

中文檢索詞彙特質研究初探

吳 美 美

臺灣師範大學社會教育學系副教授

meiwu@cc.ntnu.edu.tw

A Preliminary Study on the Chinese Search Terms

Mei-Mei Wu

Abstract

The paper reports the results of an empirical study on the Chinese search terms. Three research questions are depicted: (1) What are the characteristics of the access points used in the Chinese bibliographic database? (2) What are the characteristics of the search terms used in the Chinese bibliographic database? (3) How do the search terms match the terms within the database in terms of zero hit and/or the number of relevant items retrieved? Research methods include observation, in-depth interview, and pre- and post-questionnaires. Research results are then presented.

關鍵詞：檢索詞彙、索引詞彙、中文檢索詞彙、文獻表示法、資訊需求、Information retrieval、Search terms、Text representation、Information need

壹、研究背景及研究問題

數位化資訊檢索系統在近年來發展快速，資訊提供服務(information provision service)專業由傳統紙本媒體儲存、人工資訊組織和檢索，大步邁向線上(online)快速而便捷的儲存和檢索。然而資訊檢索系統發展之初的各種技術問題，以及使用者的生澀，無疑是系統開發者要面對，且要不

斷克服的難題。觀諸美國電腦化檢索系統的發展，從1945年Vannevar Bush提出「理想」的檢索系統模式(Memex)開始，至今研究者仍孜孜不倦地繼續在探索、鑽研各種可能實現Memex的構想，其中包括系統、使用者及人機互動的各項研究。

在中文資訊檢索系統方面，經過十年來的努力，到今日已有豐碩的成果，譬如：國家圖書館的「中華民國期刊論文索引光碟及網路版」、農資中心的「農業資料庫」、科資中心的「科學技術資料庫」、立法院的「立法資訊系統」、中央研究院「古籍全文資訊系統」，以及中央通訊社「剪報資料庫查詢系統」等。近兩年來，學術網路蓬勃發展，數位化資料庫急速增加。這些中文資料庫能否被有效地使用，有效地查檢到使用者所需要的資訊，是使用者以及系統設計者共同關心的問題。

中文資訊檢索系統的開發設計，始於八〇年代，得以直接吸取來自西方國家自六〇年代所累積的有關研究、實驗及應用索引語文的經驗，而將其研究成果，在中文檢索系統設計之中加以利用，譬如「控制詞彙」和「自然詞彙」併用的索引方法、利用切截(truncate)、布林邏輯等檢索的技術，以及系統設計直接以「終端使用者」(end user)的使用為目標等原理和技術。更具體的說，我國的數位化資料庫資訊檢索系統，其中所用的索引和檢索之技術與原理，多為西文資料庫相關技術的移植。其間不同者，就是這些資料庫所儲存的文獻資料，是為中文。在目前中文檢索系統逐漸被普遍使用，且將有更多的資訊系統陸續產生之際，國內應開始有系統的進行大規模的「使用者研究」和「使用研究」，以及系統評鑑研究，以便了解讀者和系統互相運作的情形，改善系統，使其更精益求精。換言之，我國檢索系統的研究開發，在近幾年來有長足的進步，但是在「使用研究」方面，始終未有進一步的發展。使用者對於中文資訊檢索系統的使用是否得心應手，我們尚無一定程度的了解。「使用者研究」和「使用研究」對於提供更好的「資訊供應服務」無疑是必須努力持續進行的。

從平日的觀察中發現大多數中文資訊檢索系統的使用者在進行檢索時有幾個值得研究的現象，包括：高比率的零筆檢索結果、檢索結果筆數太多，而放棄做相關判斷；這些現象深值資訊服務者關心。資訊檢索的目的在連接「文獻」和「資訊需求者」，而詞彙則是資訊檢索過程中一個重要的研究問題。藉由研究檢索詞彙，可以進一步了解使用者的詞彙和系統詞彙之間的差距，進而探索零筆結果的原因(當然也有可能是系統中，根本就沒有所要尋求的資訊)。而了解系統中的詞彙和讀者使用的詞彙之間的關係，可做為建立索引典，連結相關詞彙之依據；此外，分析資訊需求者

的資訊需求，了解其可能入手查檢資料的角度和使用的詞彙，對於系統設計，尤其文獻表示法(text representation)方面，極有幫助，例如正在研發中的自動索引、自動分類等方法，可提供有用的參考。為了促進電子化資料能夠有效的被查檢，包括儲存文件的特質(attributes)、讀者可能入手查檢資料的角度，以及較適合的文獻表示(text representation)方法等，都是重要的研究課題。

本研究期望探討以下三個研究問題：

- (一) 讀者使用中文書目檢索系統時，其「檢索點」的使用有何特性？
- (二) 讀者使用中文書目檢索系統時，其使用的「檢索詞彙」的特質為何？
- (三) 讀者輸入的「檢索詞彙」和中文檢索系統中的「(索引)詞彙」之間的對應關係如何？系統反應的檢索筆數和檢索詞彙的特質有關嗎？

貳、名詞解釋

使用者(end user) 即有資訊需求的系統使用者，在本研究中和「讀者」、「檢索者」混用。

檢索問題(search request) 使用者將「資訊需求」轉成為文字敘述之後，該文字敘述稱為檢索問題。

檢索詞彙 讀者鍵入系統的詞彙稱為「檢索詞彙」，有可能是單詞、複詞、片語或句子等形式。

索引詞彙 在本研究中，指系統中的詞彙，是任一文獻表示法的呈現，包括「控制詞彙」及「自然詞彙」。

控制詞彙(control vocabulary) 是文獻表示法(text representation)的一種，由檢索系統提供，用以描述文獻的主題特質，稱為「控制詞彙」，譬如標題表、分類表及索引典等，都是常見的控制詞彙。

自由詞彙(free-text) 相對於「控制詞彙」，「自由詞彙」是指檢索時，系統直接以資料庫中的原文來和檢索詞彙相比對，而不需受欄位的限制，因此「自由詞彙」的檢索方式，亦可稱為「全域檢索」(free-text searching)。

