

新一代資訊檢索技術在圖書館 OPAC 系統的應用

曾元顯

輔仁大學圖書資訊學系
台北縣新莊市中正路 510 號
TEL:(02)9031111 EXT. 2333
Email: tseng@blue.lins.fju.edu.tw

摘要：

圖書館的「線上公用目錄」(On-Line Public Access, OPAC)系統，是提供讀者查詢圖書館館藏資料的重要工具，也是早期資訊檢索技術主要的應用方向。本文即以輔仁大學圖書館為例，說明整合新一代資訊檢索技術於 OPAC 系統的實施現況。我們運用非營利機構可獲得的先進檢索引擎，輔以自行發展的關鍵詞擷取技術，完成具備重要性排序、近似字串、模糊搜尋、相關詞回饋、允許近似自然語言檢索字串的 OPAC 系統。初步的評估發現，這樣的系統對協助使用者簡化查詢條件、拓展檢索字彙、提昇檢索成效有相當的助益。

關鍵字：線上公用目錄、資訊檢索、模糊搜尋、近似自然語言、相關詞回饋

壹、前言

近年來電腦硬體與軟體的日益進展，以及網路上越來越多的資訊系統，如 News、Gopher、BBS、WWW (World Wide Web) 等，使得國內網際網路的使用人數，在未來二、三年內將推廣至三百萬人口 [1]。在這一波網路服務的建設中，圖書館也在加緊自動化與網路化的腳步，使讀者更方便利用圖書館的資源。圖書館的網路服務中，跟讀者最直接相關的為「線上公用目錄」(OPAC, On-line Public Access Catalog) 服務。由於館藏的查詢幾乎是一般讀者利用圖書館資源的第一個步驟，因此，各個圖書館自動化系統，無不提供 OPAC 服務，期望在使用介面上更具親和性，而且查詢方式更為便利。

在 WWW 還未出現以前，圖書館的線上公用目錄主要以「遠端登錄」(telnet)的方式，提供網路使用者時、空無礙的查詢檢索。Telnet 的優點在於整個查詢的過程中系統一直維持與使用者的連線，並維護使用者查詢的狀態，使得查詢結果的修飾，如縮小範圍、不同查詢結果的聯集、交集運算等，可以直覺的進行。然而 WWW 出現後，卻相對的暴露出以 telnet 為使用者界面的許多缺點 [2]。

此外，目前大部份的自動化系統其 OPAC 模組主要是以布林邏輯（ Boolean Logic ）運算提供書目查詢。然而一般使用者對布林邏輯的運用較不熟悉、檢索結果沒有依照符合程度排序、檢索字串要求精確無誤等原因，導致較高的檢索失敗率（ search failure ）與資訊溢檢率（ informatin overloading ）[3-5]，造成書目檢索系統不易使用 [6]。

現今的資訊檢索技術已往智慧型、自然語言檢索的方向發展 [7-9]。全球資訊網上的檢索引擎已有近似的系統出現，如 Lycos [10]、Gais [11]、Csmart [12]、Altavista [13] 等。這些系統分別具備了「模糊搜尋」（ fuzzy search ）、「相關詞回饋」（ relevance feedback ）、或容許以近似自然語言的查詢字串作為檢索條件的能力。我們希望這些新一代的資訊檢索技術，也能運用在傳統書目資料的檢索服務上 [13-15]。

本校圖書館近來正在進行自動化系統的升級工作，針對多家圖書館自動化系統的評估，大部份廠商還未能提供先進的檢索功能。然而為了達成上述的期望，在有限的人力與經費許可下，我們仍然嘗試自行發展一種彈性的架構 [16]。由於選廠決定系統的作業需要一段時程，此種架構必須能夠與任何一家廠商的系統一起運作，以確保我們不必耗費太多的努力在不同系統的整合與既有功能的開發上，而僅需專注於發展先進的檢索系統即可。

本文即以輔仁大學圖書館為例，說明整合新一代資訊檢索技術於 OPAC 系統的實施現況以及所能獲致的檢索成效。文章架構如下：下一節我們將介紹查詢模式的進展，說明資訊檢索技術的發展方向，並指出我們目前的進程；第三節簡介輔大 OPAC 雛型系統的架構與建置現況；第四節列舉一些檢索範例並討論其檢索成效；最後則是我們的結語。

