Information Mediation Using Metamodels: An Approach Using XML and Common Warehouse Metamodel

Luyin Zhao, The State University of New Jersey - Rutgers, USA
Keng Siau, University of Nebraska - Lincoln, USA

Abstract

Information mediation is an approach used to resolve heterogeneity of online information sources. This article discusses the concept of information mediation and several related research projects that use a typical mediation model. The problems and issues with existing approaches are discussed. This article then proposes the use of metamodels as an advanced architecture for information mediation, where XML is a main driving force. The new architecture is partly established on the common warehouse metamodel (CWM), a standard adopted by the Object Management Group (OMG). Illustrative examples are presented in the article to describe and discuss the new architecture.

Keywords: CWM; interoperability; information mediation; metamodel; UML; XML; XMI

Introduction

With the increasing popularity of the Internet, more and more people are switching to rich and easily accessible online information sources (Nah et al., 2005; Siau, Nah & Teng, 2002). Many online information services such as e-banking (Southard & Siau, 2004), e-education (Erickson & Siau, 2003, Siau, Sheng & Nah, 2006), e-finance (Nah et al., 2005), e-government (Siau & Long, 2006), and e-healthcare (Siau & Kam, 2006; Siau & Shen, 2006) are enabled by Internet. The availability of Internet also makes remote cooperation and collaboration such as virtual teams (Long & Siau, 2007) possible. However, an obvious problem faced by organizations and Internet users is the heterogeneous nature of online information sources, which requires interoperability solutions for the integration of information (Fileto & Medeiros, 2003). An information mediator handles the collection and integration of information from different, usually heterogeneous, information sources. These information sources can be regular databases, XML sources, record files, and e-mail systems. In other words, the informa-
A successful information mediation solution should be suitable for Internet use and be compliant with Internet oriented technologies and standards. Fortunately, the emergence of modern Internet interoperability technologies and data modeling/representation standards, especially eXtensible Markup Language (XML) which is considered a breakthrough solution for interoperability, provides such opportunities. We have seen some similar and successful solutions enabled by XML recently, such as the XTOPO methodology that uses XML document as a medium to transmit relational database on the Internet (Fong & Wong, 2004).

Metadata, simply speaking are data that define data. The concept of metadata has been successfully used in multiple areas such as hypermedia management (Suh & Lee, 2001) and data warehousing (Lee, Kim & Kim, 2001). Data interoperability is not an exception; a common method for solving interoperability problems historically is exchanging metadata based on a metamodel (meta-metadata). Common Warehouse Metamodel (CWM) (OMG, 2001; Poole, Chang, Tolbert & Mellor, 2002) is the first standard established by the Object Management Group (OMG) to enable the exchange of metadata mainly in a data warehouse domain. Furthermore, it utilizes the latest XML technology to make metadata exchangeable over the Internet. Although still being improved, this standard is one of our major motivations for designing a new metamodel-based mediation architecture to achieve better interoperability for the information mediation domain.

The rest of the article is organized as follows: the first section discusses information interoperability and mediation. The next section reviews related work in this area. The following section describes the metamodel and the common warehouse metamodel. The next section proposes the metamodel-based mediation. The next section compares the proposed metamodel-based mediation with other approaches. The last section concludes the research and identifies some future research directions.

**INFORMATION INTEROPERABILITY AND MEDIATION**

A common problem faced by a distributed information system is the requirement to integrate heterogeneous information sources, including but not limited to regular databases, file systems, Web information, and e-mail systems. Business decision makers will need to access multiple information sources to gather enough information for decision making. However the ease of accessing such multiple information sources is often impeded by information heterogeneity including, which could involve data schema (data model) difference, data format inconsistency, data semantic difference, naming inconsistency, and so forth. Information interoperability is the ability to solve heterogeneous data access problems. Typical solutions of information interoperability include the middleware-based interoperability and the mediation-based interoperability.

The basic idea of middleware-based interoperability is to encapsulate data access functions into methods and publish them using implementation independent interface definition language (IDL). This type of interoperability is at the service level because users only invoke data access methods rather than query data. Many commercial products (CORBA products, Microsoft DCOM and .NET, etc.) are designed based on this approach.

Mediation-based interoperability provides users with a (possibly converted) common data schema and a query language for querying on the common data schema. This type of interoperability works at the data level rather than the service level by middleware-based approach. More specifically, mediation-based solutions provide users with a means to send on-demand queries to heterogeneous information sources. In other words, users see a homogeneous (common) view despite the fact that the actual information sources are heterogeneous. To accomplish this, user queries issued from this view are intercepted by the mediation system and converted to query formats that
Exploring the Effects of Process Characteristics on Products Quality in Open Source Software Development
www.igi-global.com/article/exploring-effects-process-characteristics-products/3384?camid=4v1a

Managing Large Rule-Bases
*Journal of Database Management* (pp. 17-24).
www.igi-global.com/article/managing-large-rule-bases/51165?camid=4v1a