檢索詞彙的形式 傳統主題標目之形式包括有「自然語言」，以及「人工語言」。如「中文圖書標題表」●中常使用之主題標目有下列之形式：(一)自然語言，包含有「單一名詞」，譬如化學、世界主義、侵權行為

：「形容詞詞組」，譬如中國語言、民俗音樂、農業經濟；「連詞詞組」，譬如投票與開票、圖書館與讀者、圖書館與大眾傳播；以及「介詞詞組」等。(二)人工語言，包含有「限定形式的標題詞」，譬如向背(地質學)、退化(生物)、退化(病理)；以及「複分標題」，譬如石油--開採、石油--貯存、臺灣--歷史。

檢索詞彙的類型 有關檢索詞彙的特質研究，Bates, Wilde & Siegfried^{②③}以及Siegfried, Bates, & Wilde^④曾針對在Getty Art History Information Program中的人文學者進行一項為期兩年的研究計劃，深入研究其檢索Dialogue資料庫的各種行為，包括人文學者所進行研究問題和其線上檢索的關係，如檢索詞彙和檢索行為。根據Bates, Wilde & Siegfried^⑤，人文學者的資訊需求一般而言有三類：(一)查詢特定的作品或出版品；(二)查詢某一位作者之作品或出版品；(三)查詢某一主題的作品；而依檢索詞彙的種類來區分，而人文學者經常使用的詞彙有下列七種：^⑥

1. 以某作品或出版為主題檢索；
2. 各種人名--包括作者、小說中的虛構人名、神話人名、宗教人物等；
3. 地理名--包括名詞及形容詞的型式；
4. 時代--包括特定日期或時間範圍、某一時代，以及時間修飾；
5. 學科範疇；
6. 其他專有名詞；
7. 其他一般名詞。

本研究採用李宜容^⑦碩士論文中，對於中文檢索詞彙類型的定義，並依實際資料特質略予修改，共將詞彙類型分為八種：

- (一) 專有名詞，包括地名、人名、團體及組織名稱、歷史上事件、活動名稱、種族、民族名稱、語言名稱、作品名稱、朝代名稱以及動植物名稱等。依本研究之實例，專有名詞如莊子、紅樹林、李後主等。
- (二) 學派、主義(者)款目，此類款目可能尚無法明確被定義為專有名詞，但是其在學科領域中，可能是重要的主題，所代表的詞彙亦較為特定，譬如南朝文學。
- (三) 時期、年代用語，指定某特定日期或時間，譬如1812、70年代、或19世紀。
- (四) 學科名，乃指包含一廣大的研究領域，可以成為大學之一門科系者，譬如電影、音樂、考古、歷史等。

- (五) 其它專有款目：乃指未包含在上述類型之專有名詞，如檢索者所使用的主題檢索詞彙，同時包含專有名詞及上述任一種以上款目者，譬如法國女性主義。
- (六) 含有專有名詞之款目：乃指主題檢索詞彙中包含了上述之專有名詞，最常見是利用專有名詞來限定一般普通名詞之範圍，譬如客家信仰；或是專有名詞連接普通名詞，或是利用普通名詞形容專有名詞，譬如晚唐。
- (七) 普通名詞：未能包含在上述幾種詞彙類型的一般款目。
- (八) 其它：檢索者以不完整的主題檢索詞彙、或是介詞片語方式，表示其檢索問題之概念者，或是依本研究之實例，布林邏輯組合等則歸入此項。

叁、研究設計

本研究主要在探討中文書目檢索系統中，檢索詞彙和索引詞彙的問題，包括「檢索詞彙」的特性分析、「檢索詞彙」和「索引詞彙」的對應關係，以及使用者對檢索結果的相關判斷。研究的方法包括文獻探討、觀察、問卷和訪談法。首先選擇一中文書目檢索系統做為研究對象，資料的收集包括讀者的「檢索問題」、「檢索詞彙」、檢索結果，以及檢索結果的相關判斷等。資料的分析則包括內容分析和描述性統計分析。

本研究的步驟包括：

一、準備工作

- (一) 選擇一中文書目資訊系統；
- (二) 了解系統之檢索功能及設計原理；
- (三) 設計前、後問卷及觀察記錄表；
- (四) 訓練訪員；
- (五) 設計編碼簿。

二、資料收集與整理

三、資料分析

肆、研究實施

本研究以「中華民國期刊論文索引檢索系統」的使用者做為研究的主要對象，主要的原因有三：第一，選擇一較具規模的檢索系統，如中華民

國期刊論文索引系統做為研究的開始，可有許多便利之處，譬如系統的使用率較高，樣本較易獲得；第二，這個檢索系統是一個主題涵蓋範圍廣泛，為一般性的書目檢索系統，其使用者的檢索詞彙較為多樣，是研究檢索詞彙的豐富來源；第三，這個系統的使用群分佈較廣，不限於某一學科，樣本有廣泛的代表性。

本研究共收集國家圖書館該系統的使用者十三名、淡江大學圖書館該系統的使用者十八名，共計檢索者三十一名，有效查詢148次，包括以「關鍵字」為檢索點的查詢次數123次，以「書名」為檢索點的查詢次數13次，以「作者」為檢索點的查詢次數12次。在123次以「關鍵字」為檢索點的查詢中，共使用檢索詞彙135個，進一步再分析這些詞彙的特質。

一、資料的收集

本研究所收集的資料，包括系統說明文件、前後問卷、檢索者檢索過程之觀察記錄、檢索者檢索結果資料、讀者對檢索結果的相關判斷記錄。

訪談者在獲得受訪者的同意後，實際觀察並記錄讀者之系統使用的過程，包括檢索者檢索點和檢索詞彙之使用，以及檢索結果之相關判斷。資料收集的過程包括：

- (一) 在系統所在地附近等候，首先邀請並徵求使用者參加此項研究，對受訪者說明研究的目的及進行的方式，若蒙同意，請受訪者填寫檢索前問卷；
- (二) 使用者進行實際的檢索。訪談員觀察並於觀察表中，記錄使用者的檢索指令及檢索詞彙、系統的反應筆數、以及使用者的反應等檢索過程；
- (三) 進行訪談，請使用者判斷檢索結果。訪談者並將檢索過程與檢索結果存於磁片中；
- (四) 請檢索者填寫後問卷，並進行後訪談，以補充觀察記錄之不足。

