The Relevance of Reasoning and Alignment Incoherence in Ontology Matching

Christian Meilicke

KR& KM Research Group, University of Mannheim, Germany
christian@informatik.uni-mannheim.de

Abstract. Ontology matching has become an important field of research over the last years. Although many different approaches have been proposed, only few of them are committed to a well defined alignment semantics. As a consequence, the possibilities of reasoning based approaches are not yet exploited to their full extent. We argue that a reasoning based approach, backed by a well defined semantics, might not only improve ontology matching but will also be necessary to solve certain problems that hinder the progress of the whole field. In particular, we focus on the notion of alignment incoherence as a first step towards understanding and exploiting the capabilities of reasoning in ontology matching.

1 Problem Statement

Ontology matching has been identified as key element towards realizing the vision of the semantic web by bridging the gap between different conceptualizations of similar or overlapping domains. Therefore, it has become an important field of research over the last years. The core problem is the detection of semantic relations between concepts, properties or instances of two ontologies. A number of approaches have been proposed (see [3] for an overview) to solve this problem. In addition, many important side issues have been tackled, e.g. collaborative ontology matching, matcher selection, versioning of alignments, etc. Nevertheless, the role of semantics and in particular the role of reasoning in the context of ontology matching has been neglected or dealt with in an imprecise manner. This can be explained by the fact that many ontologies typically used as testcases within the matcher community are hierarchies that do not contain expressive constructs such as disjointness or property restrictions. Some of these ontologies, for example, are not modeled by knowledge engineers but based on automatically converting poorly structured knowledge bases into OWL.[1] Thus, any reasoning based approach will not exploit its full potential on most accepted testcases. Another explanation can be found in the history of ontology matching which can be seen as a further development of matching database or XML schemas. An example for a system that has originally been designed for schema matching and has been extended towards matching ontologies is the COMA++ matching system [5]. Ontologies are in most cases interpreted within a

[1] In the Ontology Alignment Initiative [1], for example, we find a large deal of ontologies that are automatically generated OWL representations of Thesauri or Web directories or have been converted from formalisms like OBO.

DL based semantics, where we find clear definitions of notions like e.g. entailment, satisfiability and inconsistency. Contrary to this, a comparable model based semantics is often not given in the context of database semantics or cannot be applied directly to the field of ontology matching.

In our work we are concerned with the role of semantics in the context of ontology matching. In particular, we focus on the problem of alignment incoherence and show that a reasoning based approach backed by a well-defined semantic can be applied in different contexts to enrich and optimize the evaluation and quality of ontology matching. However, it will turn that such an approach has to deal with a set of interrelated problems as described in the following.

2 Related Work and Motivation

Due to the lack of space, an exhaustive overview on related work cannot be given. Instead, we focus on specific aspects to clarify the need for a reasoning based approach.

Evaluation The classical measures used to evaluate ontology matching results are precision and recall. They are based on comparing an alignment $\mathcal{A}$ against a reference alignment $\mathcal{R}$ on a syntactic level. Suppose, that there exists a correspondence c in $\mathcal{A}$ not included in $\mathcal{R}$ but derivable from $\mathcal{R}$. On a pure syntactic level c has to be counted as incorrect correspondence, which is obviously an unreasonable choice. In many accepted testcases the classical syntactic approach has no negative effects, since $\mathcal{A}$ and $\mathcal{R}$ are restricted to equivalences where the deductive closure coincides with the deductive reduction. The problem has been dealt with in [2] under the notion of semantic precision an recall. Although, the problem has been precisely described, there are still some issues that have to be solved, e.g. the fact that the deductive closure of an alignment is an infinite set. Computing semantic precision and recall will become even more important, when complex correspondences (as described in [15]) come to the fore.

Semantics Lots of research is concerned with theoretical issues related to alignment semantics. In [13], for example, the authors formally investigate different alignment formalisms. Nevertheless, both lines of research - the theory of alignment semantics vs. the practically oriented task of ontology matching - are only loosely coupled. Most matching systems are not committed to a certain semantics. Neither do they ensure the coherence of generated alignments nor do they allow to generate the deductive closure with respect to a certain semantics. We find a counterpart in the procedure of the Ontology Alignment Initiative (OAEI) [1], where in none of the subtasks results are expected to agree on a certain semantics.

