Proceedings of I-SEMANTICS '08
Graz, Austria, September 3-5, 2008

Interoperability Issues, Ontology Matching and MOMA

Malgorzata Mochol
(Free University of Berlin, Königin-Luise-Str. 24-26, 14195 Berlin, Germany
mochol@inf.fu-berlin.de)

Abstract: Thought interoperability has been gaining in importance and become an essential issue within the Semantic Web community, the main challenge of interoperability and data integration is still ontology matching. With this in mind, we wish to contribute to the enhancement of (semantic) interoperability by supporting the ontology matching issue; we propose an evaluation framework for matching approaches that contributes to the resolution of the data integration and interoperability issue by creating and maintaining awareness of the link between matchers and various ontologies.
Key Words: interoperability, ontology matching, MOMA Framework
Category: I.2.4, K.6.3

1 Introduction

Over the last few years, interoperability has gained in importance and become an essential issue within the Semantic Web (SW) community. The more standardized and widespread the data manipulation tools are, including a higher degree of syntactic interoperability, the easier and more attractive using the SW approach has become. Though SW technologies can support the unambiguous identification of concepts and formally describe relationships between concepts, Web developers are still faced with the problem of semantic interoperability, which stands in the way of achieving the Web's full potential. The main problem with semantic interoperability is that the cost of its establishing, due to the need for content analysis, is usually higher than what is needed to establish syntactic interoperability [Decker et al. 00]. Semantic interoperability is necessary before multiple applications can truly understand data and treat it as information; it will thus be, according to [Decker et al. 00], a *sine qua non* for the SW. To achieve semantic interoperability, systems must be capable of exchanging data in such a way that the precise meaning of the data is readily accessible, and the data itself can be translated by any system into a form that the system understands [Heflin and Hendler 00]. Hence, a central problem in (semantic) interoperability and data integration issues in the SW vision is schema and ontology matching and mapping [Cruz and Xiao 03].
Considering these problems and the current situation in SW research, we wish to contribute to the enhancement of (semantic) interoperability by providing support to the ontology matching issue. On the one hand, the number of use cases for ontology matching justifies the great importance of this topic in the

SW [Euzenat et al. 04]. On the other hand, the development and existence of tried and tested ontology matching algorithms and support tools will be one of the crucial issues that may have a significant impact on future development, for instance, the vast SW-based information management systems. Furthermore, it has also turned out that different matching algorithms are better suited for matching different sets of ontologies. Today it takes an expert to determine the best algorithm and a decision can usually be made only after experimentation, so as a result the necessary scaling and off-the-shelf use of matching algorithms are not possible. To tackle these problems we have developed an evaluation framework – Metadata-based Ontology Matching (MOMA) Framework – that helps to resolve the data integration and interoperability issue by creating and maintaining awareness of the link between matching algorithms and various ontologies. Our approach allows for a more flexible deployment of matching algorithms (depending on the particular requirements of the application to which the matchers are to be utilized) and the selection of suitable approaches performed prior to the execution of a matching algorithm.

The remain of this paper is organized as follows: In Sec. 2 we specify the main open issues within the ontology matching domain. Then, we outline a possible solution to tackle these problems by introducing the MOMA Framework (Sec. 3); we elaborate the main use cases together with the high-level architecture and sketch the evaluation results. Sec. 4 summaries the work and provides some issues for the future work.

2 Ontology Matching Domain

Despite of the pervasiveness of ontology matching and although the development of tools to assist in the matching process has become crucial for the success of a wide variety of information management applications [Doanet al. 04], the matching process is still largely conducted by hand, in a labor-intensive and error-prone process. There is still a number of short, middle, and long-term problems that need to be resolved in order to overcome the interpretability and heterogeneity issues and to realize the vision of a fully developed SW.

No overarching matching: Many methods and tools are under development to solve specific problems in the SW however, none of these solutions can be deployed due to all the existing problems. This statement is also true in the ontology matching field, as there is no overarching matching algorithm for ontologies capable of serving all ontological sources and new approaches tackle only minor aspects of the "larger" problem in the matching domain or are mere "stop gaps" [Fürst and Trichet 05].

