

圖書館學與資訊科學論文叢刊 第七輯 胡述兆·盧荷生主編

網路資源主題檢索 機制之研究

康 芳 菁 著



漢美圖書有限公司 印行

台北·紐約·洛杉磯

G252.7
200717

港台書

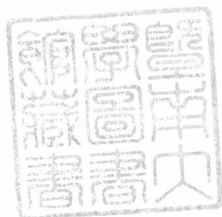
圖書館學與資訊科學論文叢刊 第七輯

胡述兆
盧荷生

主編

網路資源主題檢索 機制之研究

康 芳 菁 著



漢美圖書有限公司 印行

台北・紐約・洛杉磯

主編者弁言

「圖書館學與資訊科學論文叢刊」自 1990 年問世以來，已出版六集六十冊。由於這套叢刊涵蓋了台灣地區各大學圖書館與資訊學系博碩士論文的精華，資料新穎、內容充實，自出版以來，受到兩岸同道及海內外圖書館界的歡迎與好評。

第七輯徵選的十篇論文，六篇來自台大圖書館學研究所，四篇來自輔仁大學圖書資訊學研究所。在台大的六篇中，有三篇是博士論文，分別為陳昭珍博士的「古籍超文件全文資料庫模式之探討」，傅雅秀博士的「從科學傳播的觀點探討中央研究院生命科學專家的資訊尋求行為」，及王梅玲博士的「英美圖書館與資訊科學碩士教育之比較研究」。這三篇論文各具特色，陳文是為中國古籍之整理建立一套現代化的模式，傅文係從科學傳播的觀點探討生命科學家的資訊尋求行為，王文則係比較英美兩國的圖書館與資訊科學的碩士教育，皆係經過名家指導與專家考驗的力作，兼具學術與實用價值。其餘各篇探討的主題，包括：大學院校圖書館參考服務對網際網路的應用，網路資源主題檢索機制，機讀編目與檢索架構，圖書館導覽系統，企業識別系統 (CIS) 在圖書館的應用，資訊資源行銷，以及紙本書與電子書之比較，均係分別針對一個實際問題，作深入研究，故實用價值極高。

本輯之徵稿，係請王梅玲博士負責聯繫，特在此致謝。

胡述兆 盧荷生 謹識

1997 年 5 月於台北

主編者弁言

「圖書館學與資訊科學論文叢刊」自 1990 年問世以來，已出版六集六十冊。由於這套叢刊涵蓋了台灣地區各大學圖書館與資訊學系碩博士論文的精華，資料新穎、內容充實，自出版以來，受到兩岸同道及海內外圖書館界的歡迎與好評。

第七輯徵選的十篇論文，六篇來自台大圖書館學研究所，四篇來自輔仁大學圖書資訊學研究所。在台大的六篇中，有三篇是博士論文，分別為陳昭珍博士的「古籍超文件全文資料庫模式之探討」，傅雅秀博士的「從科學傳播的觀點探討中央研究院生命科學專家的資訊尋求行為」，及王梅玲博士的「英美圖書館與資訊科學碩士教育之比較研究」。這三篇論文各具特色，陳文是為中國古籍之整理建立一套現代化的模式，傅文係從科學傳播的觀點探討生命科學家的資訊尋求行為，王文則係比較英美兩國的圖書館與資訊科學的碩士教育，皆係經過名家指導與專家考驗的力作，兼具學術與實用價值。其餘各篇探討的主題，包括：大學院校圖書館參考服務對網際網路的應用，網路資源主題檢索機制，機讀編目與檢索架構，圖書館導覽系統，企業識別系統 (CIS) 在圖書館的應用，資訊資源行銷，以及紙本書與電子書之比較，均係分別針對一個實際問題，作深入研究，故實用價值極高。