二、資料的整理

將上述三十一個樣本的原始資料依個案編號排列，將使用者的檢索問題、檢索詞彙、檢索結果及檢索結果相關判斷內容整理之，包括：

- (一) 檢索者的檢索問題及檢索詞彙表，包括檢索者的檢索問題、查詢步驟、檢索點及檢索詞彙的使用記錄；
- (二) 檢索詞彙及相關判斷筆數，包括檢索結果、相關判斷結果，以及詞彙出現的欄位；

(三) 檢索結果為零筆之詞彙。

三、資料的分析

資料的分析以質化分析為主，包括：

- (一) 檢索者的檢索問題和檢索詞彙分析；
- (二) 檢索點特質分析；
- (三) 檢索詞彙特質分析；
- (四) 檢索詞彙與檢索結果的關係。

伍、研究結果

一、檢索問題和檢索詞彙的關係

分析檢索者的「檢索問題」和「檢索詞彙」，發現兩者之間可歸納為四種關係：(一)「檢索詞彙」反映「檢索問題」，所使用的「檢索詞彙」通常為「檢索問題」中之關鍵字，單純書目性的檢索問題屬之；(二)「檢索詞彙」多於「檢索問題」中所使用的詞彙；(三)「檢索問題」中提出的概念和詞彙多於「檢索詞彙」；以及(四)「尋找直接解答」的「檢索問題」等四種類型。這些模式是否和「檢索問題」的特質、「檢索問題」的複雜程度，以及檢索者對「檢索問題」了解的深淺程度有關，深值思考。茲將上述四種類型說明舉例如后：

(一)「檢索詞彙」反映「檢索問題」，如：

「晚明袁宏道之論文研究」，以「袁宏道」為檢索詞彙，檢得13筆資料，判斷相關資料為2筆。

「集中支付--財政部」，以「集中支付」做為檢索詞彙，檢得9筆資料，相關判斷資料亦為9筆。

檢索問題為「有關"愛情"方面的文章，做為大學生愛情觀抽樣調查的資料」，檢索者直接用「愛情」為檢索詞彙，檢得113筆資料，相關判斷資料為10筆。

(二)「檢索詞彙」多於「檢索問題」中所使用的詞彙，如：

「有關共同基金之研究」，以關鍵字「共同基金」為檢索詞彙，除了KW為檢索問題之外，並另外檢索三個外國作者名。

「有關用水習慣(台灣地區居民用水習慣)」，以「用水習慣」檢索為0筆，之後改用「自來水」，得268筆，其中判斷7筆為相關。

另有檢索問題為「組織行為」，其檢索詞彙則有：

政治行為	9	9(檢索筆數與相關筆數)
組織溝通	7	7
溝通與政治	0	0
公共關係	93	0
政治活動	13	0

「有關融資、跨國性企業之資料(包括資金從何處來、子公司融資的來源、資產負債表的比例、發行股票、借債券的情形等)」，檢索詞彙豐富，如"多國籍企業"，亦有期望找較直接答案的詞彙如"融資來源"。

「戰爭戰役理論」，使用了17個檢索詞彙。

「總動員體制」，使用了12個檢索詞彙。

「廠商存活率」，檢索詞彙中利用"集中度"一詞檢索，因該詞與檢索問題中之"存活率"有密切關係，廠商集中度愈高，其存活率愈高，其檢索詞彙包括：

存活率	21	0(檢索筆數與相關筆數)
集中度	13	13
製造業*勞動	12	12
存活*產業*廠商	6	0
存活率	21	0

「查陳後主與李後主之亡國詩」，而檢索者以該朝文學、詩詞名稱及原作者為檢索詞彙。

「有關貨幣與所得效率性檢定，亦即探討貨幣與所得之間的關係」，由於檢索者檢索目的主要在於補充資料，使用的檢索詞彙較為有關該問題的研究方法方面。

「有關"超媒體"方面的資料」。檢索詞彙中除了出現"超媒體"之外，也包括"早自習"及"自習"的詞彙，根據檢索者表示，"自習"與"超媒體"皆為教學方面的資料，只是附帶檢索，看是否兩者之間有應用關係。

「BBS在教學上的應用」，檢索詞彙則以bbs及教學為主，並考慮到以與BBS相關的"網路"做為檢索詞彙。

「南向政策與對外直接投資」，檢索即以"南向政策"與"對外直接投資"，加上與二個詞彙相近似的詞彙做為檢索詞彙，如"南進政策"，其檢索詞彙包括：

南向政策	30	7(檢索筆數與相關筆數)
南進政策	1	0
東南亞and對外直接投資	0	0
對外直接投資	6	6
「卦變」，使用的檢索詞彙包括：		
卦變	0	0(檢索筆數與相關筆數)
圖書易學	0	0
圖書易	0	0
宋代圖書易	0	0
易	7046	0
易圖	76	4

(三)檢索問題中的概念或詞彙多於檢索詞彙，如：

「有關有線電視設立過程及相關有線電視法」，僅以"有線電視"為檢索詞彙，檢得162筆資料，判斷29筆相關。

「有關大傳與選舉之間的關係」，而僅以"大眾傳播"一詞找答案，檢得258筆資料，相關判斷資料為14筆。

「台灣水筆仔觀察實錄與感想(關於水筆仔的探討研究，尤其在環境倫理方面)」，檢索詞彙以"水筆仔"及"紅樹林"查尋，分別檢得9筆與61筆資料，相關判斷資料分別為7筆與37筆。

「企業研究方法是屬於科學或藝術？」，僅以"企業研究方法"及"企業研究"做為檢索詞彙，產生0筆結果之後，又以"研究方法"檢索，得52筆檢索結果，判斷為相關資料之筆數為11筆。

