

P.O.Box 8718, Beijing 100080, China	Journal of Software, June 2007,18(6):1429-1442
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 © 2007 by <i>Journal of Software</i>

## F-Index: 一种加速Twig查询处理的扁平结构索引

周军锋, 孟小峰, 蒋瑜, 谢敏

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

周军锋<sup>1,2</sup>, 孟小峰<sup>1</sup>, 蒋瑜<sup>1</sup>, 谢敏<sup>1</sup>

<sup>1</sup>(中国人民大学 信息学院, 北京 100872)

<sup>2</sup>(燕山大学 计算机科学与技术系, 河北 秦皇岛 066004)

作者简介: 周军锋(1977—), 男, 陕西西安人, 博士生, 讲师, 主要研究领域为XML数据库. 孟小峰(1964—), 男, 博士, 教授, 博士生导师, CCF高级会员, 主要研究领域为Web数据集成, XML数据库, 移动数据管理. 蒋瑜(1981—), 男, 硕士生, 主要研究领域为XML数据库. 谢敏(1984—), 男, 硕士生, 主要研究领域为XML数据库.

联系人: 周军锋 Phn: +86-10-62515575, E-mail: zhoujf@ysu.edu.cn, <http://www.ruc.edu.cn>

Received 2006-04-20; Accepted 2006-06-30

### Abstract

How to process twig query quickly and correctly has attracted much attention in research society recently. Filtering query irrelevant elements before query execution is an important step for reducing elements scanned at query processing. As a flattened structural index, F-Index is proposed to filter out all query irrelevant index nodes, thus query irrelevant elements can be filtered out rapidly and mostly, especially when it is processing deeply nested XML documents with a complex structure. After filtering, a new efficient query algorithm based on the remaining elements is proposed to accelerate query processing. Experimental results on various datasets indicate that twig query's performance can be improved significantly by using F-Index.

Zhou JF, Meng XF, Jiang Y, Xie M. F-Index: A flattened structural index for speeding up twig query processing. *Journal of Software*, 2007,18(6):1429-1442.

DOI: 10.1360/jos181429

<http://www.jos.org.cn/1000-9825/18/1429.htm>

### 摘要

如何快速、有效地处理twig形式的查询是XML查询处理的关键问题,通过过滤与查询无关的元素可以减少查询中需要处理的元素数目,从而提高查询的执行效率.提出一种扁平结构索引F-Index,能够快速过滤所有与查询无关的索引结点,进而过滤掉查询无关的元素,在处理深度嵌套的复杂结构XML文档时具有很大的优势.提出一种新的查询算法,能够有效处理过滤后剩余元素的匹配问题.基于不同数据集的实验表明,使用F-Index进行过滤可以极大地提高查询处理的性能.

基金项目: Supported by the National Natural Science Foundation of China under Grant Nos.60573091, 60273018 (国家自然科学基金); the National Basic Research Program of China under Grant No.2003CB317000 (国家重点基础研究发展计划(973)); 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 Program for New Century Excellent Talents in University (新世纪优秀人才支持计划)

### References:

[1] McHugh J, Widom J. Query optimization for XML. In: Malcolm PA, Maria EO, Patrick V, Stanley BZ, Michael LB, eds. Proc. of the 25th Int'l Conf. on Very Large Data Bases (VLDB). Edinburgh: Morgan Kaufmann Publishers, 1999. 315-326.

[2] Michael J, Franklin MJ. Efficient filtering of XML documents for selective dissemination of information. In: Amr EA, Michael LB, Sharma C, Umeshwar D, Nabil K, Gunter S, Kyu YW, eds. Proc. of the 26th Int'l Conf. on Very Large Data Bases (VLDB). Cairo: Morgan Kaufmann Publishers, 2000. 53-64.

