

# The Study of Indexing Strategy on XML-based Temporal Data

張晏嘉、邱紹豐

E-mail: 9601130@mail.dyu.edu.tw

## ABSTRACT

The temporal data is widely utilized and studied. The researchers are interested in the storage management and indexing technique on this special type of data. XML is a newly developed data structure and exchange standard in recent years. Because of its flexible structure of the XML document and independent communication protocol, XML document is especially suitable for transmitting on the internet network and World Wide Web. Different from the traditional database management systems, the numerous characteristics of XML make the data easier to exchange and communicate. As a result, more data are stored in XML format. If the temporal data stored in XML structure, searching data must check for each data of document, which results in high cost. For temporal data, the traditional indexing techniques considered only single attribute in data clustering, making it faster to retrieve data by the indexed attribute. However, the efficiency declines dramatically when the search criteria is not indexed. In our research, we proposed a new data clustering method for improving comprehensive efficiency of data retrieval. By building the temporal data in the multidimensional matrix observes the relation among the data. Then, we try to cut matrix apart into a lot of blocks by considering all attributes and utilized the characteristics of greedy algorithm to improve the speed of the data clustering. By this approach, we can improve comprehensive efficiency of data retrieval.

Keywords : Temporal data ; XML ; Greedy Algorithm

## Table of Contents

封面內頁 簽名頁 授權書.....	iii	中文摘要.....	iv	英文摘要.....	v	誌謝.....	vi	目錄.....	vii	圖目錄.....	ix	表目錄.....	xi									
第一章 緒論.....	1	1.1 研究動機.....	1	1.2 研究目的.....	3	1.3 論文結構.....	4	第二章 相關研究.....	5	2.1 XML.....	5	2.2 時序性資料.....	7	2.3 時序性資料庫.....	8	2.4 以XML儲存時序性資料.....	12	2.5 分群方式.....	15	2.6 索引架構.....	17	
第三章 設計方法.....	22	3.1 分群策略.....	22	3.2 分群方法.....	26	第四章 實驗評估.....	36	4.1 測試環境.....	36	4.2 效能評估.....	38	第五章 結論與未來工作.....	42	參考文獻.....	44							

## REFERENCES

- [1] 林信成,陳勇任, "基於XML之網際網路資料交換雛形系統設計",教育資料與圖書館學,第三十九,第二期,頁145-160,民90年12月。
- [2] Extensible Markup Language (XML) 1.0 (Third Edition), <http://www.w3.org/TR/XML/>.
- [3] B+ Tree, [http://en.wikipedia.org/wiki/B+\\_tree](http://en.wikipedia.org/wiki/B+_tree).
- [4] A. Segev, H. Gunadhi, "Event-Join Optimization in Temporal Relational Databases", Proc.15th Conference on Very Large Data Bases, pages 205-215, 1989.
- [5] B. Salzberg and V. Tsotras, "A Comparison of Access Methods for Temporal Data", ACM Computing Surveys, vol. 31, no. 2, pages 158-221, 1999.
- [6] Cheng Hian Goh, Hongjun Lu, Beng Chin Ooi, Kian-Lee Tan, "Indexing Temporal Data Using Existing B+-Trees", Data Knowl. Eng. 18(2),pages 147-165,1996.
- [7] Christian S. Jensen and Richard T. Snodgrass, "Semantics of Time-Varying Information", Information Systems, 21(4),pages 311 – 352, June 1996.
- [8] Christian S. Jensen, James Clifford, Ramez Elmasri, Shashi K. Gadia, Patrick J. Hayes, Sushil Jajodia, "A Consensus Glossary of Temporal Database Concepts", SIGMOD Record 23(1), pages 52-64, 1994.
- [9] Christian S. Jensen, "Introduction to Temporal Database Research", <http://www.cs.aau.dk/~csj/Thesis/pdf/chapter1.pdf>.
- [10] David B. Lomet, Betty Salzberg, "Access Methods for Multiversion Data", SIGMOD Conference, pages 315-324, 1989.
- [11] Fusheng Wang, Carlo Zaniolo, "An XML-Based Approach to Publishing and Querying the History of Databases", World Wide Web 8(3),

pages 233-259 ,2005.

- [12] Fabio Grandi, Federica Mandreoli, “ The Valid Web: An XML/XSL Infrastructure for Temporal Management of Web Documents ” , ADVIS , pages 294-303, 2000.
- [13] H. Gunadhi, A. Segev, “ Efficient Indexing Methods for Temporal Relations ” , IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 3, pages 496-509, 1993.
- [14] James Clifford, Curtis E. Dyreson, Richard T. Snodgrass, Toma's Isakowitz, Christian S. Jensen, “ Now ” , The TSQL2 Temporal Query Language, pages 383-392, 1995.
- [15] J. Clifford, C. E. Dyreson, T. Isakowitz, C. S. Jensen, R. T. Snodgrass, “ On the Semantics of “ Now ” in Databases ” , ACM Trans. Database Syst. 22(2),pages 171-214 ,1997.
- [16] Mario A. Nascimento , Margaret H. Dunham, Indexing Valid Time Databases via B+-Trees, IEEE Transactions on Knowledge and Data Engineering, vol 11 , pages 929-947, November 1999.
- [17] R. Elmasri, G. Wu, V. Kouramajian, “ The Time Index and the Monotonic B+-tree ” , in A.Tansel, J. Clifford, S.K. Gadia, S. Jajodia, A. Segev, and R. Snodgrass (eds.), Temporal Databases: Theory, Design, and Implementation, Benjamin/Cummings, pages 433-456, 1993.
- [18] T. Amagasa, M. Yoshikawa, and S. Uemura. A data model for temporal XML documents. In DEXA, 2000.
- [19] T. Amagasa, M. Yoshikawa, and S. Uemura. Realizing temporal XML repositories using temporal relational databases. In CODAS, pages 63-68,2001.
- [20] T. Y. Cliff Leung, Richard R. Muntz, “ Generalized Data Stream Indexing and Temporal Query Processing ” , RIDE-TQP, pages 124-131, 1992.
- [21] T. Y. Cliff Leung, Richard R. Muntz, “ Temporal Query Processing and Optimization in Multiprocessor Database Machines ” , VLDB, pages 383-394, 1992.
- [22] T. Y. Cliff Leung, Richard R. Muntz, “ Stream Processing: Temporal Query Processing and Optimization ” , Temporal Databases, pages 329-355, 1993.
- [23] Xiaoping Ye, Yong Tang, Zhaomei Huang, Yiquan Jiang, “ Querying XML documents with temporal variables ” , Computer Supported Cooperative Work in Design , pages 536 – 540, May 2004.