Ontology Matching with CIDER: Evaluation Report for the OAEI 2008

Jorge Gracia, Eduardo Mena

IIS Department, University of Zaragoza, Spain
{jogracia,emena}@unizar.es

Abstract. Ontology matching, the task of determining relations that hold among terms of two different ontologies, is a key issue in the Semantic Web and other related fields. In order to compare the behaviour of different ontology matching systems, the Ontology Alignment Evaluation Initiative (OAEI) has established a periodical controlled evaluation that comes in a yearly event. We present here our participation in the 2008 initiative.
Our schema-based alignment algorithm compares each pair of ontology terms by, firstly, extracting their ontological contexts up to a certain depth (enriched by using transitive entailment) and, secondly, combining different elementary ontology matching techniques (e.g., lexical distances and vector space modelling). Benchmark results show a very good behaviour in terms of precision, while preserving an acceptable recall.
Based on our experience, we have also included some remarks about the nature of benchmark test cases that, in our opinion, could help improving the OAEI tests in the future.

1 Presentation of the system

In [7] we presented a system that analyzes a keyword-based user query, in order to automatically extract and make explicit, without ambiguities, its semantics. Firstly, it discovers and extracts candidate senses (expressed as ontology terms) for each keyword, by harvesting the Semantic Web. Local ontologies or lexical resources, as WordNet [6], can also be accessed. Then, an alignment and integration step is carried out in order to reduce redundancies (many terms from different ontologies could describe the same intended meaning, so we integrate them as a single sense). Finally, a disambiguation process is run to pick up the most probable sense for each keyword, according to the context. The result can be eventually used in the construction of a well-defined semantic query (expressed in a formal language) to make explicit the intended meaning of the user.

We realized that the alignment component of our system is general enough to be used for many other tasks so, based on it, we have developed an independent aligner to be evaluated in the OAEI contest[1]. The latest version of our alignment

[1] http://oaei.ontologymatching.org/2008/

service is called CIDER (Context and Inference baseD alignER), which is the subject of this study. It relies on a modified version of the semantic similarity measure described in [7].

1.1 State, purpose, general statement

According to the high level classification given in [3], our method is a *schema-based* system (opposite to others which are instance-based, or mixed), because it relies mostly on schema-level input information for performing ontology matching. As it was mentioned in the previous section, the initial purpose of our algorithm was to discover similarities among possible senses of user keywords, in order to integrate them when they were similar enough (to be later disambiguated and used in semantic query construction). Therefore, our alignment algorithm was initially applied to a previously discovered set of ontological terms, describing possible senses of a keyword.

For this study we have generalized the method, to admit any two ontologies, and a threshold value, as input. Comparisons among all pairs of ontology terms (not only the ones that could refer to a same user keyword) are established, producing as output an RDF document with the obtained alignments.

1.2 Specific techniques used

Our alignment process takes as basis a modified version of the *semantic similarity measure* described in [7]. A detailed discussion of the introduced improvements is out of the scope of this paper. However, here is a brief summary of them:

1. Addition of a *transitive entailment* mechanism during the extraction step, which has remarkably improved our results in terms of quality.
2. Enrichment of our initially naive comparisons between instances, by considering also their properties and corresponding values.
3. Optimization of the initially costly comparisons among properties of concepts, substituting their recursive focus with the use of vector space modelling. We have found that it preserves quality, while reduces time significantly.

Figure 1 shows a schematic view of the way our matcher works. O_1 and O_2 represent the input ontologies. M is the matrix of resultant comparisons among ontology terms, and A is the extracted alignment.

The first step is to extract the ontological context of each involved term, up to a certain depth. That is (depending on the type of term), their synonyms, textual descriptions, hypernyms, hyponyms, properties, domains, roles, associated concepts, etc. This process is enriched by applying a transitive inference mechanism, in order to add more semantic information that is not explicit in the asserted ontologies.

The second step is the computation of similarity for each pair of terms. It is carried out differently, depending on the type of ontology term (concept, property or individual). Without entering into details, comparisons are performed like this:

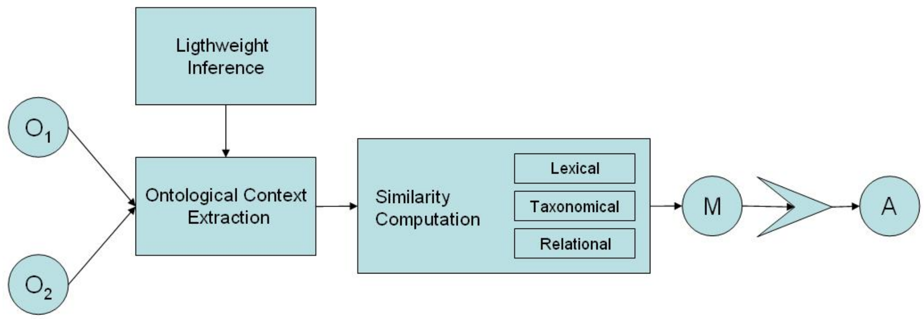

Fig. 1. Scheme of the CIDER process.

