Using AgreementMaker to Align Ontologies for OAEI 2010*

Isabel F. Cruz, Cosmin Stroe, Michele Caci, Federico Caimi, Matteo Palmonari**, Flavio Palandri Antonelli, Ulas C. Keles

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

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 2010 OAEI competition in three tracks: benchmarks, anatomy, and conference. After its successful participation in 2009, where it ranked first in the conference track, second in the anatomy track, and obtained good results in the benchmarks track, the goal in this year's participation is to increase the values of precision, recall, and F-measure for each of those tracks.

1 Presentation of the system

We have been developing the AgreementMaker system since 2001, with a focus on real-world applications [5, 8] and in particular on geospatial applications [4, 6, 7, 9–13]. However, the current version of AgreementMaker, whose development started two years ago, represents a whole new effort.

1.1 State, purpose, general statement

The new AgreementMaker system [1–3] supports: (1) user requirements, as expressed by domain experts; (2) a wide range of input (ontology) and output (agreement file) formats; (3) a large choice of matching methods depending, on the different granularity of the set of components being matched (local vs. global), on different features considered in the comparison (conceptual vs. structural), on the amount of intervention that they require from users (manual vs. automatic), on usage (standalone vs. composed), and on the types of components to consider (schema only or schema and instances); (4) improved performance, that is, accuracy (precision, recall, F-measure) and efficiency (execution time) for the automatic methods; (5) an extensible architecture to incorporate new methods easily and to tune their performance; (6) the capability to evaluate, compare, and combine different strategies and matching results; (7) a comprehensive user interface that supports advanced visualization techniques and a control panel that

* Research supported by NSF Awards IIS-0513553 and IIS-0812258.
** Additional affiliation: University of Milan-Bicocca, Italy.

drives all the matching methods and evaluation strategies; (8) a feedback loop that accepts suggestions and corrections by users and extrapolates new mappings.

In 2009 AgreementMaker was very successful in the OAEI competition. In particular, AgreementMaker ranked (a close) second among ten systems in the anatomy track. AgreementMaker also participated successfully in two other tracks: benchmarks and conference. In the former track, AgreementMaker was ranked first in terms of precision and seventh in terms of recall among thirteen systems and in the latter track AgreementMaker was ranked first with the highest F-measure (57% at a threshold of 75%) among seven competing systems.

1.2 Specific techniques used

AgreementMaker comprises several matching algorithms or *matchers* that can be used for matching (or aligning) the source and target ontologies. The matchers are not restricted to any particular domain. The architecture of AgreementMaker relies on a stack of matchers that belong to three different layers (see Figure 1). Specific configurations of the stack have been used for the benchmarks, anatomy, and conference tracks, as discussed in what follows. However, we describe first the different components in the stack: the matchers, the combination and evaluation modules, and the final alignment module.

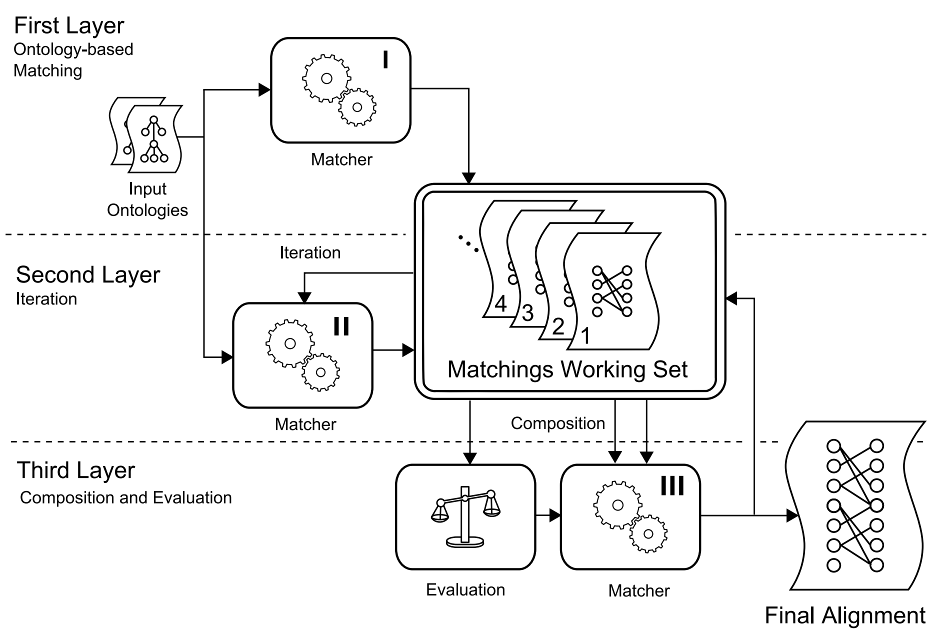

Fig. 1. AgreementMaker OAEI 2010 matcher stack.