貳、查詢模式的進展

傳統的資訊檢索模式通常採用布林邏輯模式，布林邏輯模式在實作上較為簡單，也可針對不同欄位的資料或相同欄位的多個關鍵詞做布林邏輯運算，以縮小檢索範圍。然而其最大的缺點在於檢索結果雖然都是符合條件的文件或記錄，卻無法區別出個別文件對此次檢索的重要程度。此外，目前網路通達的程度與普及速度，讓一般使用者皆可從家裡、辦公室、或任何方便的地方連線上網查詢資料，使用者如果沒有受過相當的訓練，或具備足夠的經驗，很難利用布林邏輯的方式，擬出比較有效的檢索策略，進行資料的檢索。這與過去使用者必須到圖書館查資料，而圖書館裡有館員協助的情形大不相同。所以檢索技術的發展方向是逐漸把使用者端的檢索複雜度，推向到檢索主機這一端，或是設計更便利的人機介面模式，讓使用者的檢索環境越來越簡單，但還維持一定、甚至更好的檢索效率。

資訊檢索技術雖已發展一段時日，拜網路技術的發展與計算機功能的不斷提昇，近幾年進展得更為快速，尤其在使用者端的介面方面，有多種查詢模式被提出來。綜合文獻上提出的各種模式 [17-21]，列舉重要者說明如下：

一、 Boolean model （布林邏輯查詢）：此模式製作上較為簡單、檢索速度快、可以用不同欄位資料來限定檢索範圍，對主題明確的檢索（如明確的作者名稱、標題名稱）非常有效，然一般使用者比較難以利用此種模式表達較為複雜的查詢。

二、 **Ranking**（重要性排序）：檢索結果按符合程度排序，以加快檢索結果的檢視整理與利用，此為布林邏輯模式難以達到的重要功能。

三、 **Fuzzy search**（模糊搜尋）：即容錯式、全文式、非控制字彙、近似字串 (proximity)、允許利用近似自然語言的方式表達檢索字串與條件的檢索模式。此種模式大大降低資訊檢索的複雜度，對不明確自己檢索主題的使用者幫助尤其顯著。

四、 **Relevance feedback**（相關回饋）：使用者根據系統對檢索條件的回應，將相關資料或條件回饋給系統，以導引系統搜尋的方向，來逐漸逼近自己所要的資料。例如，使用者可以根據系統的初步回應，指出哪些文件或相關詞跟他的檢索主題相關，而哪些又是完全不相關，將此訊息回饋給系統，系統根據此種訊息，再作進一步的搜尋。因此，相關回饋表現出來的常成為漸進式查詢 (progressive query) 或範例查詢 (query by example) 的模式。

五、 **Personalized service**（個人化服務）：檢索系統記錄個別使用者的資訊需求，或是把相同需求的使用者的記錄組合運用，讓使用者彼此的需求和興趣交互推薦，使得使用者查詢時，系統提供較符合個人興趣的回應，做到個別化的服務。

六、 **Information filtering**（資訊過濾）：此種模式與使用者主動查詢資料而系統被動反應相反。使用者向系統登錄自己的資訊需求、興趣或檢索條件後，由系統主動、持續的為使用者蒐集相關的資料，再定時或不定時傳給使用者，使用者只需被動接受檢索與過濾後的結果即可。因此，在一段期間內，使用者只需做一次檢索，即可獲得持續的檢索結果。

七、 **Query by voice**（語音檢索）：由文字介面轉變為較為自然的口語語音介面，減低文字打字輸入的困難度，可以配合其他檢索模式運用。

八、 **Query by dialog**（對話式查詢）：系統以文字或語音為介面同使用者對話，從對話中透露使用者的資訊需求與意圖，系統再據以檢索。此種模式強調的是使用者端運用較為精緻的人機介面技術，因此與檢索引擎的實際運作原理可以較無關係。

九、 **Query by natural language**（自然語言檢索）：對話式查詢仍由系統主導話題與使用的語句，自然語言檢索則允許使用者以不限定的自然詞語、句法與系統溝通，因此使用者的負擔更輕，但系統的介面部份必須具備相當程度的人工智慧能力，以瞭解自然語言的意義。

