

Spider: Bringing Non-Equivalence Mappings to OAEI

Marta Sabou¹ and Jorge Gracia²

¹ Knowledge Media Institute (KMi), The Open University, UK
r.m.sabou@open.ac.uk

² IIS Department, University of Zaragoza, Spain
jogracia@unizar.es

Abstract. With the large majority of existing matching systems focusing on deriving equivalence mappings, OAEI has been primarily focused on assessing such kind of relations. As the field inevitably advances towards the discovery of more complex mappings, the contest will need to reflect such changes as well. In this paper we present Spider, a system that provides alignments containing not only equivalence mappings, but also a variety of different mapping types (namely, subsumption, disjointness and named relations). Our goal is both to get an insight into the functioning of our system and, more importantly, to assess the current support for dealing with non-equivalence mappings in the OAEI contest. We hope that our observations will contribute to further enhance the procedure of the contest.

1 Presentation of the system

1.1 State, purpose, general statement

Our purpose was to investigate two concrete issues related to non-equivalence mappings.

1. *Do non-equivalence mappings bring a good addition to alignments made up of purely equivalence mappings?* We have investigated this question during OAEI'07 without being able to draw a clear conclusion. During that contest we submitted an alignment made up of only non-equivalence mappings to the FAO food task. While the expert evaluation gave a general insight in the performance of the matcher itself, due to the large size of the dataset it was impossible to draw a conclusion on whether this alignment was a useful increment to simply equivalence mappings. Also, such a study was hampered by the fact that our system contained only non-equivalence mappings. Based on these lessons, this year we have teamed up with a matcher which provides equivalent mappings only. Also, and most importantly, we have restricted our study to the smallest test of the contest, the benchmark test set. This should give us a clearer understanding on the amount and quality of non-equivalence mappings that can be discovered in addition to equivalence mappings. Such results, if positive, will motivate us (and hopefully others) in building hybrid systems which can go beyond equivalence mappings.

2. *Is the OAEI procedure capable of handling non-equivalence mappings?* Since the majority of matching systems focus on equivalence mappings, the OAEI contest is currently geared towards evaluating such mappings. However, as the field inevitably evolves towards more complex mappings, this will probably have an impact on OAEI as well. We want to assess the current support for evaluating non-equivalence mappings and, based on our experience, offer our ideas about potential future improvements.

1.2 Specific techniques used

Our system combines two concrete subsystems. First, the CIDER algorithm is used to derive equivalence mappings. Second, this alignment is extended with non-equivalence mappings derived by Scarlet.

CIDER, also described in this workshop, uses a semantic similarity measure to compare the concepts of the two input ontologies. This schema-based method combines different elementary techniques, as linguistic similarities or vector space modelling, to compare the ontological context of each of the involved terms. The discovered correspondences that score below a certain threshold are filtered out of the resultant alignment. This measure has been adapted from the authors' earlier work on word sense disambiguation [6]. More details about CIDER are provided in [2].

Scarlet [5] automatically selects and explores online ontologies *to discover relations between two given concepts*. For example, when relating two concepts labeled *Researcher* and *AcademicStaff*, Scarlet 1) identifies (at run-time) online ontologies that can provide information about how these two concepts inter-relate and then 2) combines this information to infer their relation. In [5] the authors describe two increasingly sophisticated strategies to identify and to exploit online ontologies for relation discovery. Hereby, we rely on the first strategy that derives a relation between two concepts if this relation is defined within a single online ontology, e.g., stating that *Researcher* \sqsubseteq *AcademicStaff*. Besides subsumption relations, Scarlet is also able to identify disjoint and named relations. All relations are obtained by using derivation rules which explore not only direct relations but also relations deduced by applying subsumption reasoning within a given ontology. For example, when matching two concepts labeled *Drinking Water* and *tap.water*, appropriate anchor terms are discovered in the TAP ontology and the following subsumption chain in the external ontology is used to deduce a subsumption relation: *DrinkingWater* \sqsubseteq *FlatDrinkingWater* \sqsubseteq *TapWater*.

1.3 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/>

Scarlet can be accessed online and downloaded from: <http://scarlet.open.ac.uk/>

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

Our results can be found at <http://kmi.open.ac.uk/people/marta/oaei/SPIDER.zip>.

2 Results

We have focused on test sets 3xx as these propose the comparison of real-life ontologies and contain a few non-equivalence mappings in the reference alignment. The rest of the tests in this set do not make sense for Scarlet as comparison is sought between modified versions of the same ontology.

