

天主教輔仁大學圖書資訊學系碩士班碩士論文

指導老師：曾元顯

中文互動式檢索輔助功能之效益評估

－以關聯提示詞為例

研究生：葉佳昀 撰

中華民國九十三年六月

## 摘要

互動式檢索輔助功能在資訊檢索系統的應用越來越普及，例如國內 Openfind、Tornado、WebGenie 檢索系統，都提供「關聯詞提示」的功能，以協助使用者下達準確的查詢詞，進而獲得高品質的檢索結果。本論文利用既有的關聯提示詞自動擷取與提示功能，以量化與質化的方式來探討其應用在資訊檢索系統的成效，期望本研究探討出來的結果與使用者的回饋，可以提供系統未來改進的參考，設計出符合使用者需求的檢索輔助功能。因此，本研究分別從『系統的角度』及『使用研究』兩方面進行實驗，在『系統的角度』方面，針對關聯提示詞的品質與文件量大小的關係進行探討與分析；在『使用研究』方面，利用實際的檢索任務來探討檢索者對關聯提示詞圖示功能的使用情形。各實驗結果彙整如下：1. 實驗結果發現文件量的大小的確會影響關聯詞的相關程度，但不是呈現出絕對正比的關係，因為關聯詞的品質還是會因文件資料的特性、文件的年代等其它因素所影響。2 在『使用評估方面』則以 TREC Interactive Track 的設計為原則，分別從『效能』、『效率』及『使用者滿意度』，再輔以『介面的可用性』來評估使用者實際使用互動式檢索功能的情形，從評估結果得知，在『效能』方面檢索者均反映關聯提示詞能夠提示檢索者找到『所預期不到的結果』；在『效率』方面則是認為關聯提示詞雖不能有效縮短查詢的時間，但是卻可以提高檢索結果的品質；至於在『使用者滿意度』及『介面的可用性』方面，所有檢索者均認為此關聯提示詞圖示功能的確能夠提高檢索結果的品質，但必須克服檢索者的使用『習慣』。最後，本實驗驗證了文件量與關聯詞的成效，且也試圖評估利用二維圖示的視覺化技術來展現關聯詞的可用性。

# **An Evaluation of Interactive Searches Based on Related Term Suggestion for Chinese Documents**

## **Abstract**

Interactive search-aide tools are now becoming more and more popular. Major search engine vendors, such as Openfind, Tornado, and WebGenie are providing tools such as “related terms” in response to users’ query terms. These “related terms” suggest more topic-related terms for use in a second search in the hope to lead to better results. However, such search-aide functions are rarely evaluated in a rigorous manner. Based on a self-developed technique that can extract and suggest related terms in 2-D map, the goal of this research is to evaluate this “related terms in 2-D map” in an interactive search environment. By exploiting qualitative and quantitative research methods, we designed effective approaches to evaluate how the quality of the suggested related terms varies with the quantities of the documents and how effective is it to apply this search-aide function in an interactive search environment. It is hope that our results can help system designers to improve the performance of their search-aide functions. In our research, we applied two approaches: one is system-oriented and the other is user-oriented. Some of the experiment results are summarized as follows: 1. The quality of the related terms is affected by the number of indexed documents, but the relatedness of the related terms does not always increase when more documents are indexed. For example, in our experiment, the relatedness percentage does not obviously increase when 154,720 Chinese news articles are increased to 406,593 articles. The reason is that the related term is influenced by other factors such as the timestamp and the type of the documents. 2. In our user evaluation experiment, we follow the TREC interactive-track’s discipline to evaluate search effectiveness, efficiency and user satisfaction. For the effectiveness evaluation, the experiment results showed that all searchers could find extra related documents through the use of the related terms in 2-D map. As to efficiency, although the 2-D map could not save the search time effectively, it could help the searcher to find more documents and get better results. For user satisfaction, all searchers thought that the 2-D map could help people who wanted to search something which they did not know. Finally, we tried to integrated usability evaluation to evaluate the text retrieval system which used the information visualization techniques to render the term relations.

## 目次

<b>第一章 緒論</b> .....	<b>1</b>
第一節 研究動機 .....	1
第二節 研究目的 .....	3
第三節 研究問題 .....	4
第四節 研究範圍與限制 .....	4
第五節 詞彙定義 .....	5
第六節 預期貢獻 .....	7
<b>第二章 文獻探討</b> .....	<b>8</b>
第一節 相關判斷 .....	8
第二節 檢索評估 .....	13
第三節 互動式輔助檢索之設計與評估 .....	16
<b>第三章 研究方法與設計</b> .....	<b>46</b>
第一節 研究流程與架構 .....	47
第二節 研究對象 .....	49
第三節 研究設計 .....	49
第四節 研究方法 .....	56
第五節 研究工具 .....	58
第六節 資料之分析與處理 .....	61
<b>第四章 實驗結果分析</b> .....	<b>64</b>
第一節 關聯提示詞之相關評估結果分析 .....	64
第二節 互動式檢索輔助功能評估實驗分析 .....	73
第三節 綜合分析與討論 .....	105
<b>第五章 結論與建議</b> .....	<b>111</b>
第一節 結論 .....	111
第二節 建議 .....	114
<b>參考書目</b> .....	<b>118</b>
附錄一 檢索任務(Tasks).....	122
附錄二 基本資料問卷(Entry-Questionnaire).....	124
附錄三 檢索任務後問卷(Post-Search Questionnaire).....	128
附錄四 單一檢索系統後問卷(Post-System Questionnaire).....	130
附錄五 檢索實驗後問卷(Exit-Search Questionnaire).....	131
附錄六 可用性訪談大綱及問題(Interview).....	133

## 圖目次

圖 1-1	Vivisimo 自動歸類	1
圖 1-2	ET-Map	2
圖 1-3	Openfind 即時動態相關詞查詢	2
圖 1-4	互動式檢索功能之關聯提示詞	5
圖 2-1	TREC 相關判斷結果範例	9
圖 2-2	TREC 2002 interactive track experiments 的檢索任務	20
圖 2-3	HuddleSearch 的互動式檢索結介面	24
圖 2-4	MDD 允許長字串檢索之介面	28
圖 2-5	SDD 不允許長字串檢索之介面	28
圖 2-6	實驗設計之第一群 16 位實驗對象的檢索任務分配	29
圖 2-7	Keyphind user interface	33
圖 2-8	難易度比較性問題的統計數字列表	35
圖 2-9	Phind 的瀏覽介面	37
圖 3-1	研究流程與架構	47
圖 3-2	文件標記範例	52
圖 3-3	.csv 的檔案格式	53
圖 3-4	一篇文件及其關聯詞	59
圖 3-5	WebGenie 知識檢索系統介面	60

## 表目次

表 2-1 查全率查準率和雜訊比之 2 乘 2 表格	14
表 2-2 GOV 資料集統計數字	19
表 2-3 TREC 互動式評估實驗進行方式	21
表 2-4 描述使用者搜尋行為的各項變數列表及定義	30
表 2-5 每項檢索任務完成之後的問題	34
表 2-6 最終的訪談問題	34
表 2-7 互動式輔助檢索功能的設計	41
表 2-8 互動式檢索評估方式-評估實驗設計之研究方法	42
表 2-9 非參與 TREC Interactive Track 的評估實驗設計	43
表 2-10 互動式檢索評估方式-分析方式	44
表 2-11 各輔助檢索功能的檢索成效綜合分析	45
表 3-1 CIRB011	51
表 3-2 NTCIR 中跨語言檢索的 CIRB011 & CIRB020 文件集	51
表 3-3 涵蓋面評估先做系統 A 再做系統 B 的前 8 人檢索任務分配	54
表 3-4 涵蓋面評估先做系統 B 再做系統 A 的後 8 人檢索任務分配	55
表 3-5 探索性評估的檢索任務分配	55
表 3-6 之前實驗的 30 個檢索詞彙	56
表 4-1 前 50 個提示出的詞彙被判斷為相關的相關比例	65
表 4-2 各文件集之查詢詞的平均查準率統計表	65
表 4-3 小文件集與中文集的敘述統計量表	66
表 4-4 小文件集與中文集的觀察值	66
表 4-5 小文件集與中文集的配對樣本 T 檢定表	66
表 4-6 中文件集與大文集的敘述統計量表	66
表 4-7 中文件集與大文集的觀察值	66
表 4-8 中文件集與大文集的配對樣本 T 檢定表	67
表 4-9 平均查準率呈現降低的查詢詞彙列表	68
表 4-10 查詢詞『會計師』在各文件集的 trec_eval 輸出	68
表 4-11 查詢詞『國安基金』在各文件集的 trec_eval 輸出	70
表 4-12 查詢詞的平均查準率之中文件集優於大文件集	71
表 4-13 查詢詞『國科會』在三個文件集的 trec_eval 輸出	72
表 4-14 非圖書資訊學領域之背景之受試者先從 A 系統進行實驗	73
表 4-15 圖書資訊學領域之背景之受試者先從 A 系統進行實驗	74
表 4-16 非圖書資訊學領域之背景之受試者先從 B 系統進行實驗	74
表 4-17 圖書資訊學領域之背景之受試者先從 B 系統進行實驗	74
表 4-18 具備圖書資訊學背景的電腦技能統計量表	75
表 4-19 非圖書資訊學背景的電腦技能統計量表	75

表 4-20 具備圖書資訊學背景的檢索技能統計量表	76
表 4-21 非圖書資訊學背景的檢索技能統計量表	77
表 4-22 圖資背景在檢索任務的熟悉程度	78
表 4-23 非圖資背景在檢索任務的熟悉程度	78
表 4-24 圖資背景在檢索任務的專業程度	79
表 4-25 非圖資背景在檢索任務的專業程度	79
表 4-26 圖資背景分別在兩系統之檢索結果的信心分數	80
表 4-27 非圖資背景分別在兩系統之檢索結果的信心分數	81
表 4-28 具備圖資背景的受試者在資訊量與回答之答案的信心統計表	82
表 4-29 非圖資背景的受試者在資訊量與回答之答案的信心統計表	83
表 4-30 關聯提示詞圖示之功能對檢索的幫助程度比較表	84
表 4-31 利用 A 系統執行檢索任務 1 至 4 所需的時間	85
表 4-32 利用 B 系統執行檢索任務 1 至 4 所需的時間	85
表 4-33 利用 A 系統執行檢索任務 5 至 8 所需的時間	85
表 4-34 利用 B 系統執行檢索任務 5 至 8 所需的時間	86
表 4-35 各項任務完成所需之時間敘述統計量表	86
表 4-36 具備圖書資訊學背景的受試者在 A 系統各項滿意度	88
表 4-37 具備圖書資訊學背景的受試者在 B 系統的各項滿意度	88
表 4-38 非圖書資訊學背景的受試者在 A 系統各項滿意度	89
表 4-39 非圖書資訊學背景的受試者在 B 系統各項滿意度	90
表 4-40 8 位具備圖書資訊學背景之受試者的開放式問題答案	91
表 4-41 具備圖書資訊學背景之受試者的檢索實驗後問卷統計表	92
表 4-42 具備圖書資訊學背景之受試者的次數分配表	92
表 4-43 8 位非圖書資訊學背景之受試者的開放式問題答案	93
表 4-44 非圖書資訊學背景之受試者的檢索實驗後問卷統計表	93
表 4-45 非圖書資訊學背景之受試者的次數分配表	93
表 4-46 16 位受試者的檢索實驗後問卷統計表	94
表 4-47 16 位受試者的次數分配表	94
表 4-48 A 系統資訊檢索能力與檢索結果滿意度相關表	95
表 4-49 B 系統資訊檢索能力與檢索結果滿意度相關表	96
表 4-50 A 系統資訊檢索能力與容易操作相關檢定表	96
表 4-51 B 系統資訊檢索能力與容易操作相關檢定表	97