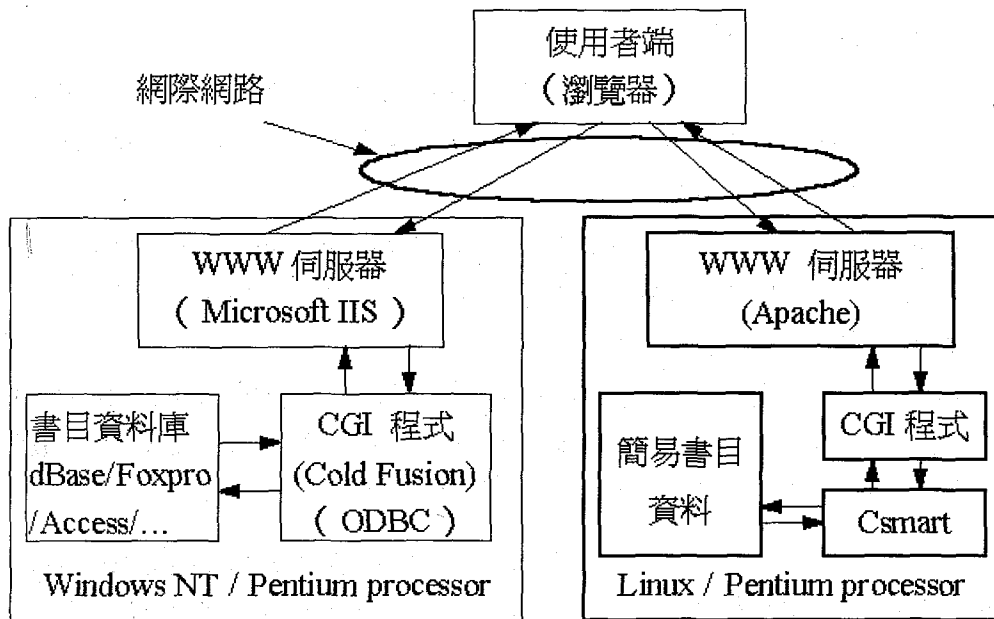
十、 **Intelligent search agent**（智慧型檢索精靈）：使用者將自己的資訊需求交代檢索精靈程式，由此精靈代替使用者到各個相關資料庫檢索資料，再回報回使用者。此種模式是以上各種模式與人機介面技術的綜合運用，使得資訊檢索可以達到時空無礙、虛擬實境的境界。

雖然「智慧型檢索精靈」在電影中已經展示過了，電影中提示到的個別科技目前也有相當的進展，但如自然語言處理等技術，距想像中實際可用並且大量普及的階段仍有距離，目前的進展僅能在某些領域內才能發揮作用 [22]。然而，若檢視其他查詢模式，有些已接近成熟實用階段，例如，目前我們建構的 OPAC 雛型系統，便具備重要性排序、模糊搜尋與相關詞回饋的功能，對協助使用者簡化查詢條件、拓展檢索字彙、提昇檢索成效有相當的助益。

參、輔仁大學 OPAC 雛型系統

一般 OPAC 的查詢模式常以兩個階段來進行：第一階段讀者下達檢索條件，系統則回應相關書目的簡略資料；第二階段讀者檢視簡易書目，點選某一筆記錄後，系統再顯示該比記錄的詳細資料。在第一階段中，讀者下達的檢索條件，目的是盡可能蒐集到所有相關的書目，要求的是較廣泛而完整的檢索範圍；第二階段則明確的限定查出某筆記錄，僅做一筆記錄詳細內容的調閱。一般 OPAC 系統所提供的布林邏輯查詢，對檢索主題與查詢範圍非常明確的讀者而言，可以有效的找出心中想要的資料（如某作者著作的書），或在第二階段有效找出某筆特定記錄。但是布林邏輯查詢對於檢索概念或範圍不明確的讀者，則常常在第一階段無法一次就蒐集到足夠的相關書目，必須經過多次的嚐試。然而由於一般讀者對布林邏輯的運用較不熟悉、檢索結果沒有依照符合程度排序、檢索字串要求精確無誤等原因，導致較高的檢索失敗率。此外，多次試驗性的查詢，也加重主機的負擔，造成系統效能降低。

為了解決查詢主機的負擔，並且讓我們發展出來的檢索系統，能在不同的軟硬體系統下皆能整合運作，我們提出分散式處理的架構。此分散式架構邏輯層次上可分為兩部主機，如下圖所示。其中一部主機安裝先進的檢索軟體，提供使用者以直覺的方式下達查詢條件，進行第一階段的查詢。另一部主機則安裝傳統的圖書館自動化系統，透過 WWW 的超連結能力，提供第二階段顯示詳細書目的查詢。此主機並可進行讀者預約登錄與編目、採訪等作業的處理。當然，在實際的運作上，只要軟體能互相搭配，上述的作業全部集中在同一部主機亦無不可。以輔大的 OPAC 系統為例，由於目前僅是過渡性的雛型系統，採用的設備是 PC 等級的電腦，因此採用雙主機的配置，以提昇書目檢索的速度。



