

Association for Information Systems AIS Electronic Library (AISeL)

AMCIS 2004 Proceedings

Americas Conference on Information Systems
(AMCIS)

December 2004

Federating Heterogeneous Information Systems Using Web Services and Ontologies

Hemant Jain

University of Wisconsin - Milwaukee

Huimin Zhao

University of Wisconsin - Milwaukee

Follow this and additional works at: <http://aisel.aisnet.org/amcis2004>

Recommended Citation

Jain, Hemant and Zhao, Huimin, "Federating Heterogeneous Information Systems Using Web Services and Ontologies" (2004).
AMCIS 2004 Proceedings. 479.

<http://aisel.aisnet.org/amcis2004/479>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Federating Heterogeneous Information Systems Using Web Services and Ontologies

Hemant Jain

University of Wisconsin - Milwaukee
jain@uwm.edu

Huimin Zhao

University of Wisconsin - Milwaukee
hzhao@uwm.edu

ABSTRACT

The objective of this research is to develop an approach to federating heterogeneous information systems using emerging semantic web technologies. This research-in-progress paper describes a system architecture, in which web services are used to resolve system level heterogeneities while ontologies are used to resolve semantics level heterogeneities across local systems. The healthcare domain has been chosen to initially evaluate the utility of the proposed approach.

Keywords

Heterogeneous information system, federated information system, web services, ontology, healthcare

INTRODUCTION

As our ability to build information systems continues to grow, so does the need to integrate the systems we have built. Legacy systems developed over time in different sections of an organization need to be integrated for strategic purposes. Business mergers and acquisitions force information systems previously owned by different institutions to be merged. Information needs to be shared or exchanged across system boundaries of cooperating enterprises. A particular example is the healthcare industry, which demands a massive cooperation of information systems operated by different organizations (Bell and Sethi, 2001).

The objective of this research is to develop an approach to federating heterogeneous information systems. While federated systems have been extensively studied for over two decades (Sheth and Larson, 1990; Bouguettaya, Benatallah and Elmagarmid, 1999), this research capitalizes upon the fruits of past research and overcomes the shortcomings of past approaches using the emerging semantic web technologies, including web services and ontologies. In this research-in-progress paper, we will review past approaches to federating heterogeneous information systems, identify major difficulties encountered by these approaches, and describe how using web services and ontologies can alleviate these difficulties.

INTEGRATING HETEROGENEOUS INFORMATION SYSTEMS

A major challenge in integrating a collection of distributed, heterogeneous, and autonomous information systems is to resolve the inter-system heterogeneities, which exist on two levels: the system level and the semantics level (Ram and Venkataraman, 1999). System level heterogeneities are due to the differences between the individual information systems with respect to their platforms, data models (e.g., pre-relational, relational, object-oriented, semi-structured, and unstructured), data types and sizes, and database management systems (DBMS). Semantics level heterogeneities are due to the differences in perspectives and the terminologies used by the individual information systems.

A traditional approach to resolving system level heterogeneities is to adopt a global (or federated) DBMS, which communicates with the diverse local systems through proprietary bridges, adapters, or common intermediaries, such as the Open Database Connectivity (ODBC). In such a federated database system, the local schemas are integrated into a unified global (or federated) schema, which is typically defined as views based on the local schemas (Sheth and Larson, 1990). This approach relies on mappings between individual database schemas and global schema. These mappings become very difficult to establish because of differences in data models, especially when some of the underlying sources are unstructured or semi-structured. Recently, the wrapper/mediator approach is gaining popularity due to its flexibility in dealing with different types of data models (Wiederhold, 1992; Cluet, Delobel, Simeon and Smaga, 1998; Park, 1999). In this approach, multiple data sources are integrated via wrappers and mediators, which are usually hard-coded programs, rather than a special DBMS. A wrapper is essentially a translator, which is used to convert the format of a data source to a chosen common data model. A mediator integrates the wrappers surrounding different data sources and resolves the semantic conflicts between the wrappers. The wrapper/mediator approach is more flexible than the traditional federated database system approach and is a

particularly attractive way to integrate semi-structured data sources on the Internet. The wrappers and mediators are often implemented as software modules. The most critical and difficult part of building a mediator/wrapper system, however, is the development and assembly of wrappers and mediators.