# 第一章 緒論

## 第一節 研究動機

資訊檢索系統發展至今可謂是處於百家爭鳴之局面，環顧各家系統無一不是為了幫助檢索者進行搜尋與過濾資訊。綜觀各家檢索系統輔助檢索者的功能，大致可從二部份著手，一是在檢索前的輔助操作，而其功能從介面上來說，可以一般文字檢索系統中常見的欄位式檢索、全文檢索及關鍵詞等功能，再輔以其他限定條件的欄位來幫助檢索者下檢索詞彙。另一種則是檢索時與使用者互動之提示功能，發展至今可以說是融合許多技術，如：自動關聯提示詞、自動分類、文件分群及自動索引典等技術，這些核心技術都可以看到應用在國內外各大檢索系統，如 Vivisimo 的自動歸類功能（見圖 1-1）、亞歷桑大學人工智慧實驗室研發的娛樂地圖文件分群技術及（見圖 1-2）國內知名搜尋引擎網擎(Openfind)的動態關聯提示詞（見圖 1-3）等系統。由此可知，互動式檢索功能的開發對檢索系統的發展將產生出很大的影響，且隨著各項資訊檢索技術的成熟，如何選擇一個能夠有效輔助檢索者的系統，更是當前資訊檢索領域中重要的課題

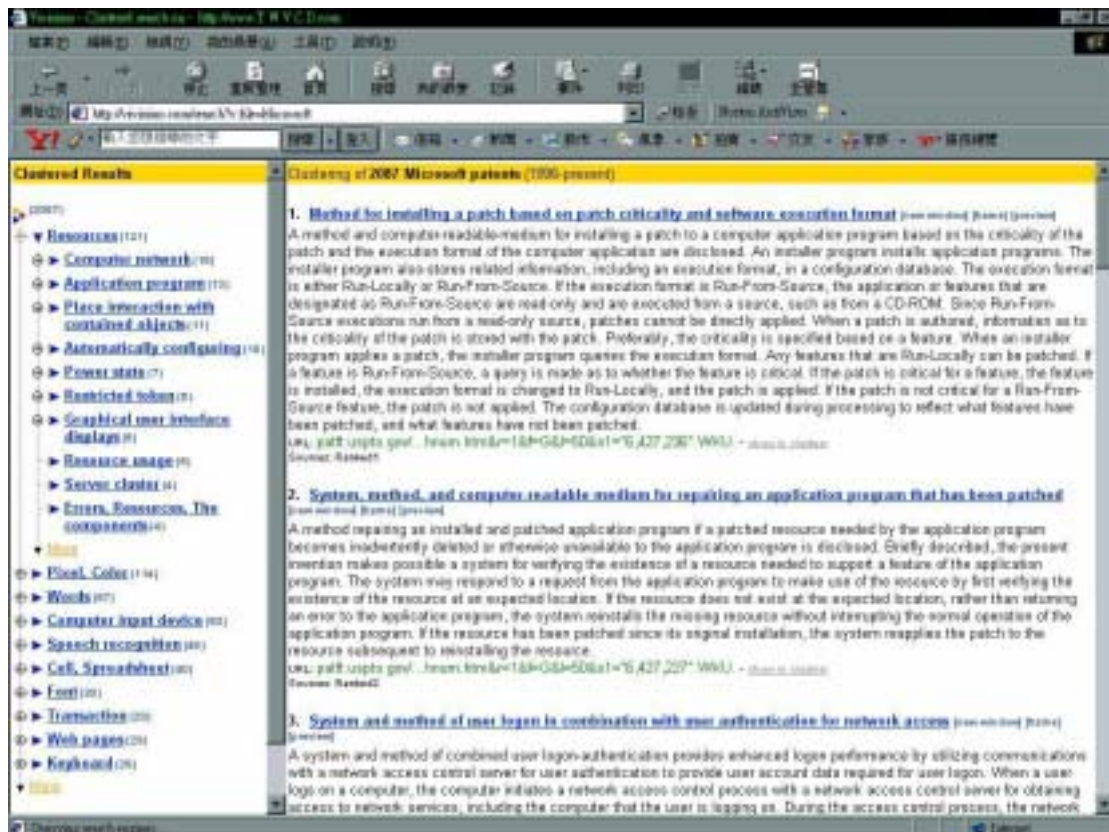


圖 1-1 : Vivisimo 自動歸類



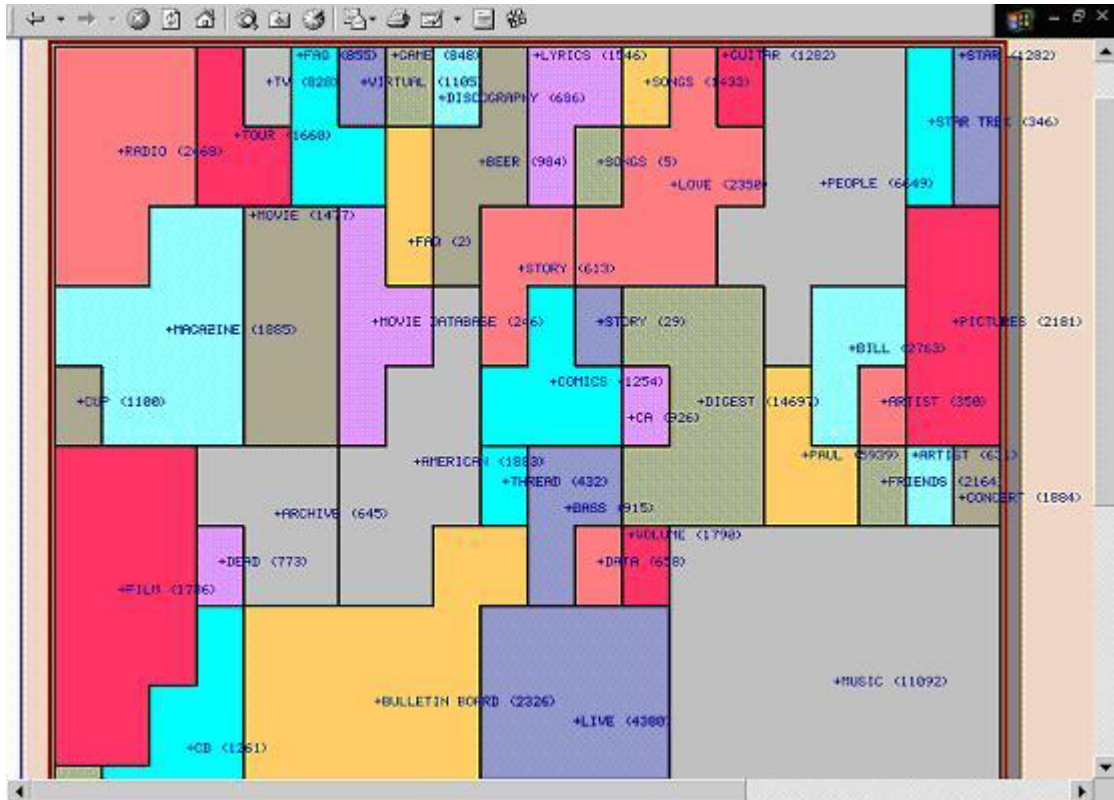


圖 1-2 : ET-Map

資料來源：<http://ai2.bpa.arizona.edu/ent/entertain1/>



圖 1-3 : Openfind 即時動態相關詞查詢

相對於本研究的互動式檢索功能即是關聯提示詞圖示，而本實驗之關聯提示詞即為曾元顯研發之互動式檢索功能，為了讓檢索系統更趨完善且符合系統發展中使用者導向的需求，特別設計一套評估的方法，解決「系統的設計者常常不是系統使用者，系統的使用挫折和不便常常無法為系統設計者所洞悉」<sup>1</sup>的問題。

## 第二節 研究目的

根據 Jansen 及 Silverstein 在 1998 年的分析報告指出，大多數的使用者在搜尋引擎所下的檢索詞彙大多不超過 2.3 個字，且國內學者卜小蝶也曾經分析台灣網路搜尋引擎的檢索詞彙記錄，初步了解台灣地區使用者具有檢索詞彙簡短、存在核心詞彙及檢索專有名詞比例高等特性<sup>2</sup>。因此，為了解決檢索者所下之檢索詞彙簡短的問題，相關詞提示(Term Suggestion)即成為相當重要的一種互動式檢索功能，而自動關聯詞提示更是利用詞彙共現(Co-Occurrence)的方式，來模擬人類建構的索引典。所以關聯提示詞的成效與品質對檢索者的檢索結果將有很大的影響。本實驗特別利用不同文件集的大小，探討是否會影響關聯提示詞的關聯程度，藉此找出除了影響詞彙提示的核心技術之外，是否有相關的因素會影響其成效。

而資訊檢索另一重要的領域就是資訊檢索系統的評估，評估的趨勢慢慢從量化評估逐漸趨向大融合，也就是說愈來愈注重檢索系統與人互動之後的評估。以往資訊檢索者在執行檢索任務時，必須屈就於系統的設計，導致心理的認知與系統的索引詞彙不符合，且查詢不到相關的文件，進而造成系統的成效不彰。此外，資訊檢索系統也因為加入「人的因素」使得評估方式變得沒有一套公用的準則。因此，本實驗嘗試設計一套符合互動式檢索評估的環境，針對互動式檢索功能中的關聯提示詞進行評估，以提供系統設計及改善的考量。

具體目的如下：

- 一、利用文獻分析探討現行國內外互動式檢索功能之資訊檢索系統的設計與評估方式，進而設計出一套適合互動式評估方法的準則與依據。
- 二、分析探討是否因為資料量(文件集)的大小而影響其自動產生之關聯提示詞的分佈。除了文件集的大小會影響其成效之外，是否還有其他因素會影響智慧型關聯詞的成效，如：新聞事件因為有短暫且集中報導的趨勢，因此其關聯詞的相關程度可能會受此影響。

<sup>1</sup> 吳美美，「試論資訊檢索理論」，王振鵠教授七秩榮慶祝壽論文集編輯小組，當代圖書館事業論集：慶祝王振鵠教授七秩榮慶論文集(台北市：中正，民 83)，頁 745。

<sup>2</sup> 卜小蝶，「網路使用者查詢語彙紀錄之分析與應用初探」，資訊傳播與圖書館學 6 卷 2 期 (民 88)，頁 49-62。

### 第三節 研究問題

面對資訊量的日益成長，各種檢索系統之互動式檢索功能的輔助，是否能夠有效幫助檢索者在浩瀚的資訊中滿足其資訊需求，而本研究以輔仁大學圖書資訊學系曾元顯研發的互動式檢索功能作為實驗的系統，探討其關聯提示詞的功能是否能夠降低檢索者的資訊負擔，實驗中將結合量化與互動式的評估方法，分別從系統的角度與使用研究來探討下列各項問題：

- 一、關聯提示詞的成效（系統角度）
  1. 文件集的大小是否會影響關聯提示詞的成效與品質？
  
- 二、互動式檢索輔助功能的評估（使用研究）？
  1. 檢索者根據指派的任務對關聯提示詞圖示功能
    - 認知情況？
    - 滿意度？
  2. 未來欲改善或美化該介面時，有哪些需要加強之處？

