

Ontology Matching with CIDER: evaluation report for OAEI 2011

Jorge Gracia¹, Jordi Bernad², and Eduardo Mena²

¹ Ontology Engineering Group, Universidad Politécnica de Madrid, Spain
jgracia@fi.upm.es

² IIS Department, University of Zaragoza, Spain
{jbernad, emena}@unizar.es

Abstract. CIDER is a schema-based ontology alignment system. Its algorithm compares each pair of ontology terms by, firstly, extracting their ontological contexts up to a certain depth (enriched by using lightweight inference) and, secondly, combining different elementary ontology matching techniques. In its current version, CIDER uses artificial neural networks in order to combine such elementary matchers.

In this paper we briefly describe CIDER and comment on its results at the Ontology Alignment Evaluation Initiative 2011 campaign (OAEI'11). In this new approach, the burden of manual selection of weights has been definitely eliminated, while preserving the performance with respect to CIDER's previous participation in the benchmark track (at OAEI'08).

1 Presentation of the system

CIDER (Context and Inference baseD alignER) is a system for ontology alignment that performs semantic similarity computations among terms of two given ontologies. It extracts the ontological context of the compared terms and enriches it by applying lightweight inference rules. Elementary similarity comparisons are performed to compare different features of the extracted ontological contexts. Such elementary comparisons are combined by means of artificial neural networks (ANNs).

CIDER was initially created in the context of a system [9] for discovering the semantics of user keywords and already participated in the OAEI'08 campaign, leading to good results [5]. In CIDER's previous version, the elementary comparisons performed during the similarity computation were combined linearly. The weights of this linear combination were manually tuned after experimentation. This was a major limitation of the approach, which hampered the flexibility of the method and the capacity for quickly adapting it into different domains. This has been solved in the current version by the addition of ANNs.

We expect to confirm that this contribution will not have a negative impact on the initial algorithm. We also expect to discover areas of potential improvement that guide us in our future exploration of this research path. For OAEI'11 campaign, CIDER has participated in the Seals-based tracks³, i.e., *benchmark*, *anatomy*, and *conference* tracks.

³ <http://oaei.ontologymatching.org/2011/seals-eval.html>

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. CIDER admits any two OWL ontologies and a threshold value as input. Comparisons among all pairs of ontology terms are established, producing as output an RDF document with the obtained alignments. In its current version, the process is enhanced with the use of ANNs. The type of alignments that CIDER obtains is *semantic equivalences*.

1.2 Specific techniques used

Our alignment process takes as basis the *semantic similarity measure* described in [9], with the improvements introduced in [5]. Briefly explained, the similarity computation is as follows (see Figure 1):

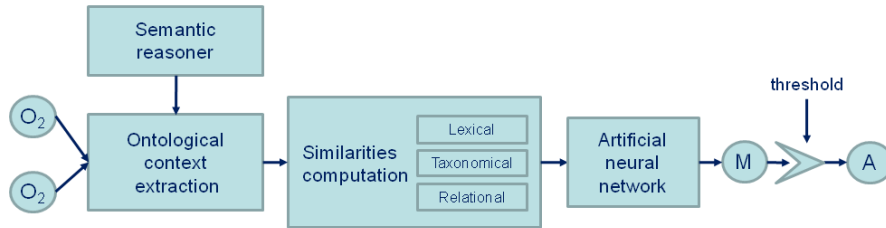


Fig. 1. Scheme of the matching process.

1. Firstly, the ontological context of each ontology term is extracted, 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 lightweight inference mechanism⁴, in order to add more semantic information that is not explicit in the asserted ontologies.
2. Secondly, several similarities between each pair of compared terms are computed: Linguistic similarity between the labels of terms (based on Levenhstein [6] similarity) as well as structural similarities, by exploiting the ontological context of the terms and using vector space modelling [7] in the comparisons. This comprises comparison of taxonomies and relationships among terms (e.g., properties of concepts).

