Monolingual and Cross-lingual Ontology Matching with CIDER-CL: evaluation report for OAEI 2013

Jorge Gracia[1] and Kartik Asooja[1,2]

[1] Ontology Engineering Group, Universidad Politécnica de Madrid, Spain
jgracia@fi.upm.es
[2] Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland
kartik.asooja@deri.org

Abstract. CIDER-CL is the evolution of CIDER, a schema-based ontology alignment system. Its algorithm compares each pair of ontology entities by analysing their similarity at different levels of their ontological context (linguistic description, superterms, subterms, related terms, etc.). Then, such elementary similarities are combined by means of artificial neural networks. In its current version, CIDER-CL uses SoftTFIDF for monolingual comparisons and Cross-Lingual Explicit Semantic Analysis for comparisons between entities documented in different natural languages. In this paper we briefly describe CIDER-CL and comment its results at the Ontology Alignment Evaluation Initiative 2013 campaign (OAEI'13).

1 Presentation of the system

CIDER-CL is the evolution of CIDER (Context and Inference baseD alignER) [7], now incorporating cross-lingual capabilities. In order to match ontology entities, CIDER-CL extracts their ontological context and enriches it by applying lightweight inference rules. Then, elementary similarity comparisons are performed to compare different features of the ontological contexts. Such elementary comparisons are combined by means of artificial neural networks (ANNs) [9] to produce a final similarity value between the compared entities. The use of ANNs saves a lot of effort on manual tuning and allows to quickly adapt the system into different domains (as far as there are reference alignments available for them).

In its current version, the aligner has been re-implemented to include more features in the comparisons and to add new metrics for similarity computation. In particular, cross-lingual capabilities have been added by including the use of Cross-Lingual Explicit Semantic Analysis (CL-ESA) [10] between entities documented in different natural languages. Further, the previous metrics for monolingual comparison (based on Vector Space Modelling [8]) have been changed by the SoftTFIDF metric [3]. CIDER-CL is not intended to be used with large ontologies, particularly in the cross-lingual case (CL-ESA computation is quite costly in terms of time).

1.1 State, purpose, general statement

According to the high level classification given in [5], 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-CL can operate in two modes: (i) as an *ontology aligner*, taking two ontologies as input and giving their alignment as output, and (ii) as a *similarity service*, taking two ontology entities as input and giving the similarity value between them as output. In the first case the input to CIDER-CL are two OWL ontologies and a threshold value and the output is an RDF file expressed in the *alignment format*[3], although it can be easily translated into another formats such as EDOAL[4].

The type of alignment that CIDER-CL obtains is *semantic equivalence*. In its current implementation the following languages are covered: English (EN), Spanish (ES), German (DE), and Dutch (NL).

1.2 Specific techniques used

In this section we briefly introduce the monolingual and cross-lingual metrics used by CIDER-CL, as well as the overall architecture of the ontology aligner.

SoftTFIDF. SoftTFIDF [3] is a hybrid string similarity measure that combines TF-IDF, a token-based similarity widely used in information retrieval [8], with an edit-based similarity such as Jaro-Winkler [11] (although any other could be used instead).

Typically, string comparisons to compute TF-IDF weights are based on exact matching (after some normalisation or tokenisation step). The idea of SoftTFIDF is to use an edit distance instead to support a higher degree of variation between the terms. In particular, we use Jaro-Winkler similarity with a 0.9 threshold, above which two strings are consider equal. SoftTFIDF measure has proved to be very effective when comparing short strings [3]. In our case, the corpus used by SoftTFIDF is dynamically created with the lexical information coming from the two compared ontologies (extracting their labels, comments, and URI fragments).

CL-ESA. For cross-lingual ontology matching we propose the use of CL-ESA [10], a cross-lingual extension of an approach called Explicit Semantic Analysis [6] (ESA). ESA allows comparing two texts semantically with the help of explicitly defined concepts. This method uses the co-occurrence information of the words from the textual definitions of the concepts using, for instance, the Wikipedia articles. In short, ESA extends a simple bag of words model to a *bag of concepts* model. Some reports [2] have demonstrated the good behaviour of CL-ESA for certain tasks such as cross lingual information retrieval.

To compare two texts in different languages semantically, Wikipedia-based CL-ESA represents the two texts as vectors in a vector space that has the Wikipedia titles (articles) as dimensions, each vector in its own language specific Wikipedia. The magnitude of each title/dimension is the associativity weight of the text to that title. To quantify this associativity, the textual content of the Wikipedia article is utilized. This weight can be calculated by using different methods, for instance, TF-IDF score.