In both of the above approaches a common global data model, which provides the semantic integrity across local systems, needs to be developed. A traditional approach to resolving semantics level heterogeneities is the bottom-up schema integration approach (Batini, Lenzerini and Navathe, 1986; Kwan and Fong, 1999; Ram and Venkataraman, 1999). Local schemas are translated into a chosen Common Data Model (also called Canonical Data Model) and then compared to identify semantic correspondences and conflicts (e.g., homonyms and synonyms). A unified schema is then defined on top of the transformed local schemas. Semantic correspondences and conflicts are resolved in the mappings between the unified schema and the local schemas. A drawback of this bottom-up approach is the instability of the unified schema. Whenever a new data source needs to be integrated into the system, the unified schema has to be modified and existing systems accessing it may be affected. Additionally, changes in local schemas may also force the unified schema to change. Recently, a top-down ontology based approach is proposed to overcome this drawback (Lee and Siegel, 1996; Park, 1999; Wache, Vögele, Visser, Stuckenschmidt, Schuster, Neumann and Hübner, 2001). This approach integrates heterogeneous data sources using a common ontology (Wache et al., 2001) that describes the common concepts and semantic relationships between concepts that can exist in an application domain. The common ontology is independent of local database schemas. Local database schemas map to the common ontology instead of to each other or to a unified schema derived from them. A major advantage of using a common ontology is that, when new data sources are added into the cooperative system the ontology remains unchanged and the existing local accesses are not affected. A difficulty of this approach is in the development of a common ontology that can cover an entire application domain (Ouksel and Ahmed, 1999). A compromise is to combine both top-down and bottom-up approaches, extending the ontology whenever new concepts found in a local data source cannot be mapped to the concepts in the ontology. As the ontologies become more comprehensive and may be standardized at domain level (e.g., Bodenreider, 2001) the need for such extensions decreases.

PROPOSED APPROACH

In our proposed approach to federating heterogeneous information systems, we adopt and extend the wrapper/mediator approach for resolving system level heterogeneities and the ontology approach for resolving semantics level heterogeneities. We overcome the shortcomings of these approaches by using the emerging service oriented architecture, web services technology (Hansen, Madnick and Siegel, 2002), and standard ontologies. Figure 1 contrasts the proposed approach to previous approaches to integrating heterogeneous information systems.

Recently, there has been increased interest in developing both general-purpose ontologies and domain-specific ontologies. For example, WordNet is a widely used general purpose ontology, the Unified Medical Language System (UMLS) is a dominating ontology for the medical domain (Bodenreider, 2001; Burgun and Bodenreider, 2001; Leroy and Chen, 2001), and Health Level Seven (HL7) provides XML-based standards for the healthcare domain (www.hl7.org).

Web services are expected to greatly enhance interoperability across systems. Based upon emerging XML-based open standards such as SOAP, WSDL, UDDI, and BPEL4WS, web services allow loosely coupled systems to be quickly built by assembling existing or new application components wrapped and published as web services. Offering a language-neutral, environment-neutral computing model, web services technology and standards are expected to make it easier to integrate heterogeneous systems through the Internet both within and across enterprises (Gottschalk, Graham, Kreger and Snell, 2002; Hansen et al., 2002).

Figure 2 outlines preliminary system architecture for federating heterogeneous information sources using web services and ontologies, using the healthcare domain as an illustrative example. The major characteristics of this architecture are briefly described as follows.

- A wrapping web service converts the format of each data source into a standard XML format and annotates each element using a concept in a general-purpose ontology such as WordNet or a domain-specific ontology such as UMLS and HL7.
- A mediating web service accepts user accesses, dynamically discovers and assembles appropriate wrapping services, translates user accesses into requests to individual wrapping services, integrates the results returned from the wrapping services, and presents the integrated results to the user.
- SOAP is used to pass XML messages between the mediating service and the wrapping services. Specifications and capabilities of the services are defined with WSDL and published in a public or private UDDI registry.

- Both general-purpose ontologies and domain-specific ones are used to provide the user with multiple ways to access the federated system. Ontologies and mappings between the concepts in ontologies and the semantic elements accessible through wrapping services are stored in a semantics repository, which is managed by a DBMS.
- Users can access the federated information system in several ways. Those who are familiar with a particular underlying information system can use their familiar terminology that is specific to that system. Professionals who are familiar with the domain can use the terminology of a domain-specific ontology. Other average users may use the terminology of a general-purpose ontology. The mediating service will translate between these different terminologies based on the annotations provided by the wrapping services. For example, in the healthcare domain, previous users of a local system, technicians who are familiar with the standard terminology of HL7, physicians who are familiar with the terminology of UMLS, and patients who can only use ordinary terms available in a general-purpose ontology such as WordNet, will all be able to access the federated information system using their familiar terms.
- Another advantage of the proposed approach is that complex business processes, not just data, can potentially be integrated using the service oriented architecture.