本輯之徵稿，係請王梅玲博士負責聯繫，特在此致謝。

胡述兆 盧荷生 謹識

1997 年 5 月於台北

致謝詞

網織紀畧綴爲繡
路出籬界行者流
搜宇寰奇咫尺間
尋蠻過夷網路留

網路的世界是如此地浩瀚、多采多姿，一路行來的懣懣、摸索到了解，箇中滋味，是甘甜、也是苦樂。

本論文承蒙指導教授陳昭珍老師悉心的指點與教導，啓發學生探討這個課題，此論文方能順利完成，于此謹致最深的敬意與謝忱。感謝盧荷生院長、毛慶禎老師、林麗娟老師、高錦雪老師、蘇媛老師、吳政猷老師於論文計畫審核期間的指導；張淳淳老師及陳雪華老師於論文口試時的諄諄賜正與建議，使本論文更臻充實，在此一併致謝。

此外，感謝好友宜慧以及遠在彰化的芳亭在研究的路程上，藉著電話常常給予關心與打氣；研究所日子中相互扶持的麗芳學姐，在論文構思、內容上，提供諸多的建議、靈感及心情的分享；同學顯強、學弟履揆提供相關的資料協助研究，研究所的同學們，在論文寫作過程不時給予支持與鼓勵；以及政緯助教、學弟國誠在電腦設備使用上的協助，使得論文得以順利完成。

再者，也謝謝家中可愛的狗兒們-吉利、莉莉、嘟嘟、卡卡以及最貼心的狗女兒妞妞，在研究所的這些日子以來，陪伴我渡過無數的夜晚；在我煩悶、憂心時，適時地扮演了開心果的角色，舒解身心的壓力。

最後，感謝父母及家人們多年來在精神與經濟上的照顧、支持與包容，使我無憂無慮地順利完成學業，謹以此論文獻給我最親愛的父母。

一路行來，太多的感謝；網路瞬間的變化，雖是難以捉摸、掌握，但那些關懷與鼓勵銘記心中，很高興在自己的生命中有過這麼一段日子。

摘要

隨著科技的日益月異，圖書館將會從傳統以印刷資料為主的館藏形式改變為以網路資源為主體的數位化圖書館，圖書館所面臨的資料將會是豐富且多樣化的網路資源。扮演知識領航員的圖書館，在傳統圖書館作業中對於資訊的組織與整理，不論是資料外觀形式的敘述、或是主題內容的分析，都已發展出一套有系統的整理方式與完整的理論基礎。而今，在 Internet 上，所有的資訊皆以數位化的電子文件方式呈現，對於這些資料量眾多且不再具備實體外觀的電子文件而言，主題檢索的功用愈顯其重要性。因此有效地利用網路資源的主題檢索機制，將網路上豐富的資源，予以明確、快速的分析及引導讀者找到其所需要的資源，便利地供讀者使用，是一個值得探討的課題。

本研究將以文獻分析法來來分析了解圖書館界所發展出來的主題檢索理論，目前所遭遇的困難、問題及其未來的發展趨勢；並就網路之特性，擬出一套分析比較各主題檢索機制的評估要項表，並根據目前網際網路上主要的主題檢索機制進行實地操作，綜合各主題檢索機制之設計架構、優缺點，以及參考相關的文獻研究，以建構出網路資源主題檢索機制之設計原則，以期提供更完善的網路資源主題檢索服務之參考。

本研究調查的網路資源主題檢索機制包括：Lycos、Harvest、World Wide Web Worm、WebCrawler、Aliweb、Infoseek、OpenText、Alta Vista、Excite 以及中文的 GAIS 網路資訊搜尋系統。調查顯示，各主題檢索機制所涵蓋的資源主要包括全球資訊網、高佛系統以及網路論壇，掌握了 Internet 的重要資源。在檢索語言方面，幾乎所有的機制都採用自然語言檢索，也有部分機制建立類似於控制詞彙的分類標目協助使用者作檢索；在檢索功能方面，多數的機制都有提供布林、切截、鄰近字查詢等基本資訊檢索功能，不過在使用的指令或是符號上卻因機制而異，容易造成使用者的困擾，至於限制、加權查詢、同義字運算等方面，卻少有機制提供。在檢索結果內容中，大多包含了資源的總筆數、題名、關鍵字、位址、摘要、資料大小等基本內容，不過卻缺少了資源的類型來供檢索者作更進一步的相關性判斷；大部分的機制在使用者的查詢層次及介面多採用單一模式及文字圖形相混合的方式，此外，在輔助說明上，各機制利用超連結的特性來顯示各種輔助訊息，雖然具有彈性，但設計及內容呈現上不是很完整，以致於對檢索者來說並沒有真正發揮「輔助」的功能。

