Expert Systems with Applications 39 (2012) 7371–7390

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An intelligent XML-based multidimensional data cube exchange

Jia-Lang Seng [a,*], Zon Wong [b]

[a] Department and Graduate School of Accounting, College of Commerce, National Chengchi University, Taipei, Taiwan
[b] IT Sector, Hua-Nan Financial Holdings Inc., Taipei, Taiwan

ARTICLE INFO

Keywords:
Data warehouse
Multidimensional data cube
Intelligent metadata management
XML
Data exchange
Generic construct
Mapping and transformation

ABSTRACT

Motivated by the globalization trend and Internet speed competition, enterprise nowadays often divides into many departments or organizations or even merges with other companies that located in different regions to bring up the competency and reaction ability. As a result, there are a number of data warehouse systems in a geographically-distributed enterprise. To meet the distributed decision-making requirements, the data in different data warehouses is addressed to enable data exchange and integration. Therefore, an open, vendor-independent, and efficient data exchange standard to transfer data between data warehouses over the Internet is an important issue. However, current solutions for cross-warehouse data exchange employ only approaches either based on records or transferring plain-text files, which are neither adequate nor efficient. In this research, issues on multidimensional data exchange are studied and an Intelligent XML-based multidimensional data exchange model is developed. In addition, a generic-construct-based approach is proposed to enable many-to-many systematic mapping between distributed data warehouses, introducing a consistent and unique standard exchange format. Based on the transformation model we develop between multidimensional data model and XML data model, and enhanced by the multidimensional metadata management mechanism proposed in this research, a general-purpose intelligent XML-based multidimensional data exchange process over web is facilitated efficiently and improved in quality. Moreover, we develop an intelligent XML-based prototype system to exchange multidimensional data, which shows that the proposed multidimensional data exchange model is feasible, and the multidimensional data exchange process is more systematic and efficient using metadata.

1. Introduction

1.1. Research motivation

The explosion of the Internet and e-business (EB) in recent years has caused a consequent explosion of information available to enterprise. The challenge confronting with business today is *information integration* (II). In order to answer specific questions or make decisions, users often need to collect and combine information from multiple sources. However, they often found that it is impossible to obtain the required information in a timely fashion due to the difficulties of accessing the different systems and due to the heterogeneity of information sources. In recent years, various models of integration have been proposed as solutions, among them stands *data warehouse*.

A data warehouse is a repository of integrated information, available for querying and analysis, and generally used for decision-making, rather for transactional operations. At the warehouse, queries can be answered and data analysis can be performed quickly and efficiently since the information is directly available, with model and semantic differences already resolved. In recent years, data warehouse systems have attracted a great deal of interest in both academic and industrial communities. However, there are still some emerging requirements to be addressed in data warehousing, especially the increasing needs of *multidimensional data cube exchange* and *metadata management* for this area.

An ideal source of data to extend a data warehouse is another data warehouse in which data is already cleaned and verified. Such integration of data warehouses could happen within a single company or across multiple companies over Intranets or Internet, due to merging or cooperation between enterprises, resulting in the increasing need for OLAP-based (online analytical processing) distributed decision-making. Thus, web-based inter-warehouse data cube exchange and integration becomes an important issue.

The multidimensional data model has proven to be adequate for explorative analysis on OLAP-manipulated information stored in a data warehouse. Various ideas for the multidimensional modeling have been proposed in recent years, yet there is not a unique and standard multidimensional data model was accepted. However, an effective data integration can only be realized with the help of

* Corresponding author.
E-mail address: seng@nccu.edu.tw (J.-L. Seng).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.069

well-organized and well-structured metadata. Lack of a common and standardized multidimensional metadata, users cannot gain support of interpreting and understanding the data and query results of data warehouses. With that in mind, the need for a common metadata or data exchange format to enable the interaction of different data warehouses over network is arising.

Metadata is generally defined as any information that may be used to support the administration and effective exploitation of data warehouses. Metadata are produced by various tools and often exist everywhere, such as directly in data dictionaries and certain tool repositories or hidden in programs, scripts, and paper documents. However, metadata are not worth much unless they are captured, stored, and consistently managed so as to be accessible by software components. Due to this, the design of an effective metadata management system is in need for data exchange. Moreover, in addition, metadata represents a uniform and consistent data exchange format between heterogeneous information sources. By following this standard structure, the data to be exchanged from different sources can be transformed into a unique format. The complexity that often arises from many-to-many data exchange scenario can be reduced and solved. Therefore, involving metadata into the data exchange process between heterogeneous data warehouses is an effective approach of resolution.

