Knowledge-Based Systems Group

Institute of Information Systems | Faculty of Informatics | Vienna University of Technology

Expose for Master Thesis

Towards an Alternative Approach for Combining Ontology Matchers

Advisor: Priv.-Doz. Dipl.-Ing. Dr. Axel Polleres

Simon Steyskal
0828067
066 937
Software Engineering & Internet Computing

June 13^{th}, 2013

.....
Priv.-Doz. Dipl.-Ing. Dr. Axel Polleres

Expose for Master Thesis

1 Problem Description

Mapping between the internal models of software systems is a crucial task in almost all data and system integration scenarios, keeping computer scientists busy for the last decades with still no standardized solution in sight.

A possible application domain is mapping of object-oriented models from different software systems as mentioned in [9]. This use case is important for infrastructure providers like Siemens, which uses various software tools related to configuration like software configuration of electronic parts in end customer deployments, each with its own underlying internal model. These tools use internal, partially overlapping representations (mostly object-oriented models) of the final products and their parts, which need to be transferred from one configuration step to the next, thus requiring alignment of the underlying models.

To solve such alignment problems, ontologies and ontology matching [4, 1, 7] seem to be a good starting points particularly for aligning object-oriented models, since ontologies can represent (and be extracted from) object-oriented models fairly naturally.

In the past decade, the field of ontology matching has matured. Many different ontology matchers participated in the last years' OAEI [1] campaigns, exposing different strengths and weaknesses. The different tracks provided by the OAEI target different matching problems and since every matching tool has its specific special techniques and features, good performance in a specific sub-track could still mean a bad performance in another one. So, intuitively, combining the results of a heterogeneous set of ontology matchers, taking the "best off-the-shelf matcher for the problem at hand" seems to be a promising approach for a "universal matcher" without the need for designing a new matcher from scratch.

Unfortunately, (i) neither deciding which matcher is best suitable for a set of given ontologies nor (ii) combining ontology matchers in a semi-automatic human-guided matching process is trivial, thus possible solutions for these issues have to be devised. The development of one such combined approach to ontology matching is the primary topic of present thesis.

2 Expected Result

Aim of this master thesis is the evaluation of different current state-of-the-art ontology matchers and the development of a prototype called *Mix'n'Match* which should be able to combine off-the-shelf ontology matchers in an iterative manner.

Our hypothesis is that it should be possible to reuse gathered knowledge from each matching round in terms of found alignments and then enrich the to be matched ontologies with this additional knowledge. Therefore it should be possible to develop a combined ontology matcher, which performs better on average than a single matcher on a number of heterogeneous ontology matching problems (from the OAEI campaign), without the need to choose a single matcher, but by iteratively combining results of different matchers, and feeding them as support into sub-

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

sequent runs of those matchers.

Apart from reference alignments, background knowledge in the form of specific SPARQL queries which represent constraints valid in object-oriented models should be used to improve the results retrieved by matching ontologies which are based on such real world object oriented models shall be clarified.

3 Method and Approach

After doing extensive research in the main topics of this master thesis like semantic web technologies and ontology matching, which will be described in the first chapters of this work, selected ontology alignment tools will be investigated to document their strengths and weaknesses and their usability for the combined matching approach described in this thesis.

The main part of this thesis is developing an approach for combining different ontology matchers and therefore be able to generate an aggregated set of alignments which are an improvement in terms of precision, recall and F-measure in comparison to the results obtained by the off-the-shelf matchers themselves. Hence the evaluation results gathered during the investigation of the single off-the-shelf matchers are taken into account to form a heterogeneous set of ontology matchers. To be able to combine those different matchers, several combination strategies will be developed and evaluated and finally implemented using Java.

Since matching ontologies based on real world schemas is another important part of this thesis, the applicability of using SPARQL for refining the alignment results will be investigated. Therefore selected SPARQL queries, representing constraints which holds for object oriented models will be constructed and integrated into the previously developed framework.

