Finding Ariadne's thread in the data mediation maze

Christian Drumm

SAP Research Center CEC Karlsruhe
SAP AG
christian.drumm@sap.com

Abstract. In this paper we will present a general scenario describing the implications that the usage of ontologies has on the data mediation problem. Based on this we will present different possible usage scenarios of ontologies in this problem area. The intention of this paper is to be a basis for discussions of the problems of data mediation using ontologies in the context of web service.

1 Introduction

The data mediation problem in the context of web services is concerned with the transformation of a source message M_S which adheres to a schema S_S into a target message M_T that adheres to a schema S_T. To solve a given mediation problem a mapping needs to be created based on the source and the target message schema. In general the creation of such a mapping in very complex, making the task of developing such transformations very strenuous and error prone [1].

Currently the usage of ontologies in the transformation creation process is seen as a possible solution to this problem. This assumption is based on the hope that by creating transformations on the semantic level rather than on the syntax level the creation of the transformations is simplified. In this paper we will present a general scenario describing the implications that the usage of ontologies has on the data mediation problem. Based on this scenario we will present different possible usage scenarios of ontologies in this problem area and their characteristics. The intention of this is to start a discussion on which facets of ontology based data mediation are the most important ones for the success of the overall approach.

The remaining of the paper is organized as follows: first we introduce the generic scenario followed by the different usage scenarios. Finally in section 3 we'll raise some questions as discussion points.

2 The data mediation maze

As stated in the introduction the generic data mediation problem is concerned with the transformation of a source message M_S into a target message M_T. To

solve this problem a mapping capable of performing this transformation needs to be generated based on the source message schema S_S and the target message schema S_T. In the presence of ontologies the step of creating a mapping is split up into several steps (see Fig. 1).

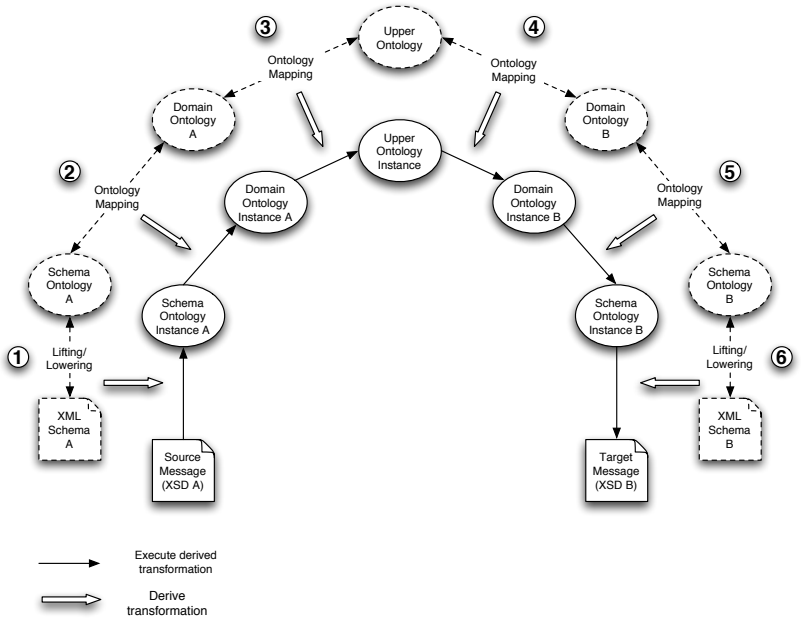


Fig. 1. The most general version of the data mediation problem

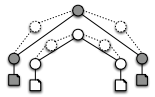
1. First the source message schema S_S needs to be lifted (see [2]) to the ontology level. The result of this step is a source schema ontology. Note that this ontology contains only the information present in the source message schema as this lifting step is mainly a syntactic transformation.
2. The next step is the mapping of the source schema ontology to the source party ontology. The source party ontology is an ontology that describes the world form the perspective of the message sender (e. g. the Daimler Chrysler ontology).
3. In the third step a mapping needs to be performed between the source party ontology and a domain ontology. This domain ontology models the view of the world that several parties have agreed on (e. g. the automotive industry ontology).
4. Now the domain ontology needs to be mapped with the target party ontology of the message receiver.