根據主題檢索的文獻分析及網路資源主題檢索機制的調查，提出幾點建議：(1)各主題檢索機制應加強其所掌握的資訊範圍的深入程度(2)圖書館應積極應用資訊組織與整理深厚的理論基礎或現有的控制詞彙工具，參予有關網路主題檢索機制或是主題目錄的設計，加強主題檢索機制檢索語言，以協助網路資源查詢更具效率；(3)加強各機制的檢索功能，以便提高檢索效率(4)檢索結果的內容應更具參考的價值(5)各主題檢索機制的介面設計應考量不同的需求，作多重模式的設計；其輔助說明的設計該應用相關的技術及考慮網路的環境，從檢索者的角度出發作更有效的處理與呈現。

A Research of Subject-oriented Search Engines on Internet

Abstract

With the coming of the networking age, the library's traditional collection "printed media" will be substituted by the digital information and libraries will have more more network resources. The library, being the knowledge navigator, has had a lot of experience and theory background on the organization of the information. However, "physical description" is emphasized more and the "subject analysis" less. For the digital information, which does not have physical appearance, the subject retrieval is becoming more and more important.

In this research, we use literature review to analyze the problems, methods and the future trends about the traditional library subject retrieval and network information. We then conclude with a checklist about the subject-oriented search engine. With this checklist, we survey those main subject-oriented search engines on Internet. We finally offer some principles for designing a good network subject-oriented search engine. We expect that these principles will provide for better network resources subject-oriented retrieval services.

Based on the research, we suggested that following five respects to improve the subject-oriented search engine : the network information scope 、 retrieval language 、 the system control function 、 the output results and the user interface.

目次

第一章 緒論

第一節 研究動機及問題陳述.....	1
第二節 研究方法與步驟.....	2
第三節 研究目的.....	2
第四節 研究範圍與限制.....	3
第五節 名詞解釋.....	4

第二章 文獻分析

第一節 主題檢索問題探討.....	7
第二節 主題檢索的改進.....	10
第三節 主題檢索機制.....	21

第三章 Internet 與網路資源主題檢索工具

第一節 Internet 發展沿革.....	30
第二節 Internet 基本功能與資源.....	32
第三節 網路資源主題檢索工具及相關計劃.....	38
第四節 主題檢索機制概況.....	46

第四章 主題檢索

第一節 檢索語言.....	65
第二節 檢索功能.....	67
第三節 使用者介面.....	71

第五章 主題檢索機制之評估

第一節 網路資源主題檢索機制之評估需求.....	76
第二節 主題檢索機制評估之要項.....	76
第三節 實證分析及結果.....	81
第四節 綜合討論.....	92

第六章 結論與建議

第一節 研究結果.....	94
第二節 主題檢索機制改進之建議.....	97
第三節 進一步研究之建議.....	100

參考書目.....	101
-----------	-----

附錄:主題檢索機制一覽表.....	109
-------------------	-----

索引.....	139
---------	-----

圖表目次

表 3-1	Internet 資源檢索工具種類及其用途.....	36
表 3-2	常用的主題資源目錄.....	40
表 3-3	常用的主題檢索機制.....	42
表 5-1	主題檢索機制之評估要項表.....	80
表 5-2	檢索範圍.....	82
表 5-3	檢索語言.....	84
表 5-4	檢索控制功能.....	87
表 5-5	檢索結果之呈現.....	89
表 5-6	使用者介面.....	90
表 5-7	輔助說明.....	91
圖 3-1	Lycos.....	47
圖 3-2	Harvest.....	48
圖 3-3	World Wide Web Worm.....	50
圖 3-4	WebCrawler.....	51
圖 3-5	Aliweb.....	52
圖 3-6	Infoseek.....	54
圖 3-7	OpenText.....	55
圖 3-8	Alta Vista.....	56
圖 3-9	Excite.....	57
圖 3-10	GAIS.....	59

