

P.O.Box 8718, Beijing 100080, China	Journal of Software, May 2004,15(5):720-729
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2004 by The Editorial Department of Journal of Software

基于区域划分的XML结构连接

王 静, 孟小峰, 王 珊

[Full-Text PDF](#) [Submission](#) [Back](#)

王 静¹, 孟小峰², 王 珊², ¹(中国科学院 计算技术研究所,北京 100080)²(中国人民大学 信息学院,北京 100872)

作者简介: 王静(1975—),女,山西襄垣人,博士生,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系统,Web数据管理,移动数据管理;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系统,数据仓库.

联系人: 王 静 Phn: +86-10-62515575, Fax: +86-10-62519453, E-mail: jwang_lq@yahoo.com.cn

Received 2003-03-02; Accepted 2003-06-26

Abstract

Structural join is the core operation in XML query processing, and catches the research community's attention. Efficient algorithm is the key of efficient query processing. There have been a number of algorithms proposed for structural join, and most of them are based on one of the following assumptions: the input element sets either have indexes or are ordered. When these assumptions are not true, the performance of these algorithms will become very bad due to the cost of sorting input sets or building index on the fly. Motivated by this observation, a structural join algorithm based on range partitioning is proposed in this paper. Based on the idea of task decomposition, this algorithm takes advantage of the region numbering scheme to partition the input sets. The procedure of the algorithm is described in detail, and its I/O complexity is analyzed. The results of the extensive experiments show that this algorithm has good performance and is superior to the existing sort-merge algorithms when the input data are not sorted or indexed. It can provide query plans with more choices.

Wang J, Meng XF, Wang S. Structural join of XML based on range partitioning. Journal of Software, 2004, 15(5):720~729.

<http://www.jos.org.cn/1000-9825/15/720.htm>

摘要

结构连接是XML查询处理的核心操作,受到了研究界的关注.高效的算法是高效查询处理的关键.目前已经提出了许多结构连接的算法,它们中的大多数都基于如下的前提条件之一:输入元素集合存在索引或者有序.当这些条件不成立时,由于对输入数据临时排序或建索引的代价,这些算法的性能会大大下降.基于这样的观察,提出了一种基于区域划分的结构连接算法.该算法基于任务分解的思想,利用区域编码的特点对输入集合进行划分.给出了详细的算法设计,并对算法的I/O复杂性进行了分析.大量的实验结果显示,该算法具有良好的性能,在输入数据无序或没有索引的情况下优于现有的排序合并算法,可以为查询计划提供更多的选择.

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 the 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 (国家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 (second edition). W3C Recommendation 6, 2000. <http://www.w3.org/TR/2000/REC-xml-20001006>

[2] Shanmugasundaram J, Tufte K, He G, Zhang C, DeWitt D, Naughton J. Relational databases for querying XML documents: Limitations and opportunities. In: Atkinson MP, Orłowska 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. 302~314.

- [3] Florescu D, Kossmann D. Storing and querying XML data using an RDBMS. IEEE Data Engineering Bulletin, 1999,22(3): 27~34.
- [4] Clark J, DeRose S, eds. XML path language (XPath) Version 1.0. W3C Recommendation 16, 1999, <http://www.w3.org/TR/1999/REC-xpath-19991116>
- [5] Chamberlin D, Clark J, Florescu D, Robie J, Simeon J, Stefanescu M. XQuery: A query language for XML. W3C Working Draft 07, 2001. <http://www.w3.org/TR/2001/WD-xquery-20010607>
- [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] Al-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] Chien SY, Vagena Z, Zhang DH, Tsostras VJ, Zaniolo C. Efficient structural joins on indexed XML documents. In: Bernstein PA, et al., eds. Proc. of the 28th Int'l Conf. on Very Large Data Bases. San Francisco: Morgan Kaufmann Publishers, 2002. 263~274.
- [10] Jiang HF, Lu HJ, Wang W, Ooi BC. XR-Tree: Indexing XML data for efficient structural joins. In: Dayal U, Ramamritham K, Vijayaraman TM, eds. Proc. of the 19th Int'l Conf. on Data Engineering. Los Alamitos: IEEE Press, 2003. 253~264.
- [11] Wang W, Jiang HF, Lu HJ, Jeffrey XY. PBiTree coding and efficient processing of containment joins. 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.
- [12] Lee YK, Yoo SJ, Yoon K. Index structure for structured documents. In: Edward AF, Gary M, eds. Proc. of the 1st ACM Int'l Conf. on Digital Libraries. New York: ACM Press, 1996. 91~99.
- [13] IBM Corporation. XML data generator. <http://www.alphaworks.ibm.com/tech/xmlgenerator>
- [14] DBLP. <http://dblp.uni-trier.de/xml/>
- [15] Schmidt AR, Waas F, Kersten ML, Florescu D, Manolescu I, Carey MJ, Busse R. The XML benchmark project. Technical Report, INS-R0103, Amsterdam: CWI, 2001.