1. Linguistic similarity between terms, considering labels and descriptions, is computed.
2. A subsequent computation explores the structural similarity of the terms, exploiting their ontological contexts and using vector space modelling in comparisons. It comprises comparison of taxonomies and relationships among terms (e.g. properties of concepts).
3. The different contributions are weighted, and a final similarity degree is provided.

After that, a matrix M with all similarities is obtained. The final alignment A is then extracted, finding the highest rated one-to-one relationships among terms, and filtering out the ones that are below the given threshold.

In terms of implementation, CIDER prototype has been developed in Java, extending the Alignment API [2]. The input are ontologies expressed in OWL or RDF, and the output is served as a file expressed in the *alignment format* [2], although it can be easily translated to other formats as well.

1.3 Adaptations made for the evaluation

As the benchmark test does not consider mappings between instances, we have not computed instance alignment for this particular test. No other adaptations have been needed.

1.4 Link to the system and parameters file

The version of CIDER used for this evaluation can be found at
http://sid.cps.unizar.es/SEMANTICWEB/ALIGNMENT/OAEI08/

1.5 Link to the set of provided alignments (in align format)

The obtained alignments for the contest can be found at
http://sid.cps.unizar.es/SEMANTICWEB/ALIGNMENT/OAEI08/results/CIDER.zip

2 Results

The following subsections describe the participation of our system in two tracks of the contest: benchmark and directory. Some remarks specific to each test are described, as well as a tentative explanation of the obtained results. Further information about the whole results of the contest can be found at [1].

2.1 Benchmark

The target of this experiment is the alignment of bibliographic ontologies. A reference ontology is proposed, and many comparisons with other ontologies of the same domain are performed. The tests are systematically generated, modifying differently the reference ontology in order to evaluate how the algorithm behaves when the aligned ontologies differ in some particular aspects. A total of 111 test cases have to be evaluated. They are grouped in three sets:

1. Concept test (cases *1xx*: 101, 102, ...), that explore comparisons between the reference ontology and itself, described with different expressivity levels.
2. Systematic (cases *2xx*). It alters systematically the reference ontology to compare different modifications or different missing information.
3. Real ontology (cases *3xx*), where comparisons with other "real world" bibliographic ontologies are explored.

We cannot provide results for benchmark cases 202 and 248-266, because our system does not deal with ontologies in which syntax is not significant at all (these cases present a total absence or randomization of labels and comments). Consequently, we expect a result with a low recall in this experiment, as the benchmark test unfavours methods that are not based on graph structure analysis (or similar techniques).

In Table 1 we show the obtained results, grouped by type of cases. We have obtained a very high precision (97%), which is in the top-three best values obtained in the contest (out of 13 participants), while recall has been lower (62%), due to the above mentioned reason. The extended results for the complete dataset has been published separately by the organizers[2].

	1xx	**2xx**	**3xx**	**Average**	**H-Mean**
Precision	0.99	0.97	0.90	0.97	0.97
Recall	0.99	0.60	0.73	0.61	0.62

Table 1. Averaged results for the benchmark dataset.

Alternatively to the official results, we have computed the precision and recall of the benchmark test excluding the cases 202 and 248-266 (and their variations

[2] http://oaei.ontologymatching.org/2008/results/benchmarks.html

248-2, 248-4, etc.), in which ontology terms are described with non expressive texts. This is an "internal" exercise, which does not let us direct comparisons with other methods in the contest, but gives us another point of view (more accurate, according to the final usage of our system) of the behaviour of our method. Results are given in Table 2.

	1xx	**2xx**	**3xx**	**Average**	**H-Mean**
Precision	0.99	0.97	0.90	0.96	0.97
Recall	0.99	0.87	0.73	0.87	0.86

Table 2. Results for the benchmark dataset omitting cases with no significant texts.

2.2 Directory

The objective of this experiment is to match terms from plain hierarchies, extracted from web directories. It consist of more than 4 thousand elementary alignments. We consider that our method cannot show all its strengths in this, because the available information is extremely sparse, lacking in semantic descriptions beyond hierarchical relationships (no instances, no properties, no comments, no synonyms, ...).

Results have been: 60% precision, 38% recall and 47% F-measure, which has been the second best result in this year competition (out of seven participants). A detailed comparison has been published by organizers[3]. We see that, even directory alignment is not the target of our system, it behaves reasonably well when matching plain hierarchies.

3 General comments

The following subsections contain some remarks and comments about the results obtained, as well as about the test cases and evaluation process.

