

# THE CONSTRUCTION OF AN XML SHOPPING ENGINE ACROSS WEBSITES - TO PURCHASE HANDSET PHONES

顧值豪、陳鴻文

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

## ABSTRACT

HTML LANGUAGE HAS BEEN WIDELY USED TO DESIGN WEB PAGES AS INTERNET GROWS FAST. IT IS OBVIOUS THAT THE LANGUAGE IS ESPECIALLY SUITABLE FOR DESCRIBING SIMPLE STATIC HOMEPAGES AND THUS ALLEVIATES THE ENORMOUS DEVELOPMENT OF BOTH WWW (WORLD WIDE WEB) AND EC (ELECTRONIC COMMERCE). IN ORDER TO MAKE CONSUMERS MORE CONVENIENT FOR ON-LINE PURCHASING, MANY WEB STORES OF EC PROVIDE AUXILIARY TOOLS SUCH AS SHOPPING ENGINES. HOWEVER, MOST DATABASES USED IN THOSE SHOPPING ENGINES ARE RESTRICTED IN AFFORDING ONLY THE INFORMATION OF MERCHANDISES WITHIN A SINGLE WEB STORE AND MUST BE MANAGED BY HUMAN IN PRIOR. MOREOVER, NO ABUNDANT INFORMATION PRODUCED BY STRING MATCHING OF CONTENTS OF WEB PAGES USUALLY BLOCKS THE PRECISION AND ACCEPTANCE OF INFORMATION RETRIEVALS IN INTERNET. A NEW WEB DESIGNING LANGUAGE XML, EXTENSIBLE MARKUP LANGUAGE, ALLOWS USERS TO DEFINE NEW TAGS WITH SPECIAL MEANINGS AND ALSO PROVIDES A WELL ORGANIZED DATA STRUCTURES OF HOMEPAGES. BY THOSE EXCELLENT PROPERTIES OF XML, AN INTELLIGENT SHOPPING ENGINE OF MORE EFFICIENCY IN RETRIEVING WEB PAGES ACROSS WEBSITES IS THUS PROPOSED TO AUTOMATICALLY ANALYZE HOMEPAGES OF IMAGES AND TEXT. THE PROPOSED SHOPPING ENGINE COMPRISES ONLINE AND OFF-LINE SUBSYSTEMS, AND EIGHT COMPONENTS: MEMBERSHIP FUNCTION INITIALIZATION, TERMINOLOGY DICTIONARY MODULE, DOMAIN KNOWLEDGE DATABASE, RETRIEVAL COMMIT MODULE, INFORMATION FILTERING MODULE, HOMEPAGE RETRIEVAL MODULE, IMAGE INDEXING MODULE AND DISPLAY MODULE. IN ADDITION, THREE DATABASES, INCLUDING INDEX DATABASE, TAG DATABASE AND HOMEPAGE DATABASE, ARE THUS DESIGNED AND CREATED. COMPARED WITH TRADITIONAL SEARCHING ENGINES FOR HTML HOMEPAGES, EXPERIMENTS ILLUSTRATE A PROMISING PERFORMANCE OF THE PROPOSED SHOPPING ENGINE. BY THE SUCCESS OF THIS RESEARCH, THE PROPOSED PROTOTYPE OF SHOPPING ENGINE IS EXPECTED EAGERLY TO EXTEND TO OTHER BUSINESS FIELDS OF ONLINE SHOPPING, SUCH AS COMMUNICATION GOODS, COSTUME AND ACCESSORIES, CARS AND JEWELS.

Keywords : XML、E-COMMERCE、SEARCH ENGINE、SH INFORMATION RETRIEVAL OPPING ENGINE、IMAGE-TEXT

## Table of Contents

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

## REFERENCES

- [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]李勁、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 CYBERNTICS, 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.w3.org/W3C/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.w3.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.w3.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 ( COMMERCENET 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/)。