PENGINTEGRASIAN DATA EXTENSIBLE MARKUP LANGUAGE (XML) KE DALAM PANGKALAN DATA HUBUNGAN MENGGUNAKAN COMMON WAREHOUSE METAMODEL (CWM)

WAN MOHD HAFFIZ BIN MOHD NASIR

Tesis ini dikemukakan
sebagai memenuhi syarat penganugerahan
ijazah Sarjana Sains (Sains Komputer)

Fakulti Sains Komputer dan Sistem Maklumat
Universiti Teknologi Malaysia

SEPTEMBER 2007

ABSTRAK

Permintaan kepada pengintegrasian data secara pantas menjadi semakin tinggi dengan semakin banyak sumber-sumber maklumat yang terdapat di dalam perusahaan moden. *Extensible Mark-up Language* (XML) telah menjadi satu piawaian baru bagi perwakilan dan pertukaran data dalam *World Wide Web* (WWW), contohnya di dalam aplikasi *Business to Business* (B2B) pada e-dagang. Ini memerlukan alatan analisis data untuk mengendalikan data XML di samping format data tradisional. Tujuan penyelidikan ini adalah bagi meningkatkan kaedah pengintegrasian data XML ke dalam data hubungan berikutan berlakunya beberapa masalah daripada hasil proses pengintegrasian seperti kehilangan data. Kaedah yang dicadangkan daripada penyelidikan ini adalah melalui pengaplikasian *Common Warehouse Metamodel* (CWM) sebagai piawaian bagi pertukaran dan pengurusan metadata yang menggabungkan satu perkongsian metamodel bagi sintak dan semantik metadata. Hasil daripada penyelidikan ini adalah satu cadangan peningkatan senibina dan pendekatan pengintegrasian melalui pengaplikasian CWM serta satu perisian antaramuka yang telah dibangunkan bagi memudahkan proses pengintegrasian sebagai penyelesaian kepada masalah kehilangan data.

ABSTRACT

The demand for rapid data integration is getting higher as more and more information sources appear in modern enterprises. Extensible Mark-up Language (XML) is fast becoming the new standard for data representation and exchange on the World Wide Web, e.g., in B2B e-commerce, making it necessary for data analysis tools to handle XML data as well as traditional data formats. The purpose of this research is to enhance the technique for XML data integration into relational data to solve integration problems such as missing data. The method that had been proposed is to apply Common Warehouse Metamodel (CWM) for metadata interchange and metadata management that incorporates a common shared metamodel for metadata syntax and semantics. The results of this research are the enhancement of the integration architecture and approach by applying CWM as well as the development of an interface software to simplify the integration process as a solution for the missing data problem.

KANDUNGAN

SENARAI JADUAL

SENARAI RAJAH

SENARAI SINGKATAN

CWM	-	Common Warehouse Metamodel
XML	-	Extensible Mark-up Language
DTD	-	Document Type Definition
MIW	-	Meta Integration Works
MIMB	-	Meta Integration Model Bridge
B2B	-	Business To Business
OLAP	-	Online Analytical Processing
ETL	-	Exctract, Transform, & Load
BI	-	Bussiness Intelligence
OMG	-	Object Management Group
UML	-	Unified Modeling Language
URL	-	Uniform Resource Locator
OIM	-	International Organization for Migration
WWW	-	World Wide Web

SENARAI ISTILAH

Sistem Pengurusan Pangkalan Data	-	*Database Management System*
Capaian semula maklumat	-	*Information retrieval*
Pertanyaan	-	*Query*
File Teks	-	*Text file*
Hubungan	-	*Relation*
Jadual	-	*Table*
Bahasa Pertanyaan Berstruktur	-	*Structured Query Language*

SENARAI LAMPIRAN

RUJUKAN

Agosta, L. (2001). Reports of the demise of metadata are premature.DM Review 3

Auth, G. dan Eitil, V.M. (2002). A Software Architecture for XML-based Metadata Interchange in Data Warehouse Systems.

Bertino E. dan Ferrari E. (2001). XML and Data Integration. IEEE Internet Computing.

Bremeau C. (2001). XML Data Movement Components for Teradata. www.metaintegration.com, July 2004.