Incoherence We already claimed that no matching system ensures the coherence of its results. One might raise the objection that system like e.g. ASMOV and Lily make both use of a debugging component (see section 10 in [1]), which is also referred to as semantic verification. To the best of our knowledge, these debugging mechanisms are pattern based, which means that the final alignment is checked for certain patterns that indicate erroneous combinations of correspondences. Non of these patterns is explicitly backed by a certain semantics and their motivation, although comprehensible, is only an ad hoc explanation. Notions like completeness and soundness are not applicable with respect to these patterns. [2]

3 Approach and Methodology

Euzenat and Shvaiko have listed the problem of reasoning with alignments as one of the ten outstanding challenges in ontology matching [14]. The previous section indicates that there exists no unique approach to cope with this challenge. Thus, we can only focus on a specific aspect from the bunch of interrelated problems. We propose an approach centered around the notion of alignment incoherence. In particular, we are concerned with the following research questions.

(1) How can the incoherence of an alignment be measured and interpreted?
(2) Is it possible to use incoherence to automatically improve the results of a matcher?
(3) Can incoherence be used to support manual alignment revision?

First of all, our approach requires to define the notion of alignment incoherence. Given a semantics $\mathcal{S}$ and an alignment $\mathcal{A}$, we defined $\mathcal{A}$ as incoherent with respect to $\mathcal{S}$, whenever there exists a satisfiable class in one of the aligned ontologies that becomes unsatisfiable due to interpreting $\mathcal{A}$ in terms of $\mathcal{S}$. This definition abstracts from the specifics of $\mathcal{S}$ and is applicable for any semantics $\mathcal{S}$ that supports the notion of satisfiability. An answer to the first question requires to elaborate a theory of alignment incoherence and to apply the resulting measures in the evaluation process. For answering the second question we have to compare an alignment generated by a matching system against a subset of this alignment where all incoherences have been resolved automatically. The third issue requires to take at least two aspects into account. On the one hand we have to compare the resulting alignment against the results of an unsupported revision. On the other hand we have to analyze in how far the effort for the human in the loop can be decreased.

4 Results

The origin of our work, starting two years ago, can be found in a paper that describes how Distributed Description Logics (DDL) can be used to reason about ontology alignments [16]. Based on the proposed framework, we used DDL as alignment semantics in our first experiments. In [10] we focused on automatically repairing alignments generated by different state of the art matching systems. In particular, we applied a greedy approach to resolve alignment incoherences by removing the correspondence with lowest confidence from minimal sets of conflicting correspondences. This approach increased the precision of the results between 2% and 19% depending on the particular matching system. In [12] we used a similar approach but imposed this choice on the user. In

[2] S-Match [4], on the contrary, employs sound and complete reasoning procedures. Nevertheless, the underlying semantic is restricted to propositional logic due to the fact that ontologies are interpreted as tree-like structures.

particular, we observed that the effort of a reasoning supported revision is in average reduced to 40% compared to a complete manual revision.[3]

Since DDL is mainly motivated by its possibilities to reason in a distributed environment, we proposed another semantic motivated by the use case of merging ontologies. This semantics can be seen as the natural translation of correspondences into DL-axioms. We first defined this semantics in a study where we applied the approach to synthetic ontologies, experimenting with a decision procedure that aims to remove a minimum number of correspondences weighted via their confidences [7] instead of the greedy approach explained above. A comparison of several methods for choosing a coherent subset of an alignment and their application to real world ontologies can be found in [6], where we observed minor improvements often based on a trade off between precision and recall. Our experiments indicate that a heuristic that aims to achieve a global optimum will result in better choices whenever we have alignments with a high number of conflicts.

Recently, we focused on the first research question and proposed four ways to measure the degree of incoherence of an alignment [8]. Even though our definitions are based on merging ontologies, they are directly applicable to any other alignment semantics that supports the notion of satisfiability. We first applied this approach to the submissions of the OAEI 2008 conference track, where we observed that in average about 15% of the correspondences in an alignment have to be removed for logical reasons.[4] In particular, it turned out that even systems with debugging components could not ensure the coherence of their alignments. Moreover, we proved that a certain way to measure the incoherence of an alignment $\mathcal{A}$ results in a strict upper bound for the precision of $\mathcal{A}$ which was an unexpected and important result of our research.

5 Conclusions and Future Work

We motivated our work by general consideration about the role of reasoning in ontology matching and argued that there are many open problems related to the issue. Thus, we decided to tackle a specific family of interrelated problem centered around the notion of alignment incoherence. Answering the research questions listed in section 3 requires diverse experimental studies. Thus, we implemented a tool that allows to reason about alignments to repair incoherent alignments and to measure its degree of incoherence. The implemented techniques, partially derived from the field of ontology debugging, require a full fledged reasoning approach that will become problematic for large ontologies. In future work we have to tackle this problem by optimizing our reasoning strategies.