Matchers can be concept-based (if they consider only one concept) or structural (if they consider a subgraph of the ontology). The concept-based matchers support the comparison of strings. They include: the Base Similarity Matcher (BSM) [7], the Parametric String-based Matcher (PSM) [2] and the Vector-based Multi-Word Matcher (VMM) [2]. BSM is a basic string matcher that computes the similarity between concepts by comparing all the strings associated with them. PSM is a more in-depth string matcher, which for the competition is set to use a substring measure and an edit distance measure. VMM compiles a *virtual document* for every concept of an ontology, transforms the resulting strings into TF-IDF vectors and then computes their similarity using the cosine similarity measure. These matchers have been extended in the AgreementMaker configuration used this year by plugging in a set of lexicons, which are used to expand the set of strings with synonyms. The extended matchers are therefore called BSM^{lex}, PSM^{lex}, and VMM^{lex}. The Advanced Similarity Matcher (ASM) is a string-based matcher that computes mappings between source and target concepts (including their properties) by comparing their local names, and providing better similarity evaluation in particular when compound terms are used. ASM outperforms generic string-based similarity matchers because it is based on a deeper linguistic analysis.

Structural matchers include the Descendants' Similarity Inheritance (DSI) matcher [7]. This matcher is based on the idea that if two nodes are similar, then their descendants should be similar. The Group Finder Matcher (GFM) is another structural matcher that filters out the mappings provided by another matcher (the input matcher). It identifies groups of concepts and properties in the ontologies and assumes that two concepts (or properties) that belong to two groups that were not mapped by the input matcher will likely have different meanings and should not be mapped. The Iterative Instance Structural Matcher (IISM) takes into account instances. Classes that have mapped individuals can then be aligned. In addition, values of the properties are also considered. The structural part of IISM is quite complex and takes into account superclasses, subclasses, properties, subproperties, cardinalities, and the range and domain of properties.

The combination and evaluation modules are used together, as follows. The Linear Weighted Combination (LWC) [2] combines its inputs (e.g., from several string matchers), using a local confidence quality measure provided by the evaluation module, in order to automatically assign weights to each result computed by the input matchers. After this step, we have a single combined set of alignments that includes the best alignments from each of the input matchers. The final alignment module is given as input a mapping cardinality (e.g., 1:1) and a threshold and outputs the best set of alignments given those two inputs [2].

Benchmarks For the benchmarks track we used the following configuration:

$$\text{IISM}(\ \text{LWC}(\text{ASM}, \text{PSM}^{lex}, \text{VMM}^{lex}, \text{BSM}^{lex})\ )$$

LWC is adopted to combine the results of four string-based matchers, namely ASM, PSM^{lex}, PSM^{lex}, and BSM^{lex}; the last three make use of two lexicons, namely WordNet and a dictionary built from the ontologies; the similarity values computed at this step are then given as input to the IISM structural matcher.

Anatomy For the anatomy track we used the following configuration:

$$\text{LWC}(\text{PSM}^{lex}, \text{VMM}^{lex}, \text{BSM}^{lex})$$

LWC is adopted to combine the results of four string-based matchers, namely PSM^{lex}, VMM^{lex}, and BSM^{lex}; the last three make use of two lexicons, namely WordNet and a dictionary built from the ontologies.

Conference For the conference track we used the following configuration:

$$\text{GFM}(\ \text{LWC}(\text{ASM}, \text{PSM})\ )$$

LWC is adopted to combine the results of two string-based matchers, namely ASM and PSM; the similarity values computed at this step are then given as input to the GFM structural matcher.

1.3 Link to the system and parameters file

The AgreementMaker system is available at http://agreementmaker.org/.

1.4 Link to the set of provided alignment (in align format)

The alignment results obtained by AgreementMaker in the OAEI 2010 are available at http://agreementmaker.org/oaei.

2 Results

In this section, we present the results obtained by AgreementMaker in the OAEI 2010 competition. It participated in three tracks: benchmarks, anatomy, and conference. 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.

2.1 Benchmarks

In this track, a source ontology is compared to 111 ontologies that describe the same domain. These ontologies can be divided into 3 categories: concept tests cases (1xx cases), systematic tests cases (2xx cases), and real ontology test cases (3xx cases). AgreementMaker employs the algorithm which we described in section 1.2 for aligning two ontologies.

