Chapter XII
A Scalable Middleware for Web Databases

Athman Bouguettaya
Virginia Tech, USA

Zaki Malik
Virginia Tech, USA

Abdelmounaam Rezgui
Virginia Tech, USA

Lori Korff
Virginia Tech, USA

ABSTRACT

The emergence of Web databases has introduced new challenges related to their organization, access, integration, and interoperability. New approaches and techniques are needed to provide across-the-board transparency for accessing and manipulating Web databases irrespective of their data models, platforms, locations, or systems. In meeting these needs, it is necessary to build a middleware infrastructure to support flexible tools for information space organization, communication facilities, information discovery, content description, and assembly of data from heterogeneous sources. In this paper, we describe a scalable middleware for efficient data and application access that we have built using the available technologies. The resulting system is called WebFINDIT. It is a scalable and uniform infrastructure for locating and accessing heterogeneous and autonomous databases and applications.

INTRODUCTION

Traditional database systems are usually deployed in *closed* environments where users access the system only via a restricted network (e.g., an enterprise's internal network). With the emergence of the World Wide Web, access has become possible from virtually anywhere, to any database with a Web interface. These Web-accessible databases, or *Web databases*, provide an elegant solution to store any data content to which a ubiquitous access is needed (Gribble,

2003). However, there is a need to provide users with a uniform, integrated view for querying the content of *multiple* Web databases.

Providing an integrated view of multiple databases is both important and challenging. Two particular challenges must be overcome: *connectivity* and *interoperability*. The Web has provided the necessary "pipes" to interconnect isolated data islands. However, to address the interoperability issue, more than a networking infrastructure is needed. The challenge remains to cope with the heterogeneity amongst the different databases as it obstructs interoperability. The need has therefore arisen for a middleware that transcends all types of heterogeneities and provides users with a uniform view of the content of Web databases (Bouguettaya, Rezgui, Medjahed, & Ouzzani, 2004).

In the context of Web databases, a middleware achieves uniform database access and interoperability. The challenge is to provide across-the-board transparency in order to allow users to access and manipulate data irrespective of platforms, locations, systems, or any other database-specific characteristics (Vinoski, 2002). To meet this challenge, we identify the following key issues:

- *Locating* relevant information sources. In Web applications, the information space is very large and dynamic. A way must be found to organize that information space in a rational and readily comprehensible manner to facilitate the location of pertinent data.
- *Understanding* the meaning, content, terminology, and usage patterns of the available information sources. Users must be *educated* about the information of interest and dynamically provided with up-to-date knowledge of database contents. Users must also be instructed as to the appropriate means of linking to information sources.
- *Querying* sources for relevant information items. Once appropriate information

sources have been found, users need to be provided with the tools necessary to access and integrate data from these information sources.

To address the previously mentioned issues, we have developed the WebFINDIT system. The major contribution of the system is providing support for achieving effective and efficient data sharing in a large and dynamic information space. WebFINDIT presents an incremental and self-documenting approach. The system processes a user query in two steps. First, querying metadata for information sources location and semantic exploration. Second, querying selected sources for actual data. WebFINDIT provides support for educating the user about the available information space. The efforts related to registering and advertising the content of information sources are minimized.

We have provided an extensible middleware for querying autonomous Web databases and applications. We have incorporated Web services in our system to provide uniform access to applications. The Web services technology has been developed to assist in the integration and interoperation of isolated, autonomous and heterogeneous sources of information and services. The participants of a Web services system do not have to worry about the operating system, development language environment or the component model used to create or access the services.

In this paper, we present a middleware framework for supporting seamless access to Web databases and applications. WebFINDIT integrates a large set of heterogeneous technologies. A key feature of the system is the large spectrum of heterogeneities being supported at all levels, including hardware, operating system, database, and communication middleware. We have presented an easy to use architecture for databases to be accessed over the Web, despite their distributed, autonomous and heterogeneous nature. WebFINDIT provides a scalable and distributed ontological

Related Content

Database-Driven Product Catalog System
Johanna Wenny Rahayu, Andrew Flahiveand David Taniar (2003). *Web-Powered Databases (pp. 144-174).*
www.irma-international.org/chapter/database-driven-product-catalog-system/31427

Scalable QSF-Trees: Retrieving Regional Objects in High-Dimensional Spaces
Ratko Orlandicand Byunggu Yu (2004). *Journal of Database Management (pp. 45-59).*
www.irma-international.org/article/scalable-qsf-trees/3315

Application of Decision Tree as a Data mining Tool in a Manufacturing System
S.A. Oke (2007). *Intelligent Databases: Technologies and Applications (pp. 117-136).*
www.irma-international.org/chapter/application-decision-tree-data-mining/24232

Performance Studies of Locking Protocols for Real-time Databases With Earliest Deadline First
Kam-Yiu Lam, Sheung-Lun Hungand Ken Chee-Keung Law (1995). *Journal of Database Management (pp. 22-32).*
www.irma-international.org/article/performance-studies-locking-protocols-real/51148

Combined Use of Conceptual Models in Practice: An Exploratory Study
Mohammad Ali Jabbari Sabeghand Jan Recker (2017). *Journal of Database Management (pp. 56-88).*
www.irma-international.org/article/combined-use-of-conceptual-models-in-practice/182869