# OMReasoner: Combination of Multi-matchers for Ontology Matching: results for OAEI 2014

Guohua Shen, Yinling Liu, Fei Wang, Jia Si, Zi Wang, Zhiqiu Huang, Dazhou Kang

College of Computer Sci. &Tech., Nanjing Univ. of Aeronautics and Astronautics, Nanjing, China
{ghshen,ylliu,fwang,jsi,zwang,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 external dictionary *WordNet* and description logic reasoner. It handles ontology matching in both literal and semantic level, and it makes use of the semantic part of OWL-DL as well as structure. This paper describes the result of **OMReasoner** in the OAEI 2014 competition in three tracks: benchmark, conference, and MultiFarm.

## 1 Presentation of the system

Ontology matching finds correspondences between semantically related entities of the ontologies. It plays a significant 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.

Many approaches focus on syntactic aspects instead of semantic ones. OMReasoner achieves the matching by means of some external dictionary and reasoning techniques. Still, this approach includes strategy of combination of (mainly syntactical) multi-matchers (e.g., EditDistance matcher) before match reasoning.

### 1.1 State, purpose, general statement

The matching process can be defined 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.

Alignments express correspondences between two entities. A correspondence must express two corresponding entities and the relation that is supposed to hold between them. Given two ontologies, a correspondence is a 5-tuple:<id,e1,e2,R,n>, where

. id is a unique identifier of the given correspondence;

. e1 and e2 are the entities of the first and the second ontology respectively;

. R is a relation (e.g., equivalence(=), more general(>), less general(<), disjointness ⊥)) holding between e1 and e2. In OAEI campaign, equivalence is mainly considered;

. n is a confidence measure (typically in the [0 1] range) for the correspondence between e1 and e2.

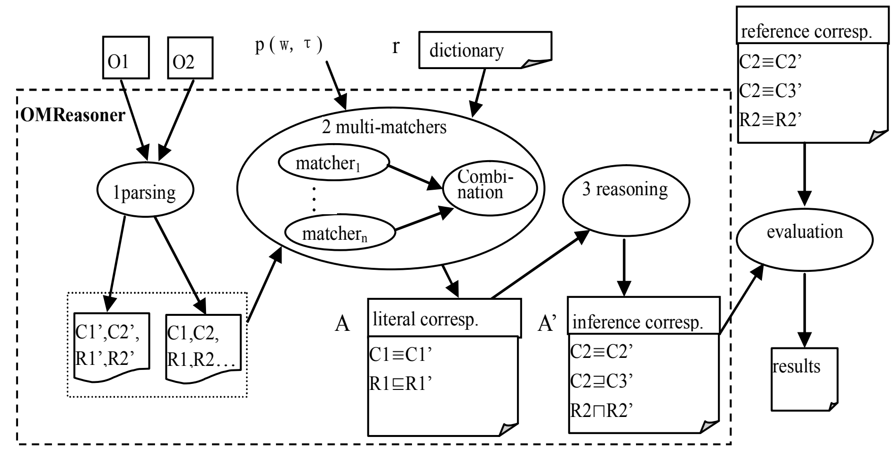


Fig.1. Ontology matching in OMReasoner

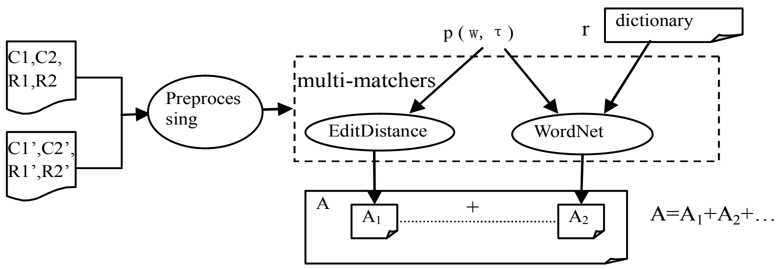


Fig.2. Instances of multi-matchers in OMReasoner

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. Meanwhile, some semantic correspondences can be achieved by using an external dictionary: WordNet. Then the multiple match results can be combined by using specific strategy.