We started our work with automatically debugging ontology alignments. It turned out that we increased the quality of an alignment in most cases. Nevertheless, we often observed a trade off between precision and recall. We assume that this is based on the fact that we worked on the final alignment generated by a matcher, e.g. extracted from a similarity matrix. The benefits of our reasoning component can be further increased

[3] A summary of both approaches and their results can be found in [11].

[4] Notice that this result clearly shows that the capabilities of reasoning in ontology matching are not yet exploited to their full extent. Detailed results are reported in [1].

if correspondences eliminated by the extraction method of the matching system, would also be available for the decision process of our component. To check this hypotheses we are currently implementing our own matching system where the input to the reasoning component is not restricted to a previously filtered set of correspondences. In addition, we extended our system to support subsumption correspondences and correspondences between properties. First results show that the property extension significantly increases the number of detectable coherence conflicts, resulting in an increased recall for removing incorrect correspondences.

As continuation of the work reported in [12], we used a component of our reasoning system to support the revision of the preliminary reference alignments of the OAEI conference track [9]. It turned out that informing the user about conflicts between correspondences often points to problems that are hard to detect without support. In future work we plan to implement a Protege-Plugin that both exploits the benefits of a reasoning based approach and Protege functionalities for visualizing context information.

References

1. C. Caracciolo, J. Euzenat, L. Hollink, R. Ichise, A. Isaac, V. Malaise, C. Meilicke, J. Pane, P. Shvaiko, H. Stuckenschmidt, O. Svab-Zamazal, and V. Svatek. Results of the OAEI 2008. In *Proc. of the ISWC 2008 Workshop on Ontology Matching*, 2008.
2. J. Euzenat. Semantic precision and recall for ontology alignment evaluation. In *Proc. of the 20th International Joint Conference on Artificial Intelligence*, 2007.
3. J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer, 2007.
4. F. Giunchiglia, M. Yatskevich, and P. Shvaiko. Semantic matching: Algorithms and implementation. *Journal on Data Semantics*, 2007.
5. S. Massmann, D. Engmann, and E. Rahm. Coma++: Results for the ontology alignment contest OAEI 2006. In *Proc. of the ISWC 2006 Workshop on Ontology Matching*, 2006.
6. C. Meilicke and H. Stuckenschmidt. Analyzing mapping extraction approaches. In *Proc. of the ISWC 2007 Workshop on Ontology Matching*, 2007.
7. C. Meilicke and H. Stuckenschmidt. Applying logical constraints to ontology matching. In *Proc. of the 30th German Conference on Artificial Intelligence*, 2007.
8. C. Meilicke and H. Stuckenschmidt. Incoherence as a basis for measuring the quality of ontology mappings. In *Proc. of the ISWC 2008 Workshop on Ontology Matching*, 2008.
9. C. Meilicke, H. Stuckenschmidt, and O. Svab-Zamazal. A reasoning-based support tool for ontology mapping evaluation. In *Demo-Paper accepted at the ESWC*, 2009.
10. C. Meilicke, H. Stuckenschmidt, and A. Tamilin. Repairing ontology mappings. In *Proc. of the 22th Conference on Artificial Intelligence*, 2007.
11. C. Meilicke, H. Stuckenschmidt, and A. Tamilin. Reasoning support for mapping revision. *Journal of Logic and Computation*, 2008.
12. C. Meilicke, H. Stuckenschmidt, and A. Tamilin. Supporting manual mapping revision using logical reasoning. In *Proc. of the 23th AAAI Conference on Artificial Intelligence*, 2008.
13. L. Serafini, H. Stuckenschmidt, and H. Wache. A formal investigation of mapping language for terminological knowledge. In *Proc. of the International Joint Conference on Artificial Intelligence*, 2005.
14. P. Shvaiko and J. Euzenat. Ten challenges for ontology matching. 2008.
15. H. Stuckenschmidt, L. Predoiu, and C. Meilicke. Learning complex ontology mappings - a challenge for ILP research. In *Proc. of the 18th International Conference on ILP*, 2008.
16. H. Stuckenschmidt, L. Serafini, and H. Wache. Reasoning about ontology mappings. In *Proc. of the ECAI 2006 Workshop on Contextual Representation and Reasoning*, 2006.