The evaluation of the approach developed in this thesis as well as the evaluation of its extension using SPARQL for refining the alignment results will be evaluated and compared to other off-the-shelf ontology matchers in the end of this thesis. To be able to do so, datasets provided by the Ontology Alignment Evaluation Initiative (OAEI) as well as from a real world use case from SIEMENS in the form of railway domain ontologies will be used.

4 State of the Art

The approach of combining matching techniques in an iterative process which reruns them till no further alignment pairs were found is not new per se. Tools like Anchor-Flood [5] or ASMOV [6] are based on such a framework but in contrast to the approach planned for this master thesis neither one of them use off-the-shelf matchers for retrieving alignments. The benefit of using whole tools instead of specific matching techniques is obvious, on the one hand it can highly increase the flexibility of the approach since single off-the-shelf matcher can easily be integrated and segregated from the matching process and on the other hand since matching tools evolve over time and perform better and better by themselves, Mix'n'Match will also benefit from this evolution based on its framework.

Besides of combining ontology matchers during the matching step, some approaches only focus on the combination of alignment sets of already matched ontologies. But these approaches doesn't rerun the matching process with the gathered additional knowledge again, although they

received very good results especially by using machine learning techniques for the combination of the alignments [3, 2, 8].

5 Connection to Software Engineering & Internet Computing

There are several lectures in the curriculum of Software Engineering & Internet Computing, which have their focus on topics related to those of this master thesis.

Topic: Semantic Web
Lecture: *Introduction to Semantic Web(188.399)*
General introduction into semantic technologies and ontologies as well as Description Logics and tool support for the Semantic Web.
Lecture: *Semantic Web Technologies(184.729)*
Deeper insights into languages, standards and technologies of the Semantic Web with focus on formal semantics. Furthermore an introduction into inference and querying on the Web, Semantic Web search and Linked Data.
Lecture: *Semi-Automatic Information and Knowledge Systems(188.387)*
Principles of ontology alignment, semantic integration and reasoning as well as an introduction to Semantic Web and ontology engineering

Topic: Model-Driven Engineering
Lecture: *Model Engineering(188.923)*
Precise explanation of different model transformation algorithms and techniques as well as the definition of UML and metamodeling.

6 Contents

References

[1] Namyoun Choi, Il-Yeol Song, and Hyoil Han. A survey on ontology mapping. *SIGMOD Record*, 35(3):34–41, 2006.

[2] AnHai Doan, Jayant Madhavan, Robin Dhamankar, Pedro Domingos, and Alon Y. Halevy. Learning to match ontologies on the semantic web. *VLDB J.*, 12(4):303–319, 2003.

[3] Kai Eckert, Christian Meilicke, and Heiner Stuckenschmidt. Improving ontology matching using meta-level learning. In *Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications*, ESWC 2009 Heraklion, pages 158–172, Berlin, Heidelberg, 2009. Springer-Verlag.

[4] Jérôme Euzenat and Pavel Shvaiko. *Ontology Matching*. Springer, 2007.

[5] Md. Seddiqui Hanif and Masaki Aono. An efficient and scalable algorithm for segmented alignment of ontologies of arbitrary size. *J. Web Sem.*, 7(4):344–356, 2009.

[6] Yves R. Jean-Mary, E. Patrick Shironoshita, and Mansur R. Kabuka. Ontology Matching with Semantic Verification. *Web Semantics*, 7(3):235–251, September 2009.

[7] Yannis Kalfoglou and W. Marco Schorlemmer. Ontology mapping: the state of the art. *Knowledge Eng. Review*, 18(1):1–31, 2003.

[8] Paulo Maio, Nuno Bettencourt, Nuno Silva, and João Rocha. Evaluating a confidence value for ontology alignment. In Pavel Shvaiko, Jérôme Euzenat, Fausto Giunchiglia, and Bin He, editors, *Proceedings of the 2nd International Workshop on Ontology Matching (OM-2007) Collocated with the 6th International Semantic Web Conference (ISWC-2007) and the 2nd Asian Semantic Web Conference (ASWC-2007), Busan, Korea, November 11, 2007*, volume 304 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2007.

[9] Erhard Rahm and Philip A. Bernstein. A survey of approaches to automatic schema matching. *The VLDB Journal*, 10:334–350, 2001.