第一章 緒論

第一節 研究動機及問題陳述

隨著電腦網路的建立與發展，網路資源已是處處可見。然而，網路資源的豐富和多樣化，一方面為資訊的檢索帶來了便利性，另一方面，如何找到適用的網路資源，也是目前我們所必須去探究的課題。圖書館在網路時代，乃扮演一個知識領航員(knowledge navigator)的角色，傳統圖書館對於印刷資料與非書媒體的處理，不論是資料外觀形式的敘述、或是主題內容的分析，都已發展出一套有系統的整理方式與完整的理論基礎(註 1)。而今，在 Internet 上，所有的資訊皆以數位化的電子文件方式呈現，對於這些資料量眾多且不再具備實體外觀的電子文件而言，主題檢索的功用愈顯其重要性。因此有效地利用網路資源的主題檢索機制，將網路上豐富的資源，予以明確、快速的分析及引導讀者找到其所需要的資源，便利地供讀者使用，是一個值得探討的課題。

本研究探討全球資訊網(World Wide Web，簡稱 WWW)上幾個重要的主題檢索工具，如 Lycos、Harvest、WWW、WebCrawler、Aliweb、Infoseek、OpenText、Alta Vista、Excite 及 GAIS 等主題檢索機制(subject-oriented search engine)，並分析圖書館學界已發展的主題檢索理論，對網路資源主題檢索機制之設計提供修正的意見，以有效地引導使用者尋檢到其合用的資源。

第二節 研究方法及步驟

本研究將根據文獻的探討，來分析了解圖書館界所發展出來的主題檢索理論、目前所遭遇的困難、問題、改進方案及其未來的發展趨勢，並就網路資源的特性，擬出一套分析比較各主題檢索機制的評估要項表（**checklist**），並根據目前全球資訊網資源的主題檢索機制進行實地操作，主要研究步驟如下：

- 一．利用各種參考工具書收集相關文獻。
- 二．研究文獻後界定研究目的及範圍。
- 三．以文獻分析法擬出一套分析網路資源各主題檢索機制之評估要項表(**checklist**)。
- 四．實地針對各主題檢索機制進行操作，將所得結果記錄下來。
- 五．研究不同主題檢索機制的特點、涵蓋範圍、優缺點等，並分析比較之。
- 六．參考相關文獻及實作結果，撰寫研究論文。

第三節 研究目的

- 一、了解目前 **Internet** 上網路資源類型及主題檢索工具之概況。
- 二、分析整理圖書館界已發展的主題檢索理論，並就網路資源的特性，提出適用於網路資源主題檢索機制的評估要項表。
- 三、分析比較 **WWW** 上各主題檢索機制之檢索範圍、檢索語言、檢索控制功能、檢索結果之呈現、使用者介面以及輔助說明之異同、優劣。

四、綜合各主題檢索機制優缺點，並參考主題檢索相關之文獻，試提出改善網路資源主題檢索機制之設計原則，以期提供更完善的網路資源主題檢索服務之參考。

第四節 研究範圍與限制

網路供應的資源十分龐雜，其範圍除電子形式的期刊、書籍、文章、摘要、索引之外，也有所謂的電腦軟體程式、遊戲等，種類相當繁多，因此，如何界定或歸類網路資源，目前尚無公認的標準或原則。本研究限於時間、人力、語言之因素，遂規範以下研究限制：

一、研究範圍

本研究所指之網路資源乃以 Internet 上資源為主，而主要的主題檢索機制是指利用 Internet 資源自動收集系統來收集網路資源，這類工具是以關鍵字或主題來檢索網路資源，但不具瀏覽的功能，大部以“英文文獻”作為檢索對象的機制為研究探討之範圍。另外，本研究擬就各主題檢索機制的邏輯架構、檢索範圍、檢索功能等作為研究主體，不包括各主題檢索機制的程式撰寫之技術層面。

二、時間限制

由於網路的情形是隨時在變動的，相同的，各主題檢索機制的面貌也是不斷地修正改變，因此本研究的評估依據乃是以民國 85 年 1 月至 5 月中旬期間當時各主題檢索機制的系統情況為主。