2.1 Results Computed by Organizers

The evaluation of the benchmark alignments consists in an automatic comparison to a manually built reference alignment. The reference alignments for cases 3xx contain mostly equivalence mappings. The alignments for cases 301, 302 and 303 also contain a few subsumption relations between the matched ontologies but these are not enough to evaluate a significant part of our alignment which contains non-equivalences. A good way to practically demonstrate this is to compare the results obtained by CIDER and Spider. As it is visible from Table 1, despite the fact that the second alignment is more complex, numerically speaking, the results are worse. Indeed, as expected, while recall increases for those cases where the reference alignments also contain subsumption relations, precision is heavily affected.

Test Set	CIDER		Spider	
	Prec.	Rec.	Prec.	Rec.
301	0.88	0.59	0.27	0.67
302	0.90	0.56	0.26	0.75
303	0.77	0.77	0.08	0.81
304	0.95	0.95	0.16	0.95

Table 1. Results computed for the 3xx tests by comparison to the reference alignments.

We think that the current evaluation should be improved to better accommodate non-equivalence mappings, because, as we will see in the next sections, such mappings can bring an important addition to alignments made up only of equivalences.

2.2 Results Obtained with Other Modalities

We have performed a manual evaluation of the non-equivalence mappings obtained for the 3xx benchmark tests. Given the simplicity of the domain, the evaluation was performed by a single person, one of the authors. Therefore we regard these results as indicative only until a more extended multi-evaluator evaluation will be performed.

Test Set	Total mappings	True mappings	True redundant	True - non redundant	False	Overall Precision	Core Precision
301	112	71	30	41	41	63%	50%
302	116	64	11	53	52	55%	50%
303	458	233	84	149	225	51%	40%
304	386	255	128	127	131	66%	50%

Table 2. Results for the manual evaluation of the 3xx benchmark tests.

For each alignment we have assessed the true and false mappings. In the case of true mappings, we differentiate between redundant and non-redundant mappings. Redundant mappings are those mappings which could be deduced by considering the source ontologies and the equivalence alignment. Formally: $mapping_r \models (O_s, O_t, A_{=})$. Obviously, one can argue that these mappings are of little interest as they could be easily deduced.

Consequently, we compute two kinds of precision values. First, the overall precision takes into account all true mappings, whether redundant or not. Second, the core precision excludes the redundant mappings and considers only the non-redundant ones.

The results are shown in Table 2. The overall precision of the alignment is in the range of 50% and 70% thus correlating to earlier findings performed in different domains [5]. If we do not take redundant mappings into account, the precision of the remaining alignment drops to an average of 50%. This shows that, on average, at least half of the mappings in the extended alignment are correct and thus bring an addition to the purely equivalence based mappings. The number of non-redundant true mappings shows the net increment that this tool brings to the equivalence based alignment. Even for small ontologies as those in the benchmark test, we were able to find novel mappings that could have not been derived from the existing ontologies.

2.3 Error Analysis

We have performed an error analysis in order to identify possible ways in which the alignment’s precision could be improved. Table 3 shows the various types of false mappings and their numbers. We have identified four types of false mappings. First, we found mappings that simply stated a false statement about the domain and which were derived from ontologies containing such incorrect domain knowledge (e.g., $Person \sqsubseteq Event$). Another class of false errors were derived due to incorrect anchoring of the source concepts into the online domain ontologies. For example, the mapping $Academic \sqsubseteq Lecturer$ is false, because in the context of the source ontology *Academic* refers to academic publications and not to a type of employees.

The largest set of false mappings were due to relations derived by inheritance from high-level, generic concepts such as *Thing*. For example, we established a mapping called *editorBook* between *Book* and *Report* because in one ontology³ the following path has been followed:

³ <http://www.aifb.uni-karlsruhe.de/WBS/meh/mapping/data/swrcla.rdf>

Test Set	False	False anchor	False generic-c	False part of	Total False
301	4	4	33	-	41
302	15	5	32	-	52
303	81	23	118	3	225
304	77	15	36	3	131

Table 3. Error analysis.

Report \sqsubseteq *Publication* \sqsubseteq *Root* and *editorBook(Book, Root)*

Indeed, due to the particular modeling followed by the swrc1a.rdf ontology, it has lead to 75 out of 118 false mappings in this category. These were mostly caused by properties which had *Root* as a domain or range and which were then inherited by the subclasses of *Root*.

Finally, some subsumption mappings were established between concepts that are in fact related by meronymy relations (e.g., *Journal* \sqsupseteq *Article*).

3 General comments

3.1 Comments on the results

The results obtained by our in-house evaluation show that it is possible to obtain alignments containing not only equivalent mappings and that the precision of the non-equivalence mappings is around 60% if we take into account redundant mappings and 50% when the redundant mappings are excluded. This results are encouraging and could be further improved as discussed in the next section.