1.2. Research issue

Data warehouse systems are the widely spread approach for the enterprise decision support. Nowadays, the idea of integrating data warehouses in a company or across companies over Internet is evolving, but the technical and semantic problems are also growing. In order to solve these problems and enable data exchange between different data warehouses, moreover, to embed into further application or services over Internet, a commonly accepted and vendor-independent standard for describing multidimensional data is required so as to exchange multidimensional data in a standardized format.

It is important to note that current solutions for cross-warehouse data exchange employ only approaches either based on records or transferring plain-text files. These approaches, neither efficient nor practical and using no metadata, result in the whole exchange process not only ambiguous but also misleading. To improve such a process, a systematic and methodological approach is necessary from further research efforts. To exchange multidimensional data between heterogeneous and distributed data warehouses, a transformation model would be necessary to enable the many-to-many generic transformation instead of the one-to-one ad hoc mapping. More specifically, the required solution is not only a local one that converts multidimensional data into XML, but also an extended solution that can exchange and integrate multidimensional data across distributed warehouses and web sites. The data exchange process can be enabled by the use of metadata. Furthermore, a consistent data management ensuring interoperability and integration of related metadata is required. Metadata management for data warehouses can provide documentation as well as control information.

This research reviews the current development of data warehouse, examines the related issues and finds out the unsolved problem inside. We adopt the Web-based system approach, focus on the multidimensional data exchange, and develop the metadata management mechanism. A recent XML standard for storing multidimensional data is elaborated, and has been extended and implemented in the method in order to fulfill the data exchange between data warehouses. Besides, a metadata management mechanism has also been extended from that standard. Thus, the objective of this research can be categorized in three parts:

(1) Applying the standard XML for data exchange to multidimensional data, enabling web-based inter-warehouse data exchange and facilitating OLAP-based distributed decision-making.
(2) Developing a generic-construct-based approach to transform data between the multidimensional data model and the XML data model, facilitating the general-purpose data exchange and integrating heterogeneous data warehouses.
(3) Providing intelligent XML-based multidimensional metadata management mechanism, improving the quality of data warehouse and facilitate metadata integration and interoperability.

2. Literature review

2.1. Multidimensional data modeling open issues

The data model for designing traditional online transaction processing (OLTP) systems is not well-suited for modeling complex queries in data warehousing and online analytic processing (OLAP) environment due to their different constructs. The transactions in OLTP systems are made up of simple, pre-defined queries, yet in the data warehousing environments, the queries tend to use joins on more tables, have a large computation time, and are ad hoc in nature (Dinter, Sapia, Hofling, & Blaschka, 1998). However, unlike the relational model, there does not exist a standard and commonly agreed upon conceptual model for OLAP. Though the notion of the data cube we reviewed above has been accepted as the underlying logical construct of data warehouses, i.e. multidimensional databases, there does not exist a precise model for data cube which means a common standard for a logical multidimensional data model is still absent (Datta & Thomas, 1999).

The lack of a formal multidimensional data model has attracted the interest of several researches for the past years. This resulted in a multitude of different proposed models, yet no data model provides a comprehensive set of structural and operational tools necessary for a flexible and extensive analysis of information stored within a data warehouse system (Hummer, Lehner, Bauer, & Schlesinger, 2009). Moreover, each proposal presents its own terminology, notation, and formalism, and none of them has been widely accepted as a standard model.

It attracts our attention that Hummer et al. (2009) address ten open issues and some solutions on multidimensional data modeling. They state the common problems reaching from defects within dimensional structures over multidimensional structure to new analytical requirements and more. Among these open problems, the one that regarding metadata management accounts for a key issue of this research. We summarize the description of that particular issue in Table 1. We can recognize from the issue that since there is not a standard multidimensional data model, a flexible and adaptable metadata representation as well as a data exchange format turns out to be necessary and important.

2.2. XML standards for multidimensional data

2.2.1. XCube

We consider that by exchanging multidimensional data between data warehouses and capturing the corresponding metadata simultaneously, the data integration can be realized effectively. Therefore, in this research, we adopt part of the "XCube" standard format to represent and exchange the multidimensional data in data warehouse system. The following is an introduction of XCube.

2.2.1.1. Introduction. Hummer et al. (2009) propose XCube, an XML-based encoding of multidimensional data with the purpose of data exchange. XCube is a collection of XML-specification