OMReasoner: Using Multi-matchers and Reasoner for Ontology Matching: results for OAEI 2012

Guohua Shen, Changbao Tian, Qiang Ge, Yiquan Zhu, Lili Liao, Zhiqiu Huang, Dazhou Kang

Nanjing University of Aeronautics and Astronautics, Nanjing, China
{ghshen,cbtian,qge,yqzhu,llliao,zqhuang,dzkang}@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 OAEI 2012 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*.

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

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.

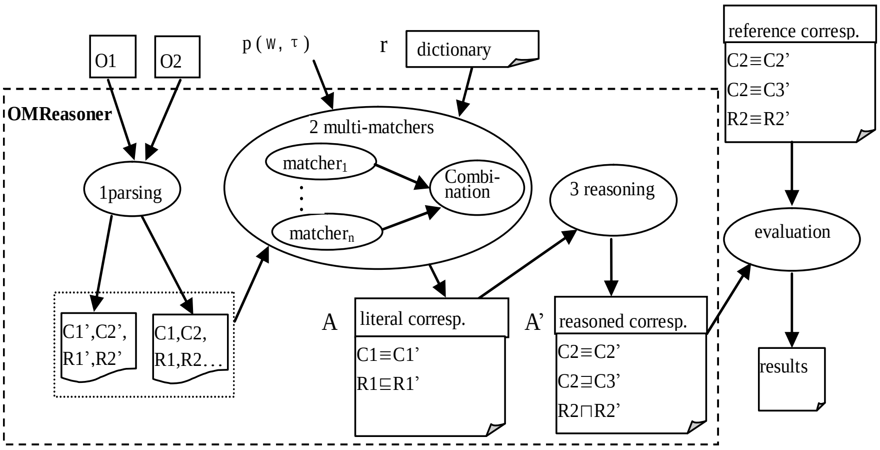

Fig.1. Ontology matching in OMReasoner

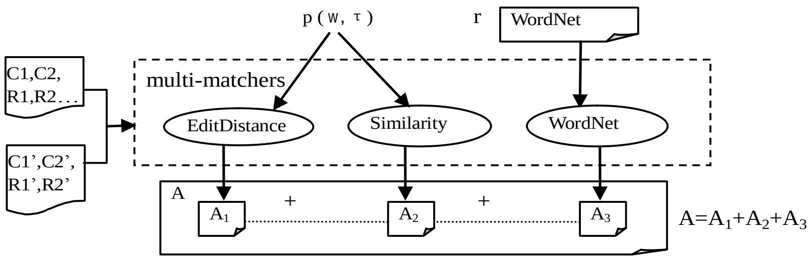

Fig.2. Instances of multi-matchers in OMReasoner

The OMReasoner achieved ontology alignment as following three steps (see Fig.1):

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.

Finally, we evaluate the results against the reference alignments, and compute two measures: precision and recall.

In OMReasoner, the framework for multi-matchers is flexible, and any new individual matcher can be included. Now, the instances of multi-matchers include *EditDistance, Similarity* and *WordNet* (see Fig.2).

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.

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

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 46 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 0516, 0.379 and 0.419 respectively. The average measures of Conference are 0.159, 0.506 and 0.266 respectively. In conclusion, the precision, recall and F-Measure are not satisfying. However, we will improve it in the future.

2.1 Benchmark

We evaluated the results against reference alignments, and obtained precision varies from 0 to 0.949, and recall varies from 0 to 1.000, F-Measure varies from 0 to 0.990. Some measures are zero, because the reference alignments are a little bit strange. For example, *aqdsq* in dataset 248 is equivalent to some class in dataset 101.