"Unused" reuse: The ontology matching field continue to pay little notice to a strategy based on reusing existing matching, merging, and aligning approaches. Consequently, the reuse of these semantic-based approaches have not yet been analyzed satisfactorily within the SW realm. Our experiences collected during the development of ontology-based applications [Bizer et al. 05, Garbers et al. 06, Niemann et al. 06] confirm previous findings in the literature that building such applications is still a tedious process, as a result of the lack of tested and proved support tools and that reusing of existing methods within new application contexts is currently not extensively discussed in depth. When implementing an application using a matching approach, the corresponding algorithm is typically built from scratch, and only small, marginal attempts to reuse existing methods are made.

"Evil" diversity: Since much time and effort have been spent on the development of new ontology alignment and matching algorithms, the collection of such algorithms is still growing. For this reason, we are all confronted with the same problem: there is an enormous amount of divergent work from different communities that claims some sort of relevance to ontology mapping, matching, alignment integration, and merging [Kalfoglou et al. 03]. Given this multiplicity, it is difficult to identify both the problem areas and the solutions. In this view, the diversity of matching approaches is a weakness rather than a strength. Part of the problem is also the lack of a comprehensive survey, a standard terminology, obscure assumptions or undisclosed technical details, and the dearth of evaluation metrics [Kalfoglou et al. 03].

"Holes" in the approaches Despite an impressive number of research initiatives in the matching field, current matching approaches still feature significant limitations [Shvaiko 04, Giuchiglia et al. 04, Melnik et al. 02, Madhavan 01]: current matching approaches, though containing valuable ideas and techniques, are tailored to particular types of ontologies and are confined to specific schema types [Do et al. 02]; they need to be customized for a particular application setting (like schema and data integration); they cannot be applied across various domains with the same effect; they do not perform well (or have not yet been tested) on inputs with heterogeneous (graph) structures or on large-sized inputs.

Lack of infrastructure: After years of extensive research and development of numerous matching approaches, it is time to deploy some of the procedures, techniques, and tools created [Zhao 07]. Thus, what is required are techniques and tools capable of handling different ontological sources [Castano et al. 04] and the requirements of the emerging applications. Furthermore, users need help in choosing an appropriate matcher or combining the most appropriate matchers for their particular use [Euzenat and Shvaiko 07].

Additional issues: Beside the problems mentioned, there are many other aspects of a general nature which need to be resolved. There is the question of what should be matched based upon what needs to be found. It is also important to avoid performing blind matching while knowing when to stop the process. To this end, it is necessary to adapt the systems, i.e. adjust it, not to the data to be processed, but to the issue that needs to be resolved with the given matcher.

Though we most definitely will not be able to “solve all these problems and save the world” in our research work, we will tackle some of these issues. We have concentrated on the selection of suitable matching approaches, which, in our opinion, is one of the main issues in the ontology matching domain. By proposing a framework that supports the selection of relevant matching algorithms suitable w.r.t the given specification while taking into account the definition of the appropriate criteria for the decision making process, we address the issues of “lack of infrastructure” and “evil diversity” and, in some measure, the problems in terms of “unused” reuse and “no overarching matching”.

3 MOMA Framework

Due to the above mentioned issues and the fact that the existing matching algorithms cannot be optimally used in ontology matching tasks, as envisioned by the SW community, we need a strategy to remedy the weaknesses and take advantages of the particularity of the various approaches in the selection of suitable matchers; we need a matcher evaluation process, which performing prior to the execution of a matching algorithm, allows a selection of suitable algorithms. Thus, we have developed a ***Metadata-based Ontology MAtching (MOMA) Framework*** which on the basis of dependencies between algorithms and the ontology types on which the former are able to process successfully, the capabilities of existing matching algorithms and factors that influence the matching tasks recommends the appropriate matchers for a given application.

