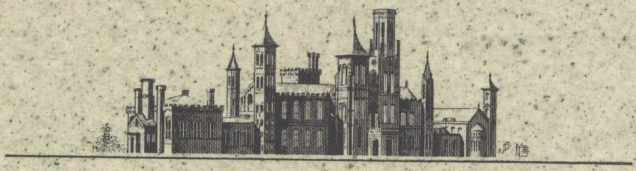
新圖書館學叢書



中文資訊檢索系統使用研究

吳美美

台灣學生書局印行

中文資訊檢索系統使用說明

[redacted]

中文資訊檢索系統使用研究

吳美美著

臺灣 學生書局 印行

國家圖書館出版品預行編目資料.

中文資訊檢索系統使用研究

吳美美著. – 初版. – 臺北市：臺灣學生，2001 [民 90]
面；公分
參考書目：面
含索引

ISBN 957-15-1065-3 (精裝)
ISBN 957-15-1066-1 (平裝)

1. 資訊儲存與檢索系統

028 90003028

中文資訊檢索系統使用研究 (全一冊)

著作者：吳美美
出版者：臺灣學生書局
發行人：孫善治
發行所：臺灣學生書局
臺北市和平東路一段一九八號
郵政劃撥帳號：00024668
電話：(02)23634156
傳眞：(02)23636334

本書局登記證字號：行政院新聞局局版北市業字第玖捌壹號

印刷所：宏輝彩色印刷公司
中和市永和路三六三巷四二號
電話：(02)22268853

定價：精裝新臺幣三四〇元
平裝新臺幣二七〇元

西元二〇〇一年四月初版

ISBN 957-15-1065-3 (精裝)
ISBN 957-15-1066-1 (平裝)

自　序

數位時代，人類儲存、使用資訊都需要藉助數位新媒體，然而是否能獲得正確且需要的資訊，除了有賴資訊儲存的正確完整，檢索是否順利，查全又查準，更是關鍵因素。數位圖書館前身是單一資訊檢索系統，最早是書目型資料庫，而後則有全文資料庫、全文檢索系統。稱爲「數位」因爲將資料數位化、稱爲「圖書館」因爲資料內容和資料類型包羅萬象，經過某種機器處理過程，標記索引，能供人檢閱。也有稱目前網際網路爲「數位圖書館」，實際上並不正確，「數位圖書館」需能提供有效獲取資訊的功能。編撰「數位圖書館資源書」(*Source Book on Digital Library*)美國維琴尼亞技術學院(Virginia Tech) Fox 教授即認爲「經過索引過程」的數位資料集合，才能稱爲數位圖書館。

數位圖書館在 1990 年代中期以來倍受囑目，數位圖書館研究是實用導向研究領域，奠基在許多相關基礎研究，如：檢索機制、自然語言處理、使用研究等。未來發展數位圖書館，必須先了解運作中的檢索系統、使用者的使用情形、資訊需求、檢索問題、檢索行爲，以及對檢索結果反應等，俾爲未來數位社會免於產生「科技剝奪」(technology censorship)鋪路。

現代政府和民眾重視資訊系統、數位圖書館、數位典藏建設。但爲避免「資訊貧」和「資訊富」差距過大，產生不平衡的社會，須從政策、教育、系統研究三管齊下，防範「數位鴻溝」和「社會鴻溝」現象。資訊獲取公平化，人民不分貧、富、智、愚都有均等

機會利用資訊科技、電腦、網路接收資訊。1993 年美國國家資訊基礎建設明定學校和公共圖書館必須接上網際網路，做爲美國國家發展目標之一。公平獲取資訊，是政府擘畫資訊服務政策所應重視的議題。其次，人人都能具備利用資訊科技獲取資訊的能力，透過教育體系全面實施，每位國民有均等機會接受資訊素養能力訓練，是資訊素養教育的課題。開發易於使用、親和特質的資訊檢索系統，了解系統使用的問題，加強改良系統功能和介面設計，增加使用率和使用效能，則是資訊檢索研究的課題。

一九八八年返國，在淡江大學教資系講授管理資訊系統，九五年在師大開授資訊檢索、索引與摘要等課程，介紹系統理論、檢索技術、資訊行爲、智慧檢索系統設計原理，與學生互動，發現中文資訊檢索系統日多，但對中文資訊檢索使用研究甚爲有限。中央研究院資訊所簡立峰博士及其研究群、中正大學吳昇教授及其研究群關注中文檢索引擎研究，受國際資訊檢索研究社群重視，但是中文資訊檢索使用研究的成果仍尚不明朗。

本書探討中西文資訊檢索研究，以及實務系統運作情形。資訊檢索研究層面，追溯資訊檢索研究的派別、探討資訊檢索研究緣起與範疇、各研究主題的重要發現。在資訊檢索實務運作層面，比較中西文資訊檢索系統發展軌跡、選擇數個中文檢索系統，利用實徵研究方法分析使用情況，包括資訊檢索問題、檢索詞彙、檢索點特質、成功與失敗檢索案例分析、使用困難等。全書分十章，緒論、資訊檢索研究範疇、中西文資訊檢索系統、使用研究、資訊檢索評鑑、研究方法、檢索問題與檢索詞彙、檢索案例分析、檢索系統使用滿意度調查、最後一章呼籲建立中文資訊檢索研究環境。