圖一：輔仁大學 OPAC 雛型系統架構

此種分散式書目檢索架構，需要從完整的書目資料分離出簡略的書目記錄，並放置在不同的主機上。然而相同資料放在不同主機需要特別注意資料一致性的問題。所幸作為第一階段查詢用的書目記錄，不牽涉到需要及時更新的資料，如讀者預約記錄，此簡略書目不需要以每分鐘或每小時的速率更新，

每天或每週的某幾天更新即可。在此情況下要維護資料的一致性較無問題，僅需在使用量最低時，如深夜或清晨，以程式自動更新即可。資料可分散的特性，是此種分散式處理架構能夠實現的先決條件。

圖一所示目前的 OPAC 雛型系統架構，都是以 Pentium 等級的個人電腦為硬體平台。一套主機安裝中文版 Windows NT 3.51 作業系統，存放完整的書目資料，主要目的是利用較為普及廉價的資料庫管理系統提供傳統的查詢檢索功能。另一套主機安裝國內學術界發展的中文全文檢索引擎，允許近似自然語言的檢索字串，提供容錯式、近似字串的模糊搜尋。目前國內對先進資訊檢索技術的研究較著名的有中正大學的 Gais 系統 [11]，以及中央研究院資訊科學研究所的 Csmart 系統 [12]。這些研究已有成果發表在網路上，供社會各界免費使用於非營利之公眾服務。我們選用 Csmart 系統，因為它非常適合我們目前的中文書目記錄。

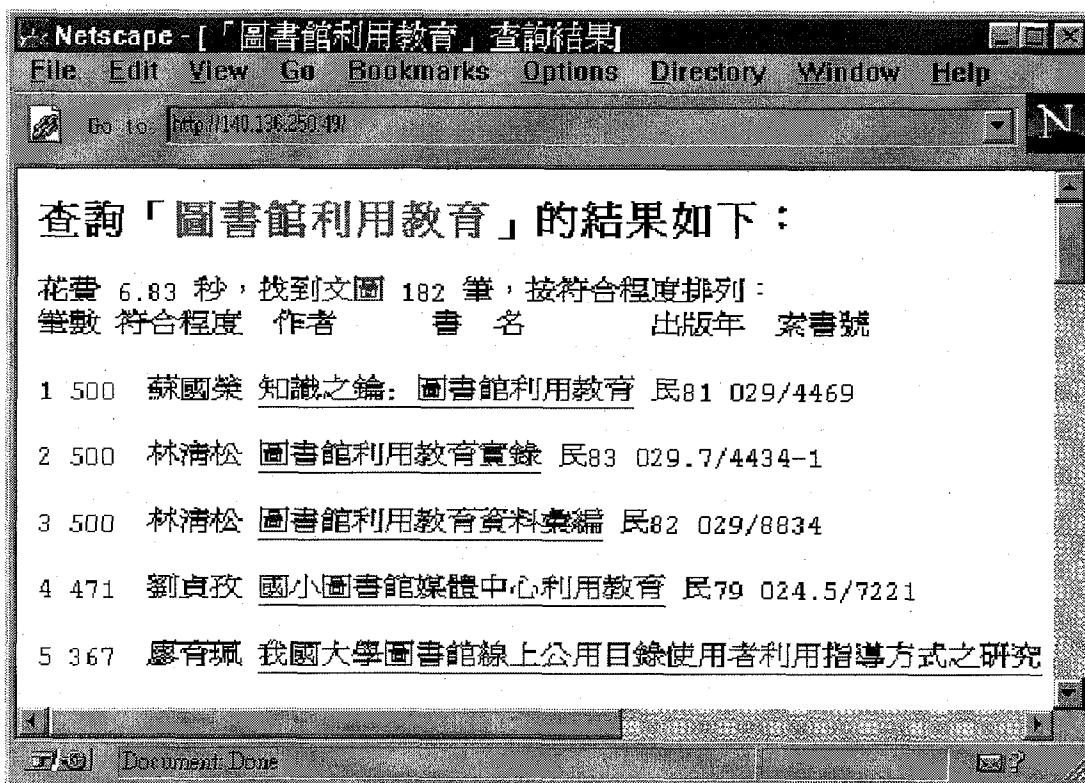
Csmart 檢索引擎能夠提供近似自然語言的查詢功能，查詢的結果會依相關程度由大到小排列順序。例如，查詢有關心理學入門的書，可以輸入「**基礎心理學導論概論原理入門**」這樣的字串。儘管這樣的字串並不是理想的檢索條件，容易找出其他不相關的書目出來，但是「心理學導論」（比對到五個字，而且為連續的詞）、「心理學基礎」、「心理學入門」（比對到五個字）等書目還是會被列在前面，而「幼兒心理學」、「犯罪心理學」等專門的書雖然也會被找出，但是會排列在後面（只比對到三個字）。Csmart 根據中文的特性，用比剛才舉的例子更為精準的方式排列查詢結果 [9, 15]，讀者可根據查詢結果的符合程度，來判斷排列在後面的資料是不是還有參考的價值，以節省檢視查詢結果的時間。此種查詢方式對不諳查詢技巧的使用者而言極為便利，想到什麼詞句，就輸入什麼字串，而不用關心應該要用 AND 邏輯，或是 OR 運算。Csmart 已開發出語音輸入的版本，將來使用者甚至能用口語來表達查詢條件。

