OMReasoner: Using Reasoner for Ontology Matching : results for OAEI 2011

Guohua Shen, Lantao Jin, Ziyue Zhao, Zhe Jia, Wenmin He, Zhiqiu Huang

Nanjing University of Aeronautics and Astronautics, Nanjing, China
{ghshen,ltjin,zyzhao,zjia,wmhe,zqhuang}@nuaa.edu.cn

Abstract. Ontology matching produces correspondences between entities of two ontologies. The **OMReasoner** is unique in that it creates an extensible framework for combination of multiple individual matchers, and reasons about ontology matching by using description logic reasoner. It handles ontology matching in semantic level and makes full use of the semantic part of OWL-DL instead of structure. This paper describes the result of **OMReasoner** in the 2011 OAEI competition in two tracks: benchmark and conference.

1 Presentation of the system

Ontology matching finds correspondences between semantically related entities of the ontologies. It plays a key role in many application domains.

Many approaches to ontology matching have been proposed: The implementation of match may use multiple match algorithms or matchers, and the following largely-orthogonal classification criteria are considered [1-3]: schema-level and instance-level, element-level and structure-level, syntactic and semantic, language-based and constraint-based.

Most approaches focus on syntactic aspects instead of semantic ones. OMReasoner achieves the matching by means of reasoning techniques. Still, this approach includes strategy of combination of (mainly syntactical) multi-matchers (e.g., EditDistance matcher, Prefix/Suffix matcher, WordNet matcher) before match reasoning.

1.1 State, purpose, general statement

The matching process can be viewed as a function f (see Fig.1).

$$A' = f(O1, O2, A, p, r)$$

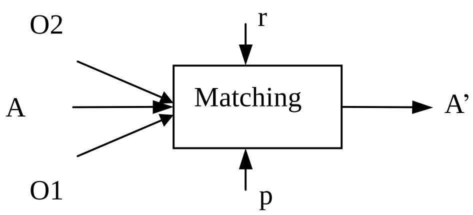

Fig. 1 The ontology matching process

Where *O1* and *O2* are a pair of ontologies as input to match, *A* is the input alignment between these ontologies and *A*' is new alignment returned, ***p*** is a set of parameters (e.g., weight *w* and threshold τ) and ***r*** is a set of oracles and resources.

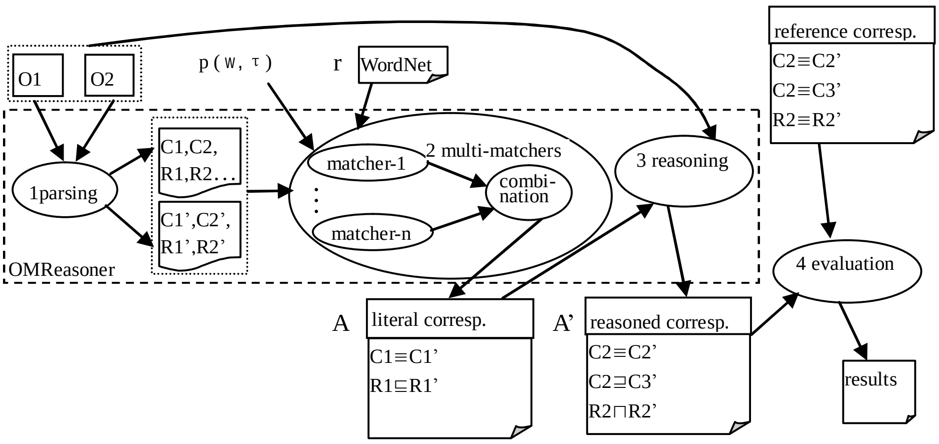

Fig 2. Ontology matching in OMReasoner

The OMReasoner achieved ontology alignment as following four steps (see Fig.2):

1. Parsing: we can achieve the classes and properties of ontologies by using ontology API: Jena.
2. Combination of multiple individual matchers: the literal correspondences (e.g. equivalence) can be produced by using multiple match algorithms or matchers, for example, string similarity measure (prefix, suffix, edit distance) by string-based, constrained-based techniques. Also, some semantic correspondences can be achieved by using some external dictionary: WordNet. Then the multiple match results can be combined by weighted summarizing method. The framework of multi-matchers combination is supported, which facilitates inclusion of new individual matchers.
3. Reasoning: the further semantic correspondences can be deduced by using DL reasoner, which uses literal correspondences produced in step 2 as input.
4. Result evaluation: the evaluation measures will be precision and recall computed against the reference alignments.