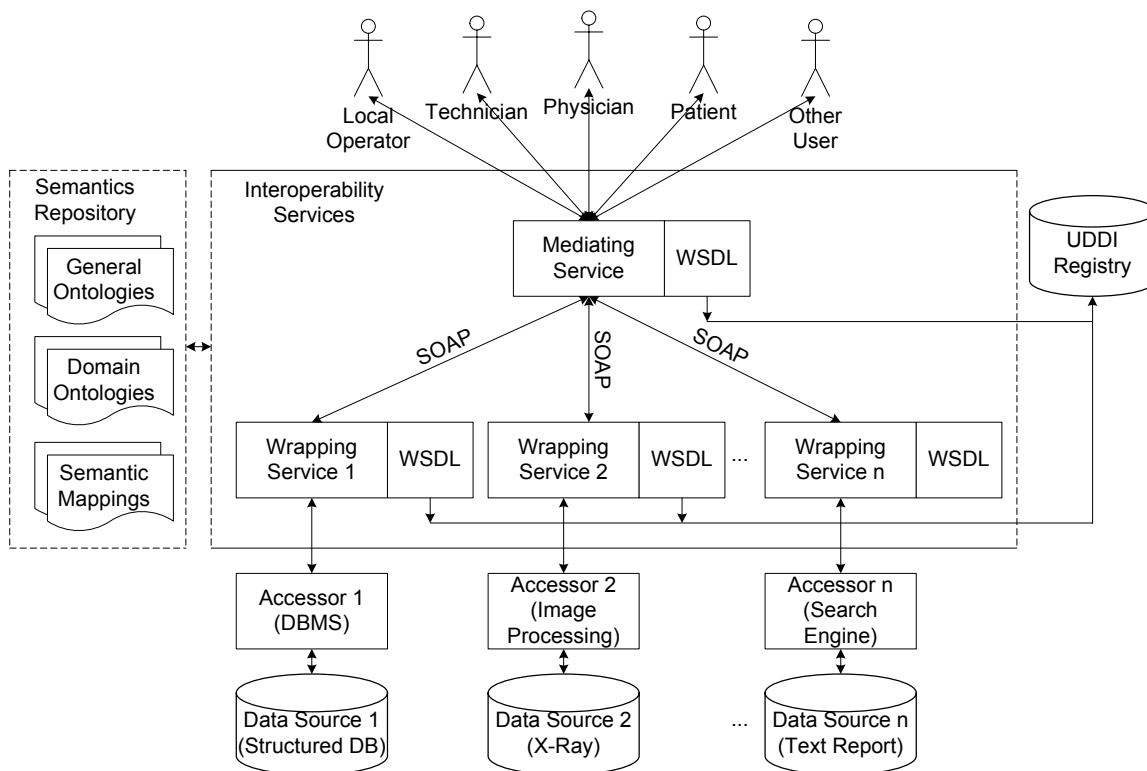


Figure 2. System Architecture of a Federated Information System

CONCLUSION AND FUTURE RESEARCH PLAN

In this paper, we have described an approach to federating heterogeneous information systems, which utilizes web services and ontologies to overcome the shortcomings of past approaches in resolving system level and semantics level heterogeneities. This research is still in an early stage. The realization of the proposed approach requires work on various related issues. Some of these issues are identified here:

- Identification of appropriate services for each of the local systems. The local systems are generally designed to fulfill particular needs in the organization. They generally do not focus on providing services or information to other systems

(except the set of reports they are designed to generate). Based on the information collected, stored and processed by a local system and the standard ontologies for the business domain it should be possible to identify a set of services that will be useful to other systems and various units inside and outside the organization. A systematic methodology for identifying these services needs to be developed. We are currently in the process of developing such a methodology.

