Chapter IX

Integrating Heterogeneous Data Sources in the Web

Angelo Brayner, University of Fortaleza, Brazil

Marcelo Meirelles, University of Fortaleza, Brazil

José de Aguiar Moraes Filho, University of Fortaleza, Brazil

Abstract

Integrating data sources published on the Web requires an integration strategy that guarantees the local data sources’ autonomy. A multidatabase system (MDBS) has been consolidated as an approach to integrate multiple heterogeneous and distributed data sources in flexible and dynamic environments such as the Web. A key property of MDBSs is to guarantee a higher degree of local autonomy. In order to adopt the MDBS strategy, it is necessary to use a query language, called the MultiDatabase Language (MDL), which provides the necessary constructs for jointly manipulating and accessing data in heterogeneous data sources. In other words, the MDL is responsible for solving integration conflicts. This chapter
describes an extension to the XQuery Language, called MXQuery, which supports queries over several data sources and solves such integration problems as semantic heterogeneity and incomplete information.

**Introduction**

The Web (World Wide Web) can be seen as a wide network consisting of the union of several local area networks (LANs) spread over the entire world. However, the local networks that constitute the Web are autonomous and capable of plugging or unplugging themselves into and from the Web at any time.

Over the last few years, the Web has been used to publish several databases. Of course, databases available on the Web are heterogeneous since they might be defined by using different data models (e.g., relational or object data model), managed by different database systems (DBSs), or running in different computational environments (regarding operating system and hardware). Furthermore, the integration of databases on the Web should be realized without interfering in the management and processing of local data. In other words, databases should be integrated preserving the local autonomy of each database. Despite the fact that heterogeneity and the autonomy of multiple databases on the Web is a reality nowadays, users (and applications) need shared access to those databases. Thus, it is possible to submit queries against several heterogeneous databases located in distinct local networks throughout the Web.

Consequently, integrating databases published on the Web has become a challenge to the database technology. Several approaches for integrating heterogeneous and autonomous data sources have been proposed since the late ’80s. In this chapter, we propose a strategy based on the multidatabase approach for integrating heterogeneous databases published on the Web. For that reason, we describe a new MultiDatabase Language (MDL), called MXQuery, since the proposed strategy uses XML (extensible markup language) as the common data model (CDM; conceptual schema) to represent the multiple data sources’ schemas. The MXQuery, which is an extension to the XQuery Language, provides constructors and operators for supporting queries over multiple heterogeneous data sources. The MXQuery solves integration problems such as semantic heterogeneity and incomplete information. Furthermore, this chapter presents an architecture to process MXQuery queries.

This chapter is organized as follows. Approaches for integrating heterogeneous data sources are studied next. Then, related work is discussed, followed by a description of the MXQuery MultiDatabase Language. Next we present in detail the features of the proposed integration strategy, and then give an overview of the query-processor architecture for the MXQuery Language. Finally, we conclude the chapter.
18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product’s webpage:
www.igi-global.com/chapter/integrating-heterogeneous-data-sources-web/31102?camid=4v1

www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Real-Time Unspecified Major Sub-Events Detection in the Twitter Data Stream That Cause the Change in the Sentiment Score of the Targeted Event

Development and Evaluation of a Methodology for Developing Marketing Websites
Tomayess Issa, Martin West and Andrew Turk (2010). *Integrating Usability Engineering for Designing the Web Experience: Methodologies and Principles* (pp. 103-123).
www.igi-global.com/chapter/development-evaluation-methodology-developing-marketing/40495?camid=4v1a

Usability Methodologies for Real-Life Voice User Interfaces
www.igi-global.com/article/usability-methodologies-real-life-voice/40345?camid=4v1a
Toward An Effective Virtual Learning Environment: From a Social Presence Perspective
www.igi-global.com/chapter/toward-an-effective-virtual-learning-environment/137404?camid=4v1a