3.2 Discussions on the way to improve the proposed system

Our current efforts focus on automatic ways for filtering out a significant part of the incorrect mappings. First, we are currently finalizing a more complex anchoring mechanism for Scarlet which goes beyond lexical comparison of strings. An initial feasibility study of improving anchoring has been presented in [1]. Second, some of the heuristics we observed could be used to build filters for excluding potentially false mappings - e.g., mappings relying on very long inheritance paths and/or containing generic concepts such as *Root*, *Thing*, *Agent*, etc.

Now that we have a better insight in the additions one can bring to the alignment based on equivalence mappings only, we will consider building a hybrid matcher which better integrates CIDER and Scarlet instead of just running them sequentially. For example, we wish to include the process of checking whether a mapping is redundant or not within the matching process itself (and not just running it on the final alignment).

3.3 Comments on the OAEI 2008 procedure

Our main conclusion related to OAEI is that it would be beneficial to extend it with support for evaluating non-equivalence mappings as well, possibly for all test cases.

The evaluation of alignments, in general, is a difficult task, with many open questions persisting even in the case of equivalence mappings. Non-equivalence mappings introduce an extra level of complexity as, unlike in the case of equivalence mappings, it is difficult (if not impossible) to manually build reference alignments for such cases. Therefore the automatic assessment of such mappings by following the model used for equivalence mappings is not feasible.

An interesting line of work described in [4] and [7] is to use logical reasoning in order to assess the quality of mappings in a given alignment. Their assumption is that mappings which introduce logical inconsistencies are likely to be incorrect and should be eliminated. We think that this work could be used for automatically assessing some of the non-equivalence mappings.

One of the shortcomings of the above mentioned methods is that they are hampered by underspecified ontologies. Also, so far, they are only able to assess the quality of subsumption and equivalence mappings and have not considered disjoint and generic, named relations.

To address these problems we envision the development of a set of methods which rely on a different paradigm. Namely, they would use the Web (or other knowledge sources, e.g., Wikipedia, online ontologies) to predict the correctness of a given mapping automatically. For example, in their recent paper [3], Gracia and Mena have shown that web-based relatedness measures can judge the correctness of a mapping almost as well as humans do. Their measure reached the same conclusion as human evaluators for 80% of a corpus of 160 mappings. This is a remarkable result given the fact that inter-evaluator agreement between humans is often as low as 70%. While the results of such evaluation might be slightly less precise than human evaluation, a key advantage is that all submitted alignments would be judged in a uniform and robust way.

3.4 Proposed new measures

Based on the lessons from our evaluation, we think that making a clear distinction between redundant and non-redundant mappings is an important issue and also a process that can be easily automated. According to this we have devised two precision values.

Having said that, we think that the measures will highly depend on the concrete evaluation procedure that will be used, so the measures we presented here might not be feasible in combination with an automated evaluation.

4 Conclusion

We have investigated two main issues related to non-equivalence mappings. First, we have shown that our system can bring an important number of non-redundant and correct non-equivalence mappings to an equivalence based alignment. Our error analysis has also shown that more can be done in order to filter out obviously false mappings.

Second, we have pointed out that, the current OAEI procedure is biased towards dealing with equivalence mappings and as such there is no suitable support for evaluating non-equivalence mappings (except the manual evaluations offered by some of the tests). We think that as the field evolves towards more complex mappings this needs to

be taken into account by OAEI. As a first step, we think it could be useful to investigate a set of methods that could be used for automatic mapping evaluation.

Acknowledgements

This work was funded by the NeOn IST-FF6-027595 and CICYT TIN2007-68091-C02-02 projects.

References

1. J. Gracia, V. Lopez, M. d'Aquin, M. Sabou, E. Motta, and E. Mena. Solving Semantic Ambiguity to Improve Semantic Web based Ontology Matching. ISWC Workshop on Ontology Matching, 2007.
2. J. Gracia and E. Mena. Ontology Matching with CIDER: Evaluation Report for the OAEI 2008. ISWC Workshop on Ontology Matching, 2008.
3. J. Gracia and E. Mena. Web-based Measure of Semantic Relatedness. In *Proc. of 9th International Conference on Web Information Systems Engineering (WISE 2008), Auckland (New Zealand)*, volume 5175, pages 136–150. Springer Verlag LNCS, ISSN 0302-9743, ISBN 978-3-540-85480-7, September 2008.
4. C. Meilicke, H. Stuckenschmidt, and A. Tamilin. Reasoning Support for Mapping Revision. *Journal of Logic and Computation*, 2008.
5. M. Sabou, M. d'Aquin, and E. Motta. Exploring the Semantic Web as Background Knowledge for Ontology Matching. *Journal on Data Semantics*, XI, 2008.
6. 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.
7. P. Wang and B. Xu. Debugging Ontology Mappings: A Static Approach. *Computing and Informatics*, 27(1), 2008.