3.1 Comments on the results

As expected, we obtained better precision than recall in the benchmark test (due to the reasons mentioned in Section 2.1). Also in the directory experiment precision was higher than recall. However, it is consistent with the fact that our alignment is targeted to be used in an automatic way, minimizing human intervention. In this conditions, precision have to be promoted over recall. That is, maybe our system does not discover all correspondences, but we have to be sure that, in case it discovers an equivalence mapping between two terms, they

[3] http://www.disi.unitn.it/~pane/OAEI/2008/directory/result/

are most likely referring to the same meaning. Otherwise their later integration would be erroneous, and the mistake would eventually be propagated to the other steps of the system.

3.2 Discussions on the way to improve the proposed system

Our method does not consider extensional information when comparing concepts, focusing only on the semantic description of the terms in the corresponding ontologies. Its inclusion could improve results in some cases where this information is available.

Additionally, although our system considers many features of ontologies, their richness vary a lot from one case to another. We consider that the addition of mechanisms to auto-adjust weights to the characteristics of ontologies (as they do in [5]) could largely benefit our method.

Finally, although our similarity measure has been much optimized, in terms of time response, the overall alignment process can still be subject of further improvement.

3.3 Comments on the OAEI 2008 test cases

We have found the benchmark test very useful as a guideline for our internal improvements of the method, as well as to establish a certain degree of comparisons with other existing methods.

On the other hand, we have missed some important issues that are not taken into account in the systematic benchmark series:

1. Benchmark tests only consider positive matchings, not measuring the ability of different methods to avoid links among barely related ontologies (only case 102 of benchmark goes in that direction).
2. For our purposes, we try to emulate the human behaviour when mapping ontological terms. As human experts cannot properly identify mappings between ontologies with scrambled texts, neither does our system. However, reference alignments provided in the benchmark evaluation for cases 202 and 248-266, do not follow this intuition. We hope this bias will be reduced in future contests.
3. Related to the latter, cases in which equal topologies, but containing different semantics, lead to false positives, are not explicitly taken into account in the benchmark.
4. How ambiguities can affect the method is not considered either in the test cases. It is a consequence of using ontologies belonging to the same domain. For example, it would be interesting to evaluate how "film" in an ontology about movies, is mapped to "film" as a "thin layer" in another ontology. Therefore it is difficult to evaluate the benefits of including certain disambiguation techniques in ontology matching [4].

3.4 Comments on the OAEI 2008 measures

Unsuitability of precision and recall measures for ontology matching evaluation is a well known problem [3]. We encourage organizers to try different measures that count all correct found correspondences, even when they are not explicit in the reference alignment.

4 Conclusion

We have presented here some results of our first participation in the OAEI 2008 contest. We have limited to two tracks: benchmark and directory, but we hope to extend our participation in the future.

Our schema-based alignment algorithm compares the ontological contexts of each pair of terms (enriched with transitive inference) by combining different elementary ontology matching techniques (comparing vocabulary, taxonomies, relations,...). Benchmark results show a very good behaviour of our system in terms of precision, while keeping an acceptable recall. It confirms the validity of the measure we have conceived, and its suitability to be applied in ontology matching tasks. It encourages us to tackle further improvements, and to extend the scope and applicability of our techniques.

We have also included, based on our experience, some considerations about the nature of benchmark test cases that, in our opinion, could help improving future contests.

Acknowledgments. This work is supported by the CICYT project TIN2007-68091-C02-02.

References

1. C. Caracciolo, J. Euzenat, L. Hollink, R. Ichise, A. Isaac, V. Malaisé, C. Meilicke, J. Pane, P. Shvaiko, H. Stuckenschmidt, O. Šváb-Zamaza, and V. Svátek. First results of the ontology alignment evaluation initiative 2008. In *In Proc. ISWC-2008 Workshop on Ontology Matching*, 2008.
2. J. Euzenat. An API for ontology alignment. In *3rd International Semantic Web Conference (ISWC'04), Hiroshima (Japan)*. Springer, November 2004.
3. J. Euzenat and P. Shvaiko. *Ontology matching.* Springer-Verlag, 2007.
4. J. Gracia, V. López, M. d'Aquin, M. Sabou, E. Motta, and E. Mena. Solving semantic ambiguity to improve semantic web based ontology matching. In *Proc. of 2nd Ontology Matching Workshop (OM'07), at 6th International Semantic Web Conference (ISWC'07), Busan (Korea)*, November 2007.
5. Y. Jean-Mary and M. Kabuka. ASMOV: Ontology alignment with semantic validation. In *Proc. of Joint SWDB-ODBIS Workshop on Semantics, Ontologies, Databases, Vienna (Austria)*, September 2007.
6. G. Miller. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11), nov 1995.
7. R. Trillo, J. Gracia, M. Espinoza, and E. Mena. Discovering the semantics of user keywords. *Journal on Universal Computer Science. Special Issue: Ontologies and their Applications*, November 2007.