Chapter 2.7

Designing Data Marts from XML and Relational Data Sources

Yasser Hachaichi
Mir@cl Laboratory, Faculté des Sciences Economiques et de Gestion, Tunisia

Jamel Feki
Mir@cl Laboratory, Faculté des Sciences Economiques et de Gestion, Tunisia

Hanene Ben-Abdallah
Mir@cl Laboratory, Faculté des Sciences Economiques et de Gestion, Tunisia

ABSTRACT

Due to the international economic competition, enterprises are ever looking for efficient methods to build data marts/warehouses to analyze the large data volume in their decision making process. On the other hand, even though the relational data model is the most commonly used model, any data mart/warehouse construction method must now deal with other data types and in particular XML documents which represent the dominant type of data exchanged between partners and retrieved from the Web. This chapter presents a data mart design method that starts from both a relational database source and XML documents compliant to a given DTD. Besides considering these two types of data structures, the originality of our method lies in its being decision maker centered, its automatic extraction of loadable data mart schemas and its genericity.

INTRODUCTION

Faced with the ever increasing economic competition, today's enterprises are hard-pressed to rely on decision support systems (DSS) to assist them in the analysis of very large data volumes. As a response to this constraint, data warehousing technologies have been proposed as a means to extract pertinent data from information systems and present it as historical snapshots used for *ad hoc* analytical queries and scheduled reporting. Indeed, a data warehouse (DW) is organized in such a way that relevant data is clustered together for an easy access. In addition,

a DW can be used as a source for building data marts (DM) that are oriented to specific subjects of analyses.

Traditionally, the data loaded into a DW/ DM is mainly issued from the enterprise's own operational information system. Thus, most currently proposed DW/DM construction approaches suppose a single, often relational data source; *cf.,* (List, Bruckner, Machacze, & Schiefer, 2002), (Golfarelli, Maio, & Rizz, 1998), (Cabibbo, L., & Torlone, R. 1998), (Moody , & Kortnik, 2000), (Prat, Akoka , & Comyn-Wattiau, 2006), (Zribi, & feki, 2007), (Golfarelli, Rizzi, & Vrdoljak, 2001), (Vrdoljak, Banek, & Rizzi, 2003), (Jensen, Møller, & Pedersen, 2001). However, due to the international competition, enterprises are increasingly forced to enrich their own data repository with data coming from external sources. Besides data received from partners, the web constitutes the main external data source for all enterprises. For instance, an enterprise may need to retrieve from the web data about the exchange rates in order to analyze the variation of the quantities of its sold products with respect to the exchange rates during a period of time.

To deal with such an open data source, a DW/ DM construction approach must, hence, overcome the main difficulty behind the use of multiple data sources: the structural and semantic heterogeneities of the sources. In fact, even though the relational data model is the most commonly used model (Wikipedia encyclopedia, 2008), a DW construction approach must now deal with other data types and in particular XML documents which represent the dominant data type on the web. On the other hand, the semantic data heterogeneity comes into play when the internal and external data sources are complementary, *e.g.*, the case of transactional data between partners. This type of heterogeneity remains a challenging problem that can be treated either at the data source level or the DW/DM level (Boufares, & Hamdoun, 2005).

This chapter deals with the structural data heterogeneity when designing a data mart. More precisely, it presents a DM design method that starts from both a relational database source and XML documents compliant to a given DTD. Besides considering these two types of data structures, our method has three additional advantages. First, it provides for a DSS development centered on decision makers: it assists them in defining their analytical needs by proposing all analytical subjects that could be automatically extracted from their data sources; the *automatic* extraction of DM schemas distinguishes our method from currently proposed ones. Secondly, it guarantees that the extracted subjects are loadable from the enterprise information system and/or the external data sources. The third advantage of our design method is its genericity: It is domain independent since it relies on the structural properties of the data sources independently of their semantics. It automatically applies a set of rules to extract, from the relational database and XML documents, all possible facts with their dimensions and hierarchies.

To achieve these advantages, our method operates in four steps. First, it structurally homogenizes the two types of data sources by converting the DTD into a relational model. Secondly, it classifies the set of relations issued from both the converted DTD and the repository of the source relational DBMS. This classification is then used to identify automatically the facts, measures, dimensions and their attributes organized into hierarchies; these identified multidimensional elements are modeled as star DM schemas. Finally, the resulting DM schemas can be manually adapted by the decision makers/designers to specify their particular analytical needs. The automatic steps of our design method allowed us to incorporate it into a CASE toolset that interactively provides for DM schema adaptation.

The remainder of this chapter is organized as follows. First we overview current DW design approaches for relational and XML data sources. Then, we illustrate our four-step DM design method through a relational data source and a

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/designing-data-marts-xml-relational/48557

Related Content

The Next Generation of Customer Relationship Management (CRM) Metrics

Timothy Shea, Ahern Brown, D. Steven White, Catharine Curranand Michael Griffin (2008). *Techniques and Tools for the Design and Implementation of Enterprise Information Systems (pp. 258-270).*

www.irma-international.org/chapter/next-generation-customer-relationship-management/30123

Microservices Data Mining for Analytics Feedback and Optimization

Kindson Munonyeand Péter Martinek (2021). *International Journal of Enterprise Information Systems (pp. 22-43).*

www.irma-international.org/article/microservices-data-mining-for-analytics-feedback-and-optimization/268361

Impact of Organisational Intranets on Profitability in Organisations

Udo R. Averweg (2007). *Managing Information Communication Technology Investments in Successful Enterprises (pp. 44-78).*

www.irma-international.org/chapter/impact-organisational-intranets-profitability-organisations/25851

Assessing Information Technology Capability versus Human Resource Information System Utilization

Ralf Burbachand Tony Dundon (2011). *Enterprise Information Systems: Concepts, Methodologies, Tools and Applications (pp. 1370-1378).*

www.irma-international.org/chapter/assessing-information-technology-capability-versus/48618

Examining the Contributing Factors for Cloud Computing Adoption in a Developing Country

Winfred Yaokumahand Rebecca Adwoa Amponsah (2017). *International Journal of Enterprise Information Systems (pp. 17-37).*

www.irma-international.org/article/examining-the-contributing-factors-for-cloud-computing-adoption-in-a-developing-country/176389