「朱博湧、吳壽山發表於科學發展月刊之論文」；以"朱博湧"為檢索詞彙，檢得5筆資料，判斷相關資料為1筆，然而因為檢索者已經擁有此筆資料，故未予圈選。

(四)尋找直接解答的檢索問題

檢索問題期望找到直接答案的，換言之，可在某些供快速檢索的參考書，如百科全書或專書中查得相關之資料者，如：

「利率與匯率之關聯性」，而僅以"利率"與"匯率"做為檢索詞彙，分別檢出609筆與654筆資料，判斷相關的筆數則分別為4筆與9筆：

利率	609	4
匯率	654	9

「莊子內的儒家人物」，僅以"莊子"一個檢索詞彙檢出382筆資料，而判斷出2筆相關資料。

「鈦金屬於飛機結構上之用途」，僅以一個「鈦」字做為檢索詞彙。

以上由研究中所發現的四種檢索問題和檢索詞彙的關係，和個別檢索者當時的知識結構(認知狀態)可能有密切的關係，值得進一步探究。上述的發現可說為資訊尋求行為的研究領域又增加一些新的思考空間。

二、檢索點的使用特質

本研究共收集31位受訪者之查詢過程，共進行有效查詢148次。在所有受訪者之查詢中，檢索者以關鍵字為檢索點之查詢次數為最多，共計123次，佔所有查詢次數的83.1%。其中檢索者僅以關鍵字為檢索點，未加布林邏輯查詢之次數計有107次(佔以關鍵字為檢索點查詢次數的87%)，檢索者以關鍵字為檢索點並結合出版日期之查詢次數有14次(佔11.4%)，檢索者以關鍵字為檢索點並結合期刊題名，以及檢索者以關鍵字為檢索點並結合期刊題名與出版日期之查詢次數則各有1次(各佔0.8%)；以題名為檢索點之查詢次數居第二位，計有13次(佔8.7%)；而以作者為檢索點之查詢次數居第三位，僅有12次(佔8.1%)(如表一)。

表一 各檢索點使用之次數及百分比

檢索點類型	檢索次數	百分比
關鍵字 (KW)	123	83.1%
題 名 (TI)	13	8.7%
作 者 (AU)	12	8.1%
合 計	148	100.0%

在各類型檢索點獲得零筆結果及判斷為零筆相關資料之分析方面，在所有受訪者之查詢中，獲得零筆結果之次數有47次，佔所有查詢次數的31.8%。其中以「關鍵字」或以「題名」為檢索點查詢，獲得零筆結果之比例皆較以「作者」為檢索點之查詢略高，各佔32.5%及30.8%；而以作者為檢索點查詢，產生零筆結果之百分比為25%(如表二)

表二 各檢索點獲得零筆結果之次數及百分比

檢索點	檢索次數	零筆結果次數	零筆結果次數/使用次數%
關鍵字 (KW)	123	40	32.5%
題 名 (TI)	13	4	30.8%
作 者 (AU)	12	3	25.0%
合 計	148	47	31.8%

至於查詢獲得相關資料之情形，以「作者」為檢索點查詢之成功率為最高，總計12次的查詢中，以「作者」為檢索點約共有9次獲得檢索結果，其中有8次判斷為有相關資料，若以每次檢索有獲得相關資料為「成功」，則以「作者」為檢索點的成功率約為66.7%；以「關鍵字」為檢索點查詢共有123次，檢索得到非零筆結果的次數為83次，判斷之後，獲有相關資料之次數為64，若以每次檢索有獲得相關資料為「成功」，則其成功率約為52.0%；以「題名」為檢索點查詢之次數共13次，計有9次獲得檢索結果，其中約有5次判斷有相關資料，若以每次檢索有獲得相關資料為「成功」，則其檢索成功率為38.5%，為三種檢索點成功率最低者(如表三)。綜合而言，在148次查詢中，判斷之後獲有相關資料者為77次，佔總查詢次數的52.0%，換言之，查詢的整體成功率為百分之五十二，亦即檢索者約有一半的機會，能檢索到相關的資料。

表三 各檢索點查獲相關資料之次數及百分比

檢索點類型	檢索次數	非零筆結果 次數	判斷之後獲有 相關資料者	判斷之後獲有相關 資料次數/檢索次 數%
作者(AU)	12	9(75%)	8	66.7%
關鍵字(KW)	123	83(67%)	64	52.0%
題名(TI)	13	9(69%)	5	38.5%
合計	148	101(69%)	77	52.0%

三、檢索詞彙之特質

從上述分析結果可以發現，受訪者最常使用的檢索點是「關鍵字」查詢，而以「關鍵字」為檢索點的成功率較以「書名」為檢索點為高，但較「作者」為檢索點的成功率稍低。檢索者使用書目檢索系統時，似乎多數會選擇以「關鍵字」為檢索點查詢，然而以「關鍵字」為檢索點查詢，則可能會因為檢索者對於某一概念所採用的「檢索詞彙」與系統中的詞彙不同，而產生零筆結果或判斷為零筆相關資料。

在本研究中，受訪者以「關鍵字」(KW)為檢索點進行查詢之次數共計123次，其中使用中文關鍵字為檢索點之查詢計有117次，使用英文關鍵字為檢索點之查詢計有6次。檢索詞彙中，以布林邏輯做連結者，詞彙另計，如以"教學*軟體"做為檢索詞彙，則以二個詞彙計算，因此，在以中文關鍵字為檢索點之117次查詢中，包括14次各使用一個布林邏輯運算元連結兩個詞彙進行查詢，以及2次各使用二個布林邏輯運算元連結三個詞彙

進行查詢，總計中文檢索詞彙共有135個。本研究主要以「關鍵字」為檢索點之中文「檢索詞彙」做為分析的對象，並探討檢索者使用「檢索詞彙」的形式，以及「檢索詞彙」之「類型」，期能對檢索者使用檢索詞彙有進一步了解。

(一) 檢索詞彙之形式及其檢索結果

依據前述「中文圖書標題表」對於標題形式之分類，以及湯廷池(民76)對中文詞語之定義，本研究將檢索者的檢索詞彙嘗試歸納舉例如后：

- 單一概念詞：
例如：共同基金、自來水、中文、加工、水筆仔
- 形容詞詞組：
例如：用水習慣、企業社會責任、食品加工
- 連詞詞組：
例如：溝通與政治、網路與教學
- 介詞詞組：
例如：國際融資的方式