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* and *WordNet* (see Fig.2).

### 1.2 Specific techniques used

1. Threshold

Threshold is necessary for many matchers (especially syntactic ones) to determine whether the similarity is regarded as equivalence. For example, the edit distance of "book" and "booklet" is 3/7 (i.e., the similarity confidence measure is 1-3/7=0.57). If the threshold is 0.55, then these two entities are equivalent (with confidence measure 0.57); else if threshold is 0.6, they are not. So, we have to tune our tool via threshold.

2. Combination of confidence measure

Each individual matcher can produce correspondences with confidence measures. All these confidence measures will be normalized before combination. OMReasoner includes following flexible strategies to combine the multiple match results:

(a) weighted summarizing algorithm (WeightSum)

The confidence can be summarized by weighted similarity algorithm (see formula 1), where $w_k$ is the weight for a specific matcher *k*, and $sim_k(e1,e2)$ is the confidence measure of similarity (mainly equivalence) by this method.

$$\mathrm{sim}(e1, e2) = \sum_{k=1}^{n} w_k \times \mathrm{sim}_k(e1, e2), \qquad \text{where } \sum_{k=1}^{n} w_k = 1.0 \tag{1}$$

(b) maximum method (Max)

The maximum confidence measure is chosen among n matchers (see formula 2) .

$$sim(e1, e2) = \max(sim_1(e1, e2), \ldots, sim_n(e1, e2)) \tag{2}$$

3. semantic matching

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.

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

OMReasoner employs DL reasoner provided by Jena. OMReasoner includes external rules to reason about the ontology matching. However, reasoning is time consuming and only contributes a little to results. In this version, reasoning is skipped.

## 2 Results：a comment for each dataset performed

In this section, we present the results obtained by OMReasoner in the OAEI 2014 competition. It participated in three tracks: benchmark, conference, and MultiFarm. Tests were carried out on a PC running Windows Server 2008 R2 Standard with Intel Core i5 processor running at 2.8 Ghz and 16 GB RAM.

### 2.1 Benchmark

In this track, the ontologies can be divided into 3 categories(see Table 1) . In group 1, the lexical information 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. In group 2 have ontologies that have flattened hierarchies, expanded hierarchies or no hierarchies at all. In group 3 the ontologies are the most challenging ones to align. This is because labels have been scrambled such that they comprise a permutation of letters of a particular length. We tune our tool by using threshold T and combination strategy S, then get the better results ( $\tau_{wd}$=0.95, $\tau_{ed}$=0.9; S=Max). The results obtained by OMReasoner in the benchmarks track are summarized in Table 2.

Table 1. The categories of the Benchmark 2014

| category | concept | systematic | | | real ontology |
|---|---|---|---|---|---|
| tests cases | 101-104 | 201-210 | 221-247 | 248-266 | 301-304 |

Table 2. Results for the Benchmark 2014

| | 101-104 | 201-210 | 210-247 | 248-266 | 301-304 | H-mean |
|---|---|---|---|---|---|---|
| precision | 0.898 | 0.675 | 0.820 | 0.637 | 0.925 | 0.791 |
| recall | 1.000 | 0.414 | 1.000 | 0.517 | 0.437 | 0.647 |
| F-measure | 0.946 | 0.491 | 0.898 | 0.555 | 0.574 | 0.694 |

### 2.2 Conference

The confidence data set consists of numerous real-world ontologies describing the domain of organizing scientific conferences. We use Combination strategy to run our system tool in Conference track. The results obtained by OMReasoner in the benchmarks track are summarized in Table 3 ($\tau_{wd}$=0.9, $\tau_{ed}$=0.8**; S=Max**).