⁴ Typically transitive inference, although RDFS or more complex rules can be also applied, at the cost of processing time.

3. Then, the different similarities are combined within an ANN to provide a final similarity degree. ANNs constitute an adaptive type of systems composed of interconnected artificial neurons, which change the structure based on external or internal information that flows through the network during a learning phase [8]. CIDER uses two different neural networks for computing similarities between classes and properties, respectively⁵.
4. Finally, a matrix (M in Figure 1) with all similarities is obtained. The final alignment (A) is then extracted from this matrix, finding the highest rated one-to-one relationships among terms, and filtering out the ones that are below the given threshold.

Figure 2 shows, as an example, the structure of the neural network for computing similarity between classes (the other one for properties follows an equivalent pattern). Without entering into the details, this corresponds to a *multilayer perceptron*, which consists of multiple layers of nodes in a directed graph, each layer fully connected to the next one. Each connection (synapse) has an associated weight. In our particular situation, the network is composed of three layers: input, hidden, and output layer (with five, three, and one neurons respectively; additionally two bias neurons are used in the input and hidden layer respectively). Each neuron in the input layer receives the value of an elementary similarity measure. Each intermediate neuron uses a sigmoid function to combine the inputs. Finally, the resultant similarity value is given by the neuron in the output layer.

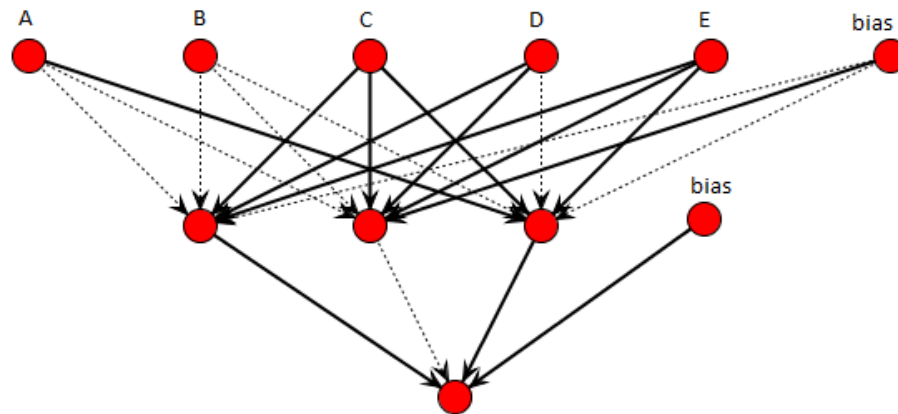


Fig. 2. Scheme of the neural network for computing similarity between classes. Highlighted connexions correspond to higher weights.

⁵ Similarity between individuals follows the approach of the previous versions, although the addition of a new ANN for that is planned as future work.

The inputs for the neural network that computes class similarity (labelled A - E in the figure) are: lexical similarity between labels, similarity of textual descriptions (e.g., `rdfs:comment`), similarity between hypernyms, similarity between hyponyms, and similarity between associated properties.

In terms of implementation, the CIDER prototype has been developed in Java, extending the Alignment API [1]. To create and manipulate neural networks we use Neuroph Studio library⁶. The input to CIDER are OWL ontologies and the output is served as a file expressed in the *alignment format* [1], although it can be easily translated into another formats.

1.3 Adaptations made for the evaluation

According to the conditions of the competition⁷ “it is fair to select the set of parameters that provide the best results (for the tests where results are known)”. Thus, we chose a subset of the OAIE’08 benchmark to train the neural networks and find suitable weights for combining the elementary matchers that CIDER uses. We used the 2008 benchmark dataset but excluding cases 202 and 248-266 (which present a total absence or randomization of labels and comments). The weights and the configuration of the neural network remained constant for the whole evaluation.

Furthermore, as the Seals-based tracks of the competition do not consider mappings between instances, we have disabled instance comparison. Finally, some minor technical adaptations were required for integrating the system into the Seals platform, such as updating some libraries (e.g., Alignment API) or changing the way some parameters are communicated.