第五節 名詞解釋

一.網際網路(Internet)： Internet 是一個以 TCP/IP 為共同通訊協定所建立的跨國際之資訊服務網路，起源於美國國防部之 ARPANET 計劃，而後快速地發展成今日橫跨五大洲的超級大網路系統(註 2)，它是一個傳輸的媒介，並不是一個真正的實體，藉由共同的通訊協定，使用者可以自由、快速地瀏覽、查詢及擷取各種資訊。

二.網路資源搜尋工具 (Networked Information Retrieval Tool，簡稱 NIR Tool)：這是指為了協助使用者瀏覽、查詢及擷取網路上各類資源所發展出來的工具，其種類並沒有明確的區分，不過若從使用者的角度來分類，可大致分為以瀏覽及以查詢為主的二個系統：(註 3)

1. 以瀏覽為主(browsing-based)：這類工具是當查詢主題不確定時，利用層級式(hierarchical)或超連結(hyperlink)的方式來瀏覽資訊，前者如階層式的 Gopher，後者如 WWW 的各種類型的資源目錄。

2. 以查詢為主(query-based)：這類工具是當查詢主題確定時，可以用某些關鍵詞或查詢命令作查詢的系統，如查詢公用軟體的 Archie、尋找人名資訊的 Netfind、WHOIS、X.500 等系統，而 Gopher 及 WWW 這二個系統不僅可以瀏覽，也可以作查詢。

三.高佛系統(Gopher)： Gopher 是明尼蘇達大學(University of Minnesota)於 1991 年 4 月所發展出來的一套網路應用軟體及協定，它是一個分散式文件傳輸處理系統(Distributed Document Delivery System)，採用 Client/Server 架構，具

有易於瀏覽、有彈性的階層式選單，是屬於瀏覽式的 NIR Tool，另外，還有一個 Veronica 系統作 Gopher 選單關鍵字查詢的輔助。(註 4)

四.全球資訊網(World Wide Web，簡稱 WWW)：它是一個廣域性分散式超媒體系統 (distributed hypermedia system)，於 1989 年柏尼斯里 (Tim Berners Lee)在歐洲粒子物理實驗室 (CERN-the European Center for Nuclear Research)所發展出來的，以超文件(hypertext)、超連結的方式來整理 Internet 資源，同時也是一個兼具瀏覽與查詢的系統，使用者藉此可以找到全世界的數據化資訊，如文字、聲音、影像、圖形等資訊，其連結情況一如蜘蛛網一樣向外擴展(Web)，是目前 Internet 上最流行，發展相當快速的系統(註 5)。

五.Spider, Worm, Robot, Wanderer, Web Crawler：這些名詞都是指一種搜尋 Internet，收集網路資源之相關訊息的一種軟體程式或電腦網路技術(註 6)。這類工具經由程式原先的設定及維護，不斷自動地搜尋整個 Internet 的資源、記錄文件或節點的連結、關鍵字等狀況，隨時探索具有價值的資訊加以收集，並做某種程度的分類編目，最後創造出一個有價值的資料庫，供予使用者來檢索(註 7)。

六.主題檢索機制(subject-oriented search engine)：它是指 WWW 中利用 Internet 資源自動收集系統，如 Spider、Robot 等，建立一個有價值的資料庫，讓使用者以相關關鍵字鍵入的方式，加上各種檢索功能的利用，如布林邏輯、切截等，來查詢有關某特定主題的網路資源(註 8)。本文將此工具稱之為「主題檢索機制」，是屬於 NIR Tool 的第二類。

註釋

1. 莊道明，「由各行業在 Internet 之應用探討未來圖書館服務之主要改變」圖書館與國家資訊基礎建設(NII)研討會會議論文集，中華圖書資訊學教育學會主辦（民 84 年 10 月 13-14 日），頁 134。
2. Harley Hahn and Rick Stout, The Internet Complete Reference(Berkeley : McGraw-Hill, 1994),pp. 2-4.
3. George H. Brett, "Networked Information Retrieval Tools in the Academic Environment:Toward a Cybernetic Library," Internet Research 3 (Fall 1993) : 26-36
卜小蝶，「Internet 資源搜尋系統與圖書館資訊服務-以 Gopher 為例」中國圖書館學會會報 53 期(民 83 年 12 月)，頁 83-105
4. Louise McGillis, "Gopher Searching Using VERONICA" The Reference Librarian 41/42(1994):25-35.
5. 陳亞寧，「World-Wide Web:網網相連的網路世界」教育資料與圖書館學 32 卷 2 期(1995)，頁 194-200.
6. Wide Web Robots, Wanderers, and Spiders.
(<http://info.webcrawler.com/mak/projects.robots.robots.html>)
7. 同上註
John December and Neil Randall, The World Wide Web Unleashed 2nd ed.(Indianapolis:Sams.net, 1995), pp. 443-445.
Stephanie Johnson, "Accidental User," Online Access 11:2(Feb. 1996):65.
8. 同上註