Table 3. Results for the Conference 2014

| test case | precision | recall | F-measure |
|---|---|---|---|
| Conference | 0.778 | 0.518 | 0.647 |

### 2.3 MultiFarm

MultiFarm track is composed of a subset of the Conference dataset, translated in eight different languages. In this track, the ontologies can be divided into 2 categories. In group 1 the alignments ontologies are the same. In group 2 the alignments ontologies are different.

Firstly, we take use of dictionary to translate different languages into English. Then, the translated English is imported in multi-matchers by using Max strategy. Finally we get the results. We tune our tool by using threshold, and the results can be seen in Table 4($\tau_{wd}$=0.8, $\tau_{ed}$=0.6; S=Max), which show that the F-Measures of the ontologies alignments in group 2 are obviously worse than those in group 1. We think the reasons are that OMReasoner is not well designed to match different ontologies which are written in completely different languages yet.

Table 4. Results for the MultiFarm 2014

| Test case | precision | recall | F-measure |
|---|---|---|---|
| Group 1: Same ontologies | 0.955 | 0.800 | 0.853 |
| Group 2: Different ontologies | 0.584 | 0.438 | 0.471 |

To choose better threshold, we compare the results (see Table 5) across several thresholds in Conference track. Still we use Max method to run our tool. From the results, we find that when threshold $\tau_{wd}$=0.9, $\tau_{ed}$=0.8, our tool performs best. So that we take use of threshold $\tau_{wd}$=0.9, $\tau_{ed}$=0.8 in Conference track. Using the same method, we get the better thresholds for Benchmark and MultiFarm track.

Table 5. Comparison results of different thresholds for the Conference 2014

| **Threshold** | | precision | recall | F-measure |
|---|---|---|---|---|
| $\tau_{wd}$ | $\tau_{ed}$ | | | |
| 0.8 | 0.8 | 0.782 | 0.508 | 0.599 |
| 0.95 | 0.8 | 0.787 | 0.466 | 0.580 |
| 0.9 | 0.8 | 0.778 | 0.518 | 0.647 |
| 0.9 | 0.9 | 0.796 | 0.476 | 0.580 |

# 3 General comments

## 3.1 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 approaches to improving our tool are listed as follows:

1. Adopt more flexible strategies in multi-matchers combination instead of just weighed sum.
2. Add some preprocessing (see Fig.2), such as eliminating specific character (e.g., '-', '_') or 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. Reexamine the use of an appropriate threshold value to optimize accuracy.

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.

## 3.2 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.

## 3.3 Comments on the OAEI test cases

The OAEI test cases involve all kinds of fields which include conference, anatomy, language, etc. The variety of tracks and the improvements introduced along the years makes the campaign very useful to test the performance of ontology aligners and analyses their strengths and weaknesses. Nevertheless, we miss blind tests cases in more tracks, which will allow a fair comparison between systems.

## 3.4 Proposed new measures

After serious discussion, we believe that OMReasoner can improve a lot. Some new ways proposed as follows:

1. Enrich the semantic dictionaries because WordNet is not a professional dictionary, which cannot obtain more comprehensive semantic concepts.
2. Take into account hierarchical ones instead of only all concepts and properties.
3. Find NCI thesaurus for anatomy track.
4. Find different languages dictionaries for MultiFarm.
5. Improve the algorithm of some matchers.
6. Include more different matchers.

## 4 Conclusions

In this paper, we presented the results of the OMReasoner system for aligning ontologies in the OAEI 2014 competition in three tracks: benchmark, conference, and MultiFarm. The combination strategy of multiple individual matchers and DL reasoner are included in our approach. This is the third time we participate the OAEI, the results are 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 6$^{th}$ International Workshop on Ontology Matching.
15. Shen, G., Tian, C., Ge, Q., Zhu, Y., Liao, L., Huang, Z., and Kang D.: OMReasoner: using multi-matchers and reasoner for ontology matching: results for OAEI 2012. *In Proceedings of* the 7$^{th}$ International Workshop on Ontology Matching.