

P.O.Box 8718, Beijing 100080, China	Journal of Software, Feb. 2007,18(2):323-331
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2007 by <i>Journal of Software</i>

XML数据物化模式的生成与优化技术

张亮, 李然, 汪卫, 施伯乐

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

张亮, 李然, 汪卫, 施伯乐

(复旦大学 计算机与信息技术系, 上海 200433)

作者简介: 张亮(1963—), 男, 湖北武汉人, 博士, 教授, 博士生导师, CCF高级会员, 主要研究领域为数字图书馆, 多媒体技术, 信息集成, 生物信息学. 李然(1973—), 男, 硕士, 主要研究领域为数字图书馆, XML管理. 汪卫(1970—), 男, 博士, 教授, 博士生导师, 主要研究领域为数据库原理, 复杂结构数据管理, 数据挖掘. 施伯乐(1936—), 男, 教授, 博士生导师, CCF高级会员, 主要研究领域为数据库理论与应用, 数字图书馆, 数据仓库, 数据挖掘.

联系人: 张亮 Phn: +86-21-65643501, Fax: +86-21-65642219, E-mail: zhangl@fudan.edu.cn, <http://www.cit.fudan.edu.cn>

Received 2004-11-22; Accepted 2006-03-31

Abstract

One way to improve the performance of XML (extensible markup language) management systems is to materialize part of the XML documents and store them aside in cache memory. In this paper, a method is presented to characterize query sets of XML data as schema graph, which is a technique to generate materialized plan based on the distribution of user's queries. Experimental results demonstrate its performance gain in XML cache management.

Zhang L, Li R, Wang W, Shi BL. A technique to generate and optimize the materialized model for XML data. *Journal of Software*, 2007, 18(2): 323-331.

DOI: 10.1360/jos180323

<http://www.jos.org.cn/1000-9825/18/323.htm>

摘要

根据用户查询的分布情况, 基于缓存以及在硬盘上对XML(extensible markup language)数据进行物化是提高XML数据存储与查询系统性能的主要方法之一. 提出了一种XML查询集合的描述方法, 即查询模式图, 并以此为基础提出了一种能够充分考虑查询优化策略的物化方案生成方法. 实验结果表明, 该方法可以快速地生成物化方案, 能够满足缓冲区管理等应用领域的需要.

基金项目: Supported by the National Natural Science Foundation of China under Grant Nos.60303008, 69933010 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2002AA4Z3430 (国家高技术研究发展计划(863))

References:

- [1] Extensible markup language (XML) 1.0. 4th ed., W3C, 2006. <http://www.w3.org/TR/xml>
- [2] 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 Computer Society, 2002. 141-154.
- [3] Deutsch A, Fernandez M, Suci D. Storing semistructured data with STORED. In: Davidson SB, Faloutsos C, eds. Proc. of the 1999 ACM SIGMOD Int'l Conf. on Management of Data. New York: ACM Press, 1999. 431-442.
- [4] Shanmugasundaram J, Tufte K, He G, Zhang C, de Witt 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.

[5] Abiteboul S. Querying semi-structured data. In: Afrati FN, Kolaitis PG, eds. Proc. of the Int'l Conf. on Database Theory. LNCS 1186, London: Springer-Verlag, 1997. 1-18.

[6] Tian F, DeWitt DJ, Chen JJ, Zhang C. The design and performance evaluation of alternative XML storage strategies. ACM SIGMOD Record, 2002,31(1):5-10.

[7] Gupta H. Selection of views to materialize in a data warehouse. In: Goos G, Hartmanis J, Leeuweneds JV, eds. Proc. of the 6th Int'l Conf. on Database Theory. LNCS 1186, Heidelberg: Springer-Verlag, 1997. 98-112.

[8] Arms WY, Write; Shi BL, Zhang L, Wang W, et al., Trans. Digital Libraries. Beijing: Publishing House of Electronics Industry, 2001 (in Chinese).

[9] Yang LH, Lee ML, Hsu W. Efficient mining of XML query patterns for caching. In: Freytag JC, Lockemann PC, Abiteboul S, Carey MJ, Selinger PG, Heuer A, eds. Proc. of the 29th Int'l Conf. on Very Large Data Bases. San Francisco: Morgan Kaufmann Publishers, 2003. 69-80.

[10] Dong X, Halevy AY, Tatarinov I. Containment of nested XML queries. In: Nascimento MA, Ozsu MT, Kossmann D, Miller RJ, Blakeley JA, Schiefer KB, eds. Proc. of the 30th Int'l Conf. on Very Large Data Bases. San Francisco: Morgan Kaufmann Publishers, 2004. 132-143.

[11] Lu Y, Zhang L, Wang W, Duan QY, Shi BL. DTD ranking in the smart XML query. Journal of Computer Research and Development, 2003,40(11):1579-1585 (in Chinese with English abstract).

[12] Jagadish HV, Lakshmanan LVS. TAX: A tree algebra for XML. In: Goos G, Hartmanis J, Leeuwen JV, eds. Proc. of the 8th Int'l Workshop on Databases and Programming Languages LNCS 2397, Heidelberg: Springer Berlin, 2002. 149-164.

[13] Shukla A, Deshpande PM, Naughton JF. Materialized view selection for multidimensional datasets. In: Gupta A, Shmueli O, Widom J, eds. Proc. of the 24th Int'l Conf. on Very Large Data Bases. San Francisco: Morgan Kaufmann Publishers, 1998. 488-499.

附中文参考文献:

[8] Arms WY, 著; 施伯乐, 张亮, 汪卫, 等, 译. 数字图书馆概论. 北京: 电子工业出版社, 2001.

[11] 路燕, 张亮, 汪卫, 段起阳, 施伯乐. XML查询中DTD的排序技术. 计算机研究与发展, 2003, 40(11): 1579-1585.