

Chapter IX

Web Data Warehousing Convergence: From Schematic to Systematic

D. Xuan Le

La Trobe University, Australia

J. Wenny Rahayu

La Trobe University, Australia

David Taniar

Monash University, Australia

ABSTRACT

This article proposes a data warehouse integration technique that combines data and documents from different underlying documents and database design approaches. The well-defined and structured data such as relational, object-oriented and object relational data, semi-structured data such as XML, and unstructured data such as HTML documents are integrated into a Web data warehouse system. The user specified requirements and data sources are combined to assist with the definitions of the hierarchical structures, which serve specific requirements and represent a certain type of data semantics using object-oriented

features including inheritance, aggregation, association, and collection. A conceptual integrated data warehouse model is then specified based on a combination of user requirements and data source structure, which creates the need for a logical integrated data warehouse model. A case study is then developed into a prototype in a Web-based environment that enables the evaluation. The evaluation of the proposed integration Web data warehouse methodology includes the verification of correctness of the integrated data, and the overall benefits of utilizing this proposed integration technique.

INTRODUCTION AND MOTIVATION

Currently, there are more and more techniques being provided to accommodate the high demand for exchanging and storing business information including Web and operational data. While the well-defined structured data are operated and stored in relational, object-oriented (Buzydowski, 1998), object relational database environments, semi-structured data in XML or unstructured documents are stored in HTML. The problem of related information being separated and stored in multiple places happens quite often within an organization. Information from these applications is extracted and further developed into business analysis tools such as OLAP and data warehousing, which aim to support data analysis, business requirements, and management decisions.

Relevant business Web data have rapidly increased in significant amounts. Recently, XML has increased in popularity and has become a standard technique for storing and exchanging information over the Internet. The data integration (Breitbart, Olson, & Thompson, 1986) in the data warehousing has certainly received a lot of attention. There are three particular articles that are very close to the work in this article. Jensen, Moller and Pedersen (2001) allow an integration of XML and relational data. Even though the

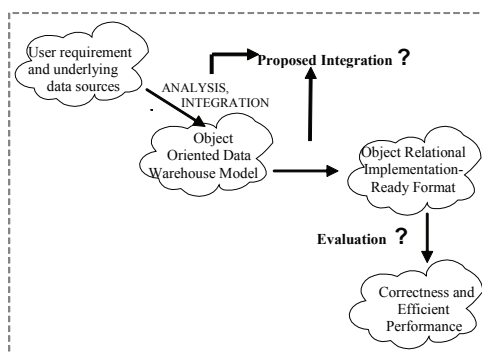
object-oriented concept is used in this model, the semantic contribution in this work lacks object-oriented features. Therefore, the semantics of data have been only partially supported. Other systems (Golfarelli, Rizzi, & Birdoljak, 1998, 2001; Huang & Su, 2001) focus on supporting Web data at the schematic level. While their initial focus is to incorporate XML data, Relational data have also been mentioned but not yet been incorporated. They mostly concentrate on the creation of a logical model.

Hence, it is clear that there is yet to be developed a standard integration technique that provides a means of handling multiple data sources being integrated into a data warehouse system (Bonifati, Cattaneo, Ceri, Fuggetta, & Paraboschi, 2001), and allowing a full capture of semantics of data in the data source models.

The purpose of this article can be summarized as follows:

- To ensure the integration technique allows a meaningful uniformed integrated object-oriented data warehouse structure.
- To ensure the integrated data and their semantics are explicitly and fully represented.
- To ensure a proposed integrated data warehouse system with consistency and high quality.

Figure 1. Integration Web data warehouse overview



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/gender-discrepancies-through-college-years/28553

Related Content

Privacy-Preserving Data Mining

Alexandre Evfimievskii and Tyrone Grandison (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 527-536).

www.irma-international.org/chapter/privacy-preserving-data-mining/20737

Repairing and Querying Inconsistent Databases

Gianluigi Greco, Sergio Greco and Ester Zumpano (2003). *Effective Databases for Text & Document Management* (pp. 318-359).

www.irma-international.org/chapter/repairing-querying-inconsistent-databases/9218

A Big Data Management and Analytics Framework for Supporting Machine Learning, OLAP, and Visualization on Big COVID-19 Data

Alfredo Cuzzocrea, Carson K. Leung, Siyuan Shang and Yan Wen (2025). *Journal of Database Management* (pp. 1-44).

www.irma-international.org/article/a-big-data-management-and-analytics-framework-for-supporting-machine-learning-olap-and-visualization-on-big-covid-19-data/386132

Accelerating Web Service Workflow Execution via Intelligent Allocation of Services to Servers

Konstantinos Stamkopoulos, Evaggelia Pitoura, Panos Vassiliadis and Apostolos Zarras (2012). *Cross-Disciplinary Models and Applications of Database Management: Advancing Approaches* (pp. 385-416).

www.irma-international.org/chapter/accelerating-web-service-workflow-execution/63675

Distinguishing Personality Recognition and Quantification of Emotional Features Based on Users' Information Behavior in Social Media

Chunnian Liu, Qi Tian and Mengqiu Chen (2021). *Journal of Database Management* (pp. 76-91).

www.irma-international.org/article/distinguishing-personality-recognition-and-quantification-of-emotional-features-based-on-users-information-behavior-in-social-media/273618