除了利用先進檢索引擎提供模糊搜尋的功能外，我們也自行發展了一套相關詞自動擷取功能 [23]。此功能可就使用者檢索的結果進行分析，自動擷取出與此次檢索主題相關的關鍵詞，以拓展使用者檢索的字彙，供進一步的搜尋參考。相關回饋在文獻上被認為對提昇檢索成效有相當大的助益 [17-19]，下一節的範例中，亦可看出實際運用時的優點。

肆、查詢範例與成效

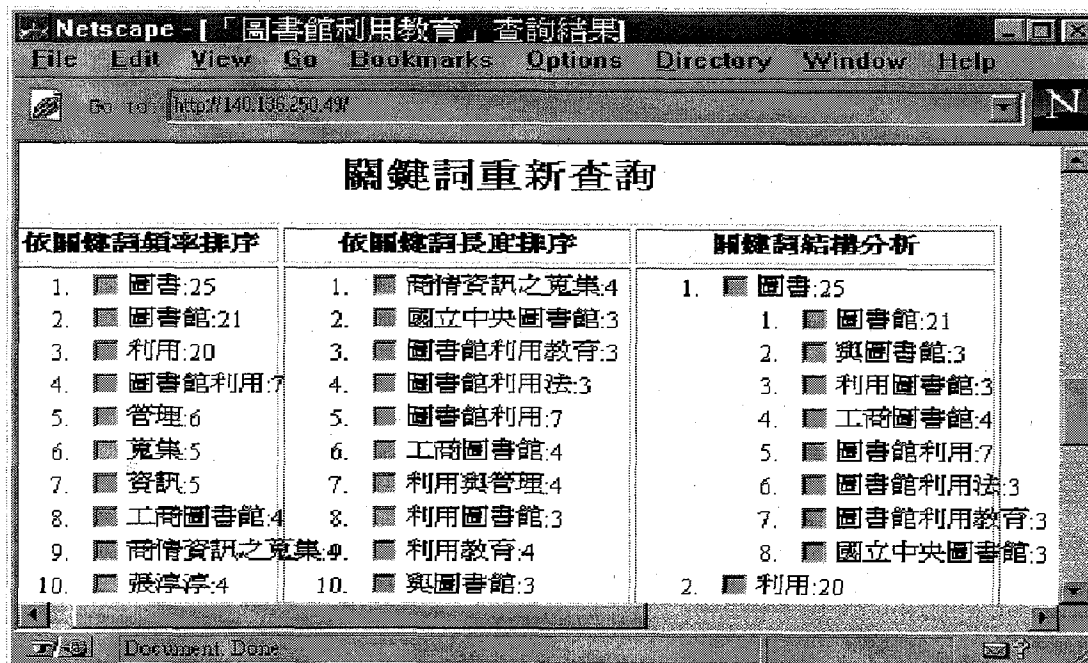
輔大 OPAC 雛型系統暫時位址為 <http://140.136.250.49/>，讀者可自行連線測試。底下所舉範例，來自對檢索記錄檔的分析，以呈現使用者實際的檢索行為。