3.1 Main Use Cases

During discussions with SW-based application developers, researchers, and experts in the ontology matching domain, we noticed there were two types of users interested in the matcher application and the utilization of relevant supportive tools. Consequently, we made a conscious decision to ensure that our MOMA Framework serves both developers/computer scientists by supporting them in their implementation and research work, and the matching providers, enabling them to utilize our matching tool in different service tasks. To this end, we have classified the MOMA users into two main groups: (i) *human matcher users* (e.g. ontology engineers, SW application developers[1]) - the process of choosing the suitable approach can occur both manually and (semi-)automatically; (ii)*machine matcher users* (e.g. service/matching providers) - in this case, the

[1] In terms of the ontology development, which is mostly not conducted by people with a high level of expertise in the ontology matching domain, there is a need to aid them in selecting and applying ontology management tools, incl. matching algorithms.

process of choosing suitable matchers is envisioned to be performed only (semi-)automatically. Considering these use cases the objective of MOMA is to supply a tool that offers methods to support the manual and (semi-)automatic detection of suitable matchers (manual and (semi)-automatic mode, respectively).

3.2 High-level Architecture

The MOMA Framework (cf. Fig. 1) consists of three main components: (i) *M*ultilevel *C*haracteristic for *MA*tching approaches (*MCMA*) - utilized to describe the matching algorithms, their incoming sources, and feasible output, together with application features in which the matching approach is to be applied; (ii) *Knowledge Base* that includes information (based on the MCMA structure) regarding existing matchers which may be selected for application and sources that are to be matched; it also contains some rule statements that describe the dependencies between the matching approaches and ontologies; (iii) *Selection Engine* that is responsible for the matcher selection which conducts manually or (semi)-automatically the matcher determination process. Therefore, in the following, we analyze MOMA w.r.t the manual and (semi-)automatic selection.

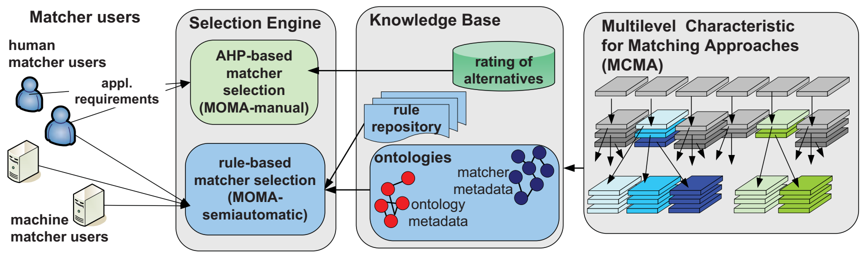

Figure 1: *MOMA Framework*

Matcher characteristic: To find suitable matching approaches for a particular application, it is important to recognize cross application needs and define a matcher characteristic that allows comparison of different approaches and the subsequent selection of suitable algorithms. For this reason, we have collected various features of matching approaches (together with input, output, costs, etc.) and targeted application, identified those that have an impact on the selection of an appropriate matching approach, and finally build a matcher characteristic - *Multilevel Characteristic for Matching Approaches (MCMA)* - that serves as the basis for the final decision regarding the suitability issue [2].

Manual approach: To allow the manual selection of matchers and thereby serve the human matcher users, we have adopted the Analytic Hierarchy Process (AHP) [Saaty 00], which uses pairwise comparisons along with a semantic and ratio scale to assess the decision maker's preferences [Guitouni and Martel 98].

[2] For the detailed description of the MCMA, the reader is referred to [Mochol et al. 06]

AHP allows decision makers to model a complex problem in a hierarchical structure in order to show the relationships of the goal objectives (find suitable matchers), sub-objectives (MCMA), and alternatives (different matchers). The crucial step in the manual selection is the comparison and weighting of sub-objectives. This means that the users of the MOMA manuell define the requirements of their application concerning their specification of the potential suitable matching approach by weighing the properties defined within the MCMA in the pairwise comparison. By reducing complex decisions to a series of pairwise comparisons and synthesizing the results, decision-makers arrive at the optimal decision based on a clear rationale [Saaty 99]. In our case, the users of the MOMA Framework obtain a list of matchers ordered by their suitability to the given context[3].

Semi-automatic approach: In order to serve the machine users, we need to provide a (semi-)automatic selection process. As a possible solution, we propose a framework based on rules and defined in the form of ontologies metadata: *ontology metadata* - additional information regarding the ontologies (based on MCMA), like size or representation language, and *matcher metadata* - information regarding existing ontology matching algorithms; to determine automatically which algorithms suit the concrete inputs, explicit knowledge is needed concerning the dependencies between these algorithms and the structures on which they operate. We have formalized this knowledge in terms of dependency rule-statements - *rule repository.* The core of the MOMA Framework within an automatic mode is the *selection engine* which is responsible for the decision making process by means of rules grouped into a rule repository; for a given set of ontologies to be matched, the selection engine must decide (concerning the ontology and matcher metadata and by firing the predefined rules) which matching algorithms are applicable w.r.t the given context.

