

# 跨越站式XML購物引擎之建構-以手機選購為例

顧值豪、陳鴻文

E-mail: 9020192@mail.dyu.edu.tw

## 摘要

伴隨著網際網路的快速起飛，HTML語言也被廣泛使用於網頁的設計。顯然地，HTML語言非常適合於單純網頁的撰寫，也因此造就了全球資訊網及電子商務的蓬勃發展。為了便利線上交易者的選購，許多電子商務商店網站亦提供了購物引擎檢索功能；然而現今購物引擎往往侷限於以人工方式，事先整理單一電子商店網站內部資料，及以純文字資訊檢索為主的方式，卻也阻礙了網路上資訊檢索的精確性。相對地，新一代的網頁語言-可延伸標記語言XML (EXTENSIBLE MARKUP LANGUAGE)，因為可以提供一個標示結構化的資料架構，同時使用者也可自訂標籤，故能適切的呈現網頁的內容涵意和架構性，所以預期將能提供較準確且更有效率的商品搜尋。本研究希望藉由XML上述結構化的特性，開發出跨網站式智慧型圖文檢索購物引擎。此系統由主題知識庫、主題詞彙庫、特徵給分演算法、初始化隸屬函數設定、檢索委託模組、網頁檢索模組及資訊攔阻模組等八個模組，再配合索引資料庫、標籤料庫和網頁料庫等子系統所組成。建置之後，網頁資料內容使用XML標籤加以辨識，再配合有效的資訊檢索流程設計，就可以輕易的依照不同的查詢條件（包括口語化表達）來加以搜尋及個人化資訊呈現。藉由本研究的成功經驗，我們也預期將此套購物引擎機制延伸至其他適合線上購物的商品上，諸如通訊商品、服飾、汽車、珠寶等。

關鍵詞：XML、電子商務、搜尋引擎、購物引擎、圖文檢索。

## 目錄

第一章 緒論--P1 第一節 研究背景--P1 第二節 研究動機--P4 第三節 研究目的--P9 第四節 研究範圍及限制--P10 第五節 研究流程--P11 第六節 論文架構--P13 第二章 文獻探討--P14 第一節 搜尋引擎的發展與比較--P14 第二節 XML網頁標記語言及特性比較分析--P20 第三節 XML基本名詞認知--P25 第四節 相關文獻探討--P28 第三章 研究方法與系統設計--P29 第一節 研究方法--P29 第二節 XML網頁的型態與特徵探討--P32 第三節 系統功能架構--P38 第四章 系統實作與效能評估--P56 第一節 開發工具與環境--P56 第二節 系統實作--P57 第三節 檢索實驗與方法--P62 第四節 檢索效能比較與分析--P63 第五章 結論及後續研究建議--P77 第一節 研究結論--P77 第二節 未來研究方向--P78 參考文獻--P80

## 參考文獻

- [1]王紹宇(民88)，類神經模糊系統之結構精簡與應用，大葉大學資訊管理研究所碩士論文。
- [2]高嘉祺(民89)，線上圖文購物引擎-以手機應用為例，大葉大學資訊管理研究所碩士論文。
- [3]陳長念、陳勤意(民89)，XHTML網頁語言新標準，知域數位科技股份有限公司。
- [4]張宇崑(民87)，搜尋引擎大搜密，頁158-162，第三波資訊。
- [5]華陽子(民87)，網路上的終極咒語-搜尋語法詳解，頁167-171，第三波資訊。
- [6]李勤、沈兆陽(民90)，SQL SERVER2000與XML整合應用，文魁資訊股份有限公司。
- [7]比爾·蓋茲(BILL GATES)、納桑·米佛德博士(NATHAN MYHRVOLD, PH.D.)和彼得·雷諾生(PETER RINEARSON)(民86)，擁抱未來，遠流出版事業股份有限公司。
- [8]BRIAN E. TRAVIS(民90)，搞懂XML看清SOAP攻佔BIZTALK(紀學勤譯)，微軟圖書，華彩代理發行。
- [9]陳錦輝、江鈞(民89)，從HTML到XML，頁1-16，文魁資訊有限公司。
- [10]GUERRIRRI E. (1999) SOFTWARE DOCUMENT REUSE WITH XML. IEEE TRANS. ON SYSTEM MAN AND CYBERNETICS, 154-161.
- [11]OUAHID H., A. KARMOUCH (1999) CONVERTING WEB PAGES INTO WELL-FORMED XML DOCUMENTS. IEEE TRANS. ON SYSTEM., MAN AND CYBERNETICS, 676-680.
- [12]DOORENBOS, R.B., O. ETZIONI, AND D. S. WELD (1996) A SCALABLE COMPARISON-SHOPPING AGENT FOR THE WORLD WIDE WEB. UNIVERSITY OF WASHINGTON, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, TECHNICAL REPORT.
- [13]BEN S. I., ET. AL. (1999) ADDING SUPPORT FOR DYNAMIC AND FOCUSED SEARCH WITH FETUCCINO. COMPUTER NETWORK VOL.31, 1653-6665.
- [14]CHIDI I. A. AND F. FOTOUHI (1999) AN ADAPTIVE REAL-TIME WEB SEARCH ENGINE. ACM SIGMOD RECORD, 12-16.