The 2xx benchmarks test cases are subdivided into 3 groups: 1) 201 to 210, 2) 221 to 247 and 3) 248 to 266. The lexical information in the ontologies in group 1 have been altered to change their labels or identifiers. This alteration includes replacing the labels or identifiers with other names that follow a particular naming convention, a random name, a misspelled name or a foreign word. However, the structure of the ontologies is not modified. The test cases in the second group have ontologies that have flattened hierarchies, expanded hierarchies or no hierarchies at all. The test cases in the third group are the most challenging ones to align. This is because the labels have been scrambled such that they comprise a permutation of letters of a particular length. In addition, the structure of the ontology has been flattened, expanded such that it has more depth or removed completely.

The results obtained by AgreementMaker in the benchmarks track are summarized in Table 1.

	101-104	201-210	221-247	248-266	301-304	H-mean
precision	0.98	0.97	0.95	0.96	0.88	0.95
recall	1.00	0.90	0.99	0.74	0.53	0.79
F-measure	0.99	0.94	0.97	0.82	0.61	0.84

Table 1. Results achieved by AgreementMaker in the benchmarks track of the OAEI 2010 competition.

2.2 Anatomy

This track consists of two real world ontologies to be matched. The source ontology describes the Adult Mouse Anatomy (with 2744 classes) while the target ontology is the NCI Thesaurus describing the Human Anatomy (with 3304 classes). Matching these ontologies is also challenging in terms of efficiency because these ontologies are relatively large. The anatomy track consists of four subtracks: subtrack 1, which emphasizes F-measure, subtrack 2, which emphasizes precision, subtrack 3, which emphasizes recall, and subtrack 4, which tests the capability of extending a partial reference alignment.

The results obtained by AgreementMaker in the anatomy track are summarized in Table 2. We show the precision, recall, and F-measure for subtrack 1, 2 and 3; subtracks 1 and 3 are also evaluated by considering the recall+ measure, which measures how many non trivial correct correspondences, not detectable by string equivalence, can be found in an alignment. Evaluation of subtrack 4 is carried out by analyzing the changes in precision, recall and f-measure when subtrack 1 is compared with subtrack 4.

Anatomy Track	Subtrack 1	Subtrack 2	Subtrack 3	Subtrack 4
precision	0.90	0.96	0.77	+0.03
recall	0.85	0.75	0.87	-0.03
F-measure	0.87	0.84	0.82	+0.00
recall+	0.63	-	0.70	-

Table 2. Results achieved by AgreementMaker in the anatomy track of the OAEI 2010 competition.

2.3 Conference

The conference track consists of 15 ontologies from the conference organization domain and each ontology must be matched against every other ontology. Since the AgreementMaker OAEI 2010 matcher stack considers only two ontologies at a time, we compute 120 alignment files, in total containing 2070 individual alignments. The results obtained are summarized in Table 3. Here we show how precision, recall, and F-measure vary depending on the threshold used for the selection of the mappings.

Threshold	0.60-0.64	0.65-0.69	0.70-0.74	0.75-0.79	0.80-0.84	0.85-0.89	0.90-0.94	0.95-1.00
precision	0.49	0.53	0.61	0.70	0.72	0.75	0.81	0.81
recall	0.64	0.60	0.54	0.52	0.47	0.47	0.46	0.46
F-measure	0.56	0.56	0.58	0.60	0.58	0.58	0.58	0.59

Table 3. Results achieved in the conference track of the OAEI 2010 competition.

2.4 Comments on the results

Benchmarks In the OAEI 2009 competition, AgreementMaker was first in terms of the precision of discovered mappings. However, in terms of recall, AgreementMaker was outperformed by six other systems (thirteen systems competed). The new matchers used in the OAEI 2010 competition address specifically the issue of the alignment of concepts that are not lexically similar. The results of this effort increased the recall by 18% at a cost of 3% in precision in comparison with last year's results. An important contribution to this result comes from the IISM matcher, which exploits instances and structural properties of the ontologies and makes the alignment process less sensitive to lexical differences. A detailed comparison between the results achieved in the 2009 and 2010 competitions in terms of the obtained change in precision, recall, and F-measure for each group of test cases, and the overall H-mean is shown in Table 4.

