

P.O.Box 8718, Beijing 100080, China	Journal of Software July 2003,14(7):1195-1202
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2003 by The Editorial Department of Journal of Software

## 基于XML的软件构件查询匹配算法研究

徐如志, 钱乐秋, 程建平, 王渊峰, 朱三元

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

徐如志<sup>1</sup>, 钱乐秋<sup>1</sup>, 程建平<sup>2</sup>, 王渊峰<sup>1</sup>, 朱三元<sup>3</sup> 1(复旦大学 计算机与信息技术系,上海 200433)2(中创软件工程股份有限公司 软件研究院,山东 济南 250014)3(上海计算机软件技术开发中心,上海 200233)

第一作者: 徐如志(1966—),男,山东泰安人,博士生,高级工程师,主要研究领域为软件复用,构件库系统,软件项目管理.

联系人: 徐如志 Telephone: 86-21-65643785, Fax: 86-21-65642826, E-mail: rzxu@fudan.edu.cn

Received 2002-11-08; Accepted 2003-01-20

### Abstract

Based on the research of unordered tree-inclusion matching, a matching algorithm for XML-based component query is proposed. This algorithm can greatly improve the recall and provide support for Boolean query while maintaining a high level precision. Moreover, by adding some constraints on the basis of features of software component and using dynamic programming, the computation of matching cost is resolved in polynomial time, so that a high efficiency for the component query is guaranteed. Furthermore, the feasibility and efficiency of the new matching algorithm in practical application to software component query are confirmed by the results of a series of experiments on a prototype system RCRS.

Xu RZ, Qian LQ, Cheng JP, Wang YF, Zhu SY. Research on matching algorithm for XML-based software component query. *Journal of Software*, 2003,14(7):1195~1202.

<http://www.jos.org.cn/1000-9825/14/1195.htm>

### 摘要

在研究无序树包含匹配的基础上,提出一种新的基于XML的软件构件查询匹配算法.该算法可以在保持较高构件查准率的前提下,显著地提高构件的查全率,并提供对布尔查询的支持.此外,通过合理地设定约束条件以及利用动态规划的方法,将计算查询匹配代价的算法时间复杂度限定为多项式级,确保构件查询具有足够的查询效率.最后,通过在构件库原型系统RCRS上进行的一系列实验,进一步证明了新的查询匹配算法在软件构件查询实际应用中的可行性和有效性.

基金项目: Supported by the National High-Tech Research and Development Plan of China under Grant No.2001AA1100241 (国家高技术研究发展计划(863))

### References:

[1] Ivar J. Software reuse: Architecture, process and organization for business success. Reading: Addison-Wesley Publishing Company, 1997. 4~15.

[2] Mili H, Mili A. Reuse based software engineering. New York: John Wiley & Sons Inc., 2002. 444~459.

[3] Frakes WB, Pole TP. An empirical study of representation methods for reusable software components. *IEEE Transactions on Software Engineering*, 1994,120(8):617~630.

[4] Gibb F, McCartan C, O'Donnell R, Sweeney N, Leon R. The integration of information retrieval techniques within a software reuse environment. *Journal of Information Science*, 2000,26(4):520~539.

- [5] Torshen S. ApproXQL: Design and implementation of an approximate pattern matching language for XML. Technical Report, B 01-02, Freie University at Berlin, 2001.
- [6] Thorsten R. A new measure of the distance between ordered trees and its applications. Research Report, 85166, Department of Computer Science, University of Bonn, 1997.
- [7] Torshen S, Naumann F. Approximate tree embedding for querying XML data. In: Proceedings of ACM SIGIR Workshop on XML and Information Retrieval. Athens, 2000.
- [8] Zhang KZ. On the editing distance between unordered labeled trees. Information Processing Letters, 1992,42(3):133~139.
- [9] Wang YF. Research on retrieving reusable components classified in faceted scheme [Ph.D. Thesis]. Shanghai: Fudan University, 2002 (in Chinese with English abstract).
- [10] Chang JC, Li KQ, Guo LF, Mei H, Yang FQ. Representing and retrieving reusable software components in JB (Jadebird) system. Electronic Journal, 2000,28(8):20~24 (in Chinese with English abstract).
- [11] Kilpelainen P. Tree matching problems with applications to structured text databases. Technical Report, A-1992-6, Department of Computer Science, University of Helsinki, 1992.
- [12] Kilpelainen P. Ordered and unordered tree inclusion. SIAM Journal on Computing, 1995,24(2):340~356.

附中文参考文献:

- [9] 王渊峰.基于剖面描述的构件检索算法研究[博士学位论文].上海:复旦大学,2002.
- [10] 常继传,李克勤,郭立峰,梅宏,杨芙清.青鸟系统中可复用软件构件的表示与查询.电子学报,2000,28(8):20~24.