依前述定義以及本研究中檢索者使用詞彙之特性，將135個檢索點為關鍵字為之檢索詞彙分析如表四。依據表四，檢索者以關鍵字為檢索點之檢索詞彙形式，以「單一概念詞」為最多，共79個，佔所有以關鍵字為檢索點之檢索詞彙的58.5%；其次為「形容詞詞組」之形式，共52個，佔所有以關鍵字為檢索點之檢索詞彙的38.5%；「連詞詞組」之形式居第三位，僅3個，佔所有以關鍵字為檢索點之檢索詞彙的2.2%；而「介詞詞組」之形式僅有1次，僅佔所有以關鍵字為檢索點之檢索詞彙的0.7%。

表四 以關鍵字為檢索點之檢索詞彙形式分析

詞彙形式	次數及百分比
單一概念詞	79(58.5%)
形容詞詞組	52(38.5%)
連詞詞組	3(2.2%)
介詞詞組	1(0.7%)
合計	135(100.0%)

以檢索詞彙的形式和檢索結果而言，在117次以中文關鍵字為檢索的查詢中，單純形容詞詞組的查詢有46次，單一概念詞有39次，對於檢索者以關鍵字為檢索點並利用布林邏輯結合其他關鍵詞、出版時間、或是題名

檢索詞彙之查詢則將其歸入「其他」項者，有28次；連詞詞組3次，介詞詞組1次，共計117次(如表五)。

在117次以中文關鍵字為檢索點的檢索結果中，「單一概念詞」之檢索詞彙能獲得資料之百分比約為82.1%，使用「形容詞詞組」之檢索詞彙檢得資料之百分比約為56.5%，而利用「連詞詞組」或是「介詞詞組」之檢索詞彙能檢得資料之百分比皆為零。至於「其它」，亦即檢索者利用布林邏輯進行查詢而能獲得資料之比例，和受訪者使用「單一概念詞」之檢索詞彙進行查詢有相同之百分比約82.1%，兩者都是檢索詞彙的各種形式之中，能檢得資料之可能性最高者。

表五 以中文關鍵字為檢索點之「檢索詞彙」形式及非零筆檢索結果之次數及百分比

詞彙形式	檢索次數	非零筆檢索結果	非零筆檢索結果 /檢索次數%
單一概念詞	39	32	82.1%
其他(如布林邏輯組合)	28	23	82.1%
形容詞詞組	46	26	56.5%
連詞詞組	3	0	0
介詞詞組	1	0	0
總數	117	81	69.2%

(二) 檢索詞彙之類型

本研究採用李宜容●的碩士論文對於中文檢索詞彙的類型，共分為八種（參考本文貳、名詞解釋），將總共135個以關鍵字為檢索點之檢索詞彙予以分析，其次數及百分比如表六。檢索者最常使用之以關鍵字為檢索詞彙之類型以「普通名詞」為最多，共111個，佔所有以關鍵字為檢索點之檢索詞彙的82.2%；而非普通名詞之使用較少，總共僅有24個，為所有以關鍵字為檢索點之檢索詞彙的17.8%。非普通名詞之中，「專有名詞」之詞彙類型有12個，為所有以關鍵字為檢索點之檢索詞彙的8.9%，「含有專有名詞」之詞彙類型共7個，佔所有以關鍵字為檢索點之檢索詞彙的5.2%，「學派、主義(者)」之款目類型有4個，佔所有以關鍵字為檢索點之檢索詞彙3%，「其它專有款目」之詞彙類型僅有1個，佔所有以關鍵字為檢索點之檢索詞彙的0.7%。至於「時間用語」、「學科名」等，在本研究中皆未被使用，其原因是否和使用者的學科特質有關，值得探究。

表六 以關鍵字為檢索點之「檢索詞彙」類型次數及百分比

詞彙類型	次數及百分比
普通名詞	111(82.2%)
專有名詞	12(8.9%)
含有專有名詞之款目	7(5.2%)
學派、主義(者)款目	4(3.0%)
其它專有款目	1(0.7%)
時期、年代	0
學科名	0
其它	0
合計	135(100%)

以檢索詞彙的類型和檢索結果而言，檢索者「利用布林邏輯等其他項」所檢得資料之百分比最高，約82.1%，檢索者以「專有名詞」為檢索詞彙能檢得資料之百分比為70%，居第二位，檢索者以「普通名詞」為檢索詞彙能檢得資料之百分比則為68.1%，居第三位，受訪者以「學派、主義」為檢索詞彙能檢得資料之百分比為50%，居第四位，至於受訪者以「含有專有名詞之款目」為檢索詞彙而檢得資料之百分比則為33.3%，居第五位(如表七)。注意表六和表七「詞彙次數」和「檢索次數」略有不同，以詞彙類型而言，「專有名詞」雖出現12次、「含有專有名詞之款目」有7次，「其他專有款目」一次，但在檢索中，多以「布林邏輯」等方式呈現，而非單獨出現，故「詞彙次數」和「檢索次數」略有不同。

表七 以「關鍵字」為檢索點之「檢索詞彙」類型及非零筆結果之次數及百分比

詞彙類型	檢索次數	非零筆檢索結果 分配次數	非零筆檢索結果 分配次數/檢索次 數%
利用布林邏輯等其它項	28	23	82.1%
專有名詞	10	7	70.0%
普通名詞	69	47	68.1%
學派、主義(者)款目	4	2	50.0%
含有專有名詞之款目	6	2	33.3%
其它專有款目	0	0	0
時期、年代	0	0	0
學科名	0	0	0
總數	117	81	69.2%

四、檢索詞彙和索引詞彙欄位的對應關係

利用「關鍵字」(KW)為檢索點的使用最多，因為以「關鍵字」(KW)為檢索點，在系統中查詢的範圍較為廣泛，以「中華民國期刊論文索引」為例，包括篇名、並列篇名、附錄及內容註，皆屬關鍵字查詢。因此分析檢索詞彙對應到資料中相同字串出現的位置，亦即索引詞彙的欄位，有助於了解檢索詞彙和索引詞彙的對應關係。「關鍵字」詞彙欄位的所在位置包括篇名、內容、附、附表及附錄等(參附錄)。