九六年至九八年連續三年出席 ACM SIGIR 年會，或發表論文、擔任論文評審、主持議程，觀察資訊檢索研究學域愈來愈年輕化，一方面欣喜，一方面覺得國內資訊檢索學域未能定型，中文數位化、數位圖書館呼聲漫天價響，中文資訊檢索基礎研究應該要有更多人參與。建立一個研究環境，讓中文資訊檢索研究者有對話、切磋的機制，是本書最後一章的期待。

本書適合閱讀對象包括對資訊檢索系統使用研究有興趣的圖書資訊學系、所學生，對中文資訊檢索系統使用有好奇的研究生，對改善檢索功能有熱情的研究者等。希望本書與您對話。

定稿成書真非易事。感謝學生書局鮑邦瑞先生協助出版。臺灣大學圖書資訊學系林珊如教授惠賜意見兼催生，黃慕萱教授於試讀版提供意見，精闢中肯，十分感謝。本書實徵研究係國科會贊助研究計畫「中文書目檢索系統的索引和檢索問題」、「中文資訊檢索系統使用者檢索行爲研究」、資策會「中文文件檢索效能評估系統建置」部分研究結果，感謝兩會贊助研究。研究助理黃育君、黃碧伶、蕭曉娟、林玉芳、劉英享、楊曉雯等俊材協同收集、整理資料，研究過程繁瑣、耗時、費神，數百匿名使用者願意參與，拓展我們對中文資訊檢索系統使用的了解，熱忱可感，特別致謝。撰寫過程數度易稿。姜杏蓉小姐協助編輯、校稿，編輯索引，謹申謝忱。至寫作期間，神在宇外，家人對我心不在焉的包容，就不稱謝了。

吳美美　謹識

民國九十年二月

於國立臺灣師範大學

目　次

圖 目 次

表 目 次

第一章 緒 論

第一節 資訊檢索研究緣起

資訊檢索作爲一個研究領域，不同年代有不同研究重點和研究群。從 1950 年代初期最早的資訊檢索研究開始算起，四十餘年間，資訊檢索研究領域被認爲內涵多重、詮釋各異。1991 年英國資訊學家 David Ellis 出版 *New Horizons in Information Retrieval* ❶，在序言中說「要爲資訊檢索研究作概括介紹，殊爲不易，當代資訊檢索研究涵蓋繁複、歧異的派別和技術，各派別有死忠贊同者和死敵，有專家也有生手。初進入此領域的人，難免迷惑不解。」(Ellis, 1996a, p. ix)。美國年輕資訊學家 Carol Hert 指出「資訊檢索研究理論體系和研究面向不一，研究重點和研究方法各異，對於研究結果的應用與解釋紛雜。」(Hert, 1997, p. 1)。數十年來資訊檢索研究乍看有似一盤散落拼圖，因此有必要將不同年代的資訊檢索研究予以接軌，以求具體呈現完整的研究領域和理論體系的演進。

資訊檢索研究有兩個主要研究派別，致力減少使用者大海撈針的不確定性。資訊檢索技術導向，從資料庫(知識庫)設計、檢索介面、檢索技術，如：相關回饋(relevance feedback)、相關度排序(ranking)、檢索問題擴展(query expansion)等層面，著手研究檢索

---

❶ Ellis, D. 1991 年出版 *New horizons in information retrieval* 為第一版，第二版 1996 年出版，改書名為 *Progress and problems in information retrieval*。

技術的改進；資訊檢索系統使用者研究導向，則從使用者、使用者和系統的互動，探討資訊檢索系統主要問題。研究課題包括使用者資訊需求、檢索問題、檢索詞彙和檢索策略研究、檢索結果的相關判斷、檢索滿意度評量，藉此提高對使用者檢索模式和使用困難的了解。

檢索技術導向研究自英國Cleverdon 進行Cranfield Test (1950s)算起，約有四十餘年歷史；資訊檢索使用者研究自 1985 年資訊檢索研究典範變遷算起，也有十餘年歷史。兩研究取向各有著重點，結合兩者的研究成果，對提昇資訊檢索研究領域十分重要。Hert (1997, pp. 6-10)也認爲資訊檢索研究有兩個研究典範：Cranfield Test 和 Salton 等研究傳統屬於「比對典範」(match paradigm)；資訊檢索使用者研究爲「認知典範」(cognitive paradigm)。Hert 認爲過去兩個典範各行其是，但是未來資訊檢索研究應在自然情境下進行、重視檢索過程研究，才能解決兩個研究典範間的爭議，對整體資訊檢索研究領域更有助益。

一部資訊檢索研究史如同一部資訊檢索評鑑史。Ellis (1996b)分析歷年來資訊檢索評鑑研究測量指標的困難，認爲資訊檢索評鑑研究用「相關」(relevance)作爲評鑑量標(measurement)正是問題所在。Ellis 將資訊檢索評鑑研究分爲原型階段、概率相關階段、專家系統等三個階段，各階段都有對評鑑量標的爭議。資訊檢索評鑑研究三個階段，可視爲資訊檢索研究三次典範變遷，惟典範間並非後者取代前者的關係。

原型階段　最早的資訊檢索研究比較何種資訊組織方法最有利於檢索相關資料。1953 年檢索評鑑研究指出以「相關」作爲