1.4 Link to the system and parameters file

The version of CIDER used for this evaluation (v0.4c) can be found at Seals platform: <http://www.seals-project.eu/>. More information can be found at <http://sid.cps.unizar.es/SEMANTICWEB/ALIGNMENT/>

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

The resultant alignments will be provided by the Seals platform: <http://www.seals-project.eu/>

2 Results

At the time of writing this, the preliminary results of this year competition are available at [2], although the organisers will make public additional results in the future. From the tracks in which CIDER participated, CIDER was not able to produce results on the *anatomy* tests. The reason is that CIDER’s current implementation is not optimised

⁶ <http://neuroph.sourceforge.net/>

⁷ <http://oei.ontologymatching.org/2011/>

to be used with large ontologies, and the execution of the test gave a timeout before completion. Actually only six, out of the 16 participants in the anatomy track, were able to complete the evaluation. We plan to improve CIDER’s performance for large ontologies in a near future.

In the following paragraphs, we summarize the results obtained in the benchmark and conference tracks (the detailed results can be found in [2]), as well as some complementary experiments that we run locally.

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.

As in our previous participation, we point out that our system is not intended to deal with ontologies in which syntax is not significant at all, as it is the case for benchmark cases 202 and 248-266 (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 these benchmark tests do not favour methods that are not based on graph structure analysis or similar techniques.

In addition to the traditional test data, a new benchmark data set (Benchmark2) has been provided this year. This uses the EKAW conference ontology⁸ as basis and, same as in the Benchmark data set, different systematic variations are explored, resulting in 103 test cases.

Finally, there was a “blind” benchmark data set that was used for providing the final results in the competition. This test, consisting of 102 alignments, resulted in 0.89 precision, 0.58 recall and 0.70 F-Measure for CIDER, which is *above the average results in the competition* (0.66 F-Measure, ranging from 0.32 to 0.77). Besides the usual precision/recall, other extended metrics were considered for this test in this year’s competition [2]. For instance, *weighted precision/recall* were computed taking into account the score that the tool assigned to each correspondence. For this weighted metrics CIDER obtained: 0.91 precision, 0.54 recall, and 0.68 F-measure (being the *4th best in the competition*, out of 16 participants).

In Table 1 we show the results of evaluating CIDER with the different benchmark tracks: Benchmark (2010), Benchmark2 (2011), and “blind” Benchmark (2011). Additionally, and for comparison purposes, we also run the 2008 benchmark data with the

⁸ <http://nb.vse.cz/svabo/oei2011/data/ekaw.owl>

version of CIDER submitted for evaluation (v0.4) and compared it to the results obtained by CIDER at OAEI08 (v0.1). Table 2 shows the results. Baseline results (edit distance) are also included.

	Benchmark	Benchmark2	Benchmark(blind)
Precision	0.87	0.74	0.89
Recall	0.66	0.58	0.58
F-Measure	0.75	0.65	0.70

Table 1. Averaged results of CIDER for the benchmark datasets.

	baseline(edna)	CIDER v0.1	CIDER v0.4
Precision	0.56	0.97	0.88
Recall	0.60	0.62	0.69
F-Measure	0.58	0.76	0.77

Table 2. H-mean results for the OAEI08 benchmark dataset.

2.2 Conference

This track consisted of aligning several ontologies from the conference domain, which resulted in 21 alignments. To evaluate the resultant alignments, various F-measures were computed: F1 (harmonic mean between precision and recall), F2 (promotes recall), and F0.5 (promotes precision).

The results given by CIDER were: F1-Measure = 0.53, F2-Measure = 0.48, and F0.5-Measure = 0.61, which place our system in intermediate positions in this track. We expect that the use of more “real world” training data in CIDER will improve the results in this track in future contests. The effect of the threshold selection in these results has been nicely illustrated in the figures provided by the organisers⁹, and shows that performance can be dramatically improved with a suitable selection of the threshold.