- Dynamic service discovery and assembly. The specifications and capabilities of the services provided by each local system (identified above) can be described in a directory/repository similar to UDDI registries. In this case a user request for integrated information needs to be broken down into individual service requests using semantic mappings between the standard ontologies and the services. These service requests can then be used in a dynamic service discovery process to identify appropriate services provided by the local systems. Development of generalized techniques for mapping user requests to individual service requests based on standard ontologies is an interesting research problem. Similarly, development of effective dynamic service discovery algorithm is an active area of research. Another challenging problem is the assembly of information returned by various services into a coherent answer to the user request.

We are planning to address some of these issues in the future. We are planning to use several information systems in the healthcare domain as example to help develop these concepts and evaluate the utility of the approach. While we have chosen to initially realize and evaluate the approach in the healthcare domain, we plan to apply the approach in other domains, if our initial evaluation generates encouraging results.

Successful realization of the proposed architecture can have significant impact on practice. It can change the orientation of future local application development to incorporate definition of services provided by the application. It can also impact Enterprise Application Integration (EAI), a multi billion-dollar industry.

REFERENCES

1. Batini, C., Lenzerini, M., and Navathe, S. B. (1986) A Comparative Analysis of Methodologies for Database Schema Integration, *ACM Computing Surveys*, 18, 4, 323 - 364.
2. Bell, G. and Sethi, A. (2001) Matching Records in a National Medical Patient Index, *Communications of the ACM*, 44, 8, 83 - 88.
3. Bodenreider, O. (2001) Medical Ontology Research - A Report to the Board of Scientific Counselors of the Lister Hill National Center for Biomedical Communications, National Library of Medicine. Available online at <http://etbsun2.nlm.nih.gov:8000/publis-ob-offi/pdf/2001-MOR-BoSC.pdf>.
4. Bouguettaya, A., Benatallah, B., and Elmagarmid, A. (1999) An Overview of Multidatabase Systems: Past and Present, in: A. Elmagarmid, M. Rusinkiewicz, and A. Sheth, Eds. *Management of Heterogeneous and Autonomous Database Systems*. Morgan Kaufmann, pp. 1-32.
5. Burgun, A. and Bodenreider, O. (2001) Mapping the UMLS Semantic Network into General Ontologies, *Proceedings of AMIA Symposium*, pp. 81-85.
6. Cluet, S., Delobel, C., Simeon, J., and Smaga, K. (1998) Your Mediators Need Data Conversion, *SIGMOD Record*, 27, 2, 177-188.
7. Gottschalk, K., Graham, S., Kreger, H., and Snell, J. (2002) Introduction to Web Services Architecture, *IBM Systems Journal*, 41, 2, 170-177.
8. Hansen, M., Madnick, S. E., and Siegel, M. (2002) Data Integration Using Web Services, *Proceedings of DIWeb*, pp. 3-16.
9. Kwan, I. and Fong, J. (1999) Schema Integration Methodology and Its Verification by Use of Information Capacity, *Information Systems*, 24, 5, 355-376.
10. Lee, J. L. and Siegel, M. D. (1996) An Ontological and Semantical Approach to Source-receiver Interoperability, *Decision Support Systems*, 18, 2, 145-158.
11. Leroy, G. and Chen, H. (2001) Meeting Medical Terminology Needs—the Ontology-Enhanced Medical Concept Mapper, *IEEE Transactions on Information Technology in Biomedicine*, 5, 4, 261-270.
12. Ouksel, A. M. and Ahmed, I. (1999) Ontologies are not the Panacea in Data Integration: a Flexible Coordinator to Mediate Context Construction, *Distributed and Parallel Databases*, 7, 1, 7-35.

13. Park, J. (1999) Facilitating Interoperability among Heterogeneous Geographic Database Systems: A Theoretical Framework, a Prototype System, and Evaluation, PhD Dissertation, University of Arizona, Tucson, AZ.
14. Ram, S. and Venkataraman, R. (1999) Schema Integration: Past, Present and Future, in: A. Elmagarmid, M. Rusinkiewicz, and A. Sheth, Eds. Management of Heterogeneous and Autonomous Database Systems, Morgan Kaufmann, pp. 119-156.
15. Sheth, A. P. and Larson, J. A. (1990) Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases, *ACM Computing Surveys*, 22, 3, 183-235.
16. Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., and Hübner, S. (2001) Ontology-based Integration of Information - a Survey of Existing Approaches, *Proceedings of the IJCAI Workshop on Ontologies and Information Sharing*, pp. 108--117.
17. Wiederhold, G. (1992) Mediators in the Architecture of Future Information Systems, *Computer*, 25, 3, 38-49.