- [15]FLORESCU D., D. KOSSMANN, AND I. MANOLESCU (2000) INTEGRATING KEYWORD SEARCH INTO XML QUERY PROCESSING COMPUTER NETWORK VOL.33 119-135.
- [16]GOLDFARB C.E. (1990) THE SGML HANDBOOK. Y. RUBINSKY, ED. OXFORD UNIV. PRESS. (THIS VOLUME CONTAINS THE FULL ANNOTATED TEXT OF ISO 8897 (WITH AMENDMENTS)).
- [17]BRAY, TIM, ET. AL. (1998)EXTENSIBLE MARKUP LANGUAGE(XML)1.0, W3C RECOMMENDATION. ([HTTP://WWW.W3.ORG/TR/1998/REC-XML-19980210](http://www.w3.org/TR/1998/REC-XML-19980210)) [18]ADLER, SHARON, ET. AL.(1997) A PROPOSAL FOR EXTENSIBLE STYLE LANGUAGE(XSL). INSO CORPORATION, PROVIDENCE, RI. ([HTTP://WWW.W3.ORG/TR/NOTE-XSL.HTML](http://www.w3.org/TR/NOTE-XSL.HTML))
- [19]PEMBERTON S. AND M. ALTHEIM (1999) XHTML1.0: THE EXTENSIBLE HYPERTEXT MARKUP LANGUAGE. TECHNICAL REPORT, W3C. ([HTTP://WWW.W3C.ORG/WD-HTML-IN-XML](http://www.w3c.org/wd-html-in-xml)) [20]BOS B. AND C.LILLEY (1999) CASCADING STYLE SHEETS LEVEL2(CSS2) SPECIFICATION. TECHNICAL REPORT, W3C. ([HTTP://WWW.W3C.ORG/TR/REC-CSS2](http://www.w3c.org/TR/REC-CSS2)) [21]MARTIN D. (2000) XML-BASED RENDERING LANGUAGES. O'REILLY & ASSOCIATES, INC. [HTTP://WWW.XML.COM/PUB/A/2000/04/05/STYLE/INDEX.HTML](http://www.xml.com/pub/a/2000/04/05/style/index.html) [22]A PROPOSAL FOR XSL , W3C PROPOSAL , [HTTP://WWW.W3C.ORG/TR/NOTE-XSL.HTM](http://www.w3c.org/TR/NOTE-XSL.HTM).
- [23]網路燈塔 , [HTTP://WWW.HAIYAN.COM/STEELK/NAVIGATOR/B5INDEX.HTM](http://www.haiyan.com/steelk/navigator/b5index.htm).
- [24]HARVARD CONFERENCE ON THE INTERNET AND SOCIETY , [HTTP://CYBER.LAW.HARVARD.EDU/CYBERCON98/ASP/SPLASH.ASP](http://cyber.law.harvard.edu/cybercon98/asp/splash.asp).
- [25]MICROSOFT XML DEVELOPER CENTER , [HTTP://MSDN.MICROSOFT.COM/XML](http://msdn.microsoft.com/xml).
- [26]FIND資訊工業策進會-網際網路趨勢研究中心 , [HTTP://WWW.FIND.ORG.TW/USAGE.ASP](http://www.find.org.tw/usage.asp).
- [27]YAHOO ! , [HTTP://WWW.YAHOO.COM](http://www.yahoo.com).
- [28]KIMO , [HTTP://WWW.KIMO.COM.TW](http://www.kimo.com.tw).
- [29]ALTA VISTA , [HTTP://WWW.ALTAVISTA.DIGITAL.COM](http://www.altavista.digital.com).
- [30]PCHOME , [HTTP://IMAGE.PCHOME.COM.TW](http://image.pchome.com.tw).
- [31]GAIS , [HTTP://GAIS.CS.CCU.EDU.TW](http://gais.cs.ccu.edu.tw).
- [32]仲琦八爪魚 , [HTTP://GAIS.HT.NET.TW](http://gais.ht.net.tw).