	101-104	201-210	221-247	248-266	301-304	H-mean
precision 09	0.86	0.73	0.76	0.47	0.83	0.70
precision 10	0.98	0.97	0.95	0.96	0.88	0.95
recall 09	0.86	0.73	0.76	0.47	0.86	0.70
recall 10	1.00	0.90	0.99	0.74	0.53	0.79
F-measure 09	0.92	0.71	0.86	0.45	0.83	0.70
F-measure 10	0.99	0.94	0.97	0.82	0.61	0.84

Table 4. Comparison of the results achieved by AgreementMaker in the 2009 and 2010 OAEI benchmarks track.

Anatomy In comparison with the results achieved by AgreementMaker in the OAEI 2009 competition, the experimental results obtained this year show that the system significantly improved with respect to precision, recall, F-measure, and recall+. A major contribution to these results comes from the exploitation of lexical resources to improve string-based and vector-based matchers. A comparison between the results achieved in the two competitions in terms of precision, recall, F-measure and recall+ for subtracks 1, 2 and 3 is shown in Table 5 (except for recall+ that is not evaluated on subtrack 2). Remarkably, our algorithms for retrieving non trivial mappings significantly improved, as shown by the gain of 0.15 in recall+. Instead, we do not present the comparison with the results obtained in 2009 on subtrack 4, because this year we did not exploit any specific algorithm for propagating mappings available in the input alignment.

Anatomy Track	Subtrack 1	Subtrack 2	Subtrack 3
precision 09	0.87	0.97	0.51
precision 10	0.90	0.96	0.77
recall 09	0.80	0.68	0.82
recall 10	0.85	0.75	0.87
F-measure 09	0.83	0.80	0.63
F-measure 10	0.87	0.84	0.82
recall+ 09	0.49	-	0.55
recall+ 10	0.63	-	0.70

Table 5. Comparison of the results achieved by AgreementMaker in the 2009 and 2010 OAEI anatomy track.

Conference In comparison with the results achieved in OAEI 2009, AgreementMaker significantly improved on precision, recall, and F-measure for thresholds above 0.75 as shown in the graph represented of Figure 2.4, providing more stable results. Remarkably, the new matchers used for the conference track, namely ASM and GFM, can be used on real-world ontologies, since they are based on generic lexical and structural features. Moreover, ASM can be easily adapted to different string-based similarity metrics, and can be extended by adopting a lexicon.

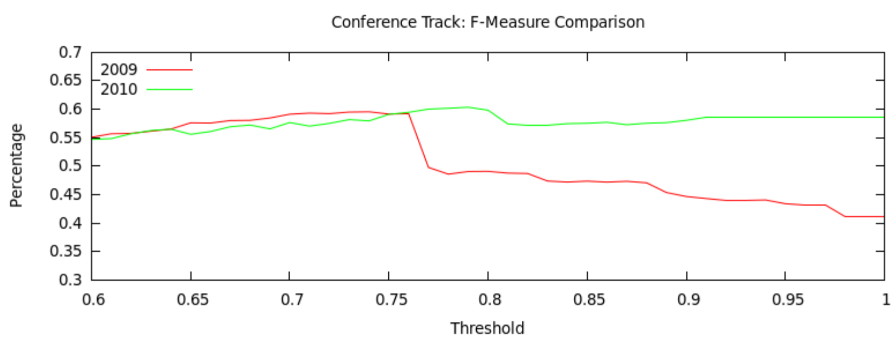

Fig. 2. F-measure comparison for the 2009 and 2010 OAEI conference track results.

3 Conclusions

In this paper we presented the results of the AgreementMaker system for aligning ontologies in the OAEI 2010 competition in the three tracks in which it participated: benchmarks, anatomy, and conference. It was our goal to improve on the results obtained by AgreementMaker in 2009. To meet this goal, we developed several new matching methods, which could be readily integrated into the AgreementMaker system because of its modularity and extensibility. Our results (which we compare with last year's results) amply demonstrate that we have met our goal.

Acknowledgements

Thanks to Catia Pesquita and Francisco M. Couto from Facultade de Ciencias da Universitade de Lisboa, Portugal, for the feedback on the AgreementMaker system, and for useful insights on biomedical ontologies, which significantly increased our understanding of the problem and our results in the anatomy track.

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. pages 168–174, 2002.
7. 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.
8. 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.
9. Isabel F. Cruz, William Sunna, and Anjli Chaudhry. Semi-Automatic Ontology Alignment for Geospatial Data Integration. volume 3234 of *Lecture Notes in Computer Science*, pages 51–66. Springer, 2004.
10. 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.
11. Isabel F. Cruz, William G. Sunna, and Kalyan Ayloo. Concept Level Matching of Geospatial Ontologies. 2005.
12. 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.
13. 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.