從使用者檢索的記錄檔中，我們選擇：「**圖書館利用教育**」此項條件重作查詢，其部份檢索結果如下：



在 182 筆結果中，前三筆資料與檢索字串完全吻合，其字串符合程度最高，分數為 500 分，排列也最前面，其後四、五筆字串符合分數稍低，但也幾乎一樣相關。相較之下，若以布林邏輯條件查詢：「**圖書館 AND 利用 AND 教育**」則前四筆雖然不一定排列在一起，仍然可找到，但第五筆書目資料就要換用不同的邏輯條件找尋了。

除了允許近似字串、模糊搜尋的條件之外，相關詞自動擷取功能可就此次的查詢結果，擷取出相關的關鍵詞，其部份結果如下：



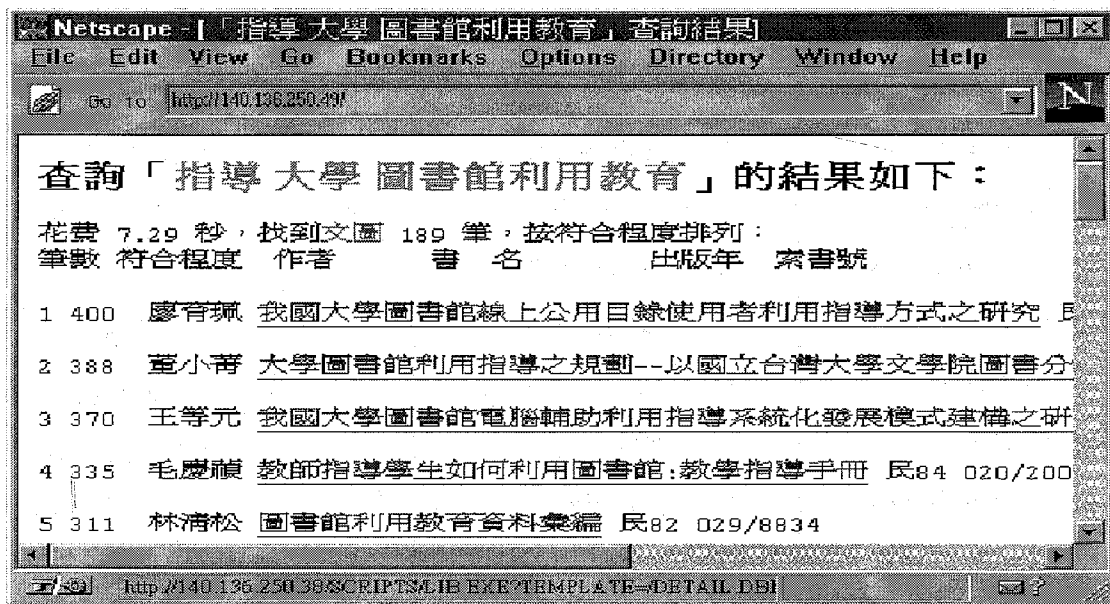
相關詞擷取結果分別按詞頻、詞長、以及廣義詞、狹義詞等方式排列，每個詞後面並標示該詞出現的次數，以便利使用者瀏覽。相關詞是以程式自動產生，由於電腦程式不瞭解文字意義，所以會有擷取錯誤的情形發生。所幸，此種錯誤發生的機率不高，擷取出來的相關詞相當理想。此例中，完整的相關詞如下表所示：