1.2 Specific techniques used

OMReasoner includes summarizing algorithm to combine the multiple match results. The combination can be summarized over the n weighted similarity methods (see formula 1), where w_k is the weight for a specific method, and $sim_k(e1,e2)$ is the similarity evaluation by the method.

$$sim(e1,e2) = \sum_{k-1}^{n} w_k sim_k(e1,e2) \tag{1}$$

OMReasoner uses semantic matching methods like WordNet matcher and description logic (DL) reasoning.

WordNet[1] is an electronic lexical database for English, where various senses (possible meanings of a word or expression) of words are put together into sets of synonyms. Relations between ontology entities can be computed in terms of bindings between WordNet senses. This individual matcher uses an external dictionary: WordNet to achieve semantic correspondences.

Another important matcher uses edit distance, which is a measure of the similarity between two words. Based on this value, we calculate the morphology analogous degree by using some math formula.

All the results of each individual matcher will be normalized before combination.

OMReasoner employs DL reasoner provided by Jena. OMReasoner includes external rules to reason about the ontology matching.

2 Results: a comment for each dataset performed

There are 21 alignment tasks in benchmark data set and 21 alignment tasks in conference data set. We test the data sets with OMReasoner and present the results in Table 1, Table 2, Fig 3 and Fig 4. The average measures (precision, recall and F-measure) of Benchmark are 0.359, 0.754 and 0.473 respectively. The average measures of Conference are 0.136, 0.801 and 0.220 respectively. In conclusion, the precision, recall and F-measure are not satisfying, however, it is our first endeavor for ontology matching and we will improve it in the future.

2.1 Benchmark

We evaluated the results against reference alignments, and obtained precision var-ies from 0.071 to 0.636, and recall varies from 0.254 to 1.0, F-measure varies from 0.121 to 0.745.

Label	Dataset	Prec.	Rec	f-Measure
B1	101-101	0.299	0.361	0.327
B2	101-103	0.308	0.722	0.431
B3	101-104	0.369	0.732	0.491
B4	101-224	0.338	0.907	0.493

[1] http://wordnet.princeton.edu/

B5	101-225	0.288	0.887	0.435
B6	101-228	0.308	0.970	0.467
B7	101-230	0.442	0.972	0.608
B8	101-232	0.420	0.918	0.576
B9	101-233	0.438	0.970	0.603
B10	101-236	0.338	1.000	0.506
B11	101-237	0.204	0.591	0.304
B12	101-238	0.186	0.732	0.296
B13	101-239	0.621	0.931	0.745
B14	101-240	0.636	0.758	0.692
B15	101-241	0.606	0.848	0.707
B16	101-246	0.586	0.931	0.719
B17	101-247	0.636	0.848	0.727
B18	101-301	0.085	0.254	0.127
B19	101-302	0.071	0.405	0.121
B20	101-303	0.146	0.417	0.216
B21	101-304	0.224	0.671	0.336