---

[3] http://alignapi.gforge.inria.fr/format.html

[4] http://alignapi.gforge.inria.fr/edoal.html

For implementing CL-ESA, we followed an information retrieval-based approach by creating a Lucene inverted index of the Wikipedia extended abstracts that exist in all the considered languages i.e., EN, ES, NL, and DE. To create the weighted vector of concepts, the term is searched over the index of the respective languages to retrieve the top associated Wikipedia concepts and the Lucene ranking scores are taken as the associativity weights of the concepts to the term. We used DBpedia URIs [1] as the pivot between cross-lingual Wikipedia spaces and to identify a Wikipedia concept no matter the language.

Scheme of the Aligner. Briefly explained, the alignment process is as follows (see Figure 1):
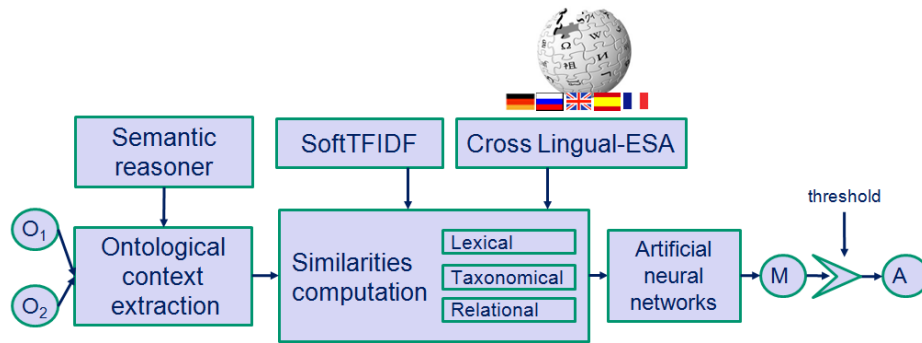

Fig. 1. Scheme of the matching process.

1. First, the ontological context of each ontology term is extracted. This process is enriched by applying a lightweight inference mechanism[5], in order to add more semantic information that is not explicit in the asserted ontologies.
2. Second, similarities are computed between different parts of the ontological context. In particular, ten different features are considered: *labels, comments, equivalent terms, subterms, superterms, direct subterms, direct superterms* (both for classes and properties) and *properties, direct properties*, and *related classes* (for classes) or *domains, direct domains*, and *ranges* (for properties).
3. Third, the different similarities are combined within an ANN to provide a final similarity degree. CIDER-CL uses four different neural networks (*multilayer perceptrons* in particular) for computing monolingual and cross-lingual similarities between classes and properties, respectively.
4. Finally, a matrix (M in Figure 1) with all similarities is obtained. The final alignment (A) is then extracted from it, finding the highest rated one-to-one relationships among terms and filtering out the ones below the given threshold.

---

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

Implementation. Some datasets used in OAEI campaigns are open and the reference alignments available for download. We have used part of such data to train our system. In particular, we chose a subset of the OAEI'11 benchmark track to train our neural networks for the monolingual case. We used the whole dataset but excluding cases 202 and 248-266, which present a total absence or randomization of labels and comments (however their variations, 248-2, 248-4, etc., were not excluded). Also the reference alignments of the conference track, which are also open, were added to the training data set.

The use of the benchmark track for adjusting the ANNs is motivated by the fact that it covers many possible situations and variations well, such as presence or absence of certain ingredients (labels, comments, etc.) or the effect of aligning at different granularity levels (flattened/expanded hierarchies), etc. Further, we add also data of the conference track to include training data coming from "real world" ontologies.

For the cross-lingual case, we trained the neural networks with a subset of the ontologies of the OAEI'13 Multifarm track (in EN, ES, DE, and NL): cmt, conference, confOf, and sigkdd. Comparisons were run among the different ontologies in the different languages, excluding comparisons between the same ontologies. Due to the slow performance of CL-ESA, we decided to perform an attribute selection analysis to discover which features have more predictive power. As result, we limited the system to compute these features for classes: *labels, subterms, direct superterms, direct subterms*, and *properties*; while for properties they were limited to: *labels, subterms*, and *ranges*.

CIDER-CL has been developed in Java, extending the Alignment API [4]. To create and manipulate neural networks we use Weka[6] data mining framework. For SoftTFIDF we use SecondString[7] and for CL-ESA we use the implementation developed by the Monnet project[8], which is available in GitHub as open source[9].

1.3 Adaptations made for the evaluation

The weights and the configuration of the neural networks remained constant for all the tests and tracks of OAEI'13, as well as the threshold. In particular we selected a threshold of 0.0025. The intention of such a small value was to promote recall over precision (while filtering out some extremely low values). Therefore, later filtering can be made to perform a threshold analysis as the organisers of some OAEI tracks do (e.g., conference track).

