Chapter II

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.
A Framework for Analysis, Design and Management of Complex Large-Scale Interconnected Open Sociotechnological Systems
www.igi-global.com/article/framework-analysis-design-management-complex/3900?camid=4v1a

Classification and Ranking Belief Simplex
www.igi-global.com/chapter/classification-ranking-belief-simplex/11242?camid=4v1a