . Improvment of our algorithms capability to correctly discern relevant concept information allowed us to increase precision by over 5% and was achieved by combining more similar matching algorithms first and using those combined results for the final combination.

3.4 Conference Track

	Manual Config. 2010	Manual Config. 2011	Automatic Config. 2011
Precision	0.53	0.83	0.71
Recall	0.62	0.45	0.62
F-Measure	0.58	0.56	0.64

Table 4. Results achieved by AgreementMaker in the 2011 OAEI conference track.

Conference Results The conference track consists of 15 ontologies from the conference organization domain and each ontology must be matched against every other ontology. We ran our algorithms on all the matching tasks and evaluated precision, recall and F-measure using the 21 provided reference alignments. We then computed an average of these measures, summarized in Table 4. Precision, recall and F-measure obtained using the automatic configuration method introduced in section 2.1 are compared with the results achieved with the manual configuration of the system used in OAEI 2010 and in OAEI 2011. AgreementMaker significantly improved on F-measure by optimizing the trade-off between precision and recall.

Conference Track Comments Some of the matching algorithms that we used in the OAEI 2010 competition underwent minor changes and improvements this year. As a consequence, when we manually defined a configuration of the system, we were able to achieve a significant gain in precision in this particular track but at the cost of a lower recall (which explains why F-measure decreases in the Conference 2011 track when a specific configuration is manually selected). However, the capability to automatically configure the system depending on the input ontologies, allows to achieve a very good trade-off between precision and recall, with a sizable gain of 6% on the average F-measure with respect to the best results we obtained last year.

4 Instance Matching

Differently from our 2009 and 2010 participations, we also entered the Instance Matching track at OAEI 2011. While our matchers were previously developed specifically for working at the schema level, we adapted our system to deal with instances. We decided to focus on the Data Interlinking sub-track, which consists of recreating the links from New York Times data to Freebase, DBPedia, and GeoNames.

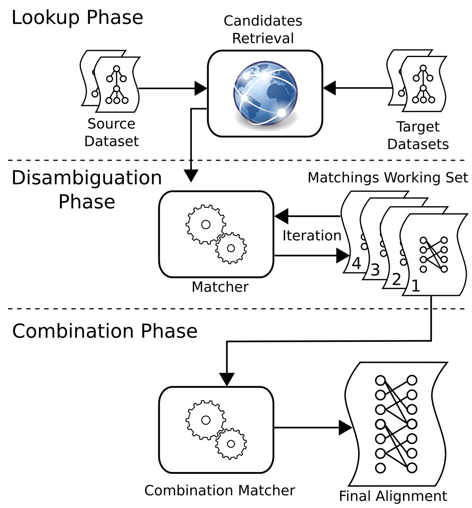

Fig. 3. Instance matching configuration.

We found this track particularly interesting and challenging, since it entails the following new problems: *a*) datasets are very large and not easy to wholly retrieve and work with, *b*) endpoints and APIs have to be queried online in order to get up-to-date information, *c*) these services provide data in different formats, and *d*) the source datasets (New York Times) do not have a schema associated with them so we cannot rely on traditional ontology matching to create schema level mappings.

All of the tasks of this track of the competition are characterized by the presence of a large amount of data. Therefore every instance in the source cannot be compared with every other in the target. We faced the problem of deciding how to reduce the number of comparisons, trying to minimize the loss in recall. Our solution consists of doing a lookup using the label of the instance, the type (when provided), and querying against an index which returns a reasonable number of candidate target instances.

We think this choice is very appropriate for several reasons: *a*) many SPARQL endpoints and APIs implement indexing which permits to get very fast answers to keyword lookups, *b*) the online version of these Knowledge Bases is always richer and more up to date with respect to downloadable versions, and *c*) we can query multiple Knowledge Bases at the same time in a parallel fashion.

Once we query the online service and obtain the results, we compute a similarity between the source instance and the candidate instances. These values are then used to rank the candidates and eventually select the best one. Reusing some of the techniques we have implemented for the other ontology matching tracks, we use different matchers

Source	Target	Types	Precision	Recall	F-Measure
NYT	Freebase	People	0.966	0.950	0.958
NYT	Freebase	Locations	0.884	0.811	0.846
NYT	Freebase	Organizations	0.873	0.735	0.798
NYT	GeoNames	Locations	0.902	0.797	0.846
NYT	DBPedia	People	0.977	0.801	0.881
NYT	DBPedia	Locations	0.790	0.612	0.690
NYT	DBPedia	Organizations	0.840	0.667	0.744
Average			0.890	0.768	0.823
Harmonic Mean			0.886	0.754	0.815

Table 5. Results of AgreementMaker in the Instance Matching track.

that compare several features about the instances, and then combine their outputs in order to give a final alignment.

The main features we use for comparison are: *a*) labels using a substring similarity, *b*) comments and other literals using a Vector Space Model approach, *c*) RDF Statements considering property-value pairs, and *d*) the score values returned by the lookup services (e.g. Freebase API, Apache Lucene score).