### 第四節 研究範圍與限制

由於國內外互動式檢索功能種類繁多，且應用在文字檢索系統的展現形式也不盡相同，因此，本實驗只針對關聯提示詞圖示之功能進行成效評估，而評估的方式為了模擬一般的檢索環境，特別利用不同任務的指派來進行檢索者的使用評估，期望利用上述的方式來收集檢索者與系統間互動後的回饋。以下是本研究的範圍與限制：

本研究的範圍與限制如下：

- 一、本研究所提及的互動式檢索功能，主要是以輔仁大學曾元顯研發的關聯提示詞之功能為主。
- 二、研究中的研究對象受限於各領域不同而會產生判斷相關詞的偏差，且本研究的目的之一是在探討關聯提示詞的相關程度，因此為了讓實驗的相關因素有一致性，且能夠看出其檢索行為的特性與傾向，我們只將實驗對象分為一般使用者(非圖書資訊學系領域)和主要使用者(具有圖書資訊學領域的人)。
- 三、由於實驗之資料集以新聞資料為主，且受限於年代的限制，所以在檢索任務的設計必須遷就於資料集的特性，無法涵蓋所有的事件。

四、研究對象中資訊檢索經驗的限制,主要是是否具有圖書資訊學領域背景來分類,因為圖書資訊學領域的學生曾經修過資訊檢索的相關課程,對線上檢索系統有一定程度的認知。

## 第五節 詞彙定義

### 一、互動式檢索功能 ( Interactive Search ):

本研究之互動式檢索功能定義,為檢索者在檢索過程中或結束後一個輔助檢索者對檢索詞彙做進一步修正或擴展的功能,而檢索者可以根據系統自動提示之詞彙重新查詢以提高檢索之效能。此外,在 TREC Interactive-Track 上參加評估的系統之互動式檢索功能,包括:自動歸類、自動摘要及自動關聯提示詞等。本研究之互動式檢索功能即為關聯提示詞圖示,關聯提示詞則利用 2D 的視覺化技術來展示與檢索詞最為相關之 20 個關聯詞,關聯詞與檢索詞的距離則為兩詞彙間的關連程度,關連程度愈高則與檢索詞的距離愈短,此功能也可以顯示出關聯詞的關聯名次。如(圖 1-4),當檢索者輸入檢索詞彙「阿扁」時,系統會自動提示關聯名次為第一名的「陳水扁」、第二名「民進黨」及第三名「許信良」等關聯提示詞。而圖 1-4 中關聯提示詞圖示正是本實驗互動式檢索功能的一種。



圖 1-4：互動式檢索功能之關聯提示詞

## 二、詞彙擴展(Query Expansion)：

詞彙擴展對資訊檢索領域來說，是一種幫助檢索者在檢索過程中可以修正檢索詞彙且降低檢索者在檢索系統中繁雜的步驟，如：重新查詢等。而詞彙擴展依產生來源的方式，粗略可分為人工和自動的擴展方式，「人工」就是傳統圖書館中索引典的建構及主題分析等；「自動」又可分為「檢索詞擴展」(Query Expansion)與「相關詞提示」(Term Suggestion)，前者擴展的詞是系統自動加入，後者則是透過與使用者的互動所產生。

## 三、關聯提示詞(Related Terms)：

本研究之關聯提示詞有別於一般人工建構的索引典，是採取自動建構的方法，即是計算單一文件中任意兩個詞彙個別出現與共同出現的句子數，再套用資訊檢索常用的相似度計算公式 Dice，求得兩詞彙之間的關聯度。計算出每一篇文件重要詞彙之間的關聯度後，累積關聯強度超過某個門檻值的關聯詞，即可完成整個文件資料庫的關聯詞庫。<sup>3</sup>

而在本實驗的操作型定義，就是經由系統主動提示檢索者相關的資訊，再經由檢索者從提示出的建議回饋給系統，經過一連串的與系統互動之後達到輔助檢索者表達其資訊需求，而關聯提示詞即是其中的一種互動式輔助檢索功能。

## 四、TREC Interactive Track：

TREC 除了測試檢索結果之外，也將檢索中人與系統互動的過程考慮在內；TREC 每年均發展出不同的測試程序，嘗試找出一個最佳的評估互動系統之方法，以作為系統互動設計的參考，例如讓使用者在某特定的時間中，儘可能檢索出最多的相關文章。TREC 至今似乎仍未發展出理想的評估方式，有些人則認為這些測試方法並沒有考慮到使用者的特質和認知對檢索結果所可能產生的影響。<sup>4</sup>

## 五、測試集(Test Collection)：

所謂的測試集，是指在規範環境中測試系統效能的機制，包括測試

<sup>3</sup> 曾元顯，「共現索引典之自動建構、評估與應用」，台灣大學圖書資訊學系四十週年系慶研討會，(民90年11月)，頁87-105。

<sup>4</sup> 江玉婷，陳光華，「TREC 現況及其對資訊檢索研究之影響」，政治大學圖書與資訊學刊 29 期(民國88年5月)，頁36-59。

問題(Queries)、測試文件集(Document Set)及相關判度(Relevance Assessment)等。<sup>5</sup>

## 第六節 預期貢獻

在傳統的資訊檢索系統中，使用者經常扮演一個被動的角色，使用者與系統之間的互動常顯得不足，也正因為如此，本實驗特別利用自動關聯提示詞與檢索者互動，藉由此互動之功能來修正及擴展檢索詞彙，提高使用者的檢索成效，降低資訊溢檢率與檢索的失敗率，且在增進檢索效能之餘，期望利用評估的方式反映出與檢索者互動的情形。再者，也可作為進一步發展線上檢索系統之評估方式的改進參考，具有實用與研究的價值。

此外，本研究嘗試結合量化之相關判斷及互動式之評估研究方法，對本實驗系統之關聯提示詞的互動式功能進行評估，希望藉由上述研究所獲得的分析結果，使本研究之互動式檢索功能更趨完善，進而提升資訊服務的品質。是故，研究者希望透過本研究的發現與分析，能夠達到以下的貢獻：

- 一、瞭解關聯詞的品質與文件量的相關程度
- 二、探討影響自動關聯提示詞建構之相關因素。
- 三、透過使用研究(Use Study)，瞭解相關詞提示(Term Suggestion)之效用及可能之改進方法(對哪些人有效，對哪些檢索任務有效)。
- 四、提供進一步研究互動式檢索功能評估方式的相關領域研究者之參考。

---

<sup>5</sup> 同前註，頁 2。

## 第二章 文獻探討

本章節的文獻分析將以本研究相關之實驗設計與研究方法為主軸，依序分析如下：第一節針對資訊檢索仰賴的量化評估方法-「相關的判斷」進行分析，且利用文獻的分析方式，簡短介紹 TREC 相關判斷的方法；第二節則針對影響互動式資訊檢索系統評估的因素做廣泛探討與分析；第三節則針對現今有提供互動式檢索功能的檢索系統，以文獻分析的方式將參與 TREC Interactive Track 的互動式檢索系統做個綜合的分析比較及介紹；再者，針對這些系統建構完成後，探討其設計的評估方式。

### 第一節 相關判斷

在資訊檢索領域中，資訊檢索系統的評估對系統的設計、發展與研究一直有顯著的關係。從 1950 年代末期 Cleverdon 進行的 Cranfield 計畫開始，他開啟了以測試集(Text Collection)配合測量基準(Measures)的評估模式，而測試集則包括測試問題(Queries)、測試文件集(Document Set)及相關判斷(Relevance Assessment)等。<sup>6</sup>測量基準主要是依據查詢問句與檢索出文件集間的相關程度，將不相關的文件過濾掉。因此，採用查準率與查全率來作為測量的基準，其規範化的評估模式亦成為後世學習的標竿。

#### 一、 TREC 的相關判斷：

由此可見，相關判斷一直是資訊檢索領域採用的評估方式，也是測試集在建構過程中最為困難 最花時間的，但它卻是最重要的一部分。早期的Cranfield 研究規模雖不大，但其相關判斷卻進行了50 餘萬次。對於 TREC 這樣的大型測試集來說，要將每個主題在文件集中進行詳盡的相關判斷，所須耗費的工程可見一般。 TREC 採用二元化的尺度，也就是只分為相關與不相關，而相關判斷主要是根據主題的 Narrative 欄位進行。對相關與否的評估原則，為只要文件部分與主題相關即可(即使只有數句)，並不要求文件的每個部分皆與主題相關。

TREC 的文件組之文件數量十分巨大，若要逐一將文件組中每篇文件與每個查詢主題做詳盡的相關判斷，所費之時間與人力是非常驚人的。故TREC採用 Pooling的方法，來輔助相關判斷的進行。此方法是參與測試評估之資訊檢索，其檢索結果以文件與查詢主題的相關程度排序後，以數個系統，針對某查詢主題，在文件組內做查詢，並抽取各系統查詢結果的前 n 篇(通常為100篇)文件，合併成一個 Pool。去除 Pool 中重覆之文件後，此 Pool 就作為該查詢主題之相關文件候選組，送回給該查詢主題之建構者進行相關判斷。在 Pool 中各文件是依文件編號排序，而非依各系統所給與的相關程度排序，主要是為了避免影響

---

<sup>6</sup> 同註 4，頁 2。

相關判斷之結果。而未被列入查詢主題相關文件候選組中的文件，則一律假設為與查詢主題不相關。<sup>789</sup>

在查詢主題之建構者對各查詢主題相關文件候選組完成相關判斷，做出相關判斷結果後，可列表整理查詢主題與文件之相關關係，此列表在結構上包含下列四個欄位(見圖 2-1)：<sup>10</sup>

- 1 Topic：查詢主題編號。
- 2 Iteration：目前不使用，此欄位內容均為 0。
- 3 Document#：文件之編號。
- 4 Relevancy：值為 0 時，表示該文件與主題不相關，若為 1 則表示相關。

1	0	AP880212-0161	0
1	0	AP880216-0139	1
1	0	AP880216-0169	0
1	0	AP880217-0026	0
1	0	AP880217-0030	0

圖 2-1：TREC 相關判斷結果範例

資料來源：[http://trec.nist.gov/data/qrels\\_eng/index.html](http://trec.nist.gov/data/qrels_eng/index.html)

TREC 相關判斷結果，如同圖 2-1 之標準答案，指出每個查詢主題，有哪些文件是與其存在相關的關係。故資訊檢索系統的檢索結果，可與相關判斷結果來做比較，以查準率等指標來評估系統之成效。由此可見，縱使在現今倡導使用者評估(效用)的趨勢，「相關的判斷」依然是評估檢索系統的唯一準則。

<sup>7</sup> 同註 4，頁 43-44。

<sup>8</sup> N.J. Belkin, C. Cool\*, D. Kelly, G. Kim, J.-Y. Kim, H.-J. Lee, G. Muresan, M.-C. Tang, X.-J. Yuan (2002). "Rutgers Interactive Track at TREC 2002." Retrieved December 4, 2003, from <http://trec.nist.gov/pubs/trec11/papers/rutgers.belkin.pdf>

<sup>9</sup> Ellen M. Voorhees and Donna Harman. "Overview of the Ninth Text Retrieval Conference (TREC-9)," Retrieved December 4, 2003, [http://trec.nist.gov/pubs/trec9/papers/overview\\_9.pdf](http://trec.nist.gov/pubs/trec9/papers/overview_9.pdf).

