整合式數位內容搜尋系統之設計與實作

陳韋任、邱紹豐

E-mail: 364924@mail.dyu.edu.tw

摘 要

數位內容檔案的應用程式都能提供功能不一的關鍵字搜尋功能，但是當檔案數量及類型增加，搜尋時就可能需要重複開啟多次應用程式。常用的應用程式幾乎都只提供簡單的關鍵字搜尋，而且無法在搜尋字串內加入邏輯運算子。在本研究中我們實做了一個搜尋平台，整合數種數位內容的搜尋功能的模組，使用者透過單一介面，可以輸入包含邏輯運算子及括號的搜尋字串，完成多個檔案且多種檔案類型的搜尋。要支援搜尋字串內加入括號以改變邏輯運算式的權重，會遇到搜尋字串內的關鍵字處理順序的問題，在此研究中，我們採取的方法是將搜尋的字串轉成二元邏輯樹，透過中序走訪(Inorder traversal)演算法和後序走訪(Postorder traversal)演算法走訪二元邏輯樹，處理邏輯運算子及括號。最後，搜尋時如檔案數量及類型不是單一筆，會有數種不同格式的輸出結果，我們設計一個封裝器(Wrapper)，透過標準的Schema將輸出結果統一轉成XML，而應用程式讀取XML，將結果輸出至使用者介面。

關鍵詞 : 整合式搜尋平台、關鍵字搜尋、搜尋模組

目錄

參考文獻

[1] Agrawal. S, Chaudhuri. S, Das. G, DBXplorer: “ A system for keyword-based search over relational database ”, Proceedings of the 18th International Conference on Data Engineering, San Jose, pp. 5-16, March, 2002.

[2] Bhalotia. G, Hulgeri. A, Nakhe. C, Chakrabarti. S, Sudarshan. S, “ Keyword searching and browsing in databases using BANKS ”, Proceedings of the 18 th International Conference on Data Engineering, San Jose, pp. 431-440, March, 2002.

[3] D. Florescu, I. Manolescu, “ Integrating Keyword Search into XML Query Processing ”, 9th WWW Conf., 2000.

[4] Blakeley, J.A. “ Universal data access with OLE DB ”,Compcon '97. Proceedings, IEEE, pp.2-7, Feb, 1997.

[5] Hassan, M.,Alhajj, R., Ridley, M.J., Barker, K, “ Database selection and keyword search of structured databases: powerful search for naive users ”,Information Reuse and Integration, 2003. IRI 2003. IEEE International Conference ,pp.175-182,Oct. 2003 [6] Hristidis. V, Papakonstantinou. Y, DISCOVER “ Keyword search in relational databases ”, Proceedings of 28th International Conference on Very Large Data Bases, Hong Kong, pp 670-681, August, 2002.

[7] Bruno Lowagie, “ iText in Action, ” Manning Publications; Second Edition edition,November 22, 2010.

[8] Sahil Malik, “ Pro Ado.net 2.0, ” Apress,September 20, 2005 [9] Bruno Lowagie, itext, http://itextpdf.com/ [10]IKVM, http://www.ikvm.net/ [11]Microsoft, OLE Compound Document, http://msdn.microsoft.com/en-us/library/windows/desktop/ms693383(v=vs.85).aspx [12]Microsoft, ODBC 的基本概念(ODBC), http://msdn.microsoft.com/zh-tw/library/thzzea08(v=vs.90).aspx [13]Microsoft, Microsoft OLE DB (OLE DB), http://msdn.microsoft.com/en-us/library/windows/desktop/ms722784(v=vs.85).aspx [14]Microsoft, ADO.NET 概觀(ADO), http://msdn.microsoft.com/zh-tw/library/h43ks021(v=vs.80).aspx [15]Microsoft, Introducing the Office (2007) Open XML File Formats, http//msdn.microsoft.com/en- us/library/aa338205(v=office.12).aspx [16]Microsoft Office Word 2007, http://office.microsoft.com/zh-tw/word-help/RZ010066490.aspx? section=29 [17]Apache, PDFBox, http://pdfbox.apache.org/ [18]Apache, POI, http://poi.apache.org/ [19]W3C, A Query Language for XML,(XML-QL), http://www.w3.org/TR/NOTE-xml-ql/ [20]W3C, Extensible Markup Language (XML), http://www.w3.org/XML [21]W3C, An XML Query Language (XQuery), http://www.w3.org/TR/xquery [22]W3C, XML Path Language (XPath), http://www.w3.org/TR/xpath