3.3 Evaluation

The evaluation process started with the expert-based evaluation of the MCMA, which resulted in refinement of the preliminarily defined characteristic and, in turn, in a revised MCMA, which has been used within both matcher selection approaches. The further evaluation was dedicated to the accuracy of MOMA predictions and was connected with the usage of MOMA framework in real-world situations. We conducted the evaluation on the basis of the test cases from the Ontology Alignment Evaluation Initiative (OAEI)[4] which aims to establish a consensus for the evaluation of alignment approaches by setting up an evaluation campaign and benchmark tests to assess the strengths and weaknesses of the alignment approaches. The application of the AHP-based MOMA Framework to

[3] For more details regarding AHP-based selection, the reader is referred to [Mochol et al. 06, Mochol et al. 07]

[4] `http://oaei.ontologymatching.org`

the OAEI case studies showed that it produces very relevant results which can serve as a direct basis for the reuse of existing matchers in new ontology-based applications (cf. [Mochol et al. 07]). The evaluation of the rule-based MOMA Framework attested to the fact that the (semi-)automatic matcher selection, which in comparison to the manual approach acts on the much less detailed information, delivers very promising results which can serve as a basic module for further examination of algorithms (cf. [Mochol and Jentzsch 08]).

4 Conclusions and Future Work

In this paper, we propose the MOMA Framework that takes into account the capabilities of existing matchers and suggests appropriate approaches for individual cases. Our framework contributes to data integration and interoperability by maintaining awareness of the link between matching algorithms and a wide variety of ontologies. It is the first step towards the reuse of existing ontology matching approaches that contributes to the more optimal utilization of ontology matching tasks as envisioned by the SW community, tackles the issues of matchers heterogeneity, exploits the valuable ideas embedded in current matching approaches, and supports developers by giving them recommendations regarding suitable matcher solutions. The future work will be mainly dedicated to the development of the web service-based MOMA access and the (semi-)automatical utilization of the recommended matchers in the particular application.

Acknowledgements

This work has been partially supported by the "InnoProfile - Corporate Semantic Web" project funded by the German Federal Ministry of Education and Research (BMBF) and the BMBF Innovation Initiative for the New German Länder - Entrepreneurial Regions.

References

[Bizer et al. 05] Bizer, C., Heese, R., Mochol, M., Oldakowski, R., Tolksdorf, R., Eckstein, R.: "The Impact of Semantic Web Technologies on Job Recruitment Processes"; In Proc. of the 7th Int. Tagung Wirtschaftsinformatik, (2005), 1367-1383.

[Castano et al. 04] Castano, S., Ferrara, A., Montanelli, S.: "Methods and Techniques for Ontology-based Semantic Interoperability in Networked Enterprise Contexts"; In Proc. of the 1st CAiSE INTEROP Workshop On Enterprise Modelling and Ontologies for Interoperability, (2004), 261-264.

[Cohen 98] William W. Cohen: "Integration of Heterogeneous Databases Without Common Domains Using Queries Based on Textual Similarity"; In Proc. 17th International Conference on Management of Data (SIGMOD), (1998), 201-212.

[Cruz and Xiao 03] Cruz, I. F., Xiao, H.: "Using a Layered Approach for Interoperability on the Semantic Web"; In Proc. of the 4th Int. Conference on Web Information Systems Engineering (WISE 2003), (2003), 221-232.

[Doanet al. 04] Doan, A.-H., Madhavan, J., Domingos, P., Halevy, P.: "Ontology Matching: A Machine Learning Approach"; Handbook on Ontologies, Springer, (2004), 385-516.

[Do et al. 02] Do, H. H., Melnik, S., Rahm, E.: "Comparison of Schema Matching Evaluations"; In Proc. of the 2nd International Workshop on Web Databases, vol. 2593 of Lecture notes in Computer Science, (2002), 221-237.