<sup>10</sup> "Text Retrieval Conference (TREC) Data – English Relevance Judgments Files List," Retrieved December 4, 2003, from [http://trec.nist.gov/data/qrels\\_eng/index.html](http://trec.nist.gov/data/qrels_eng/index.html).



## 二、 相關判斷應用於資訊檢索領域：

### (一) 影響相關判斷之因素：

一般而言，影響相關判斷的五大因素，仍以 Saracevic 所提出的五大因素為架構，分別是：1. 文件(包含文件與文件的陳述)，2. 檢索問題的陳述，3. 判斷情境，4. 判斷的尺度，5. 判斷者等。<sup>11</sup>其後，學者所提出的影響因素大都不脫離上述五大因素的範疇內，只是分析的方式及研究的變數會有所增減。下列綜合國內外學者先前之研究，對五大因素做一個簡略的陳述：

#### 1. 文件(包含文件與文件的陳述)：

大致上來說，較特定及專指的文件內容，較容易產生出相同的相關判斷結果。但是，也可能受到下列研究變數影響，分別是：主題、內容差異、難易程度、學科的艱深程度、資訊量、資訊濃縮程度、文字屬性等變數。

#### 2. 檢索問題的陳述：

檢索問題的陳述愈清楚，相關判斷的結果愈一致；而影響檢索問題的陳述之變數，分別為：主題、內容重點、難易程度、專指性及資訊量、含混性等。

#### 3. 判斷的情境：

影響資訊檢索系統的判斷情境通常包含下列變數，如：時間壓力、文件的排序、資料的筆數、文件寬度、控制組的存在、判斷一致的壓力、專指性及相關的定義等。其中又以文件的排序為最多學者所研究，Eisenberg 和 Barry 指出相關程度由高至低或由低至高排列時，讀者會產生不同的判斷結果，這種效果稱為「賭注效應」(hedging effect)。<sup>12</sup>而國內學者曾元顯在其關聯詞相關性評估的實驗中，也曾經利用「詞頻」、「時間」、「強度」的排序方式，來探討其關聯詞的分佈情形。

#### 4. 判斷的尺度：

<sup>11</sup> 黃慕萱，資訊檢索中「相關」概念之研究（台北市：台灣學生，民 85 年），頁 71-78。

<sup>12</sup> Michael Eisenberg and Carol Barry, "Order Effects: A study of the Possible Influence of Presentation Order on User Judgments of Document Relevance," Journal of the American Society for Information Science 39:5 (1988) : 293-300.

採用不同的種類之測量尺度，例如二元化(相關、不相關)、順序尺度(相關、部分相關、不相關)，及開放式之連續尺度，會產生出不同之相關判斷結果。

#### 5. 判斷者：

不同背景的相關判斷者，會產生出不一致的判斷結果，就算相同背景的相關判斷者也會因為其他因素而產生出不一致之判斷結果。而這些因素包含：知識(經驗)、智力、認知狀態、偏見、判斷經驗、判斷態度、相關概念、使用者導向、預期值及錯誤傾向等。

由上述五項因素可得知，相關判斷雖然是較具主觀性的決定，但是，許多非主觀性的因素也會影響讀者之相關判斷，故良好的相關判斷應該是結合主題相關、情境相關及心理相關的各項層面，設計出一個判斷相關的決策過程。

### 三、「相關判斷者」與「文章排序」之研究：

本實驗的設計特別針對不同之相關判斷者及不同文件集之排序方式，對互動式檢索系統的影響做探討。因此將列舉相關之研究簡述之。

前一節曾經提及「文件排序」會影響相關判斷的結果，如國內學者曾元顯在其關聯提示詞評估中，將自動產生的關聯提示詞，分為「詞頻」、「時間」、「強度」的三種排序方式，並就每個查詢詞彙，檢視前 50 個提示詞，判斷與其查詢詞的相關性。而實驗結果也指出，判斷的相關情況為強度佔 69%，其次是時間排序為 59%，依篇數排序比例為 48%。由此可知，不同的排序方式會造成不同詞彙分佈的結果。是故，如果只是單純為了降低文章排序對相關判斷之影響，不宜將檢索結果依其相關程度排列，最好將檢索結果隨機呈現。

此外，從經濟學的角度上來說，文件排序對相關判斷的影響，非常符合「邊際效益遞減之原則」(diminishing returns principle)；如果前面幾篇文章的相關程度很高，後面等同之相關文章的相關程度往往會被低估，就好比天天吃大魚大肉，數星期之後，人們對其喜好的程度一定會降低。<sup>13</sup>在上述研究中，充分顯示出相關判斷的多樣性和不確定性，也因為資訊檢索系統的使用者導向之趨勢，讓許多系統研發者嘗試找出最適合使用者心目中的文件排序方式，畢竟，相關的決定是由終端使用者主觀的判斷所產生，文件排序的方式應該還是由使用者來決定才是最適當的方法。

---

<sup>13</sup> 同註 11，頁 78-79。

另外一個「判斷者」對相關判斷的影響層面，國內學者黃慕萱曾經在其著作中指出早期的研究可以說是非常缺乏，較知名的為 Barhydt 和 Figueiredo 分別在 1976 及 1978 年所做的研究，而 20 多年只被引用過 2 次可以得知此問題不受重視。不過值得一提的是 Barhydt 在其結論中指出：「使用者才是相關判斷的主人」。時至今日，Barhydt 的論點才重新被大家所重視。

Janes 在 1994 年的研究，比較其他判斷者和資訊需求者間相關判斷之差異，但由於資訊需求者是最具資格進行相關判斷的人，因此研究方向擺在比較提供資訊服務的館員和資訊需求者之間相關判斷結果之異同。在此研究中，Janes 將次判斷者的族群分為館員、較具經驗之圖資系學生及初入學的圖資系學生。他發現次判斷者都傾向給予較高的相關判斷之結果，尤其是初入學的圖資系學生。在相關判斷的一致性上，館員的表現比較具經驗之圖資系學生為佳。其研究指出判斷者的專業訓練和工作經驗都有助於提高相關判斷的品質。<sup>1415</sup>

#### 四、 小結：

綜合各學者研究顯示，影響相關判斷的結果有許多「情境相關」的因素，這些因素包含：檢索者之電腦經驗及資訊檢索經驗、對檢索主題的熟悉程度、檢索問題答案的認知(過與不及的資訊量)等，而這些研究更指出影響相關判斷之結果的主要因素，以「有無資訊檢索經驗」的為最大宗。因此，在利用相關判斷來評估互動式資訊檢索系統，往往需搭配許多情境因素的問題，藉此了解使用者在其他經驗與相關判斷的互動層面。雖然相關判斷的演變已經由早期的「主題相關」，歷經「邏輯相關」、「情境相關」至今的「心理相關」過程，但是在預測相關判斷之模式，還是缺乏一套公用之準則，探究其原因，可以發現相關判斷的過程及其影響的因素，是一個動態的有機體，就好比現實環境中，任何一件事情與其他事件或多或少都存在一種微妙的關係，而這種關係會牽一髮而動全身。

總而言之，無論是 TREC 的評估環境或者是相關研究中都顯示，大家都是在評估的環境中摸索，企圖建立一套相關判斷的評估準則，也因為資訊檢索系統是一個動態的有機體，在利用相關判斷作為評估系統的成效之餘，也應該多方面思考，是否有其他因素會影響一個「動態」的資訊檢索系統，畢竟資訊檢索系統最終的目的是讓檢索者方便使用，而系統應該朝著「因人而異」的設計方向來努力。

<sup>14</sup> 同註 11，頁 81。Saracevic, op. cit.,p.125

<sup>15</sup> Joseph W. Janes, "Other People's Judgments: A comparison of User's and Other's Judgments of Document Relevance, Topicality, and Utility," Journal of the American Society for Information Sciences 45:3 (April 1994):160-171.

## 第二節 檢索評估

資訊檢索是一種人機互動的過程，所以在進行系統評估時，應該是將人的因素以及系統的設計一併評估；因此，有效地評估系統中每一個單一檢索之績效，才有可能真正做好系統評估的工作。

檢索評估的目的不外乎是了解檢索結果滿足使用者資訊需求的程度，因此必須從許多面向去探討，其中較多人所提及的面向包含：檢索的品質、檢索效率、檢索系統本身的設計及檢索者之檢索技巧和知識背景等。<sup>16</sup>下列文獻將從與本實驗較相關之系統評估的測量值、檢索品質及檢索系統本身的設計三方面來分析之。

### 一、 系統評估的測量值

在資訊檢索系統評估中我們需先了解三個重要的測量值，分別為查全率(Recall Ratio)、查準率(Precision)及雜訊比(Noise Ratio)，這三個測量值是評估資訊檢索系統中重要的指標，利用最具代表性的 2 乘 2 表格(見表 2-1)來說明下列三項測量值，分述如下：

- a. 代表相關文章被檢出的筆數。
- b. 代表不相關文章被檢出的筆數。
- c. 代表未被檢出之相關文章筆數。
- d. 代表正確回絕之不相關文章筆數。

#### (一) 查全率(recall ratio)：

指檢出之相關文獻佔所有相關文獻的比例。公式如下：

$$\text{查全率} = \frac{a}{a + c} = \frac{\text{檢索所得之相關文章筆數}}{\text{資料庫中所有相關文章筆數}}$$

#### (二) 查準率(precision ratio)：

指相關文獻於檢索出文章中所佔的比例。公式如下：

---

<sup>16</sup> 同註 11，頁 107-116。

$$\text{查準率} = \frac{a}{a+b} = \frac{\text{檢索所得之相關文章筆數}}{\text{檢索所得之所有書目筆數}}$$

(三) 雜訊比(noise ratio) :

檢出之不相關文獻筆數與資料庫中所有不相關文獻筆數之比值。

$$\text{雜訊比} = \frac{b}{b+d} = \frac{\text{檢索所得之不相關文章筆數}}{\text{資料庫中所有不相關文章筆數}}$$

表 2-1：查全率、查準率和雜訊比之 2 乘 2 表格

	相關	不相關	總數
檢索到	a	b	a+b
未檢索到	c	d	c+d
總數	a+c	b+d	a+b+c+d

資料來源：S.E. Robertson, "The Parametric Description of Retrieval Test," Journal of Documentation 25:1(1969).P.3.

