

Chapter 6.5

Business Information Integration from XML and Relational Databases Sources

Ana María Feroso Garcia

Pontifical University of Salamanca, Spain

Roberto Berjón Gallinas

Pontifical University of Salamanca, Spain

ABSTRACT

This chapter introduces different alternatives to store and manage jointly relational and eXtensible Markup Language (XML) data sources. Nowadays, businesses are transformed in e-business and have to manage large data volumes and from heterogeneous sources. To manage large amounts of information, Database Management Systems (DBMS) continue to be one of the most used tools, and the most extended model is the relational one. On the other side, XML has reached the de facto standard to present and exchange information between businesses on the Web. Therefore, it could be necessary to use tools as mediators to integrate these two different data to a common format like XML, since it is the main data format on the Web. First, a classification of the main tools and systems where this problem

is handled is made, with their advantages and disadvantages. The objective will be to propose a new system to solve the integration business information problem.

INTRODUCTION

Nowadays, business information systems are suffering numerous changes. Business is transforming in e-business. Business has to manage large volumes of data in addition to data from heterogeneous sources.

When business gets into the Web, it has to manage new types of Web data, such as XML (W3C, 2005c) or HyperText Markup Language (HTML), for example, to exchange information with other businesses or publish its information on the Web. However, business information also

continues to be stored in traditional databases, such as the relational one.

To manage large amounts of information, DBMS continue to be one of the most used tools, and the most extended model is the relational one, although the object-oriented model or data warehouse is being used more and more.

On the other hand, *XML* has reached the de facto standard to present and exchange information between businesses on the Web. In fact, every day it is more common to use XHTML to publish information in the Web. XHTML is the HTML language version adapted to the more restrictive rules of XML.

In this context, where organizations have to manage a lot of data, and from heterogeneous sources like conventional database systems or the new Web data sources, it could be necessary to use tools as mediators to integrate these different data to a common format to manage both types of data together. This common format could be *XML*, because it has reached the main format language on the Web.

Therefore, because almost all organizations eventually get into the Web, where XML is emerging as a more effective means of describing the semantic content of World-Wide Web (WWW) documents, and because the main amount of business information continues to be stored in relational databases, it would be useful to manage a relational database from a Web perspective using XML; that is, not only to use XML language as an interface to pass information between systems that interact with each other on the Web, but also to employ an XML-based language to access to the databases of these systems.

To manage data from all these heterogeneous sources and formats, databases should adapt to the new market demands; allowing, for example, the storing of these new types of Web data such as XML, together with its traditional database data, or changing the way to work with the data by using them as knowledge, like in the data warehouse.

In this chapter, we are going to study how in business, and especially when it is converted to e-business, it is necessary to integrate information from different sources and manage it together. We center on two main data sources: the conventional relational database system, because most business data continues to be stored in them; and XML data, because it is the most extended data format for exchanging and publishing information on the Web.

The aim is to integrate these two types of data to obtain at the end all the necessary data in the same format: *XML*. Later, these data can be used, for example, to be stored in a data warehouse or exchange information on the Web.

Therefore, the main goal of this chapter is to study the different alternatives for storing and managing these two types of data together. First, we will make a classification of the main tools and systems where this problem is dealt with. Afterwards, we will study in detail the main features of each system. Finally, once we have studied the advantages and disadvantages of the different systems analyzed, we will propose a new system to solve the integration business information problem.

This chapter should be considered within the managerial dimension in business integration, showing the most recent contributions to information integration of data from relational databases and XML sources, and proposing at the same time a solution to this dimension.

RELATIONAL DATABASES AND XML

Introduction

XML (W3C, 2005c) has reached the standard to present and exchange information on the Web. There are several causes to justify this expansion: XML is a markup language like HTML, but where the user can define his or her own marks.

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/business-information-integration-xml-relational/8014

Related Content

An Efficient Concurrency Control Algorithm for High-Dimensional Index Structures

Seok Il Song and Jae Soo Yoo (2006). *Advanced Topics in Database Research, Volume 5* (pp. 249-272).

www.irma-international.org/chapter/efficient-concurrency-control-algorithm-high/4396

Issues in Mobile Electronic Commerce

Asuman Dogac and Arif Tumer (2002). *Journal of Database Management* (pp. 36-42).

www.irma-international.org/article/issues-mobile-electronic-commerce/3275

Distributed Data Warehouse for Geo-spatial Services

Iftikhar U. Sikder and Aryya Gangopadhyay (2003). *ERP & Data Warehousing in Organizations: Issues and Challenges* (pp. 132-145).

www.irma-international.org/chapter/distributed-data-warehouse-geo-spatial/18559

Fuzzy Inclusion Dependencies in Fuzzy Databases

Awadhesh Kumar Sharma, A. Goswami and D. K. Gupta (2008). *Handbook of Research on Fuzzy Information Processing in Databases* (pp. 658-684).

www.irma-international.org/chapter/fuzzy-inclusion-dependencies-fuzzy-databases/20372

Hierarchies

Elaheh Pourabbas and Maurizio Rafanelli (2003). *Multidimensional Databases: Problems and Solutions* (pp. 91-115).

www.irma-international.org/chapter/hierarchies/26966