關鍵詞重新查詢

依關鍵詞頻率排序	依關鍵詞長度排序	關鍵詞結構分析
1. <input type="checkbox"/> 圖書:24	1. <input type="checkbox"/> 商情資訊之蒐集:4	1. <input type="checkbox"/> 圖書:24
2. <input type="checkbox"/> 圖書館:20	2. <input type="checkbox"/> 國立中央圖書館:3	1.1. <input type="checkbox"/> 圖書館:20
3. <input type="checkbox"/> 利用:19	3. <input type="checkbox"/> 圖書館利用教育:3	1.2. <input type="checkbox"/> 與圖書館:3
4. <input type="checkbox"/> 圖書館利用:7	4. <input type="checkbox"/> 圖書館利用法:3	1.3. <input type="checkbox"/> 工商圖書館:4
5. <input type="checkbox"/> 管理:6	5. <input type="checkbox"/> 圖書館利用:7	1.4. <input type="checkbox"/> 圖書館利用:7
6. <input type="checkbox"/> 蒐集:5	6. <input type="checkbox"/> 工商圖書館:4	1.5. <input type="checkbox"/> 圖書館利用法:3
7. <input type="checkbox"/> 資訊:5	7. <input type="checkbox"/> 利用與管理:4	1.6. <input type="checkbox"/> 圖書館利用教育:3
8. <input type="checkbox"/> 工商圖書館:4	8. <input type="checkbox"/> 利用教育:4	1.7. <input type="checkbox"/> 國立中央圖書館:3
9. <input type="checkbox"/> 商情資訊之蒐集:4	9. <input type="checkbox"/> 與圖書館:3	2. <input type="checkbox"/> 利用:19
10. <input type="checkbox"/> 張淳淳:4	10. <input type="checkbox"/> 圖書館:20	2.1. <input type="checkbox"/> 利用教育:4
11. <input type="checkbox"/> 利用教育:4	11. <input type="checkbox"/> 張淳淳:4	2.2. <input type="checkbox"/> 利用與管理:4
12. <input type="checkbox"/> 大學:4	12. <input type="checkbox"/> 圖書:24	2.3. <input type="checkbox"/> 圖書館利用:7
13. <input type="checkbox"/> 利用與管理:4	13. <input type="checkbox"/> 利用:19	2.4. <input type="checkbox"/> 圖書館利用法:3
14. <input type="checkbox"/> 國立:4	14. <input type="checkbox"/> 管理:6	2.5. <input type="checkbox"/> 圖書館利用教育:3
15. <input type="checkbox"/> 圖書館利用法:3	15. <input type="checkbox"/> 蒐集:5	3. <input type="checkbox"/> 管理:6
16. <input type="checkbox"/> 與圖書館:3	16. <input type="checkbox"/> 資訊:5	3.1. <input type="checkbox"/> 利用與管理:4
17. <input type="checkbox"/> 國立中央圖書館:3	17. <input type="checkbox"/> 大學:4	4. <input type="checkbox"/> 蒐集:5
18. <input type="checkbox"/> 圖書館利用教育:3	18. <input type="checkbox"/> 國立:4	4.1. <input type="checkbox"/> 商情資訊之蒐集:4
19. <input type="checkbox"/> 資料:3	19. <input type="checkbox"/> 資料:3	5. <input type="checkbox"/> 資訊:5
20. <input type="checkbox"/> 目錄:3	20. <input type="checkbox"/> 目錄:3	5.1. <input type="checkbox"/> 商情資訊之蒐集:4
		6. <input type="checkbox"/> 張淳淳:4
		7. <input type="checkbox"/> 大學:4
		8. <input type="checkbox"/> 國立:4

		8.1. <input type="checkbox"/> 國立中央圖書館:3 9. <input type="checkbox"/> 資料:3 10. <input type="checkbox"/> 目錄:3 10.1.
--	--	---

從查詢記錄檔中得知，該使用者隨即以滑鼠點選系統提供的相關詞「指導 大學 圖書館利用教育」作為新的查詢條件，亦即將「圖書館利用教育」的主題集中在大學圖書館上，並且由於圖書館文獻中，「圖書館利用教育」與「圖書館利用指導」的文字常交互出現，因此也將「指導」一詞列在新的查詢條件中，以輔助原查詢條件的不足。此項查詢部份結果如下：



上圖中有關大學圖書館的利用教育或利用指導相關的書目資料即排列在前面，供使用者參考。

伍、結語

傳統 OPAC 模組主要是以布林邏輯 (Boolean Logic) 模式提供書目查詢。然而一般使用者對布林邏輯的運用較不熟悉、檢索結果沒有依照符合程度排序、檢索字串要求精確無誤等原因，導致較高的檢索失敗率。目前網路通達的程度與普及速度，讓一般使用者皆可從家裡、辦公室、或任何方便的地方連線上網查詢資料，這與過去使用者必須到圖書館查資料，而圖書館裡有館員協助的情形大不相同。所以檢索技術的發展方向是逐漸把使用者端的檢索複雜度降低。本文期望新近的資訊檢索技術也能應用到圖書館的書目檢索上，因此運用非營利機構可獲得的檢索引擎，輔以自行發展的關鍵詞擷取技術，完成具備重要性排序、近似字串、模糊搜尋、相關詞回饋、允許近似自然語言檢索字串的 OPAC 系統。初步的評估發現，對協助使用者簡化查詢條件、拓展檢索字彙、提昇檢索成效有相當的助益。

