Rule-based Mapping in Ontology-based Mediators

(Extended Abstract)

Gunter Saake, Kai-Uwe Sattler
{*saake*|*kus*}*@iti.cs.uni-magdeburg.de*
University of Magdeburg
P.O. Box 4120, 39016 Magdeburg, Germany

Stefan Conrad
conrad@cs.uni-duesseldorf.de
University of Düsseldorf
Universitätsstr. 1, 40225 Düsseldorf, Germany

Integrating data from heterogeneous sources on the Web is an important topic of interest within the database community. Here, several issues arise such as autonomy, heterogeneity as well as scalability and adaptability with regard to a great number of – possibly changing – data sources. Suitable solutions range from simple meta search engines over materialized approaches to mediator systems which answer queries on a global schema by decomposing them, forwarding the subqueries to the source systems and combining the results into a global answer.

In mediator systems of the first generation integration is achieved mainly on a structural level. Data from the diverse sources are combined based on structural correspondences such as membership in classes of the same structure or the existence of common attributes. This works well in more or less homogeneous domains. In scenarios characterized by rather disjunct domains this approach leads to a great number of global classes that again requires detailed domain knowledge in order to be able to formulate the resulting more complex queries.

An alternative is the explicit modeling and usage of domain knowledge in form of semantic meta data, i.e. a vocabulary, a taxonomy, a concept hierarchy, or even an ontology. Similar efforts are known from the Semantic Web community, where a knowledge-based processing of Web documents is to be achieved by adding a semantic layer containing meta data. First results of this work mainly include models and languages for ontologies, e.g. RDF Schema (RDFS), DAML+OIL, and OWL as well as the corresponding technologies. Because of the strong relationships a combination of Semantic Web and mediator approaches seems very promising.

However, a special requirement from data integration is to define a mapping from the ontology layer to the source data, i.e., to specify how a data source supports a certain concept from the ontology both in a structural as well as in a semantic way. This information is necessary for query rewriting and decomposition and has to be provided as part of the registration of a source.

Specifying the mapping by hand is an expensive and error-prone process, especially for complex schemas and/or ontologies. Schema matching approaches as described e.g. in [1, 2] try to reduce the effort by comparing schemas of different sources and identify matchings based on structural correspondences and – to a certain degree – by exploiting information about the actual data.

In our talk, we argue that these approaches can be improved by using declarative rules which

- can be used during matching, even if correspondences are "hidden" due to different names of classes and attributes,
- can deal with subclass hierarchies which are often used for modeling ontologies,
- can be (semi-)automatically derived or refined by analyzing data of already mapped sources.

We present the logic-based foundations of this approach, discuss the relationships to schema matching and sketch ideas of using rules for improving the mapping process. Finally, we discuss the application of this approach for specifying mappings for our ontology-based mediator system YACOB [3] that has been developed for providing integration and query facilities in databases on cultural assets that were lost or stolen during World War II.

References

[1] E. Rahm, and P. A. Bernstein. A survey of approaches to automatic schema matching. *The VLDB Journal*, 10(4):334–350, 2001.

[2] M. A. Hernandez, R. J. Miller, and L. M. Haas. Clio: A Semi-Automatic Tool For Schema Mapping. In Proc. ACM SIGMOD Conference 2001, Santa Barbara, CA, USA, 2001.

[3] K. Sattler, I. Geist, R. Habrecht, and E. Schallehn. Konzeptbasierte Anfragevererarbeitung in Mediatorsystemen. In *Proc. BTW'03 - Datenbanksysteme für Business, Technologie und Web, Leipzig, 2003, GI-Edition, Lecture Notes in Informatics*, pages 78–97, 2003.