P.O.Box 8718, Beijing 100080, China	Journal of Software, May 2005,16(5):827-837
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2005 by The Editorial Department of Journal of Software

以目标节点为导向的XML路径查询处理

王 静, 孟小峰, 王 宇, 王 珊

王 静1, 孟小峰2, 王 宇2, 王 珊2
(北京大学 信息科学技术学院,北京 100871)
作者简介: 王静(1975－),女,山西襄垣人,博士,主要研究领域为数据库与知识库,XML数据管理;孟小峰(1964—),男,博士,教授,博士生导师,主要研究领域为数据库与知识库,Web数据管理;王宇(1973—),女,博士生,主要研究领域为数据库与知识库,XML数据管理;王珊(1944—),女,教授,博士生导师,主要研究领域为数据库与知识库,数据仓库.
联系人: 王 宇 Phn: +86-10-62515575, Fax: 86-10-62519453, E-mail: sandpiperwy@yahoo.com.cn

Received 2003-12-25; Accepted 2004-06-10

Abstract
XML query languages take complex path expressions as their core. To facilitate path expression processing, the processing strategy based on path decomposition and structural join operation needs to be investigated more deeply. In this paper, a target node aimed at path expression processing framework for XML data is proposed. This approach makes use of the extended basic operations to reduce the number of join operations. In the procedure of path decomposition and query plan selection, target node in the query tree is utilized to avoid the transfer of the intermediate results. In addition to decomposition rules and strategies, a set of extended basic operations and implementation algorithms are proposed.

Wang J, Meng XF, Wang Y, Wang S. Target node aimed path expression processing for XML data. *Journal of Software*, 2005,16(5):827-837.
DOI: 10.1360/jos160827
http://www.jos.org.cn/1000-9825/16/827.htm

摘要
XML查询语言将复杂路径表达式作为核心内容.为了加速路径表达式处理,基于路径分解和结构连接操作的处理策略需要更深入的研究.以目标节点为导向的XML路径查询处理框架被提了出来.该方法利用了扩展基本操作来减少连接操作的数目.在路径分解和查询计划选择的过程中,利用查询树中的目标节点来避免中间结果的传递.除了分解规则和策略以外,提出了一组扩展的基本操作和实现算法.初步的实验结果显示,该方法具有良好的性能.它为路径查询处理提供了更多的选择.

基金项目：Supported by the National Natural Science Foundation of China under Grant Nos.60073014, 60273018 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2002AA116030 (国家高技术研究发展计划(863)); the Key Project of Ministry of Education of China under Grant No.03044 (国家教育部科学技术重点项目); the Excellent Young Teachers Program of Ministry of Education of China (教育部优秀青年教师资助计划)

References:

[1] Bray T, Paoli J, Sperberg-McQueen CM, Maler E, eds. Extensible markup language (XML) 1.0 (2nd Edition). W3C Recommendation, 2000. http://www.w3.org/TR/2000/REC-xml-20001006

[2] Clark J, DeROse S, eds. XML Path language (XPath) Version 1.0. W3C Recommendation, 1999. http://www.w3.org/TR/1999/REC-xpath-19991116

[3] Chamberlin D, Florescu D, Robie J, Simeon J, Stefanescu M, eds. XQuery: A query language for XML. W3C Working Draft, 2001. http://www.w3.org/TR/2001/WD-xquery-2001215

[4] Goldman R, McHugh J, Widom J. From semisturctured data to XML: Migrating the lore model and query language. In: Proc. of the 2nd Int'l Workshop on the Web and Databases (WebDB'99). 1999. http://www-rocq.inria.fr/~cluet/WEBDB/lore.ps

[5] McHugh J, Widom J. Query optimization for XML. In: Atkinson MP, Orlowska ME, Valduriez P, Zdonik SB, Brodie ML, eds. Proc. of the 25th Int'l Conf. on Very Large Data Bases. San Francisco: Morgan Kaufmann Publishers, 1999. 315-326.

[6] Zhang C, Naughton J, DeWitt D, Luo Q, Lohman G. On supporting containment queries in relational database management systems. In: Timos S, ed. Proc. of the 2001 ACM SIGMOD Int'l Conf. on Management of Data. New York: ACM Press, 2001. 425-436.

[7] Li QZ, Moon B. Indexing and querying XML data for regular path expressions. In: Apers PMG, Atzeni P, Ceri S, Paraboschi S, Ramamohanarao K, Snodgrass RT, eds. Proc. of the 27th Int'l Conf. on Very Large Data Bases. San Francisco: Morgan Kaufmann Publishers, 2001. 361-370.

[8] AI-Khalifa S, Jagadish HV, Koudas N, Patel JM, Srivastava D, Wu YQ. Structural joins: A primitive for efficient XML query pattern matching. In: Agrawal R, Dittrich K, Ngu AHH, eds. Proc. of the 18th Int'l Conf. on Data Engineering. Los Alamitos: IEEE Press, 2002. 141-152.

[9] Wang W, Jiang H, Lu H, Yu X. PBiTree coding and efficient processing of containment join. In: Dayal U, Ramamritham K, Vijayaraman TM, eds. Proc. of the 19th Int'l Conf. on Data Engineering. Los Alamitos: IEEE Press, 2003. 391-402.

[10] Chan C-Y, Felber P, Garofalakis M, Rastogi R. Efficient filtering of XML documents with XPath expressions. In: Bernstein PA, et al, eds. Proc. of the 28th Int'l Conf. on Very Large Data Bases. San Francisco: Morgan Kaufmann Publishers, 2002. 354-379.

[11] Wang J, Meng XF, Wang S. SUPEX: Schema guided path index for XML data. Computer Science, 2002,29(8A):25-38 (in Chinese with English abstract).

[12] Wang J, Meng XF, Wang S. Structural join of XML based on range partitioning. Journal of Software, 2004,15(5):720-729 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/15/720.htm

[13] Wu YQ, Patel J, Jagadish HV. Estimating answer sizes for XML queries. In: Jensen CS, et al, eds. Proc. of the 8th Int'l Conf. on Extending Database Technology. Prague: Springer-Verlag, 2002. 590-608.

[14] Wang W, Jiang H, Lu H, Jeffrey XY. Containment join size estimation: Models and methods. In: Halevy AY, Ives ZG, Doan A, eds. Proc. of the 2003 ACM SIGMOD Int'l Conf. on Management of Data. San Diego: ACM Press, 2003. 145-156.

[15] Wu YQ, Patel J, Jagadish H. Structural join selection for XML query optimization. In: Dayal U, Ramamritham K, Vijayaraman TM, eds. Proc. of the 19th Int'l Conf. on Data Engineering. Los Alamitos: IEEE Press, 2003. 443-454.

[16] Nicolas B, Nick K, Divesh S. Holistic Twig joins: Optimal XML pattern matching. In: Franklin MJ, et al, eds. Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. Madison: ACM Press, 2002. 310-321.

附中文参考文献:

[11] 王静,孟小峰,王珊.SUPEX:一种基于模式的XML路径索引.计算机科学, 2002,29(8A):25-38.

[12] 王静,孟小峰,王珊.基于区域划分的XML结构连接.软件学报,2004,15(5):720-729 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/15/720.htm