5. The final steps are to map the target party ontology onto the target schema ontology which has been created using lifting from the target message schema.

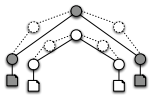
Once all the necessary mappings have been created the transformation that need to be executed to transform a given source message into a target message can easily be derived. This general ontology-based data mediation scenario is depicted in Fig. 1.

2.1 Paths through the maze

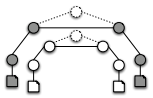
Based on the general scenario describe in the previous section four different usage scenarios can be constructed. These scenarios differ mainly in the number and the kinds of ontologies that are involved in the mapping process and will be briefly described in the following.



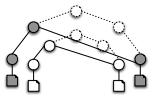
Full scenario (S_1): This is the scenario described in the previous section. Two party ontologies and the domain ontology are involved in this scenario and each of the six mapping steps needs to be performed.



Central ontology scenario (S_2): When the two parties that need to communicate have agreed on one ontology, the steps three and four of the general schema are not necessary anymore as the domain ontology and the two party ontologies are now identical. The source schema ontology is mapped directly onto the agreed central ontology and the agreed central ontology is mapped onto the target schema ontology again.



Two party ontology scenario (S_3): If there is no domain ontology available (e. g. because the two partners don't belong to the same domain) the two party ontologies need to be mapped directly. Instead of the steps three and four on single mapping step between the two party ontologies needs to be performed.



One party ontology scenario (S_4): If only one party ontology is available (e. g. because the second party has not developed an ontology yet) the situation is again different. W.l.o.g we assume that only the source party ontology is available. Now steps three and four can't be performed and the source party ontology needs to be mapped directly onto the target schema ontology instead.

From the previous discussion it is easy to see, that steps one and six, the lifting/lowering steps, are present in each of the four scenarios. This is due to the fact that the message schemas always need to be converted into ontologies first. Furthermore it is obvious that the next step after the lifting is always an ontology mapping step. It is used to align the source schema ontology with the ontology the describes the world-view of the source party. This is either the source domain ontology or the domain ontology depending on the scenario. How complex this mapping step is, depends heavily on how the message schemas and the ontologies were created. There are two kinds of scenarios for the creation of the schemas and the ontologies:

– The "Perfect world" scenarios
– The "Real world" scenarios.

In the "perfect world" scenarios the ontologies are build first. Once the ontologies are in place, the message schemas used for data exchange are derived form them. In these scenarios the mappings between message schemas and party ontologies are easier to handle as there are direct correspondences between the message elements and the concepts in the ontologies. However in the real world scenarios the message schemas are developed independently[1] from the ontologies. This results in a more complex lifting step as there are no direct correspondences between the message elements and the concepts in the ontology.

3 Where is Ariadne's thread?

As described in section 2 using ontologies splits the data mediation problem up into several steps. This fragmentation of the data mediation problem does only make sense if high quality mappings can be created more easily and quickly than with currently available solutions. To achieve this goal it is necessary to provide a high degree of automation for each of the involved steps. The main result of the previous discussion is, that there needs to be a high degree of automation for the first two steps of the general data mediation scenario as these steps are always present in the derived scenarios. However there are still a number of interesting, unsolved questions regarding the data mediation problem left:

– Is it possible to prove that there is an advantage in using ontologies to solve the data mediation problem? Are there situations where mapping the schemas directly is better?
– Which of the paths should be taken in a given situation (e.g. scenario S_3 could also be solved using the approach of S_4)?
– Which parts of the overall process can be fully automates which only partly?

This questions need to be addressed in future research in order to get a better understanding of the overall problem area.

References

1. Do, H.H., Rahm, E.: COMA - a system for flexible combination of schema matching approaches. In: Proc. 28th VLDB Conference. (2002)
2. Maedche, A., Motik, B., Silva, N., Volz, R.: MAFRA - a mapping framework for distributed ontologies. In: Proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management EKAW. (2002)

Acknowledgments

I want to thank M. Altenhofen for the inspiring discussions and his useful remarks.

[1] This happens due to several reasons e. g. the message schemas where developed before the ontologies, the message schemas are specified in a standard, ...