The Study of Extensible Index System on XML Data

陳嘉進、邱紹豐

E-mail: 9511436@mail.dyu.edu.tw

ABSTRACT

XML has emerged as the standard for data exchange, and processing. Many query systems have been developed to retrieve required information from the XML data source. However, the difficulty in processing range query is not solved, since the indexing system in such environment is not fully matured yet. In this thesis, we developed an indexing system that incorporates R-tree and data guide, called range data guide, or RDG. In addition to indexing data paths, the system also allows indexing on ranges of data.. The idea of the system is to embed R-tree at the leaf nodes in the data paths in Data Guide. If the users request is not a range query, the regular Data Guide provides indices on data paths. Otherwise, the R-tree at the leaf nodes allow users to access a certain range of data. We have developed the structure of the index and the related algorithms, i.e., insertion, deletion, and update. The experimental results show that our proposed system offers performance improvement over other traditional techniques.

Keywords : Semi-Structured Data ; index XML ; range index ; access data

Table of Contents

REFERENCES

[1] J. Roy, A. Ramanujan.,“ XML: data's universal language, ” IT Professional, Volume 2, Issue 3, pages 32-36, May-June 2000.

[2] W3C,“ Extensible Markup Language (XML), ” http://www.w3.org/XML/, 2003.

[3] A. Zisman.,“ An overview of XML, ” Computing and Control Engin. J., Volume 11, pages 165-167, Aug. 2000.

[4] W3C,“ Overview of SGML Resources, ” http://www.w3.org/MarkUp/SGML/, Nov. 1995.

[5] S. Abiteboul, P. Buneman and D. Suciu,“ Data on the Web, ” Morgan Kaufmann Publishers, 2000.

[6] W3C,“ Extensible Stylesheet Language (XSL) Version 1.0, ” http://www.w3.org/TR/xsl/, Oct. 2001.

[7] W3C,“ Mathematical Markup Language (MathML?) 1.01 Specification, ” http://www.w3.org/TR/REC-MathML/, 1999.

[8] Peter Murray-Rust, Henry Rzepa,“ Chemical Markup Language (CML?), ” http://www.xml-cml.org/, 2002.

[9] Per-Ake Larson “ XML Data Management : Go Native or Spruce up Relational Systems?, ” ACM SIGMOD 2001 Santa Barbara, May 21-24, 2001.

[10] Yannis Papakonstantinou, Hector Garcia-Molina, Jennifer Widom,“ Object Exchange Across Heterogeneous Information Sources, ” In Proc. of the 11th International Conference on Data Engineering, pp. 251-260. Mar. 1995.

[11] S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman and J. Widom,“ The TSIMMIS Project: Integration of Heterogeneous Information Sources, ” 16th Meeting of the Information Proc. Society of Japan, pp.7-8. Oct. 1994.

[12] H. Garcia-Molina, Y. Papakonstantinou, D. Quass, A. Rajaraman, Y. Sagiv, J.Ullman, and J. Widom,“ The TSIMMIS Approach to Mediation: Data Models and Languages, ” In Proc. of 2th International Workshop on Next Generation Information Technologies and Systems, pp. 185-193. Jun. 1995.

[13] J. McHugh, S. Abiteboul, R. Goldman, D. Quass, and J. Widom,“ Lore: A database management system for semistructured data, ”

Technical report, Stanford University Database Group, 1997.
[14] S. Abiteboul, D. Quass, J. McHugh, J. Widom, J. L. Wiener. “ The Lorel query language for semistructured data, ” International Journal on Digital Libraries, Volume 1, Issue 1, pages 68-88, 1997.
[15] D. Quass, A.Rajaraman, Y. Sagiv, J. Ullman, and J.Widom, “ Querying semistructured heterogeneous information, ” In Proc. of the Fourth International Conference on Deductive and Object-Oriented Databases, pp. 319-344. Dec. 1995.
[16] R. Goldman and J. Widom, “ DataGuides: Enabling query formulation and optimization in semistructured databases, ” In Proc. of the 23th International Conference on VLDB, pp. 436-445. Aug. 1997.
[17] V. Christophides, S. Cluet and G. Moerkotte, “ Evaluating Queries with Generalized Path Expression, ” In Proc. Of the ACM SIGMOD International Conference on Management of Data, pp. 413-422. June. 1996.
[18] J. McHugh and J. Widom, “ Query optimization for semistructured data, ” Technical report, Stanford University Database Group, 1997.
[19] J. McHugh and J. Widom, “ Query Optimization for XML, ” The VLDB Journal, pp. 315-326, Sep. 1999.
[20] Alin Deutsch, Mary Fernandez, Daniela Florescu, Alon Levy and Dan Suciu, “ XML-QL: A Query Language for XML, ” http://www.w3.org/TR/NOTE-xml-ql/, 19-August-1998.
[21] R. Goldman and J. Widom, “ Approximate DataGuides, ” In Proc. of the Workshop on Query Processing for Semistructured Data and Non-Standard Data Formats, pp. 436-445. Jan. 1999.
[22] R. Kaushik, P. Shenoy, P. Bohannon and E. Gudes, “ Exploiting Local Similarity for Indexing Paths in Graph-Structured Data, ” 18th ICDE, pp.129-140, 2002.
[23] B. Cooper, N. Sample, M. J. Franklin, G. R. Hjaltason, and M. Shadmon, “ A Fast Index for Semistructured Data, ” In Proc. of the 27th VLDB, pp. 341-350, Sep. 2001.
[24] F. Rizzolo, A. Mendelzon, “ Indexing XML Data with ToXin, ” In Proc. of the 4th International Workshop on the Web and Databases , pp. 49-54, May 2001.
[25] 邱紹豐, 林錡嵐, “ 半結構性資料索引之研究, ” Communication of IICM, Vol. 6, No. 1, pp. 99-112, March 2003.
[26] A. Guttman, “ R-trees: A Dynamic Index Structure for Spatial Searching, ” In Proc. of the 1984 ACM SIGMOD International Conference on Management of Data, pp. 47-57, 1984.