4.1 Instance Matching Results

The Data Interlinking sub-track of the Instance Matching at OAEI 2011 competition is composed of seven tasks. The source dataset is always the New York Times, while there are 3 different targets: Freebase, GeoNames, and DBPedia. The results obtained by AgreementMaker in the Instance Matching track are summarized in Table 5, showing precision, recall and F-measure for every matching task.

4.2 Comments about the Instance Matching results

The results are very good and encouraging. Both the average and the H-mean F-measure are over 81%. Our system performs slightly better in Freebase than in DBPedia matching, because the lookup service of the former returns fewer and more precise candidates. Therefore, the disambiguation task is easier when working with Freebase data. On GeoNames the result is very good thanks to the use of some shared properties (*geo:long*, *geo:lat*) between the datasets.

5 Conclusions

In this paper we presented the results of the AgreementMaker system for aligning ontologies in the OAEI 2011 competition in the four tracks in which it participated: benchmarks, anatomy, conference, and the data interlinking sub-track. We believe that while our results are very good already, ongoing research can lead to further improvements. The tracks of the OAEI have always been a challenge and have been a relevant measure of quality among matching systems. In order to uphold this standard, we believe that current matching tasks should be expanded to encompass the changing nature of the ontologies being used on the Semantic Web. More specifically, matching large ontologies (more than 50,000 concepts) and focusing on more linked open datasets are important directions to explore in the near future.

References

1. Isabel F. Cruz, Flavio Palandri Antonelli, and Cosmin Stroe. AgreementMaker: Efficient Matching for Large Real-World Schemas and Ontologies. *PVLDB*, 2(2):1586–1589, 2009.
2. Isabel F. Cruz, Flavio Palandri Antonelli, and Cosmin Stroe. Efficient Selection of Mappings and Automatic Quality-driven Combination of Matching Methods. In *ISWC International Workshop on Ontology Matching*. CEUR-WS, 2009.
3. Isabel F. Cruz, Flavio Palandri Antonelli, and Cosmin Stroe. Integrated Ontology Matching and Evaluation. In *International Semantic Web Conference (Posters & Demos)*, 2009.
4. Isabel F. Cruz and Afsheen Rajendran. Exploring a New Approach to the Alignment of Ontologies. In *ISWC Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data*, volume 83 of *CEUR-WS*, pages 7–12, 2003.
5. Isabel F. Cruz and Afsheen Rajendran. Semantic Data Integration in Hierarchical Domains. *IEEE Intelligent Systems*, March-April:66–73, 2003.
6. Isabel F. Cruz, Afsheen Rajendran, William Sunna, and Nancy Wiegand. Handling Semantic Heterogeneities Using Declarative Agreements. In *ACM Symposium on Advances in Geographic Information Systems (ACM GIS)*, pages 168–174, 2002.
7. Isabel F. Cruz, Cosmin Stroe, Michele Caci, Federico Caimi, Matteo Palmonari, Flavio Palandri Antonelli, and Ulas C. Keles. Using AgreementMaker to Align Ontologies for OAEI 2010. In *Proceedings of the International Conference on Biomedical Ontology*, volume 689. CEUR-WS, 2010.
8. Isabel F. Cruz and William Sunna. Structural Alignment Methods with Applications to Geospatial Ontologies. *Transactions in GIS, special issue on Semantic Similarity Measurement and Geospatial Applications*, 12(6):683–711, 2008.
9. Isabel F. Cruz, William Sunna, and Anjli Chaudhry. Ontology Alignment for Real-World Applications. In *National Conference on Digital Government Research (dg.o)*, pages 393–394, 2004.
10. Isabel F. Cruz, William Sunna, and Anjli Chaudhry. Semi-Automatic Ontology Alignment for Geospatial Data Integration. In *International Conference on Geographic Information Science (GIScience)*, volume 3234 of *Lecture Notes in Computer Science*, pages 51–66. Springer, 2004.
11. Isabel F. Cruz, William Sunna, Nalin Makar, and Sujan Bathala. A Visual Tool for Ontology Alignment to Enable Geospatial Interoperability. *Journal of Visual Languages and Computing*, 18(3):230–254, 2007.
12. Isabel F. Cruz, William G. Sunna, and Kalyan Ayloo. Concept Level Matching of Geospatial Ontologies. In *GIS Planet International Conference and Exhibition on Geographic Information*, 2005.
13. Isabel F. Cruz and Huiyong Xiao. Data Integration for Querying Geospatial Sources. In John Sample, Kevin Shaw, Shengru Tu, and Mahdi Abdelguerfi, editors, *Geospatial Services and Applications for the Internet*, pages 113–137. Springer, 2008.
14. Anika Gross, Michael Hartung, Toralf Kirsten, and Erhard Rahm. Mapping Composition for Matching Large Life Science Ontologies. In *Proceedings of the International Conference on Biomedical Ontology*, pages 109–116, 2011.
15. William Sunna and Isabel 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*, pages 82–97. Springer, 2007.
16. Samir Tartir, I. Budak Arpinar, Michael Moore, Amit P. Sheth, and Boanerges Aleman-Meza. OntoQA: Metric-Based Ontology Quality Analysis. In *IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*, volume 9, 2005.