從表 2-1 中可以得知，查準率就是指相關文獻在檢出的文章中所佔的比例；查全率是指檢出之相關文獻佔所有相關文獻的比例；雜訊比則是檢出之不相關文獻筆數與資料庫中所有不相關文獻筆數之比值。系統設計者如何在查準率與查全率兩者之間取得一個平衡點，既能精確的找到所需的文章又能不遺漏任何相關的文章，一直是大家所努力的方向。換言之，高查準率、高查全率及低雜訊率，就是系統設計者持續研發的目標。

## 二、 檢索品質

一般來說，評估檢索品質的優劣最常用的測量值就是查準率和查全率，簡而言之，查準率就是測量資訊檢索系統拒絕不相關文獻的能力，其測量的本質就是檢索結果之正確程度；而查全率就是測試系統檢索到所有相關文獻的能力，其測量的本質就是檢索結果之完整程度。是故，高查準率和高查全率通常意謂著高品質檢索，但是，查準率和查全率本身還受其他因素的影響，如：檢索者所下之檢索詞彙的能力、系統索引詞彙的詳盡及專指度以及一些檢索技巧等。因此，在評估系統的優劣中，往往可以發現必須設法控制一些變數，否則無法在不同的條件之下直接比較查準率及查全率，如：系統所收錄的資料內容及時間等。<sup>17</sup>

<sup>17</sup> 同註 11，頁 108。

### 三、 檢索效率

資訊系統評估的學派發展至今，檢索效益的評估已不再只是利用查準率和查全率的單一測量值，而是慢慢結合效用派理論來做面的評估，如：衡量單次檢索之效率，可以從檢索所花費的時間及金錢來衡量。一般而言，資訊檢索系統的評估大都以完成單次檢索任務所需的時間，來評斷系統和系統間的優劣；而金錢上的衡量，則是因為倡導使用者付費的觀念，所以付愈少的金錢得到愈多的資料，應該是使用者樂於見到的一面。但是，此評估方式應該與其他測量值一併評估，才能得到客觀的結果，其原因是在單篇文章花費少、費時少的狀況下，還是有可能遺漏重要相關文獻或是檢索出過多不相關文獻的情形。因此，最好的評估方式，應該是結合上述的查準率與查全率是較妥當的方法。<sup>18</sup>

### 四、 檢索系統本身

國內學者黃慕萱教授曾經在其著作中，強調評估檢索系統的本身可以從其親和力、回應時間及顯示格式三方面進行評估，時至今日，由於資訊科技技術的發展一日千里，在「回應時間」方面，寬頻網路及電腦處理速度的進步，大部分資訊檢索系統都可以有效縮短回應的時間；至於在現今文字檢索系統中，另一個影響回應時間的因素，是系統如何在動輒幾 Giga Byte 的資料量中，快速的回應使用者。以往檢索系統受制於電腦硬體及技術的限制，所以無法處理大量的資料，現今檢索系統不但是建製索引的速度快，更是將回應的時間縮短至 1 秒之內。因此，此一技術的進步讓使用者能夠在資訊爆炸的時代裡，滿足自己的資訊需求。

而「顯示格式」方面，資訊檢索結果的展現形式不再像以往只注重顯示的項目及順序，反而有許多人利用資訊視覺化的相關技術來呈現文字式資料，如：亞歷桑納大學人工智慧實驗室利用二維自我組織地圖 (SOM-Based Algorithm) 的資訊視覺化 (Information Visualization) 設計來瀏覽檢索結果，以及 Marti Hearst 利用像素直方圖 (Pixel Bar Chart) 的資訊視覺化設計，來進行文字探勘 (Text Mining)。從上述研究可以發現，顯示格式最終的目的就是讓使用者利用最直覺的方式，來適應二維的螢幕空間，並降低資訊視覺混淆 (Visual Cluttering) 的產生。一般來說，資訊視覺化應用在顯示格式最廣泛的學科，應該是以航空、氣象及地理性資料較具規模。或許是因為資料格式的多樣性，所以顯示格式必須呈現與文字式檢索系統不同的地方；不過，值得注意的是，在傳統的文字式檢索介面部分，是否適合利用視覺化的技術來呈現在二維的螢幕，是值得後續研究者持續關注且探討的問題。

---

<sup>18</sup> 同註 11，頁 109-110。

至於在「親和力」方面，除了傳統的錯誤提示功能之外，現今檢索系統更是注重與人互動之後的結果；因此，互動式檢索功能的設計逐漸成為目前檢索系統的核心技術，這些核心技術包含：自動分類、文件自動歸類、與自動關聯詞提示等，最終的目的就是模擬人類對表達欲檢索主題之替代物的心智模式，嘗試利用系統的回饋與使用者互動，讓系統不再只是一昧的被動輸出結果而已，而是主動的提供「檢索策略」或「檢索敘述」，進而輔助檢索者滿足其資訊需求。

## 五、 小結

檢索評估的目的不外乎在每一次檢索的歷程中，評估其成效進而累積成系統評估依據；而評估的面向，也從點的評估逐漸趨向面的評估；從設計者的評估融合使用者的評估。因此，一個良好的檢索系統，應該是經過使用者長期互動之後對系統所回饋的產物，畢竟系統設計的最終目的就是滿足使用者的資訊需求。

因此，從上述 TREC 的評估環境特別設計互動式檢索評估的實驗，可以發現資訊檢索效益不僅是利用查準率與查全率來表達其成效；更希望藉由其他面向來評估，如：人的因素，檢索者的知識狀態、心理認知及新穎程度(Novelty)等，其最終的目的就是將有可能影響到檢索系統的因素，通通列入評估考量。但是，人是易受外在環境影響的生物，也正因為這個因素，讓使用者的評估還沒有一個所謂的共通原則。是故，檢索系統的評估應該是因地制宜，這樣才能應付多樣化的系統。

## 第三節 互動式輔助檢索之設計與評估

資訊檢索系統之互動式檢索功能的研發上，還是以商業化的全文檢索系統最為蓬勃；其中不乏國內外的檢索系統廠商、圖書館自動化公司及知識管理系統等單位投入此一領域的發展，且許多學術研究單位也相繼與業界建教合作以開發各種輔助檢索的互動式檢索功能。而在圖書館的領域中，也有許多資訊科學的專家加入此一領域共同研發更方便的線上資料庫之互動式檢索功能，本研究選擇國內外知名的檢索系統及符合本研究之目的的互動式檢索功能做介紹及分析。

### 一、 互動式檢索的設計與評估

#### (一) 「互動式」輔助檢索應用於檢索系統

資訊檢索系統的檢索結果，其品質大都依賴於使用者所下的檢索詞彙，當檢

索系統的索引詞彙和檢索者所下的辭彙符合時，便能夠產生出良好的檢索結果，但此方式往往是被動地猜測檢索系統的索引詞彙，造成早期的系統往往都成效不彰，或影響一般使用者日後利用類似的檢索系統。從這些現象可以發現如何提高使用者檢索語句的品質，即為資訊檢索系統研發者所努力的地方，而資訊檢索系統與使用者「互動」的目的可解決「詞彙不匹配」的問題。一般來說，在檢索系統的互動通常是藉由檢索詞的「擴展」(Expansion)或關鍵詞的回饋來達成。而檢索詞的「擴展」又可進一步細分為「檢索詞擴展」(Query Expansion)與「相關回饋」(Relevance Feedback)，此種額外擴展的詞彙可能是系統自動加入，或是透過與使用者的互動所產生，前者可稱為「檢索詞擴展」(Query Expansion)，後者可稱為「相關詞提示」(Term Suggestion)。另一種提升檢索成效的方式是「相關回饋」(Relevance Feedback)，亦即就目前找到的文件中，挑取重要的特徵，回饋給系統，以期找到更多的相關資料。此種特徵若是文件本身，則可稱為相關文件回饋 (Document Relevance Feedback)；若為相關詞，則稱為相關詞回饋 (Term Relevance Feedback) 若由檢索到的相關文獻之主題標目去進行檢索，則是專業資訊人員檢索時常用的檢索策略，也可說是相關回饋的一種方式<sup>19</sup>。國內學者曾元顯曾歸納出相關回饋有以下的優點：

1. 讓使用者免於選擇檢索語彙與設計查詢條件的細節，允許建構有用的檢索條件而不用對檢索環境及資料庫有深入瞭解；
2. 把檢索的過程拆解成一步步較明確的步驟，可以逐漸縮小所要檢索的主題之範圍；
3. 提供一個控制查詢的修改過程，終端使用者僅需最少的訓練就可有效且合理的進行檢索。

從上述可以發現檢索者與檢索系統「互動」的方式，包含檢索詞的擴展、相關詞提示、相關回饋以及自動文件分類等技術，最終的目的就是希望檢索者只需最小的努力就能夠得到最滿意的檢索效果，因此，一些學者也強調檢索系統發展的最終原則，就是讓使用者能夠以最少的「互動」就可以獲得有效而合理的檢索。

## (二) 「互動式」的評估方式

資訊檢索系統從早期的系統導向轉變到現今的使用者導向，此改變也逐漸影響到系統研發者在建構系統中考慮使用者或檢索者互動的意見及使用情形。我們可以從國際知名文件檢索的會議中(Text Retrieval Conference，簡稱 TREC)特別增設此一實驗及評估環境中可以發現，而此實驗名稱即為「Interactive Track」。以下將簡短介紹「Interactive Track」的評估環境及實驗目的。

<sup>19</sup> 曾元顯、林瑜一，「模糊搜尋、相關詞提示與相關詞回饋在 OPAC 系統中的成效評估」，中國圖書館學會會報 61 期 (民 87 年 12 月)，頁 103-125。








(一) 目的：

TREC 為了使其測試機制亦能對資訊檢索中一些特殊的議題及新的檢索技術進行有效的評估，TREC-3 開始在Routing Task及Adhoc Task之外，非正式地進行一些其他的測試；自TREC4 起，正式產生一些特殊的測試項目，與會者可以選擇參加一個或多個的測試項目，而每個 Track 都有其測試程序與指導原則；互動式檢索(Interactive Track)則是從 TREC-4(1997) Interactive Track開始至今的 TREC-2002 Interactive Track，在這幾年的發展中最大的目的是將檢索中人與系統的互動記錄下來，以作為系統互動式設計的參考。

(二) 簡介<sup>20</sup>：

在互動式(Interactive)檢索中除了測試檢索結果之外，也將人與系統互動的過程考慮在內。TREC 每年均發展出不同的測試程序以因應不同的資訊檢索實驗環境，其大致上以檢索者在特定的時間、特定的檢索環境及特定的測試資料集檢索到檢索者認為相關的文件集，再利用量化的統計方式與質化的輔助來評估系統。而此一模式雖然無法適用至每一個資訊檢索的系統，但也提供了一套評估互動式檢索的準則。

在此以 2002 年的互動式檢索(Interactive Track)為例，其實施原則強調承襲 Interactive Track 2001 的觀察法研究方法，再以各檢索任務來進行各項的答詢問題。其實驗的資料集是用 TREC 的.GOV 的網站資料，資料特性如下<sup>21</sup>：

-  以.GOV 的網頁資料為主。
-  每 10000 個網頁或文字檔為一個單位。
-  除了純文字檔之外還包括從 pdf 及 ps 檔案格式擷取下來的文字檔。
-  所有文件儘量切截至 100K 以下，將原本所有的文件量從 35G 減到 18G。
-  有檔案格式的限定。

<sup>20</sup> “Draft Guidelines for TREC 2002 Interactive Track”, Retrieved December 4, 2003, from <http://www-nlpir.nist.gov/projects/t11i/guidelines.html>

<sup>21</sup> “TREC Web Corpus : .GOV”, Retrieved December 4, 2003, from, <http://www.ted.cmis.csiro.au/TRECWeb/govinfo.html>.