31個個案中，讀者以「關鍵字」查詢共計123次，其中以中文關鍵字查詢為117次。117次以中文關鍵字查詢到的筆數共有17,711筆，平均每次查詢得到約152筆資料。其中除去讀者放棄判斷，以及數據缺乏(data missing)之資料外，讀者進行「相關判斷」者共計65次，獲得檢索筆數共計4,390筆，其中共獲得相關筆數382筆。進一步計算，檢索者會進行判斷的檢索筆數平均約為67.5筆，而每次檢索獲得相關筆數平均約有5.8筆。此外，進一步比較其「檢索詞彙」和「索引詞彙」的對應情形，發現「檢索詞彙」與「索引詞彙」在「篇名」中對應者有4,236次；「內容」33次；「部份內容」15次；「附」34次；「附錄」及「附表」各一次。

而就讀者進行「相關判斷」，判斷相關之資料筆數382筆中，「檢索詞彙」與「索引詞彙」相對應之情形而言，以「篇名」對應到的有373次；「內容」10次；「部份內容」2次；「附」3次。值得注意的是，以「篇名」對應出來的資料，且為使用者判斷為「相關」者，100次對應中，僅有8次被判定為相關的機會；而以「內容」做為索引詞彙，則十次的對應中，有三次獲得相關資料的機會(如表八)，「內容」做為索引詞彙，其表現可說最佳。以此而言，書目資料庫的設計，仍需要重新考慮其「文獻表徵」(text representation)的方法，「內容」相較於「篇名」，顯然其「文獻表徵」的效果好得多。

「檢索詞彙」與「索引詞彙」相對應的情形以「篇名」為最高，這個現象無庸置疑，因為在系統設計中「關鍵字」索引的欄位設計中，「篇名」為最主要詞彙的來源，其他欄位，如：「內容」、「部份內容」、「附」、「附錄」、「附表」等，並不是每一筆書目資料皆具備。事實上，在4390筆資料中，未在「篇名」欄位中出現的僅為54次。而在判斷為相關的382筆資料中，未在「篇名」欄位出現者亦僅有9筆。在「內容」中對應到的相關筆數為10筆，其中就有7筆是獨特的，未在「篇名」中出現的。

表八 「檢索詞彙」和「索引詞彙」的對應情形

	所有「檢索詞彙」和「索引詞彙」在各欄位對應到的次數	判定相關文獻，其「檢索詞彙」和「索引詞彙」在各欄位對應到的次數	判定為相關文獻，其「檢索詞彙」和「索引詞彙」對應到的次數 / 所有「檢索詞彙」和「索引詞彙」對應到的次數
內容	33	10	30.3 %
部份內容	15	2	13.3 %
篇名	4,236	373	8.8 %
附	34	3	8.8 %
附錄	1	0	0
附表	1	0	0

陸、綜合討論

一、三次檢索中有一次零筆資料的可能

在關鍵字、作者和題名等三種檢索點中，關鍵字的使用次數最多，而其失敗率也最高，譬如在總計123次的關鍵字檢索中，出現零筆檢索結果的次數共有40次，約為32.5%，亦即每三次檢索中，即有一次可能獲得零筆資料的機會。在總計13次以題名為檢索點的檢索次數中，零筆結果有4次(約30.8%)，獲得零筆檢索結果的可能性亦大約為三分之一。以作者做為檢索點比對到資料的可能性則較高，統計12次的檢索中，零筆結果有3次(25%)，換言之，以作者為檢索點的檢索次數中，每4個檢索有一次得零筆結果的可能。

二、以作者為檢索點查獲資料的機會大；以題名為檢索點查獲資料的機會最低

進一步探討檢索者對檢索結果的相關判斷，則發現以作者為檢索點的查詢，檢索者反應有查詢到相關資料者，9筆中有8筆(88.9%)，以關鍵字查詢者次之，在83個有檢索結果的查詢中，64個查詢獲有相關資料(77.1%)，而以題名為檢索點，在9個結果中，僅有5個查詢是檢索者反應獲有相關資料(55.6%)。

三、詞彙形式中以「單一概念詞」，詞彙類型中以「普通名詞」使用最多

在檢索詞彙的形式方面，根據分析，檢索者使用最多的詞彙形式為「單一概念詞」（135個詞彙中，有79個），其次為「形容詞詞組」（135個詞彙中，有52個）。但是若以檢索次數來計，則總計117次以關鍵字為檢索點的查詢次數中，使用「形容詞詞組」有46次，使用「單一概念詞」的檢索次數則有39次，前者獲得檢索結果的次數為26次，約為56.5%，後者獲得檢索結果的次數有32次，約為82.1%。此外，「單一概念詞」多以「布林邏輯組合」出現，共計28次的檢索中，有23次獲有資料，約是82.1%。連詞詞組和介詞詞組各使用3次和1次，但4次的檢索結果，都獲得零筆資料。

檢索詞彙的類型方面，普通名詞使用最多，在135個詞彙中共有111個，其餘為專有名詞。

四、檢索詞彙和索引詞彙的主要對應來源是篇名欄位但相關判斷的主要來源則為「內容注」

在以關鍵字為檢索點的檢索詞彙對應到資料庫中詞彙的所在位置方面，仍以比對到篇名為最大部分，輔以附錄、內容等。而檢索者判斷為相關的資料，多以內容比例最高。中華民國期刊論文索引目前無摘要，亦無所謂的主題或類別描述語（descriptor），其文獻表示（text representation）僅有題名、作者、內容注等。對於相關文獻的可能被檢索出來的條件實在嫌不夠充分。如同一個人僅有姓名和座號，而無其他之描述語，要在一群人中將之識別出來，非得準確地知道其姓名或座號不可。因此之故，吾人認為詮釋學實有必要用於文獻表示，以增加資料被檢索的可能性。在未來圖書館員或索引員，以及其他資料組織者的訓練中，詮釋能力的培養應予以加強。