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

Label	O1-O2	Prec.	Rec	f-Measure
B1	101-101	0.919	0.588	0.754
B2	101-103	0.919	0.588	0.754
B3	101-104	0.919	0.588	0.754
B4	101-202	0	0	0
B5	101-204	0	0	0
B6	101-204	0.917	0.567	0.739
B7	101-205	0.133	0.062	0.207
B8	101-206	0.540	0.278	0.527
B9	101-207	0.551	0.278	0.527
B10	101-208	0.917	0.567	0.739
B11	101-210	0.600	0.310	0.555
B12	101-221	0.919	0.588	0.754
B13	101-222	0.914	0.570	0.741
B14	101-223	0.919	0.588	0.754
B15	101-224	0.919	0.588	0.754
B16	101-225	0.919	0.588	0.754
B17	101-228	0.868	1.000	0.990
B18	101-230	0.949	0.514	0.690
B19	101-232	0.919	0.588	0.754
B20	101-233	0.868	1.000	0.990
B21	101-236	0.868	1.000	0.990
B22	101-237	0.914	0.570	0.741
B23	101-238	0.919	0.587	0.716
B24	101-239	0.853	1.00	0.9211
B25	101-240	0.868	1.00	0.929
B26	101-241	0.868	1.00	0.929
B27	101-246	0.794	0.931	0.857
B28	101-247	0.868	1.00	0.929
B29	101-248	0	0	0
B30	101-249	0	0	0
B31	101-250	0	0	0
B32	101-251	0	0	0
B33	101-252	0	0	0
B34	101-253	0	0	0
B35	101-254	0	0	0
B36	101-257	0	0	0
B37	101-258	0	0	0
B38	101-259	0	0	0
B39	101-260	0	0	0
B40	101-261	0	0	0
B41	101-262	0	0	0
B42	101-265	0	0	0
B43	101-266	0	0	0
B44	101-301	0.800	0.203	0.324
B45	101-302	0.833	0.3125	0.455
B46	101-304	0	0	0

Table.1. Match results in the Benchmark track

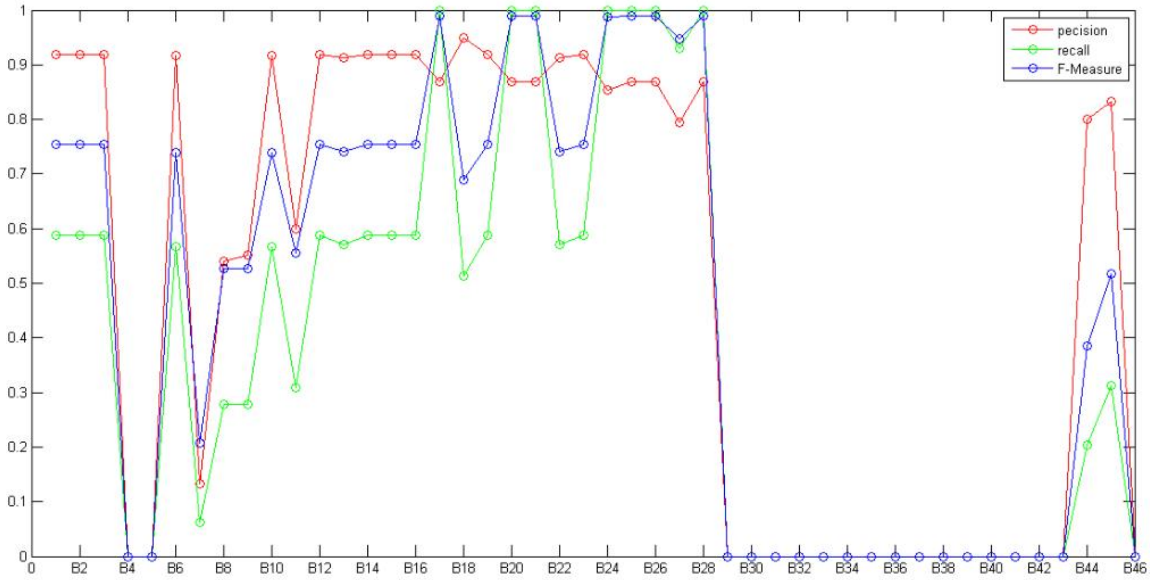

Fig.3. Comparison of match results in Benchmark

2.2 Conference

We evaluated the results against reference alignments, and obtained precision varies from 0.083 to 0.281, and recall varies from 0.296 to 1.000, F-Measure varies from 0.113 to 0.509.