表 2-2 : GOV 資料集統計數字

文件數量(Documents)	1247753 (1.25 million)
文字/網頁 text/html	1053372 (mime types reported by server)
Application/pdf	131333
純文字檔 text/plain	43754
Application/msword	13842
Application/postscript	5673
Bundles:	4613
Total size:	19455030550 = 18.1G (without 100k limit was 35.3G)
Average bunsizе:	4217435 = 4.0M
Average docsize:	15592 = 15.2k (higher due to pdf+word+ps)
Doc truncation length:	100kb (turned out to be "roughly 100k")
Docs without words:	55 (e.g. due to pdf to text producing much text followed by truncation)

上述的資料集中，大部分的參加者都使用 Panoptic Search Engine 來建置索引及檢索，而檢索結果則是利用 CGI 程式將檢索詞彙傳送至檢索系統，其顯示的方法則是以 xml 的格式來顯示。如下：

[trec.panopticsearch.com/gov/padre-sw\\_xml.cgi?collection=gov&query=bush](http://trec.panopticsearch.com/gov/padre-sw_xml.cgi?collection=gov&query=bush)

### (三) 檢索任務(Tasks) :

在 2002 年中的互動式檢索也同樣建構了八個檢索任務，其任務性質與 2001 年的實驗設計類似，且規定每個檢索者對每個檢索任務至少給予 10 分鐘的檢索時間，當檢索者按下查詢鈕開始檢索時，其時間就開始計算，在時間終了前主測者可以彈性的給予 5 至 10 分鐘的額外時間來完成檢索任務。其任務的內容大致如下：

- 找尋關於個人健康的相關資訊。
- 找尋關於美國法律、規章、條文及政策的指引。
- 擬定旅遊計畫的資訊。
- 給予特定的主題來收集相關資料。

圖 2-2 是 2002 年 Interactive Track 的實際檢索任務，從圖中可以看出檢索任務可以分為三大類：政府規章及法條、健康資訊、及專案企劃及旅遊資訊。由於檢索的資料集是以美國政府網站為主，因此為了配合實驗環境，可以看出所詢問的問題具有專指性。所以，如果實驗的環境是以中文的資料為

主，必須針對實驗系統及資料訂定符合的檢索任務，而這也是評估檢索系統中，系統具備多樣性及彈性的特點，也是資訊系統發展至今卻缺乏一套公用的評估準則之最大原因。

<p><b>Government Regulation</b></p> <ul style="list-style-type: none"><li>• You are travelling from the Netherlands, and want to bring some typical food products as gifts for your friends. What are three kinds of food products from the Netherlands that you are not allowed to bring into the US? (GR1)</li><li>• You are concerned with privacy issues related to electronic information and would like to know what laws have been passed by the US Congress regarding these issues. Identify three such laws. (GR2)</li></ul> <p><b>Health or Project</b></p> <ul style="list-style-type: none"><li>• A friend has a private well which is the family's only source of drinking water. Locate a US publication, which contains guidelines for the maintenance of safe water standards for private well use. (HP1)</li><li>• You are not sure about the safety of genetically engineered foods, and would like to find more information and research on this topic. Name four potential types of safety problems that have been raised. (HP2)</li><li>• Name/find three research programs/projects that investigate the treatment/causes of dwarfism. (HP3)</li><li>• You are interested in learning more about what measures the US government has taken since 2001 to prevent Mad-Cow Disease. Identify three such measures. (HP4)</li></ul> <p><b>Travel</b></p> <ul style="list-style-type: none"><li>• You are planning a cycling expedition along the Silk Road in Central Asia. Find a website that is a good source information about health precautions should you take. (T1)</li><li>• You are planning to travel to the northeast territories of India and wonder if there are any problems/restrictions for tourists. Find a website that is a good source of information about such problems/restrictions. (T2)</li></ul>
---

圖 2-2 : TREC 2002 Interactive Track Experiments 的檢索任務

#### (四) 檢索任務的制定原則(Guidelines) :

在 2002 互動式檢索評估的實驗中，可以依據下列原則及根據自己的實驗環境來制定特定的檢索任務。

- 1 每一個接受指派任務的檢索者至少須符合高中以上的學歷及一定程度的生活經驗。
- 2 每個檢索任務所詢問的問題中，必須包含回答此一問題的動機，尤其是非美國籍的民眾，其原因可能是實驗的資料內容為美國政府網站的資料為主，以降低此問題因國籍背景不同而產生的偏差。
- 3 每個問題必須明確的指出檢索者可以答題的方向。如 number/kind of items, intended use。
- 4 當檢索任務有需要延長或停止時，需統一規定每個檢索任務停止或延長所需的時間。

**(五) 實驗的設計(Experimental Design) :**

在實驗設計方面承襲 Trec-9 Interactive Track 的方法，可以比較兩個系統間的成效；參加檢索任務的人員至少有 16 人且每個檢索者須完成八個檢索任務，而 8 個檢索任務可分為兩群，一群檢索 A 系統再則檢索 B 系統，另一群則先檢索 B 系統再檢索 A 系統，其目的是降低使用者的使用先後順序所造成的誤差（見表 2-3）。而實驗設計的架構大致如下<sup>22</sup>：

- 1 互動式的檢索任務-利用問答式的方法。
- 2 共有八個檢索任務/問題。
- 3 最少須有 16 個參加者。
- 4 查詢的資料主要是線上新聞及電子文件。
- 5 需有檢索者的基本資料問卷。
- 6 每個實驗設計需有簡短的摘要以供其它參賽的研究團隊進行評估。

AB：表示先做 A 系統再做 B 系統。

BA：表示先做 B 系統再做 A 系統。

**表 2-3：TREC 互動式評估實驗進行方式**

檢索任務 受測者	檢索任務 1、3、5、7 (群一)	檢索任務 2、4、6、8 (群二)
第一組(8人) AB	4人	4人
第二組(8人) BA	4人	4人

**(六) 檢索者的任務—問卷調查(Searcher Tasks) :**

每個檢索者在檢索實驗的過程中，會給予一些特定的問卷，其目的是收集使用者在互動的過程中產生的回應，原則上可分為四大部分，第一部分是在檢索實驗前的基本資料及相關經驗的問卷(Entry-Questionnaires)；第二部分是在每個檢索任務完成後的問卷(Post-search Questionnaires)，其目的主要是調查此檢索者對此檢索任務的熟悉程度；第三部分則是詢問在同一檢索系統完成各項檢索任務後的觀感(Post-system Questionnaires)，最後部份則是在所

<sup>22</sup> “TREC-9 Interactive Track Guidelines”, Retrieved December 4, 2003, from, <http://www-nlpir.nist.gov/projects/t9i/spec.html>

有實驗都完成之後，兩系統間的比較及使用者對系統之回饋的調查問卷 (Exit-search Questionnaires)<sup>23</sup>。

### (七) 評估方式(Evaluation)：

檢索系統將承襲之前的評估方式，從效能(Effectiveness)、效率(Efficiency)及使用者滿意度(User Satisfaction)三方面來評估。「效能」即檢索者對每個檢索任務完成的程度，「效率」則是每次檢索時完成所需的時間，「使用者滿意度」則可以問卷的方式收集檢索者的背景、對資訊檢索的熟悉度、及檢索前應具備的能力和檢索後對各項功能的滿意度等。

上述經由問卷所收集的資料再利用社會科學統計軟體做進一步的分析。原則上，在 TREC 的互動式檢索評估中，評估方式大多以量化的數據為主，而此一因素也是以往資訊檢索系統評估的缺點，因為量化的分析只能得到概約的數據，無法完全表達真實環境中系統和使用者的互動情形。因此，在參與過 TREC Interactive Track 的分析報告中，也指出此一因素是作為日後評估檢索系統最大的關鍵。

在此，我們將以TREC Interactive Track 2001已經進行過的系統為範例做個簡單的報告，參與TREC互動式檢索評估的研究團隊包括：

1. CSIRO的研究持續在關注網站資料的組織議題上，以.gov網站的資料內容為例，它利用組織機構來將網站內容分類，來比較傳統的文件條列式排序方法。
2. Oregon Health & Science University嘗試以不同的資訊載體進行檢索，如掌上型的電腦。
3. Rutgers University利用不同介面來呈現檢索結果及允許使用者輸入較長的詞彙，期望利用最有效的互動方式來提高檢索效益。
4. University of North Carolina Chapel Hill則利用視覺化的3D空間來呈現檢索結果，以比較與傳統的文字式檢索介面的差異。
5. 以及University of Glasgow的自動歸類與自動摘要的互動式檢索功能等<sup>24</sup>。

上述的互動式檢索功能不僅包含關聯提示詞，還包括文件歸類、自

<sup>23</sup> 同註 23。TREC-9 Interactive Track Guidelines , <http://www-nlpir.nist.gov/projects/t9i/spec.html> , 2003。

<sup>24</sup> William Hersh, Oregon Health & Science University, Portland, OR, USA."TREC 2002 Interactive Track Report". Retrieved December 4, 2003, from <http://trec.nist.gov/pubs/trec11/papers/INTERACTIVE.OHSU.pdf>

動摘要及利用資訊視覺化的呈現方式來達到與查詢者的互動。在此，將選擇與本研究相關之互動式檢索功能來探討，期望利用文獻分析的方式，以「互動式檢索功能的設計」、「評估實驗設計」及「分析方式」三個面向來探討，且歸納出最適合本實驗之評估方式。

## (八) 案例(Approaches)：

### 1 Glasgow-HuddleSearch<sup>25</sup>：

#### 1) 互動式檢索功能的設計：

是由蘇格蘭的 Glasgow 大學的電腦科學學系所研發出新的文件階層式歸類方法，它可以動態的方式來組織文件且橫跨傳統階層式文件組織方式的限制，簡單來說，就是不再侷限於單一分類的缺點。而其互動式的功能是利用一些視覺化的設計結合階層式歸類與文件摘要的功能，讓檢索者自己判斷與其相關的文件。

文中提及系統的文件歸類功能不同於平面式(Flat Cluster)的系統如Vivisimo<sup>26</sup>，WiseNut<sup>27</sup>，和Grouper，HuddleSearch將文件歸類成階層式的結構以呈現檢索結果，其文件之分群以資料夾(Folder)來呈現；在最上層的資料就是最一般性(Generality)的文件，愈下層的資料夾所代表的文件則愈具專指性(Specificity)。而資料夾的標題就是代表此類文件的類目，資料夾上的數字則是此類文件的數量（見圖2-3）。

在互動式輔助檢索的設計方面，除上述的文件式階層歸類及摘要之外，其系統還利用一些視覺化的設計與檢索者互動，如當檢索者將滑鼠移動到資料夾上，便會顯示出此類文件的自動摘要，此一功能是擷取此類所有文件的自動摘要，可以減少檢索者做不必要的連結。如下圖2-3右邊的放大鏡，就是此類文件的摘要。

<sup>25</sup> Richard Osdin, Iadh Ounis and Ryen W. White. "Using Hierarchical Clustering and Summarization Approaches for Web Retrieval :Glasgow at the TREC 2002 Interactive Track.", Retrieved December 4, 2003, from <http://trec.nist.gov/pubs/trec11/papers/glasgow.int.pdf>