DAMA (2005). *"Metadata Based Impact and Lineage Analysis Across Heterogeneous Metadata Sources".* Wilshire Meta-Data Conference. 24-25 May 2005.

Do, H. H., Rahm, E. (2000). On metadata interoperability in data warehouses. Technical Report 1- 2000, Institute Informatics, University Leipzig.

Haag, S., Cummings, M., dan McCubbery, D.J (2002). Alain Pinsonneault,Richard Donvan: *Managements Information System for the Information Age*, Third Canadian Edition, McGraw-Hill Ryerson.

Halevy, A.Y, Ashish N., Bitton, D., Carey, M.J., Draper, D., Pollock, J., Rosenthal, A., dan Sikka, V. (2005). "Enterprise information integration: successes, challenges and controversies". *SIGMOD 2005*, 778-787.

Holzner, S., (1997). XML Complete. McGraw-Hill

Inmon, W.H., dan Hackathorn R.D. (2001). *Using the Data Warehouse*, John Wiley & Sons.

Jensen, M. R., Møller, T.H., dan Pedersen, T.B. (2001a). Specifying OLAP Cubes On XML Data. *Tech Report R-01-5003,* Department Of Computer Science, Aalborg University.

Jensen, M. R., dan T. H. Møller (2001b). Constructing OLAP Cubes From XML Data. *Tech Report R-02-5003,* Department Of Computer Science, Aalborg University.

Kimball, R. dan Ross, M.(2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*, John Wiley & Sons.

Koshafian, S. dan Abnous, R. (1995). Object Orientation – Concepts, Analysis, and Design, Languages, Databases, Graphical User Interfaces, Standard. 2nd ed. Wiley.

Lahiri, T et. al. Ozone (1999). Integrating Structured and Semi structured Data, *Proceedings of the Seventh International Conference on Database Programming Languages*, Kinloch Rannoch, Scotland.

Lenz, H., (1997). Summarizability in OLAP and Statistical Databases, *Proceedings of the Ninth International Conference on Statistical and Scientific Database Management*, 39-48.

Lenzerini, M. (2002). "Data Integration: A Theoretical Perspective". *PODS 2002*, 243-246.

Mimno, P. (2002). Succesful Real-Time Business Analytics: A Data Warehousing Strategy. White Paper. Informatica Carporation.

Pedersen, et. al.(2000). Extending OLAP Querying To External Object Databases, *Proceedings of the Ninth International Conference on Information and Knowledge Management*, ms. 405-413.

Pledge, K. dan McGarry, J. (2001). Data Warehousing for Actuaries. Versi (1.2): 3-12.

Poole, J., Chang, D., Tolbert, D., dan Mellor, D. (2003). Common Warehouse Metamodel Developer's Guide. John Wiley & Sons Inc.

Poole, J. (2000). The Common Warehouse Metamodel as a Foundation for Active Object Models in the Data Warehouse Environment. Position paper to ECOOP 2000 workshop on Metadata and Active Object-Model Pattern Mining – Cannes, France.

Pyle dan Dorian (2003) *Business Modeling and Data Mining*. Morgan Kaufmann,

Rafanelli, M. 1990). STORM: A Statistical Object Representation Model, *Proceedings of the Fifth Conference on Statistical and Scientific Database Management*, 14-29.

Shanmugasundaram, et. al (1999). Relational Databases for Querying XML Documents: Limitations and Opportunities, *Proceedings of the Twenty-Fifth International Conference on Very Large Databases*, Edinburgh, Scotland.

Shukla, A. (1996).Storage Estimation for Multidimensional Aggregates in the Presence

of Hierarchies, *Proceedings of Very Large Databases*, pp. 522-531.

Silicon Integration Initiative (2000). *The Electronic Component Information Exchange QuickData Architecture*.

Thomsen, E., (1997). *OLAP Solutions: Building Multidimensional Information Systems*, John Wiley & Sons, Inc.

W3C (2001). World Wide Web Consortium, *The XML Query Algebra*, W3C Working Draft,http://www.w3.org/TR/query-algebra, Dec. 4 2000.

Ziegler, P. dan Dittrich, K.R (2004). "Three Decades of Data Integration - All Problems Solved?". *WCC 2004*, 3-12.