Table.1. Match results in the Benchmark track

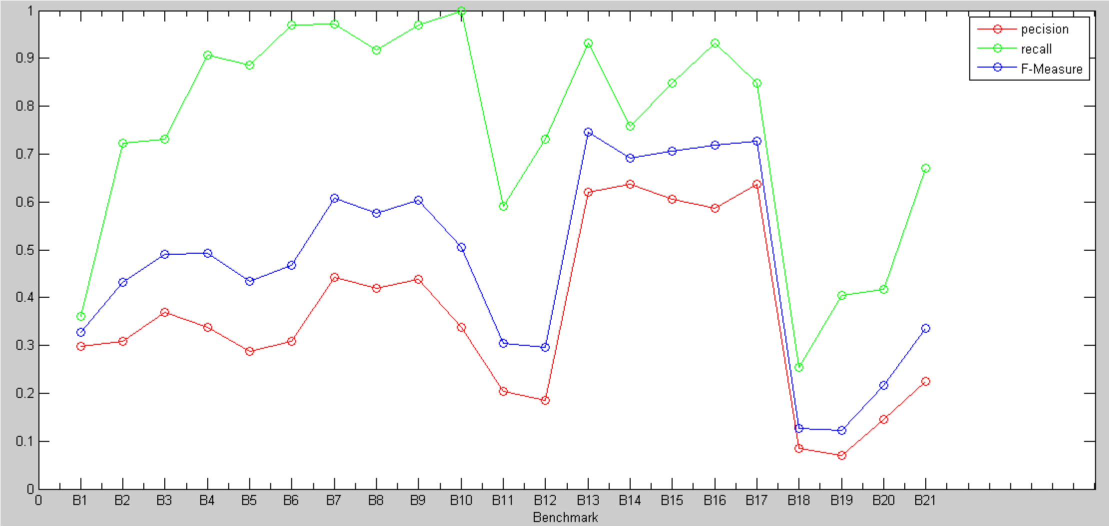

Fig.3. Comparison of match results in Benchmark

2.2 Conference

We evaluated the results against reference alignments, and obtained precision varies from 0.030 to 0.409, and recall varies from 0.526 to 1.0, F-measure varies from 0.058 to 0.495.

Label	Dataset	Prec	Rec	f-Measure
C1	iasted-sigkdd	0.072	0.800	0.132

C2	ekaw-iasted	0.030	0.800	0.058
C3	ekaw-sigkdd	0.033	0.933	0.064
C4	edas-ekaw	0.119	0.652	0.201
C5	edas-iasted	0.088	0.706	0.156
C6	edas-sigkdd	0.120	0.800	0.209
C7	confOf-edas	0.189	0.526	0.278
C8	confOf-ekaw	0.125	0.947	0.221
C9	confOf-iasted	0.071	0.800	0.131
C10	confOf-sigkdd	0.114	0.857	0.202
C11	cmt-Conference	0.158	0.667	0.255
C12	cmt-confOf	0.409	0.625	0.495
C13	cmt-edas	0.162	0.692	0.263
C14	cmt-ekaw	0.159	0.667	0.257
C15	cmt-iasted	0.069	1.000	0.129
C16	cmt-sigkdd	0.297	0.917	0.449
C17	conference-confOf	0.182	0.933	0.304
C18	conference-edas	0.126	0.706	0.215
C19	conference-ekaw	0.142	1.000	0.249
C20	conference-iasted	0.056	0.929	0.106
C21	conference-sigkdd	0.143	0.867	0.245