<sup>26</sup> <http://vivisimo.com/>

<sup>27</sup> <http://www.wisenut.com/>

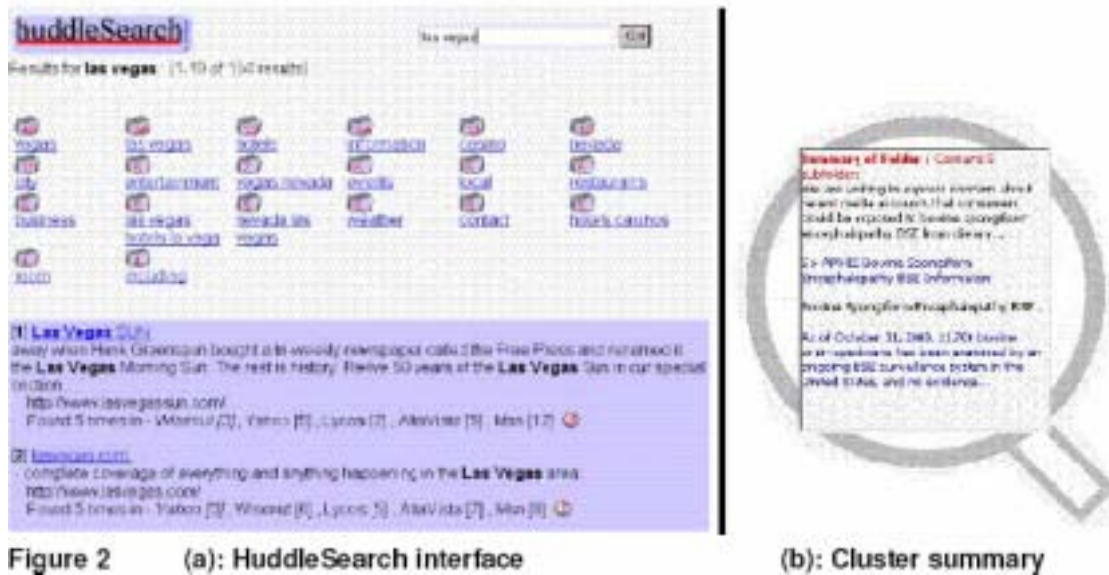


圖 2-3 : HuddleSearch 的互動式檢索結介面

## 2) 評估實驗設計：

### (1) 實驗系統：

以 HuddleSearch 的系統為實驗之系統，並利用 TREC 上所提供的 Panoptic Search Engine 作為對照之系統，分別以系統 X 代表 HuddleSearch，系統 Y 代表 Panoptic Search Engine；此用意主要是讓受試者在未得知系統的名稱下進行實驗，以降低使用者的判斷誤差。

### (2) 研究方法：

首先，以「系統實作法」建構出一個有自動歸類及自動摘要的互動式檢索輔助功能，再利用「問卷調查法」收集查詢者的背景資料、檢索實驗前在電腦或檢索方面的相關經驗，最後利用檢索後問卷收集檢索者對此功能的滿意程度。

### (3) 實驗對象：

受試者共有 16 位，學歷皆為碩士程度以上，其學科背景為不同領域的人，在文章中強調實驗對象背景的差異，主要是因為如此才符合一般的網路搜尋行為的檢索者。

**(4) 檢索(電腦)經驗：**

在檢索經驗方面，統計出大部分的受試者在使用搜尋引擎的經驗平均值為 5.7 年以上，其使用的目的皆以學術性質為主，且大部分的受試者常用的搜尋引擎都包括 Google。

**(5) 檢索任務：**

其檢索任務大致依循 TREC 2002 Interactive Track Experiments 的檢索任務(見圖 2-2)，而 16 位受試者必須在 10 分鐘內完成上述 8 個檢索任務，但是為了避免使用前後產生的影響及誤差，必須將 8 個檢索任務分為兩群，也就是說一群先利用系統 X 檢索再利用系統 Y 檢索，另一群檢索任務則反過來先利用系統 Y 檢索再利用系統 X 檢索。如此才可以降低誤差的產生(見表 2-3)。

**(6) 資料集：**

在實驗資料集方面，沿用 TREC 的 .GOV 網站資料，其特性可參考(表 2-2).GOV 資料集統計數字。

**3) 分析方式：**

實驗結果的評估方式則依據 TREC 上的準則從下列三方面來評估：

**(1) 檢索任務完成之效能(Task Completion)：**

利用 TREC 的檢索後問卷(Post-task Questionnaire)，收集檢索者對自己在時間限制內所完成之檢索任務的程度。文中提及檢索者對檢索結果的滿意程度的不同，其實也正符合真實世界中每個檢索者心理認知的差異。因此，所謂判斷受試者是否完成檢索任務，是以檢索者在時間內斷定以自己找到所要的資訊即表示完成檢索任務；最後，再以統計學中的配對樣本 T 檢定(Paired T-test)來檢定兩實驗之系統是否有顯著的差異。而結果顯示，有提供自動歸類及利用資訊視覺化呈現的自動摘要之檢索功能和對照之系統並無顯著的差異，也就是說此功能在實驗中並沒有達到預期的效果。但是，就兩個系統在檢索任務執行的成功率(效能)來說，有自動歸類及自動摘要之功能還是優於一般的檢索引擎；推究其原因，未達到統計學上的標準，可能



是實驗的對象不夠多。

### (2) 檢索任務完成之效率(Task Times) :

檢索任務的時間利用系統中使用者檔的紀錄(User Log)來記錄，當受試者無法在規定的 10 分鐘內完成檢索任務，主測者可以額外再給此受試者 1 分鐘的時間來完成。

經過問卷收集到的資訊可以統計出兩個系統在每個檢索任務所花的時間，以及平均一個系統在執行一個檢索任務所花的時間。最後，再利用雙因子變異數分析(Two-Way ANOVA)來檢定系統和任務之間的顯著性是否達到統計學的檢定水準。

研究結果顯示，比較各檢索任務的效率，再利用 Huddle Search 來執行檢索任務，其時間的確是有逐漸降低的趨勢，但是，也會隨著檢索任務的特性不同而產生出相反的結果，如：檢索任務愈專指性時間花費會愈多。最後，利用雙因子變異數分析檢定出其檢索效率的確和檢索任務及系統間的功能差異有很大的關連，換句話說，「檢索任務」和「系統的功能」的確是影響檢索效率的兩大因素。

### (3) 使用者滿意度(User Satisfaction) :

在使用者滿意度方面也是利用 TREC 的檢索後問卷(Post-task Questionnaire)來收集查詢者對互動式檢索功能設計的滿意度及自己在檢索任務完成的滿意程度，且最後也是利用統計學中的雙因子變異數分析(Two-Way ANOVA)來檢定系統和任務之間滿意程度的顯著性；也就是說，確定上述兩項因素影響查詢者的滿意度是顯著性的。

經由上述的工具分析之後，結果顯示，利用雙因子變異數分析，使用者在不同系統執行之檢索任務的滿意程度上，呈現出相同的結果，而此一數據經由雙因子變異數的檢定並不具統計學上的信度。

## 2 Rutgers University<sup>28</sup>

### 1) 互動式檢索功能的設計 :

<sup>28</sup> N.J. Belkin, C. Cool, D. Kelly, G. Kim, J.-Y. Kim, H.-J. Lee, G. Muresan, M.-C. Tang, X.-J. Yuan. "Rutgers Interactive Track at TREC 2002.", Retrieved December 4, 2003, from <http://trec.nist.gov/pubs/trec11/papers/rutgers.belkin.pdf>

Belkin 等承續其 TREC 2001 Interactive Track 分析出的結果，歸納出二個重要的假設。

**假設一：資訊檢索系統應該降低與查詢者的互動，才能提高查詢者的使用者滿意度。**

**假設二：資訊檢索系統必須允許查詢者輸入較長的辭彙來檢索。**

為了驗證上述兩項假設，作者分別針對“假設一”及“假設二”各設計兩種不同的介面，利用TREC Interactive Track的分析方式來評估。在“假設一”中利用Panoptic Search Engine設計兩種不同的介面，第一個介面還是以傳統的文章排序方式和標題來呈現文件，且每頁顯示20筆資訊(SDD系統)；另一種介面則是一個頁面顯示4筆文件標題且將文件內容顯示出來(MDD系統)，如此一來查詢者不用再點選文章連結就可以判斷是否為相關資訊。

在“假設二”中作者除了以原本 Panoptic Search Engine 允許的短字串搜尋之外，另外再建構出可以長字串來搜尋的介面；利用這兩種介面來進行評估的實驗及對照系統，以驗證“假設二”中的假設。

## 2) 評估實驗設計：

### (1) 實驗系統：

大致上可分為兩種系統，實驗之系統稱為 MDD，也就是可以允許長字串檢索及一個頁面呈現四個檢索文件的介面(見圖 2-4)；而對照之系統稱為 SDD，為不允許長字串檢索及傳統式一頁呈現 20 筆文章標題的介面(見圖 2-5)。