五、系統特別的檢索功能很少被使用

值得一提的是「中華民國期刊論文索引」檢索系統的檢索功能，包括以下四種功能：（1）完全符合（exact match）：在檢索欄位中輸入檢索字串，之後加上"\$"符號。例如：在作者欄位中輸入"王平\$"三字，檢索結果只會出現作者為"王平"的著作，而不會出現"王平安"或"王平洋"。（2）向後切截（partial match, truncate）：又稱為「前面完全符合」，在獨立檢索欄位中輸入任何文字串並緊接一個"?"符號，其欄位中自第一個字起與檢索值相同的資料都會被檢索出來。例如：在篇名中輸入"經濟?"三字，可以找到「經濟.....」等文章，但查不到「台灣經濟發展

計畫」一文。(3) 逐字檢索：除了「分類號」和「類名」外，在獨立檢索欄位中，輸入任何字串，後面不加任何符號，均視為逐字檢索。例如：在作者欄中輸入「林文」二字，可以找到「林文」、「林文謙」、「角林文雄」等人的著作。(4) 近似檢索 (similarity)：除了「分類號」和「類名」外，在獨立檢索欄位中，輸入任何字串，後面加「^」者，均視為近似檢索。「近似檢索」與「逐字檢索」的主要差別在於，「近似檢索」並不要求檢索字串中每一個字都必須緊密相連，只要在檢索字串的三倍長度內出現，且字序相同即會被檢索出來。例如：在篇名欄中輸入「中研院^」，可以找到篇名有「中研院」、「中央研究院」、「中央科學研究院」、「中央經濟研究院」等字串的文章。

除了以上檢索功能外，系統仍然提供了瀏覽功能。使用者若對有不確定的檢索值，系統提供有各檢索點索引檔的瀏覽功能，使用者可輸入部分索引值或不輸入任何字元，按下瀏覽功能鍵，即可連續瀏覽相關索引值及收錄筆數，再挑選一或多筆適當的索引值後直接選入查詢。此外，也有提供布林邏輯運算的功能。但是從研究分析中，123次的檢索中，沒有一次是用到前述「\$」、「?」、「^」的檢索功能。這個現象深值系統設計者注意。

六、學科背景不同其詞彙使用的類型亦有所不同

本研究使用李宜容●的詞彙類型表，予以歸類，發現有某些檢索詞彙的類型並未在檢索「中華民國期刊論文索引」中出現。譬如時期、年代、學科名字，這個現象，是否表示檢索詞彙的使用，仍依學科類別有所不同？Bates等人進行詞彙研究，揭示學科背景不同的學者，其詞彙使用的類型有所不同。Bates等人利用七種詞彙的類型來分析Saracevic & Kantor (1988) 在NSF (National Science Foundation) 所進行的研究，發現人文學者所使用的詞彙類型和社會學者以及自然科學學者可謂涇渭分明 (如表九)。後者 (社會及自然科學學者) 多利用一般名詞檢索，而前者 (人文學者) 以作者名、地理、學科名詞等檢索較多。當然上述兩者的問題總數相當懸殊：38和150，其研究結果只能提供參考。但在本研究的結果中，亦發現類似的結果，亦即不同學科的檢索者，有可能使用不同的詞彙類型，這項發現有助於了解不同學科檢索者的資訊搜尋行為，值得進一步探究。

表九 NSF和Getty詞彙研究之比較⑩

	NSF 社會科學		NSF 自然科學		NSF 合計		Getty 合計	
	次數	%	次數	%	次數	%	次數	%
問題總數	22	58	16	42	38	100	150	100
作品名做為主題	1	5	0	0	1	3	8	5
作者名做為主題	0	0	0	0	0	0	74	49
地理名詞	3	14	0	0	3	8	37	25
時代名詞	1	5	0	0	1	3	6	17
學科名詞	0	0	0	0	0	0	35	23
其他專門名詞	3	14	1	6	4	11	11	7
其他一般名詞	22	100	16	100	38	100	85	57

柒、建議

一、加強人文和科技的對話

資訊檢索系統的研究，在技術研究方面約有四十年，在行為研究方面亦有將近十年，兩者的研究應到了互相對話的時候。為了檢索系統的易於使用、能夠普及使用，這兩類型的研究結果是否能互相為用，換言之，人文和科技的銜接點在那裏？系統研究和使用者研究又在那裏對話？事實上，系統的電腦記錄(transaction log)有無限的資料(data)可提供分析，而使用者研究(user study)的結果，可以提供親和性系統無限寶貴的資源。舉例而言，系統設計的特別的檢索功能很少被使用，又如本研究檢索者002的檢索問題為「有關共同基金之研究」，檢索者鍵入三個英文人名，卻有二個查詢獲得零筆回應。假若研究者可以研究歸納使用者的困難類型，便可以在介面中加入聰慧的回應，例如：「本資料庫(知識庫)的範圍為.....，您現在鍵入的英文人名，建議您查.....系統」，未來並可進一步將該資料庫用超文獻的方式串接，或者也可以在系統設計中整體考慮規劃中英文資料庫合併查詢。

總而言之，系統的使用者研究在於關心使用者的需求，包括資訊需求的滿足，以及同時而來的系統互動產生的相關問題。人文和科技的對話，系統研究和使用者研究的銜接，我想應始於此。

二、研究方法的改進建議

本研究利用觀察、訪談、檢索結果和檢索詞彙的比對等方法，進行資料的收集和分析的工作，基本上是十分費時、費工、費人力。雖然研究者十分期望一次能同時進行數個檢索系統的使用和使用者研究，以便比較，進一步獲得使用者的檢索行為模式，然而經過一年餘的試驗，兩位研究助理的辛勞工作，事實說明，這是一件要有耐心的長期性工作。中文資訊檢索系統的使用者研究是十分重要的，因為它是未來完善電子圖書館發展的基礎。唯其重要，在研究方法上，便要檢討改進。以下從問卷設計、人員訓練、電腦記錄(transaction log)等分項檢討：