Table.2. Match results in the Conference track

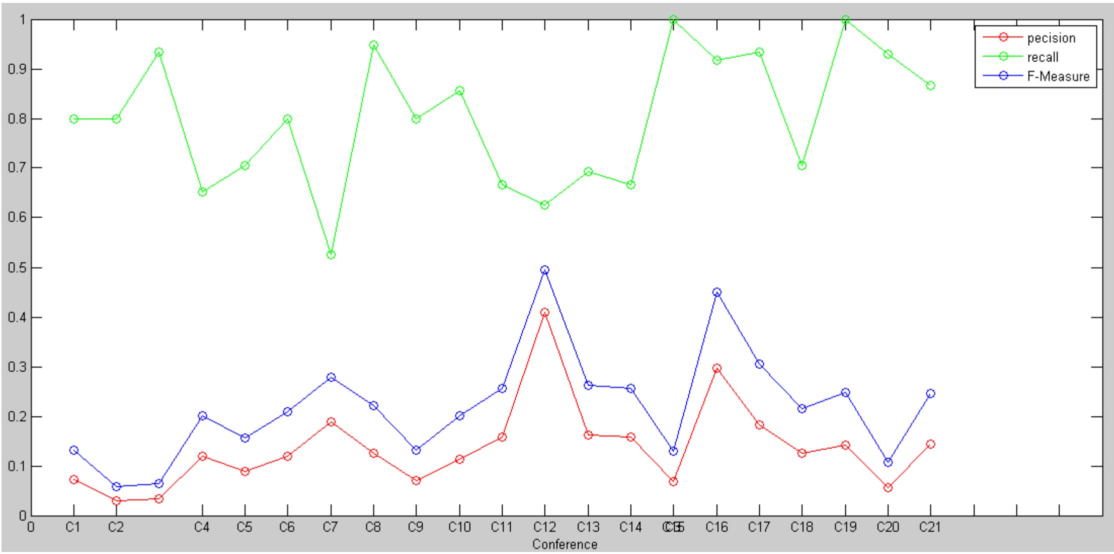

Fig.4. Comparison of match results in Conference

3 General comments

3.1 Comments on the results

The precision of results is not good enough, because only a few individual matchers are included.

The measures in Benchmark are better than those in Conference. The major reason is that the structure similarity of ontology is not considered in our tool.

3.2 Discussions on the way to improve the proposed system

The performance of inference relies on the literal correspondences heavily, so the more accurate results which are exported from multi-matchers will greatly enhance the results of our tool.

Some probable approaches to improve our tool are listed as follow:

1. Adopt more flexible strategies in multi-matchers combination instead of just weighed sum.
2. Add some pre-processes, such as separating compound words, before words are imported into matchers.
3. Take comments and label information of ontology into account, especially when the name of concept is meaningless.
4. Improve the algorithm of some matchers.
5. More different matchers can be included.

Another problem in our tool is that we ignore structure information among ontology at the present stage. And we will improve in the future.

3.3 Comments on the OAEI procedure

OAEI procedure arranged everything in good order, furthermore SEALS platform provides a uniform and convenient way to standardize and evaluate our tool.

4 Conclusions

In this paper, we presented the results of the OMReasoner system for aligning onltologies in the OAEI 2011 competition in two tracks: benchmark and conference. The combination strategy of multiple individual matchers and DL reasonor are included in our approach. This is the first time we participate the OAEI, so it is not sophisticated and the results need to be improved.

References

1. Rahm, E. and Bernstein, P.: A survey of approaches to automatic schema matching. *The VLDB Journal*, ,10(4): 334--350(2001).
2. Shvaiko, P. and Euzenat, J.: A survey of schema-based matching approaches. *Journal on Data Semantics (JoDS)* IV, 146--171(2005).
3. Kalfoglou, Y. and Schorlemmer, M.: Ontology mapping: the state of the art. *The Knowledge Engineering Review Journal*, 18(1):1--31, (2003).
4. Shvaiko, P.: Iterative Schema-based Semantic Matching. PhD, University of Trento, (2006)
5. Jian, N., Hu, W., Cheng, G. et al: Falcon-AO: Aligning Ontologies with Falcon. *In: Proceedings of the K-CAP Workshop on Integrating Ontologies* (2005)
6. Do, H. and Rahm, E.: COMA- a system for flexible combination of schema matching approaches. *In: Proceedings of the International Conference on Very Large Databases*, 610--621.(2002)
7. Giunchiglia, F., Shvaiko, P., and Yatskevich, M.: S-Match: an algorithm and an implementation of semantic matching. *In: Proceedings of the European Semantic Web Symposium*, 61--75.(2004)
8. Kalfoglou, Y. and Schorlemmert, M.: If-map: an ontology mapping method based on information flow theory. *In: Proceedings of ISWC'03, Workshop on Semantic Integration*, (2003)
9. Bouquet, P., Serafini, L., and Zanobini, S.: Semantic coordination: A new approach and an application. *In: Proceedings of the International Semantic Web Conference*, 130--145.(2003)
10. Baader, F., Calvanese, D., McGuinness, D., et al.: The description logic handbook: theory, implementations and applications. *Cambridge University Press*, (2003)
11. Ehrig, M., Sure, Y.: Ontology mapping - an integrated approach. *In Proceedings of the European Semantic Web Symposium (ESWS)*, 76--91, (2004)
12. RacerPro User Guide. http://www.racer -systems. com, 2005
13. Do, H., Melnik, S., Rahm, E.: Comparison of Schema Matching Evaluations. *In: Proceedings of the 2nd Intl. Workshop on Web Databases*, Erfurt, Germany:,221--237(2002)