未來除了希望能對系統的成效做更進一步的評估外，也期望持續發展資訊檢索技術，以提供使用者更多簡便有效的功能。

誌謝

本文的完成，感謝中央研究院資訊科學研究所簡立峰博士與李明哲先生提供 Csmart 軟體與諮詢，並感謝輔仁大學圖書館與資訊中心提供的書目資料、軟硬體設備、以及人力上的支援。

參考資料

- [1] 楊世緘，"NII 建設全力推動網際網路普及與應用"，
<http://www.dynavan.net.tw/www/magazine/data/00kq/0000007.htm>
- [2] 曾元顯，"WWW 技術在線上公用目錄的應用"，中國圖書館學會會訊，4 卷 4 期（103），85 年 12 月 31 日。Also available at <http://blue.lins.fju.edu.tw/~tseng/papers/wwwopac.htm>
- [3] Larson, R.R., "Evaluation of Advanced Retrieval Techniques in an Experimental Online Catalog," Journal of the American Society for Information Science, 1992, pp 34-53.
- [4] Hildreth, C.R., "Beyond Boolean: Designing the Next Generation of Online Catalogs," Library Trend, 1987 (35), pp. 647-667.
- [5] Cooper, W.S., "Getting Beyond Boole," Information Processing & Management, 1988 (23), pp. 243-248.
- [6] Borgman, C.L., "Why Are Online Catalogs Hard to Use? Lesson Learned from Information-Retrieval Studies," Journal of the American Society for Information Science, 1983, pp.387-400.
- [7] Patrick Hoffman, "Text Information Retrieval on the WEB - with emphasis on search engines, indexing, querying and visualization", <http://www.cs.uml.edu/shootout/papers/irsrch.html>
- [8] Gary Mooney and Robert John, "A Fuzzily Intelligent Information Retrieval Assistant," <http://www.dmu.ac.uk/~gjmooney/paris1.htm>
- [9] Lee-Feng Chien, "Fast and Quasi-Natural Language Search for Gigabytes of Chinese Texts," ACM SIGIR '95, 1995
- [10] Welcome to Lycos, <http://www.lycos.com/>
- [11] GAISWWW 亞太 WWW 資源搜尋引擎，<http://gais.cs.ccu.edu.tw/www2-adv.html>
- [12] Csmart:網路中文資源檢索系統，<http://csmart.iis.sinica.edu.tw/>
- [13] DoszKocs, T.E., "CITE NLM: Natural Language Searching in an Online Catalog," Information Technology and Libraries, 1983 (2), pp. 364-380.
- [14] Larson, R.R., "Evaluation of Advanced Retrieval Techniques in an Experimental Online Catalog," Journal of the American Society for Information Science, 1992 (43), pp. 34-53.
- [15] 卜小蝶，"Fuzzy Search 技術在中文 OPAC 的應用"，中國圖書館學會會報，第 56 期，85 年 6 月，頁 77—86。
- [16] 曾元顯，"架構在 WWW 上的分散式線上公用目錄系統"，海峽兩岸圖書館事業研討會論文集，86 年 5 月 26-28 日，頁 263-277。
- [17] Gerard Salton, "Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer" Addison-Wesley, 1989.

- [18] William B. Frakes, Ricardo Baeza-Yates, "Information Retrieval: Data Structure and Algorithm," Prentice Hall, 1992.
- [19] Keith, "Information Retrieval" <http://www.dcs.gla.ac.uk/Keith/>
- [20] Christos Faloutsos and Douglas Oard, "A Survey of Information Retrieval and Filtering Methods," The URL of this paper can be found at <http://www.cs.jhu.edu/~weiss/papers.html>
- [21] Hsinchun Chen and Jinwoo Kim, "GANNET: Information Retrieval Using Genetic Algorithms and Neural Nets," IEEE Transactions on Neural Networks, 1994. Also available at <http://ai.bpa.arizona.edu/papers/gannet93/gannet93.html>
- [22] Victo Zue, '人機對談介面', 中國時報 47 版節錄, 86 年 6 月 10 日。
- [23] Yuen-Hsien Tseng, "Fast Keyword Extraction for Chinese Documents in a Web Environment," submitted to Information Retrieval WorkShop for Asian Languages -1997.