Some minor technical adaptations were needed to integrate the system into the Seals platform, like solving compatibility issues with the libraries used by the Seals wrapper.

1.4 Link to the system and parameters file

The version of CIDER-CL used for this evaluation (v1.1) was uploaded to the Seals platform: http://www.seals-project.eu/ . More information can be found at CIDER-CL's website http://www.oeg-upm.net/files/cider-cl .

---

[6] http://www.cs.waikato.ac.nz/ml/weka/

[7] http://secondstring.sourceforge.net/

[8] http://www.monnet-project.eu/

[9] https://github.com/kasooja/clesa

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

For OAEI'13 campaign, CIDER-CL participated in all the Seals-based tracks[10]. In the following, we report the results of CIDER-CL for benchmark, conference, anatomy, and multifarm tracks. For the other tracks, the system was not fit for the type of evaluation (e.g., interactive track) or could not complete the task (e.g., library). Details about the test ontologies, the evaluation process, and the complete results for all tracks can be found at the OAEI'13 website[11].

2.1 Benchmark

This year, a blind test set was generated based on a seed ontology of the bibliographic domain. Out of the 21 systems participating in this track, CIDER-CL was within the three best systems in terms of F-measure. In particular, the obtained results were:

$$Precision(P)=0.85, Recall(R)=0.67 \text{ and } F\text{-Measure}(F)=0.75$$

Compare to the F=0.41 of edna, a simple edit distance-based baseline. In addition, confidence-weighted measures were also computed for those systems that provided a confidence value. In almost all cases the results were worse, as it was also the case of CIDER-CL: P=0.84, R=0.55, and F=0.66

Also the time spent in the evaluation was calculated. CIDER-CL took 844±19 seconds, which was slower than most of the systems (the median value was 173 sec) although still far from the slowest one (10241±347 sec).

2.2 Conference

In this track, several ontologies from the conference domain were matched, resulting in 21 alignments. In this case the organisers explored different thresholds and selected the best achievable results. This test is not blind and the participants have the reference alignments at their disposal before the evaluation phase.

Two reference alignments were used in this track: the original reference alignment (ra1) and its transitive closure (ra2). Two baselines (edna and string equivalence) were computed for comparison. Notice that the results for CIDER-CL in this track are merely illustrative and should not be taken as a proper test, due to the fact that part of the training data of its neural networks came from the conference track reference alignments (i.e., training and test data coincide partially).

Out of the 25 systems participating in this track (some of them were variations of the same system), CIDER-CL performance was close to the average. The results were:

---

[10] http://oaei.ontologymatching.org/2013/seals-eval.html
[11] http://oaei.ontologymatching.org/2013

test ra1 (original): P = 0.75, R = 0.47, and F = 0.58 with threshold= 0.14
test ra2 (entailed): P = 0.72, R = 0.44, and F = 0.55 with threshold= 0.08

CIDER-CL was in the group of systems that performed better than the two baselines for ra2 and between the two baselines for ra1. The results for ra1 illustrates an improvement with respect to the results obtained by its previous version (CIDER v0.4) for the same test at OAEI'11 (F=0.53). The runtime was also registered: CIDER-CL took less than 10 minutes for computing the 21 alignments. The other systems ranged from 1 minute to more than 40.

2.3 Anatomy

This year, the current version of CIDER-CL completed the task and gave results for the first time. In fact, in previous editions of OAEI, CIDER gave time-outs and the tool did not finish the task, due to the big size of the involved ontologies. The results are:

$$P = 0.65, R = 0.73, F = 0.69, R+ = 0.31$$

These results are below the average of the overall results (F-Measure ranging from 0.41 to 0.94, with a median value of 0.81). An "extended recall" (R+) was also computed, that is, the amount of detected non-trivial correspondences (that do not have the same normalized label). For this metric CIDER-CL behaved better than the median value (0.23). In terms of running time, CIDER-CL was the third slowest system (12308 sec) in this track, after discarding those that gave time-out.

2.4 Multifarm

This track is based on the alignment of ontologies in nine different languages: EN, DE, ES, NL, CZ, RU, PT, FR, and CN. All pairs of languages (36 pairs) were considered in the evaluation. A total of 900 matching tasks were performed. There were 21 participants in this track, 7 of them implementing specific cross-lingual modules as it was the case of CIDER-CL.