第二章 文獻分析

根據本研究所牽涉之問題，將文獻分析分成：主題檢索問題探討、主題檢索的改進、主題檢索機制等三方面敘述如下：

第一節主題檢索問題探討

目錄研究一直是圖書館界感興趣的課題，它也是讀者欲充分利用圖書館館藏所必須使用的工具。目錄一方面可以顯示圖書館的收藏，另一方面也可以提供眾多不同的檢索點來檢索其所需要的資料。由已知的著者、書名來檢索目錄並獲得良好的結果，在現今的目錄使用來說已不是個問題，然而，在面對未知的資訊時，我們如何去尋檢資料就成了一個很重要、值得探討的問題。

Cutter 曾闡明圖書館目錄的目的是要讓讀者以著者、書名、主題來檢索（註 1），在面對眾多、變化多端的印刷媒體、網路資訊時，用主題來檢索所需要的資訊在整個資訊檢索來說是相當重要的。所謂的主題檢索是指讀者利用分類號、主題標目或其他含有主題意義的詞彙（如書中的關鍵字）所進行的查尋館藏資料的過程，亦即包含系統的（systematic）、主題的（topic）、以及自然的（natural）三種檢索途徑（註 2）。在傳統目錄的研究中，ALA 的研究調查了三十九個圖書館，包括學術圖書館、公共圖書館、專門圖書館以及中學圖書館，發現在公共圖書館中，讀者較常做主題檢索，而在學術圖書館及專門圖書館，讀者較常做已知作品檢索，整體而言，已知作品檢索占 48%，主題檢索佔 52%（註 3）。密西根大學(University of Michigan)的研究則發現無論大學或公共圖書館的讀者較常利用著者/書名檢索佔

49.5%、主題檢索佔 41.6%，而在大學中，年級越高則越少做主題檢索（註 4）。耶魯大學(Yale University)的研究顯示 56%的目錄使用是已知作品的檢索，33% 是主題檢索。（註 5）

在線上目錄使用研究方面，我們可以由下列的調查發現：

美國圖書館資源委員會（Council on Library Resources）曾在 1982 年做了大規模的線上目錄使用研究，包括二十九個學術、公共、州及聯邦圖書館的十六種線上目錄，它發現主題檢索佔了 59%，而已知作品檢索佔 41%（註 6）

Steinberg 及 Metz 指出，維吉尼亞技術學院(Virginia Polytechnic Institute)的線上目錄使用，主題檢索佔 69%（註 7）在國家醫學圖書館(National Library of Medicine)及國會圖書館(Library of Congress)所做的調查中也發現較多的目錄使用者做主題檢索（註 8）

根據美國圖書館資源委員會的一項調查指出，讀者做主題檢索時最常遇到的困難之一是找不到正確的標題詞彙（註 9）。

另外，Pritchard 詢問美國國會圖書館 SCORPIO 線上目錄的使用者，一半以上的使用者回答在做主題檢索時，他們所採用的字彙是自己隨意做決定的，而非來自國會標題表（註 10）。

Markey 在 Syracuse University 所做的調查也有同樣的發現，她也指出讀者用自己的字彙來檢索，有 65%的檢索是失敗的；即使讀者查看了國會標題表，有些讀者還是不知如何去使用標題表，也不明瞭"X"、"XX"或"See"、"See Also"的意義。（註 11）

除了讀者不會使用國會標題表造成檢索失敗外，國會標題表本身有若干缺點。1983 年 Mandel 及 Herschman 指出國會標題表