Ontology Driven Data Mediation in Web Services

Meenakshi Nagarajan, Kno.e.sis Center, Wright State University, USA
Kunal Verma, Accenture Technology Labs, USA
Amit P. Sheth, Kno.e.sis Center, Wright State University, USA
John A. Miller, LSDIS Lab, University of Georgia, USA

ABSTRACT

With the rising popularity of Web services, both academia and industry have invested considerably in Web service description standards, discovery, and composition techniques. The standards-based approach utilized by Web services has supported interoperability at the syntax level. However, issues of structural and semantic heterogeneity between messages exchanged by Web services are far more complex and crucial to interoperability. It is for these reasons that we recognize the value that schema/data mappings bring to Web service descriptions. In this article, we examine challenges to interoperability; classify the types of heterogeneities that can occur between interacting services, and present a possible solution for data interoperability using the mapping support provided by WSDL-S, a key driver behind SAWSDL. We present a data mediation architecture using the extensibility features of WSDL and the popular SOAP engine, Axis 2.

Keywords: Axis 2.0; data mediation; mapping; matching; message-level heterogeneities; METEOR-S; ontology; SAWSDL; Web service interoperation

INTRODUCTION

The emergence of Web services and service oriented architectures is leading to new innovative enterprise solutions based on composition of Web services to realize business and scientific processes. So far, much of the research has focused on discovery (Universal Description, Discovery, and Integration), composition (Medjahed, Bouguettaya, & Elmagarmid, 2003; Sivashanmugam, Miller, Sheth, & Verma, 2004b; Zeng, Benatallah, Dumas, Kalagnanam, & Sheng, 2003), and execution (business process execution language for Web Services) of Web services. One of the biggest stumbling blocks in the grand vision proposed by SOA is data heterogeneity between interoperating services. By data or message level heterogeneities, we refer to incompatible formats of messages exchanged by the services. This is not a new problem. Since the inception of federated databases (Sheth, 1990), interoperability among databases with heterogeneous schemas has been a well researched issue (Litwin, 1986; Sheth, 1998). In this article, we discuss mes-
sage level heterogeneities in the Web services domain and present an approach for resolving these heterogeneities. This work was done as a part of the METEOR-S project ("METEOR-S: Semantic Web Services and Processes"), which defines four kinds of semantics for Web services: data, functional, nonfunctional, and execution semantics and utilizes these semantics in the complete lifecycle of Semantic Web processes (publishing, discovery, composition, execution, and monitoring).

Typically enterprise systems are developed over several periods of time by diverse organizations and not necessarily with the same structures and vocabularies. This leads to substantial heterogeneity in syntax, structure, and semantics when it comes to interoperation between these systems. For example, one system may encode performance as grades A-F, while another may use scores ranging from 1-100. A recent approach to interoperate between such systems exposed as Web services has been semantically representing the functional capabilities of the services and then using semantic discovery techniques to find and compose these services into a process. A common fallacy of such an approach is the assumption that a semantic match ensures interoperation.

To appreciate this, consider the case of a process that uses two Web services with heterogeneous message schemas (i.e., the input and output message schemas are incompatible) and the output of the first service is supplied as an input to the second service. The process of resolving these heterogeneities and transforming one message format to another is also referred to as data mediation. A simple solution to achieve data mediation between the services is to manually create a mapping from the first service’s output to the second service’s input (this is the proposed solution of most enterprise integration products in Web services). However, this mapping would have to be created every time services in the process are changed or upgraded, potentially making the number of generated mappings very large. An alternate solution to this problem (which is the approach we use) is mapping the inputs and outputs of the services to a conceptual model and using those mappings for interoperating between the services.

In this article, we classify impediments to data interoperability among Web services by adapting previous work on semantic interoperability in databases (Kashyap, 1996). Our approach uses the support for data mapping provided by SAWSDL (SAWSDL), an upcoming candidate recommendation for semantic annotations for WSDL by a W3C chartered working group. Semantic Annotations for WSDL (SAWSDL) is an effort to define mechanisms by which semantic annotations can be added to WSDL components. Many of the concepts in SAWSDL are based on an earlier effort WSDL-S (WSDL-S), a W3C submission. This work was conducted as a part of WSDL-S before the formation of the SAWSDL group. In this article, we will limit our discussions of SAWSDL as relevant to this work.

The aim of this work is to provide a solution to the problem of Web service interoperation by making incremental changes to Web services tools. Since SAWSDL builds upon existing Web services standards (WSDL), it also allows us to use the extensibility support provided by Axis 2 to implement the data mediation. This article has the following contributions:

- We present a comprehensive, practical approach for resolving data heterogeneities between Web services using semantic domain models, that is, Ontologies.
- We adapt previous work on schema and database integration to compile different kinds of heterogeneities one might encounter during the interoperation of Web services.
- We present a data mediation architecture that is built using the extensible elements of existing Web service standards (WSDL) and tools (Axis 2).

The rest of this article is organized as follows. We motivate the need for resolving data mediation in Web services with an example, classify the possible data or message-level