The Study of Extensible Index System on XML Data

陳嘉進、邱紹豐

E-mail: 9511436@mail.dyu.edu.tw

ABSTRACT

XML has emerged as the standard for data exchange, and processing. Many query systems have been developed to retrieve required information from the XML data source. However, the difficulty in processing range query is not solved, since the indexing system in such environment is not fully matured yet. In this thesis, we developed an indexing system that incorporates R-tree and data guide, called range data guide, or RDG. In addition to indexing data paths, the system also allows indexing on ranges of data. The idea of the system is to embed R-tree at the leaf nodes in the data paths in Data Guide. If the users request is not a range query, the regular Data Guide provides indices on data paths. Otherwise, the R-tree at the leaf nodes allow users to access a certain range of data. We have developed the structure of the index and the related algorithms, i.e., insertion, deletion, and update. The experimental results show that our proposed system offers performance improvement over other traditional techniques.

Keywords : Semi-Structured Data ; index XML ; range index ; access data

Table of Contents

封面內頁 簽名頁 授權書.......................................iii 中文摘要......................................iv 英文摘要.......................................v 誌謝..........................................vi 目錄.........................................vii 圖目錄........................................ix 表目錄........................................xi 1. 前言...................................1 1.1 研究動機...............................1 1.2 研究目的...............................2 1.3 本論文內容與架構........................3 2. 相關研究...............................4 2.1 XML與Object Exchange Model............4 2.2 XML查詢系統............................9 2.3 OEM索引機制...........................11 2.4 範圍性索引架構:R-tree...............13 3. 範圍性資料嚮導.........................16 3.1 範圍性資料嚮導設計目的..................16 3.2 範圍性資料嚮導基本原理..................17 3.3 範圍性資料嚮導概述......................18 3.4 方法..................................20 3.4.1 Search演算法..........................20 3.4.2 Insertion演算法........................21 3.4.3 Deletion演算法........................23 3.5 範例說明..............................24 3.5.1 RDG新增索引路徑範例....................24 3.5.2 RDG查詢資料範例........................26 3.5.3 RDG刪除資料範例........................27 4. 效能評估與實驗.........................29 4.1 效能評估..............................29 4.2 實驗..................................33 5. 結論..................................42 參考文獻....................................44

REFERENCES