The organisers divided the results in two types: comparisons between different ontologies (type i) and comparisons between the same ontologies (type ii). The result summary published by the organisers aggregates the individual results for all the language pairs. In the case of CIDER-CL this hampers direct comparisons with other systems, owing to the fact that CIDER-CL only covers a subset of languages (EN, DE, ES, NL) and non produced alignments in other languages penalised the overall results. For this reason we have filtered the language specific results to consider only such subset of languages. The averaged results for CIDER-CL are:

type i (different ontologies): P = 0.16, R = 0.19, F = 0.17
type ii (same ontologies): P = 0.82 , R = 0.16, F = 0.26

For type ii, CIDER-CL got the 4th best result overall in terms of F-Measure and the 3rd best result in the set of systems implementing specific cross-lingual techniques (the results for such systems ranged from F = 0.12 to F = 0.44 for the referred subset of

languages). On the other hand, for type i CIDER-CL was in 8th position out of the 21 participants, although in the last place among the set of systems implementing cross-lingual techniques (F-measure of the other techniques ranged from 0.17 to 0.35).

3 General comments

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

3.1 Comments on the results

CIDER-CL obtained good results for the benchmark track (third place out of 21 participants). This shows that our system performs well for domains in which the system could be trained with available reference data. Also that SoftTFIDF is suitable for ontology matching. In contrast, the results for the anatomy track were relatively poor. This shows that creating a general purpose aligner based on our technique is not immediate. Adding more training data from other domains would help to solve this.

The results from the multilingual track are rather modest, but the fact that even the best systems scored low illustrates the difficulty of the problem. We consider that the use of CL-ESA is promising for cross-lingual matching, but it will require more study and adaptation to achieve better results.

3.2 Discussions on the way to improve the proposed system

More reference alignments from "real world" ontologies will be used in the future for training the ANNs, in order to cover more domains and different types of ontologies. Regarding the cross-lingual matching, there is still room for continuing improving the use of CL-ESA to that end. We plan also to combine this novel technique with other ones such as machine translation.

Time response in CIDER-CL is still an issue and has to be further improved. In fact CIDER-CL 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 2013 test cases

The variety of tracks and the improvements introduced along the years makes the campaign very useful to test the performance of ontology aligners and analyse their strengths and weaknesses. Nevertheless, we miss blind tests cases in more tracks, which would allow a fair comparison between systems.

4 Conclusion

CIDER-CL is a schema-based alignment system that compares the ontological context of each pair of terms in the aligned ontologies. Several elementary comparisons are

computed and combined by means of artificial neural networks. Monolingual and cross-lingual metrics are used in the matching.

We have presented here some results of the participation of CIDER-CL at OAEI'13 campaign. The results vary depending on the track, from the good results in the bench-mark track to the relatively limited behaviour in anatomy, for instance. We confirmed that the proposed technique, based on ANNs, is suitable in conjunction with SoftTFIDF metric for monolingual ontology matching. The use of CL-ESA metric for cross-lingual matching is promising but requires more study.

References

1. C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. DBpedia - a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165, Sept. 2009.
2. P. Cimiano, A. Schultz, S. Sizov, P. Sorg, and S. Staab. Explicit versus latent concept models for cross-language information retrieval. In *Proceedings of the 21st international jont conference on Artifical intelligence*, IJCAI'09, pages 1513–1518, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.
3. W. W. Cohen, P. Ravikumar, and S. E. Fienberg. A comparison of string distance metrics for name-matching tasks. In *Proc. Workshop on Information Integration on the Web (IIWeb-03) @ IJCAI-03, Acapulco, Mexico*, pages 73–78, Aug. 2003.
4. J. Euzenat. An API for ontology alignment. In *3rd International Semantic Web Conference (ISWC'04), Hiroshima (Japan)*. Springer, November 2004.
5. J. Euzenat and P. Shvaiko. *Ontology matching*. Springer-Verlag, 2007.
6. E. Gabrilovich and S. Markovitch. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *In Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 1606–1611, 2007.
7. J. Gracia, J. Bernad, and E. Mena. Ontology matching with CIDER: Evaluation report for OAEI 2011. In *Proc. of 6th Ontology Matching Workshop (OM'11), at 10th International Semantic Web Conference (ISWC'11), Bonn (Germany)*, volume 814. CEUR-WS, Oct. 2011.
8. 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.
9. M. Smith. *Neural Networks for Statistical Modeling*. John Wiley & Sons, Inc., New York, NY, USA, 1993.
10. P. Sorg and P. Cimiano. Exploiting wikipedia for cross-lingual and multilingual information retrieval. *Data Knowl. Eng.*, 74:26–45, Apr. 2012.
11. W. E. Winkler. String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. In *Proceedings of the Section on Survey Research*, pages 354–359, 1990.