具延展性之 XML 資料索引系統之研究

陳嘉進、邱紹豐

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

摘 要

可擴展標示語言(Extensible Markup Language, XML)資料已經被大量應用於交換資料及處理資訊上，但隨著此技術廣泛地被採用，相對的資料量也愈來愈龐大，衍生的問題即是資料的資訊查詢問題。目前XML索引系統在範圍性的查詢資料時，採用的方法大多是兩階段的查詢方式；在第一階段，使用索引系統查詢出符合路徑表示式(Path Expression)的資料路徑(Data Path)集合，而在第二階段時，再一一比對每個資料路徑所表示的資料數值(Value)是否為所求，以查詢出符合範圍的資料路徑。以此傳統的查詢方式，大多的資源花費在比對資料數值上，為了使XML索引系統能擁有更好的查詢效率，在本論文中介紹範圍性資料嚮導(Range DataGuide, RDG)。其基本原理是把相同路徑的葉節點集合建立範圍性索引系統，將原本應用於傳統關聯式資料庫的範圍性索引技術應用於索引XML資料，降低比對資料數值的次數，並使索引路徑與比對資料數值皆由XML索引系統來完成，提升XML資料在範圍性查詢時的效率。 在本論文中我們提供了索引系統的結構及其相關功能之演算法，如新增、刪除、更新索引等，提供快速存取XML資料的機制。

關鍵詞 : 半結構性資料 ; XML索引 ; 範圍性索引 ; 資料存取

目錄

參考文獻

[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.