3 General comments

The following subsections contain some remarks and comments about the results obtained and the evaluation process.

⁹ See <http://nb.vse.cz/~svabo/oaei2011/eval.html#reference>

3.1 Comments on the results

As it is shown in Table 2, a direct comparison between the current and previous version of CIDER shows that the addition of ANNs does not have a negative effect on the algorithm but, on the contrary, leads to slightly better results. Such results indicate also that the new approach leads to a better recall, at the cost of precision.

The results for benchmark tracks (Table 1), although acceptable, are hampered by the presence of test cases with ontologies lacking lexical information or that has been randomly generated.

With respect to the conference track, the results are influenced by the fact that our ANNs only used open data from the benchmark track for training. More reference alignments from “real world” ontologies will be used in the future for training the ANNs, in order to cover different domains and different types of ontologies.

3.2 Discussions on the way to improve the proposed system

Although the obtained results are acceptable, we consider that there is still room for further improvements. In fact, the addition of ANNs for similarity computation in CIDER is in a preliminary stage and has to be further studied. We have to use more “real” data for training, and alternative configurations for our multilayer perceptrons has to be studied. On the other hand, time response in CIDER is still an issue and has to be further improved. Also, CIDER works well with small and medium sized ontologies but not with large ones. Partitioning and other related techniques will be explored in order to solve this.

3.3 Comments on the OAEI 2011 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. Some of them coincide with the ones we already reported in 2008:

1. Benchmark tests only consider positive matchings, not measuring the ability of different methods to avoid links among barely related ontologies.
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 describing different things, 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 whether “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].

4 Conclusion

CIDER is a schema-based alignment system that compares the ontological contexts (enriched with transitive inference) of each pair of terms in the aligned ontologies. Several elementary ontology matching techniques are computed and combined by means of artificial neural networks. We have presented here some results of the participation of CIDER at OAEI'11 contest, particularly in the Seals-based tracks (benchmark, anatomy, and conference). The results on the benchmark track are good and constitute our starting point for testing future improvements. We confirmed that the addition of artificial neural networks keeps the performance and, furthermore, eliminates the necessity of tuning the weights manually.

Acknowledgments. This work is supported by the CICYT project TIN2010-21387-C02-02, the Spanish Ministry of Science and Innovation within the Juan de la Cierva program, and the EC within the FP7 project DynaLearn (no. 231526).

References

1. J. Euzenat. An API for ontology alignment. In *3rd International Semantic Web Conference (ISWC'04), Hiroshima (Japan)*. Springer, November 2004.
2. J. Euzenat, A. Ferrara, W. R. van Hage, L. Hollink, C. Meilicke, A. Nikolov, D. Ritzé, F. Scharffe, P. Shvaiko, H. Stuckenschmidt, O. Šváb-Zamaza, and C. Trojahn. First results of the ontology alignment evaluation initiative 2011. In *In Proc. of 6th Ontology Matching Workshop (OM11), at International Semantic Web Conference (ISWC11), Bonn, Germany, 2011*.
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. J. Gracia and E. Mena. Ontology matching with CIDER: Evaluation report for the OAEI 2008. In *Proc. of 3rd Ontology Matching Workshop (OM'08), at 7th International Semantic Web Conference (ISWC'08), Karlsruhe, Germany*, volume 431, pages 140–146. CEUR-WS, October 2008.
6. V. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory*, 10(8):707–710, 1966. Original in *Doklady Akademii Nauk SSSR* 163(4): 845–848 (1965).
7. V. V. Raghavan and M. S. K. Wong. A critical analysis of vector space model for information retrieval. *Journal of the American Society for Information Science*, 37(5):279–287, 1986.
8. M. Smith. *Neural Networks for Statistical Modeling*. John Wiley & Sons, Inc., New York, NY, USA, 1993.
9. 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.