未來相關性的研究，在前後問卷中應加強檢索者的資訊需求部分，譬如資訊需求的原因、內容、資訊搜尋的階段等，檢索判斷的原因及根據等。訪員的經驗培養乃十分重要，對研究了解的程度會影響其訪談的深度，在本研究中，訪員備忘及訪談前的訓練都發揮作用，在資料分析的部分則從做中學，詞彙的分析和統計分析能力的培養是從事此類研究必不可少的素養。碩士訓練以上的研究生具備相關知識，有足夠的研究好奇，樂於探討學習，可望勝任，我們的學科在培養有上述特質的人材方面，仍需努力。在資料收集方面，本研究僅用半人工式的電腦記錄的磁帶分析法(transaction log)的方式，予以下載印出，在詞彙分析方面，亦用人工的方法予以彙集、分析、歸類、計算，主要原因在於只用「電腦記錄」的磁帶分析法，對於使用者的檢索問題及背景資料無法掌握。目前電腦或許仍無法做詞彙分析和歸類的工作，但在未來的研究中，應考慮邀請系統相關人員協助，譬如撰寫線上問卷，補助傳統電腦記錄磁帶分析的限制，再輔以訪談，如此或可減輕研究人員之繁瑣工作。

至於資料的分析尤其費時，譬如檢索詞彙的形式類型分析，檢索問題的歸類工作都是極其智力工作(intellectual work)，無法由電腦取代，仍需由高智力的人力來執行。

捌、未來的研究課題

中文資訊檢索系統大別為二種，傳統圖書館書目資料庫概念所發展出來的書目、索引等檢索系統，如各大學的OPAC、教育論文摘要、中華民國期刊論文索引、農學、科技、立法資訊系統等，另如自1985年有中文全文系統，由各該學門的人員及電腦系統人員開發而成，不受傳統圖書館書目資料庫的一套規範所束縛，1990年以後，更有網路檢索系統出現，書目

和全文這兩套不同設計理念的系統，在檢索方面的異同為何？換言之，是否可共同檢索引擎(search engine)？值得進一步探究。

此外，檢索者的個人特質和檢索問題、檢索行為乃至檢索判斷，在中文環境中，都是一值得研究的領域。譬如，在本研究中，從檢索策略可以看得出來，031和032是較為有經驗的系統使用者，因為他們能夠使用較複雜的檢索指令，如kwpd，kwjtpd等。反過來說，大多數的檢索者都傾向於使用最簡單的檢索策略，這些檢索行為和檢索者的個人特質、知識狀況和資訊需求有助於資訊系統的設計，並可做為圖書館員進行圖書館利用指導的教材及參考。

感謝國科會之贊助。本文為國科會專題研究報告：中文書目檢索系統的索引和檢索問題：從「邏輯的相關」和「適切的相關」來看文獻表徵的方法。(國科會專題研究報告，計畫編號：NSC 85-2413-H-003-013)之部分內容。

附 註

- ①國家圖書館編，中文圖書標題表。台北市：編者印行，(民84)。
- ②M. J. Bates, D. N. Wilde & S. Siegfried, " An analysis of search terminology used by humanities scholars: the Getty Online Searching Project report no.1" *Library Quarterly*, 63(1):1-39, 1993.
- ③M. J. Bates, D. N. Wilde & S. Siegfried, " Research practices of humanities scholars in an online environment: the Getty Online Searching Project report no.3" *Library and Information Science* 17:5-40, Winter 1995.
- ④S. L. Siegfried, M. J. Bates & D. N. Wilde, " A profile of end-user searching behavior by humanities scholars: the Getty Online Searching Project report no.2" *Journal of the American Society for Information Science*, 44(5):273-291, 1993.
- ⑤同註②。
- ⑥同註②，頁7。
- ⑦李宜容，人文及社會學科讀者使用線上公用目錄檢索詞彙之研究。淡江大學教育資料科學研究所碩士論文，(民85)。
- ⑧同註⑦。
- ⑨同前註。

⑩同註⑨，頁17。

參考文獻

- 李宜容(民85)。人文及社會學科讀者使用線上公用目錄檢索詞彙之研究。
私立淡江大學教育資料科學研究所圖書館學與資訊科學組，碩士論文。
- 吳美美(民85)。中文書目檢索系統的索引和檢索問題：從「邏輯的相關」
和「適切的相關」來看文獻表徵的方法。(國科會專題研究報告)計畫編
號: NSC 85-2413-H-003-013。
- 吳碧娟，宋美珍(民84)。「中華民國期刊論文索引」光碟系統之開發與應
用。國立中央圖書館館刊 28(1): 3-26。
- 湯廷池(民76)。漢語詞法句法論集。台北市：台灣學生。
- 國家圖書館編(民84)。中文圖書標題表。台北市：編者印行。
- Bates, M. J., Wilde, D. N. & Siegfried, S.(1993). An analysis of search ter-
minology used by humanities scholars: the Getty Online Searching
Project report no.1 *Library Quarterly* 63(1): 1-39.
- Bates, M.J., Wilde, D.N. & Siegfried, S. (1995) Research practices of hu-
manities scholars is an online environment : the Getty Online Searching
Project report no. 3. *Library and Information Science Research* 17:5-40.
- Brown (1995). By any other name: Accounting for failure in the naming of
subject categories. *LISR* 17: 347-385.
- Fidel, R.(1991a). Searchers' selection of search keys: I. The selection rou-
tine. *Journal of the American Society for Information Science* 42(7): 490-
500.
- Fidel, R.(1991b). Searchers' selection of search keys: II. Controlled vocabu-
lary and free-text searching. *Journal of the American Society for Infor-
mation Science* 42(7): 501-514.
- Fidel, R.(1991c). Searchers' selection of search keys: III. Searching styles.
Journal of the American Society for Information Science 42(7): 515-527.
- Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T.(1987).
The vocabulary problem in human-system communication.
Communications of the ACM 30(11): 964-971.
- Gomez, L. M., Lochbaum, C. C. & Landauer, T. K.(1990). All the right

words: Finding what you want as a function of rightness of indexing vocabulary. *Journal of the American Society for Information Science* 41(8): 547-559.

Saracevic, T., Kantor, P., Chamis, A. Y., & Trivison, D.(1988). A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science* 39(3): 161-176.

Saracevic, T., & Kantor, P.(1988). A study of information seeking and retrieving. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science* 39(3): 197-216.

Siegfried, S.L., Bates, M.J. & Wiled, D.N. (1993). A profile of end-user searching behavior by humanities scholars : the Getty Online Searching Project report no.2. *Journal of the Amevican Society for Information Science* 44(5) : 273-291.