Label	O1-O2	Prec.	Rec	F-Measure
C1	cmt-edas	0.190	0.615	0.360
C2	cmt-ekaw	0.146	0.545	0.282
C3	cmt-iasted	0.251	1.000	0.489
C4	cmt-sigkdd	0.281	0.750	0.509
C5	edas-ekaw	0.179	0.414	0.332
C6	edas-iasted	0.112	0.455	0.219
C7	edas-sigkdd	0.120	0.400	0.232
C8	ekaw-iasted	0.083	0.600	0.165
C9	ekaw-sigkdd	0.191	0.727	0.363
C10	iasted-sigkdd	0.172	0.667	0.331
C11	cmt-conference	0.149	0.412	0.219
C12	cmt-confOf	0.172	0.313	0.222
C13	conference-confOf	0.212	0.467	0.292
C14	conference-edas	0.111	0.368	0.171
C15	conference-ekaw	0.138	0.296	0.188
C16	conference-iasted	0.068	0.333	0.113
C17	conference-sigkdd	0.186	0.533	0.276

C18	confOf-edas	0.214	0.409	0.281
C19	confOf-ekaw	0.136	0.300	0.188
C20	confOf-iasted	0.095	0.444	0.157
C21	confOf-sigkdd	0.129	0.571	0.211

Table.2. Match results in the Conference track

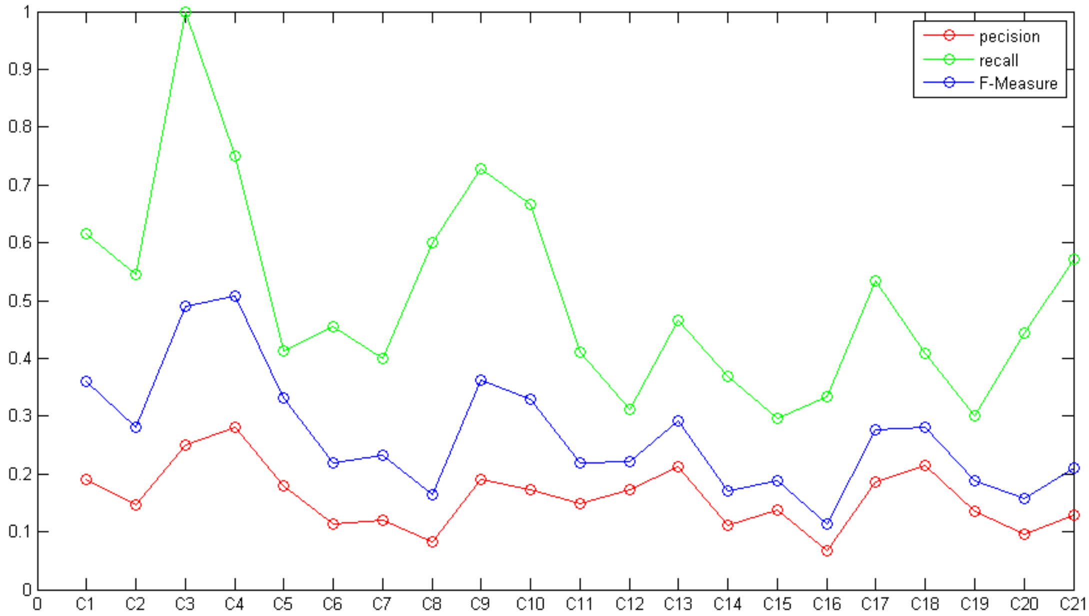

Fig.4. Comparison of match results in Conference

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 more accurate results which are exported from multi-matchers will greatly enhance the results of our tool.

Some probable approaches to improving 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 it in the future.

3.3 Comments on the OAEI 2012 procedure

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

4 Conclusion

In this paper, we presented the results of the OMReasoner system for aligning onltologies in the OAEI 2012 competition in two tracks: benchmark and conference. The combination strategy of multiple individual matchers and DL reasoner are included in our approach. This is the second time we participate the OAEI, the results is still not satisfying and we will improve it in the future.

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)
14. Shen, G., Jin, L., Zhao, Z., Jia, Z., He, W. and Huang, Z. : OMReasoner: using reasoner for ontology matching: results for OAEI 2011. *In Proceedings of* the 6th International Workshop on Ontology Matching.