

P.O.Box 8718, Beijing 100080, China	Journal of Software Sept 2003,14(9):1615-1620
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
<a href="http://www.jos.org.cn">http://www.jos.org.cn</a>	Copyright © 2003 by The Editorial Department of Journal of Software

# XML数据的路径表达式查询优化技术

吕建华, 王国仁, 于 戈

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

吕建华, 王国仁, 于 戈 (东北大学 信息科学与工程学院, 辽宁 沈阳 110004)

第一作者: 吕建华(1977—), 男, 河北承德人, 博士生, 主要研究领域为XML数据库, 索引技术, 查询处理技术.

联系人: 吕建华 Telephone: 86-24-83681250; E-mail: dbgroup@mail.neu.edu.cn

Received 2002-08-12; Accepted 2002-10-14

## Abstract

Path expression is one of the core components of most XML query languages, and many evaluation methods for path expression queries are proposed recently. However, there are few researches on the issue of path expression optimization. In this paper, two kinds of path expression optimizing principles are proposed, named path shorten and path complementing, respectively. The path shorten principle reduces the querying cost by shortening the path expressions with the knowledge of XML schema. While the path complementing principle tends to substitute the user queries with the equivalent lower-cost path expressions. The experimental results show that these two techniques can work on most path expression queries and largely improve the efficiency of path expression query processing.

Lü JH, Wang GR, Yu G. Optimizing path expression queries of XML data. *Journal of Software*, 2003,14(9): 1615~1620.

<http://www.jos.org.cn/1000-9825/14/1615.htm>

## 摘要

路径表达式作为XML数据查询语言的核心部分,关于它的计算方法的研究成果已有很多,然而针对路径表达式本身进行优化的研究却相对较少.提出了两种针对路径表达式的优化策略:路径缩短策略和补路径策略,从而提高了XML路径查询效率.路径缩短策略根据XML文档模式信息,将路径表达式查询长度缩短,从而简化查询本身以降低需要的查询代价;而补路径策略则试图使用代价更小的等价路径表达式来替换原始查询.经过对实验数据的分析,这两种优化策略对于绝大多数路径表达式查询可以应用,并可大幅度地改进路径表达式的查询性能.

基金项目: Supported by the National Natural Science Foundation of China under Grant Nos.60173051, 60273079 (国家自然科学基金); the Foundation of Teaching and Research Award Program for Outstanding Young Teachers in Higher Education Institution of China (教育部高等学校优秀青年教师教学和科研奖励基金); the Foundation for University Key Teacher by the Ministry of Education of China (教育部高等学校骨干教师资助计划)

## References:

- [1] Zhou A, Lu H, Zheng S, Liang Y, Zhang L, Ji W, Tian Z. VXMLR: A visual XML-relational database system. In: Franklin MJ, Moon B, Ailamaki A, eds. Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data. Madison: ACM, 2002. 719~720.
- [2] McHugh J, Widom J. Query optimization for XML. In: Atkinson MP, Orłowska ME, Valduriez P, Zdonik SB, Brodie ML, eds. Proceedings of the 25th International Conference on Very Large Data Bases. Edinburgh: Morgan Kaufmann Publishers, 1999. 315~326.
- [3] Li Q, 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. Proceedings of the 27th International Conference on Very Large Data Bases. Roma: Morgan Kaufmann Publishers, 2001. 361~370.

- [4] Chung C, Min J, Shim K. APEX: An adaptive path index for XML data. In: Franklin MJ, Moon B, Ailamaki A, eds. Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data. Madison: ACM, 2002. 121~132.
- [5] Lu H, Wang G, Yu G, Bao Y, Lü J, Yu Y. Xbase: Making your gigabyte disk queriable. In: Franklin MJ, Moon B, Ailamaki A, eds. Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, Madison: ACM, 2002. 630.
- [6] Lü J, Wang G, Yu JX, Yu G, Lu H, Sun B. Performance evaluation of a DOM-based XML database: Storage, indexing and query optimization. In: Meng XF, Su JW, Wang YJ, eds. Advances in Web-Age Information Management, the 3rd International Conference, WAIM 2002. Lecture Notes in Computer Science 2419, Springer-Verlag, 2002. 13~24.
- [7] Gardarin G, Gruser JR, Tang Z. Cost-Based selection of path expression processing algorithms in object-oriented databases. In: Vijayaraman TM, Buchmann AP, Mohan C, Sarda NL, eds. Proceedings of the 22th International Conference on Very Large Data Bases. Mumbai: Morgan Kaufmann Publishers, 1996. 390~401.
- [8] Schmidt A, Waas F, Kersten ML, Carey MJ, Manolescu I, Busse R. XMark: A benchmark for XML data management. In: Bernstein PA, Loannidis YE, Ramakrishnan R, Papadias D, eds. Proceedings of the 28th International Conference on Very Large Data Bases. Hong Kong: Morgan Kaufmann Publishers, 2002. 974~985.
- [9] Wang GR, Yu G, Zhang B. Selectivity estimation in object-oriented databases. Chinese Journal of Computers, 1998,21 (supplement):171~177 (in Chinese with English abstract).
- [10] Bohme T, Rahm E. Multi-User evaluation of XML data management systems with XMach-1. In: Proceedings of the 1st VLDB Workshop on Efficiency and Effectiveness of XML Tools, and Techniques. Hong Kong, 2002. 148~159.

附中文参考文献:

- [9] 王国仁,于戈,张斌.面向数据库系统中谓词选择的估算.计算机学报,1998,21(增刊):171~177.