- [33]EXCITE , [HTTP://WWW.EXCITE.COM](http://www.excite.com).
- [34]LYCOS , [HTTP://WWW.LYCOS.COM](http://www.lycos.com).
- [35]HOTBOT , [HTTP://WWW.HOTBOT.COM](http://www.hotbot.com).
- [36]INFOSEEK , [HTTP://WWW.GO.COM](http://www.go.com).
- [37]中央ARCHIE , [HTTP://ARCHIE.NCU.EDU.TW](http://archie.ncu.edu.tw).
- [38]DOWNLOAD , [HTTP://WWW.DOWNLOAD.NET](http://www.download.net).
- [39]SHAREWARE , [HTTP://SHAREWARE.CNET.COM](http://shareware.cnet.com).
- [40]PC WORLD , [HTTP://WWW.PCWORLD.COM](http://www.pcworld.com).
- [41]ZDNET , [HTTP://WWW.ZDNET.COM](http://www.zdnet.com).
- [42]台灣電子地圖服務網 , [HTTP://WWW.MAP.COM.TW](http://www.map.com.tw).
- [43]EARTHAMAPS , [HTTP://WWW.EARTHAMAPS.COM](http://www.earthamaps.com).
- [44]VICINITY , [HTTP://MAPS.YAHOO.COM](http://maps.yahoo.com).
- [45]MAPBLAST , [HTTP://WWW.MAPBLAST.COM](http://www.mapblast.com).
- [46]WEB CRAWLER , [WWW.WEBCRAWLER.COM](http://www.webcrawler.com).
- [47]ETFORECASTS , [HTTP://WWW.ETFORECASTS.COM](http://www.etforecasts.com).
- [48]NII中華民國國家資訊基本建設產業發展協進會 [HTTP://WWW.NII.ORG.TW](http://www.nii.org.tw).
- [49]蕃薯藤網路調查 , [HTTP://SURVEY.YAM.COM/SURVEY2000](http://survey.yam.com/survey2000).
- [50]上海圖書館資源 , 搜尋引擎的現況與應用 [HTTP://WWW.ISTIS.SH.CN/ISTIS/DLIB/TSQB.HTM](http://www.istis.sh.cn/istis/dlib/tsqb.htm).
- [51]HISTORY OF ARPANET , [HTTP://WWW.DEI.ISEP.IPP.PT/DOCS/ARPA.HTML](http://www.dei.isep.ipp.pt/docs/arpa.html).
- [52]昇陽(SUN)電腦公司 , [HTTP://WWW.SUN.COM.TW/](http://www.sun.com.tw/).
- [53]CN ( COMMERCE NET TAIWAN ) 研究報告 , [HTTP://WWW.NII.ORG.TW/CNT/INFO/REPORT/20000601.HTM](http://www.nii.org.tw/cnt/info/report/20000601.htm).
- [54]蕃薯藤 , [HTTP://SURVEY.YAM.COM](http://survey.yam.com).
- [55]W3C PROPOSAL , [HTTP://WWW.W3C.ORG](http://www.w3c.org).
- [56]GOING FROM HTML TO XML , [HTTP://MSDN.MICROSOFT.COM/XML/ARTICLES/HTMLTOXML.ASP](http://msdn.microsoft.com/xml/articles/htmltoxml.asp).
- [57]CLEAN UP YOUR WEB PAGES WITH HTML TIDY , [HTTP://WWW.W3.ORG/PEOPLE/RAGGETT/TIDY/#DOWNLOAD](http://www.w3.org/people/raggett/tidy/#download).
- [58]DOCUMENT OBJECT MODEL (DOM) , [HTTP://WWW.W3.ORG/DOM/](http://www.w3.org/dom/).
- [59]MEGGINSON TECHNOLOGIES , [HTTP://WWW.MEGGINSON.COM/SAX/](http://www.megginson.com/sax/).