- [3] Gottlob G, Koch C, Pichler R. Efficient algorithms for processing XPath queries. In: Stéphane B, Akmal BC, Mong LL, Jeffrey XY, Zoé L, eds. Proc. of the 28th Int'l Conf. on Very Large Data Bases (VLDB). Hong Kong: Morgan Kaufmann Publishers, 2002. 95-106.
- [4] Halvreson A, Burger J, Galanis L, Kini A, Krishnamurthy R, Rao AN, Tian F, Viglas SD, Wang Y, Naughton JF, DeWitt DJ. Mixed mode XML query processing. In: Johann CF, Peter CL, Serge A, Michael JC, Patricia GS, Andreas H, eds. Proc. of the 29th Int'l Conf. on Very Large Data Bases (VLDB). Berlin: Morgan Kaufmann Publishers, 2003. 225-236.
- [5] Lu SC, Meng XF, Lin C, Wang Y. Navigation implementation for XQuery in OrientX. Journal of Computer Research and Development, 2004,41(10):1815-1822 (in Chinese with English abstract).
- [6] Zhang C, Naughton J, DeWitt D, Luo Q, Lohman G. On supporting containment queries in relational database management systems. In: Aref WG, ed. Proc. of the ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD). ACM, 2001. 425-436.
- [7] Li Q, Moon B. Indexing and querying XML data for regular path expressions. In: Peter MGA, Paolo A, Stefano C, Stefano P, Kotagiri R, Richard TS, eds. Proc. of the 27th Int'l Conf. on Very Large Data Bases (VLDB). Rome: Morgan Kaufmann Publishers, 2001. 361-370.
- [8] Al-Khalifa S, Jagadish HV, Koudas N, Patel JM, Srivastava D, Wu Y. 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 (ICDE). San Jose: IEEE Computer Society, 2002. 141-152.
- [9] Chien SY, Vagena Z, Zhang D, Tsotras VJ, Zaniolo C. Efficient structural joins on indexed XML documents. In: Stéphane B, Akmal BC, Mong LL, Jeffrey XY, Zoé L, eds. Proc. of the 28th Int'l Conf. on Very Large Data Bases (VLDB). Hong Kong: Morgan Kaufmann Publishers, 2002. 263-274.
- [10] Jiang H, Lu H, Wang W, Ooi BC. XR-Tree: Indexing XML data for efficient structural joins. In: Umeshwar D, Krithi R, Vijayaraman TM, eds. Proc. of the 19th Int'l Conf. on Data Engineering (ICDE). Bangalore: IEEE Computer Society, 2003. 253-264.
- [11] Lam F, Shui WM, Fisher DK, Wong RK. Skipping strategies for efficient structural joins. In: Yoonjoon L, Jianzhong L, Kyuyoung W, Doheon L, eds. Proc. of the Database Systems for Advances Applications (DASFAA). Jeju Island: Springer-Verlag, 2004. 196-207.
- [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] 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 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/827.htm>
- [14] Bruno N, Koudas N, Srivastava D. Holistic twig joins: Optimal XML pattern matching. In: Michael JF, Bongki M, Anastassia A, eds. Proc. of the 2002 ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD). Madison: ACM, 2002. 310-321.
- [15] Jiang HF, Wang W, Lu H, Yu JX. Holistic twig joins on indexed XML documents. In: Johann CF, Peter CL, Serge A, Michael JC, Patricia GS, Andreas H, eds. Proc. of the 29th Int'l Conf. on Very Large Data Bases (VLDB). Berlin: Morgan Kaufmann Publishers, 2003. 273-284.
- [16] Moro MM, Vagena Z, Tsotras VJ. Tree-Pattern queries on a lightweight XML processor. In: Klemens B, Christian SJ, Laura MH, Martin LK, Per-Ake L, Beng CO, eds. Proc. of the 31st Int'l Conf. on Very Large Data Bases (VLDB). Trondheim: ACM, 2005. 205-216.
- [17] Goldman R, Widom J. DataGuides: Enabling query formulation and optimization in semistructured databases. In: Matthias J, Michael JC, Klaus RD, Frederick HL, Pericles L, Manfred AJ, eds. Proc. of 23rd Int'l Conf. on Very Large Data Bases (VLDB). Athens: Morgan Kaufmann Publishers, 1997. 436-445.
- [18] Milo T, Suciu D. Index structures for path expressions. In: Catriel B, Peter B, eds. Proc. of the Int'l Conf. on Database Theory (ICDT). Jerusalem: Springer-Verlag, 1999. 277-295.
- [19] Kaushik R, Bohannon P, Naughton JF, Korth HF. Covering indexes for branching path queries. In: Michael JF, Bongki M, Anastassia A, eds. Proc. of the 2002 ACM SIGMOD Int'l Conf. on Management of Data. Madison: ACM (SIGMOD), 2002. 133-144.
- [20] Barta A, Consenc MP, Mendelzon AO. Benefits of path summaries in an XML query optimizer supporting multiple access methods. In:

Klemens B, Christian SJ, Laura MH, Martin LK, Per-Ake L, Beng CO, eds. Proc. of the 31st Int'l Conf. on Very Large Data Bases (VLDB). Trondheim: ACM, 2005. 133-144.

[21] Chen T, Lu JH, Ling TW. On boosting holism in XML twig pattern matching using structural indexing techniques. In: Fatma -, ed. Proc. of the ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD). Baltimore: ACM, 2005. 455-466.

[22] Wang W, Wang HZ, Lu HJ, Jiang HF, Lin XM, Li JZ. Efficient processing of XML path queries using the disk-based F&B index. In: Klemens B, Christian SJ, Laura MH, Martin LK, Per-Ake L, Beng CO, eds. Proc. of the 31st Int'l Conf. on Very Large Data Bases (VLDB). Trondheim: ACM, 2005. 145-156.

[23] Hou S, Jacobsen HA. Predicate-Based filtering of XPath expressions. In: Ling L, Andreas R, Kyuyoung W, Jianjun Z, eds. Proc. of the 22nd Int'l Conf. on Data Engineering (ICDE). Atlanta: IEEE Computer Society, 2006. 53-64.

[24] Miklau G. XMLData repository. 2002. <http://www.cs.washington.edu/research/xmldatasets>

附中文参考文献:

[5] 陆世潮,孟小峰,林灿,王宇.OrientX中的XQuery的导航式处理.计算机研究与发展,2004,41(10):1815-1822.

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

[13] 王静,孟小峰,王宇,王珊.以目标节点为导向的XML路径查询处理.软件学报,2005,16(5):827-837. <http://www.jos.org.cn/1000-9825/16/827.htm>