Chapter 2.9
Goal–Oriented Requirement Engineering for XML Document Warehouses

Vicky Nassis
La Trobe University, Melbourne, Australia

Tharam S. Dillon
University of Technology, Sydney, Australia

Wenny Rahayu
La Trobe University, Melbourne, Australia

R. Rajugan
University of Technology, Sydney, Australia

ABSTRACT

eXtensible Markup Language (XML) has emerged as the dominant standard in describing and exchanging data amongst heterogeneous data sources. The increasing presence of large volumes of data appearing creates the need to investigate XML document warehouses (XDW) as a means of handling the data for business intelligence. In our previous work (Nassis, Rajugan, Dillon, & Rahayu, 2004) we proposed a conceptual modelling approach for the development of an XDW with emphasis on the design techniques.

We consider important the need of capturing data warehouse requirements early in the design stage. The elicitation of requirements and their use for data warehouse design is a significant and, as yet, an unaddressed issue. For this reason, we explore a requirement engineering (RE) approach, namely the goal-oriented approach. We will extract and extend the notion of this approach to introduce the XML document warehouse (XDW) requirement model. In order to perform this, we consider organisational objectives as well as user viewpoints. Furthermore, these are related to the XDW particularly focussing on deriving

dimensions, as opposed to associating organisational objectives to the system functions, which is traditionally carried out by RE.

INTRODUCTION

Data Warehouses

Data warehousing (DW) is an approach that has been adopted for handling large volumes of historical data for detailed analysis and management support. Transactional data in different databases is cleaned, aligned and combined to produce data warehouses. Since its introduction in 1996, the eXtensible Markup Language (XML) has become the *defacto* standard for storing and manipulating self-describing information (meta-data), which creates vocabularies to assist in information exchange between heterogeneous data sources over the Web (Pokorny, 2002). The purposes for which XML is used include electronic document handling, electronic storage, retrieval and exchange. It is envisaged that XML will also be used for logically encoding documents for many domains. Hence, it is likely that a large number of XML documents will populate the would-be repository and include several disparate XML transactional databases.

There are several distinctions instigated among data warehouses (DW) and XML data warehouses. In general, a data warehouse has two major parts:

1. **Physical store:** A database that contains all the information gathered from different sources. Operational data can be imported from various database types and once collected, the data is then structured in a uniform manner and stored in the DW. XML is well suited for representing semi-structured data. XML data originate from heterogeneous databases and are unstructured, or they have incomplete, irregular and recurrent changed structure.

2. **Logical schema:** A conceptual model that maps to the data in the physical store. This includes the following components:
    - *Class:* Contains a collection of logically grouped attributes. For example class Customer contains attributes that describe this class.
    - *Attributes:* A structure that stores data, for instance the attribute Full_Name of class Customer stores the full name of each customer.
    - *Relation:* The connection between the classes. For example the class Customer is connected to class Product. In an XML data warehouse, at the logical level the components that appear differ but operate in the same manner as in conventional data warehouses. There are *XML documents* instead of classes, which have a logical collection of *elements.* As a result, data become more dimensional, forming a tree structure as opposed to atomic data records found in traditional DWs. *Relationships* also exist amongst the XML documents. In regards to data retrieval, querying the data fields of the DW fact table is a method to obtain the necessary data. In an XML DW, querying is performed on a fact repository containing XML documents, therefore a query language with XML specific syntax such as XQuery would be appropriate.

The need for managing large amounts of XML document data raises the necessity to explore the data warehouse approach through the use of XML document marts and XML document warehouses. Our purpose is to capture and represent fully XML-type semantics. The following discussion regarding the existing work on dimensional

Related Content

Database Systems for Big Data Storage and Retrieval

Venkat Gudivada, Amy Aponand Dhana L. Rao (2018). *Handbook of Research on Big Data Storage and Visualization Techniques (pp. 76-100).*

www.irma-international.org/chapter/database-systems-for-big-data-storage-and-retrieval/198757

Enabling Information Sharing Across Government Agencies

Akhilesh Bajajand Sudha Ram (2005). *Advanced Topics in Database Research, Volume 4 (pp. 341-366).*

www.irma-international.org/chapter/enabling-information-sharing-across-government/4382

OOXKSearch: A Search Engine for Answering XML Keyword and Loosely Structured Queries Using OO Techniques

Kamal Tahaand Ramez Elmasri (2009). *Journal of Database Management (pp. 18-50).*

www.irma-international.org/article/ooxksearch-search-engine-answering-xml/4123

A Knowledge Integration Approach for Organizational Decision Support

Kee-Young Kwahk, Hee-Woong Kimand Hock Chuan Chan (2007). *Journal of Database Management (pp. 41-61).*

www.irma-international.org/article/knowledge-integration-approach-organizational-decision/3370

Providing Approximate Answers Using a Knowledge Abstraction Database

Soon-Young Huhand Jung-Whan Lee (2001). *Journal of Database Management (pp. 14-24).*

www.irma-international.org/article/providing-approximate-answers-using-knowledge/3261