[Decker et al. 00] Decker, S., Melnik, S., van Harmelen, F., Fensel, D., Klein, M. C. A., Broekstra, J., Erdmann, M., Horrocks, I.: "The Semantic Web: The Roles of XML and RDF"; IEEE Internet Computing, 4, 5, (2000), 63-74.

[Euzenat et al. 04] Euzenat, J. et al.: "State of the art on current alignment techniques"; Technical Report Deliv. 2.2.3, Knowledge Web EU NoE, (2004).

[Euzenat and Shvaiko 07] Euzenat, J., Shvaiko, P.: "Ontology Matching"; Springer Verlag, ISBN-3540496114, (2007).

[Fürst and Trichet 05] Fürst, F., Trichet, F.: "Axiom-based Ontology Matching"; In Proc. of the 3rd Int. Conf. on Knowledge Capture, ACM Press, (2005), 195-196.

[Guitouni and Martel 98] Guitouni, A., Martel, J.-M.: "Tentative guidelines to help choosing an appropriate MCDA method"; Europ. Journal of Operational Research, 109, (1998), 501-521.

[Garbers et al. 06] Garbers, J., Niemann, M., Mochol, M.: "A personalized hotel selection engine"(Poster); In Proc. of the 3rd Europ. Sematnic Web Conference, (2006).

[Giuchiglia et al. 04] Giuchiglia, F., Shvaiko, P.: "Semantic Matching"; Knowledge Engineering Review Journal, 18, 3, (2004), 265-280.

[Hu et al. 06] Hu, W., Cheng, G., Zheng, D., Zhong, X., Qu, Y.: "The Results of Falcon - AO in the OAEI 2006 Campaign"; In Proc. of the Int. Workshop on Ontology Matching (OM-2006) colloc. with ISWC2006, (2006), 124-133.

[Heflin and Hendler 00] Heflin, J., Hendler, J.: "Semantic Interoperability on the Web"; In Proc. of Extreme Markup Languages 2000, (2000), 111-120.

[Kalfoglou et al. 03] Kalfoglou Y., Schorlemmer, M.: "Ontology mapping: the state of the art"; The Knowledge Engineering Review (KER), 18, 1, (2003), 1-31.

[Madhavan 01] Madhavan, J., Bernstein, P.A., Rahm, E.: "Generic Schema Matching with Cupid"; In Proc. of the 27th VLDB Conference, (2001), 48-58.

[Melnik et al. 02] Melnik, S., Garcia-Molina, H., Rahm, E.: "Similarity Flooding: A Ver- satile Graph Matching Algorithm and Its Application to Schema Matching"; In Proc. of the 18th International Conference on Data Engineering (ICDE02), (2002).

[Mochol and Jentzsch 08] Mochol, M., Jentzsch, A.: "Towards a rule-based matcher selection"; 6th International Conference on Knowledge Engineering and Knowledge Management Knowledge Patterns (EKAW2008), (to appear 2008).

[Mochol et al. 06] Mochol, M., Jenztsch, A., Euzenat, J.: "Applying an Analytic Method for Matching Approach Selection"; In Proc. of the Int. Workshop on Ontology Matching (OM-2006) colloc. with the ISWC2006, (2006).

[Mochol et al. 07] Mochol, M., Jentzsch, A., Euzenat, J.: "Towards a methodology for selection suitable matching approaches - A Case Study"(Poster); In Proc. of the 4th European Semantic Web Conference (ESWC2007), (2007).

[Niemann et al. 06] Niemann, M., Mochol, M., Tolksdorf,R.: "Improving online hotel search - what do we need semantics for?"; In Proc. of the Semantics 2006, (2006).

[Shvaiko 04] Shvaiko, P.: "Iterative schema-based semantic matching"; Technical Report DIT-04-020, University of Trento, (2004).

[Saaty 99] Thomas L. Saaty: "How to Make a Decision: The Analytic Hierarchy Process"; European Journal of Operational Research, 48, 1, (1990), 9-26.

[Saaty 00] Thomas L. Saaty: "How to Make a Decision"; Models, Methods, Concepts & Aplications of the Analytic Hierarchy Porcess, Kluver Int. Series, (2000), 1-25.

[Zhao 07] Zhao, H.: "Semantic matching. Accros Heterogenous Data Sources"; Communication of the ACM, 50, 1, (2007), 45-50.