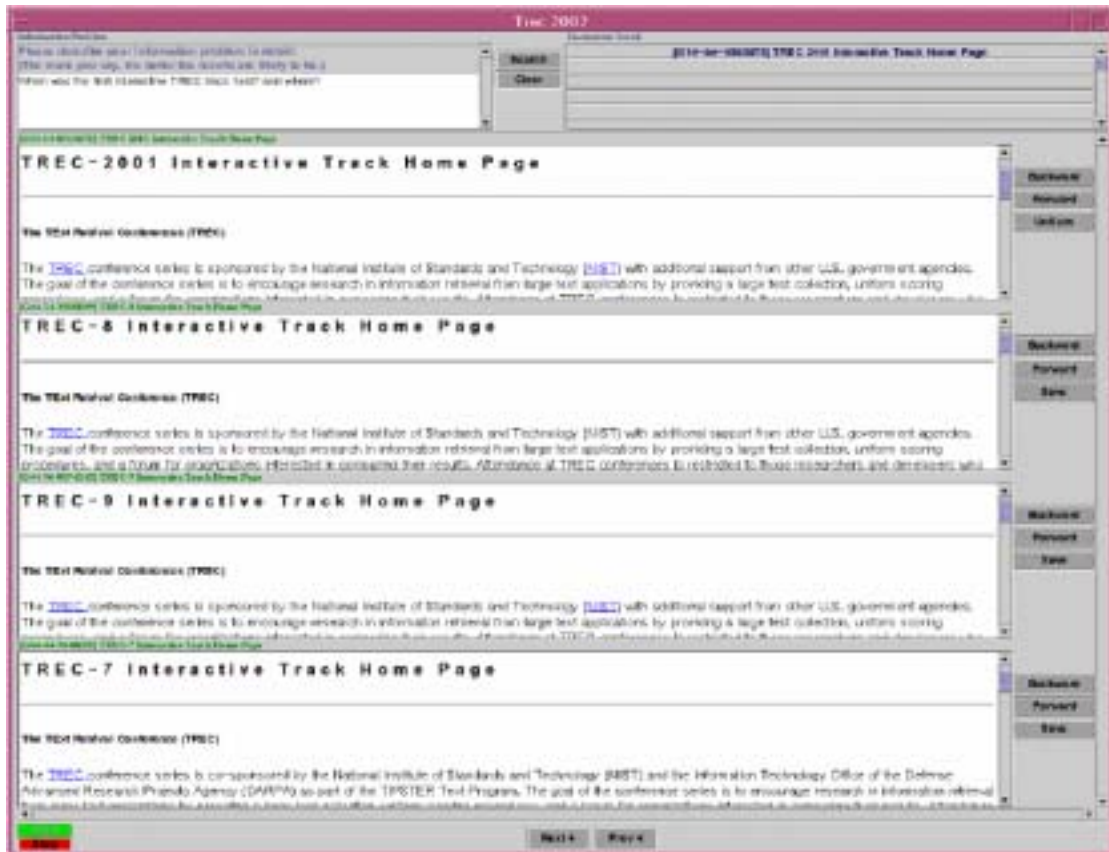


圖 2-4 : MDD 允許長字串檢索之介面

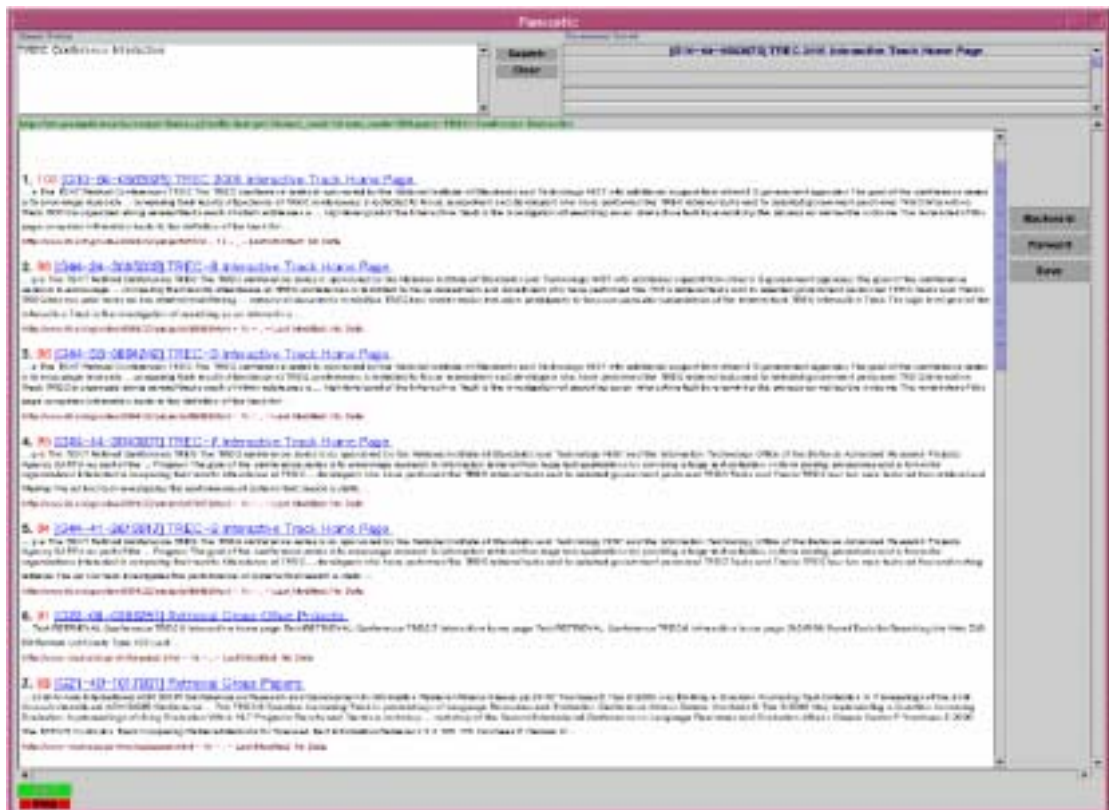


圖 2-5 : SDD 不允許長字串檢索之介面

**(2) 研究方法：**

在進行實驗之前先利用「問卷調查法」收集使用者的背景資料、檢索經驗及互動式檢索輔助功能的滿意程度。在實驗過程中利用「放聲思考法」讓檢索者將實驗中的各種反應直接表達出來，且利用錄影的方式將查詢的檢索畫面也記錄下來，最後完成每個檢索任務時，輔以「訪談法」將查詢者對互動式檢索輔助功能應用在資訊檢索系統的觀感，利用錄音機的方式將查詢者的意見記錄下來。

**(3) 實驗對象：**

參與此次實驗的人員共有 32 位 Rutgers 's SCILS(School of Communication, Information and Library Studies)的學生，其中 26 位是女生；六位是男生，年齡介於 28-37 歲，學歷大部分都是碩士生。

**(4) 檢索(電腦)經驗：**

由於大部分的受測者學歷皆為碩士，且學科背景又是相關研究領域，因此都有豐富的檢索及電腦經驗。

**(5) 檢索任務：**

檢索任務也是 TREC 2002 Interactive Track Experiments 的檢索任務(見圖 2-2), 32 位受試者必須在 10 分鐘內完成上述 8 個檢索任務，但是為了避免使用前後所產生的影響及誤差，也將受試者分為兩群各 16 位受試者(見圖 2-6)，也就是說一群先檢索系統 A 再檢索系統 B，另一群受試者則反過來先檢索系統 B 再檢索系統 A。如此才可以降低誤差的產生。

Subject	Block 1 System: Topics	Block 2 System: Topics
1	SDD: 4-7-5-8	MDD: 1-3-2-6
2	MDD: 3-5-7-1	SDD: 8-4-6-2
3	MDD: 1-3-4-6	SDD: 2-8-7-5
4	MDD: 5-2-6-3	SDD: 4-7-1-8
5	SDD: 7-6-2-4	MDD: 3-5-8-1
6	SDD: 8-4-3-2	MDD: 6-1-5-7
7	MDD: 6-1-8-7	SDD: 5-2-4-3
8	SDD: 2-8-4-5	MDD: 7-6-3-4
9	MDD: 4-3-5-8	SDD: 1-3-2-6
10	SDD: 3-5-7-1	MDD: 8-4-6-2
11	SDD: 1-3-4-6	MDD: 2-8-7-5
12	SDD: 5-2-6-3	MDD: 4-7-1-8
13	MDD: 7-6-2-4	SDD: 3-5-8-1
14	MDD: 8-4-3-2	SDD: 6-1-5-7
15	SDD: 6-1-8-7	MDD: 5-2-4-3
16	MDD: 2-8-1-5	SDD: 7-6-3-4

圖 2-6：實驗設計之第一群 16 位實驗對象的檢索任務分配

**(6) 資料集：**

在實驗資料集方面，沿用 TREC 的 .GOV 網站資料，其特性可參考(表 2-2) .GOV 資料集統計數字。

**3) 分析方式：**

分析方式主要還是承襲 TREC 互動式評估的原則，再依據自己的實驗環境做適當的調整。其分析的變數包含：使用者的搜尋行為 ( Search Behavior )、藉由每次檢索時所擷取下來文件的滿意度來評估其效能 ( Performances ) 及利用正確性(Correctness)及完整性(Completeness)來評估擷取下來文章的相關度，大致上分析如下：

**(1) 使用者的搜尋行為 ( Search Behavior )：**

在評估使用者的搜尋行為可以從下列幾個因素探討：Pages seen、Unique pages seen、Pages viewed、Unique pages viewed、Number of Documents saves、Number of final saved documents、Number of iterations、Means query length 及 Unique query length 等，而上述各項變數的定義如 ( 表 2-4 )。

**表 2-4：描述使用者搜尋行為的各項變數列表及定義**

因素 ( Variables )	定義(Definition)
Pages seen	在檢索過程中，每一頁所回覆檢索的款目 ( 標題 ) 總數 ( 只適用於SDD介面 )。
Unique pages seen	每一頁去除回覆後重覆的檢索款目 ( 標題 )。
Pages viewed	檢索者瀏覽檢索結果的文章內容的總數。
Unique pages viewed	檢索者去除重複之檢索結果內容的頁數。
Number of documents saved	檢索過程中被標記為儲存的文件總數。
Number of final saved documents	檢索結束時被標記為儲存的文件數量。
Number of iterations	檢索過程中所使用的檢索詞是從開始檢索持續使用到檢索結束後的總數。
Mean query length	檢索詞的平均長度 ( 不包含停用詞 )
Unique query length	在每次檢索中，皆有出現的檢索詞總數。

資料來源：N.J. Belkin, C. Cool, D. Kelly, G. Kim, J.-Y. Kim, H.-J. Lee, G. Muresan, M.-C. Tang, X.-J. Yuan.“Rutgers Interactive Track at TREC 2002.”, Retrieved December 4, 2003, from <http://trec.nist.gov/pubs/trec11/papers/rutgers.belkin.pdf>

從表 2-4 中可以歸納出三個因素來代表使用者搜尋行為，分別為：

- a. Pages：統計每一頁顯示的標題款目，此項因素包含標題、文章內容及扣除重覆的顯示款目等；
- b. Documents：統計檢索過程中及經過濾重複的文件之後被儲存的檢索結果文件內容；
- c. Query：統計檢索者在檢索開始及修正檢索詞的過程中，所使用的檢索詞之變化。

上述各項因素利用敘述統計 ( Descriptive Statistics ) 中的標準差及平均數將各項變數的分數計算出來，利用這些數字來作為相關推論的依據。分析結果顯示，MDD 系統不僅有效的降低使用者參與系統的互動，且讓互動顯得更有效率，所以其研究的假設皆可以成立。

雖然上述的分析結果在統計學上還不夠嚴謹，但是，如果研究者只是想對某個現象做描述，這些數字的確可以顯示出某些訊息給系統的設計者。

#### (2) 使用者滿意度分析(User Satisfaction)：

利用李克特氏七點式量表來評估使用者對每一次檢索的滿意程度。分析結果顯示，對兩系統在檢索結果的滿意程度上，分別為：MDD 為 4.65，SDD 為 3.95，換句話說，從統計學上的數字分析可以得知，使用者對 MDD 系統提供的功能，整體滿意度是較高的。

#### (3) 正確性(Correctness)：

評估每一次檢索者得到的檢索結果之正確性，也是利用統計學的平均數及標準差來評估兩系統在擷取下來之文章的正確性，其數據顯示 MDD 系統還是高於 SDD 系統。

#### (4) 完整性(Completeness)：

評估每一次檢索者得到的檢索結果是否為完整找到相關文章，也就是判斷有多少的檢索結果的文章內容被使用者擷取下來，且搭配文章的正確性 ( Completeness ) 來評估此次檢索的效能。在此項評估中 MDD 系統的完整性也是高於 SDD 系統。

## 二、 使用者滿意度評估

除了 TREC 有互動檢索評估的設計外，國外不少學者及研究團隊也有自行設計互動式檢索評估的環境，針對研發之功能進行評估。以下列舉與本研究相關之互動式檢索功能的實驗介紹及分析。

互動式檢索功能的目的是幫助使用者做更簡易的檢索及查詢得到所需的文章，因此為了得到使用者對檢索系統的回饋及反應，須針對檢索者的滿意度做進一步的分析與比較，所以從 TREC 上的檢索後問卷(Post-task Questionnaire)的設計及相關文獻，可以發現「使用者滿意度」的評估方式幾乎是不可或缺的。以下列舉有評估互動式檢索系統的實驗設計之方法與比較：

### 一、 Keyphind (Keyphrase find) :

#### (一) 評估實驗設計：

Keyphind(見圖2-7)是由加拿大Saskatchewan大學裡的人機互動實驗室<sup>29</sup>(HCI the interaction lab)針對自行研發之檢索系統所做的可用性評估，其核心技術結合視覺化設計(Presentation)和利用自動歸類(Cluster)的方式，自動產生出片語及關鍵詞，有別於一般傳統人工分類的方式，系統會依檢索詞彙將資料庫內的文件重新歸類。因此，其分類的方式依文件變動而改變，可說是動態性的分類方式。

在實驗的資料集方面，總共利用26,000篇科學性的技術報告作為實驗的文件，而這些文件是紐西蘭數位圖書館所發行Computer Science Technical Reports (CSTR)期刊的一部分，文件集數量在現今的實驗環境中雖然不是很多，但作者也強調其系統的核心技術足以應付大量的文件集。

至於評估檢索系統 Keyphind 方面，挑選 10 人的受測群體，並於受測前對此受測群體做兩系統的檢索方法介紹，每個受試者對兩檢索系統有相當程度的了解。當前置作業完成時，每個受測者被要求對新 (Keyphind)、舊(全文檢索系統)檢索系統分別做三次檢索任務的實驗；而在檢索任務執行的過程中，受測者的言行舉止均會被錄影機記錄下來；且為了了解受測者在實驗中資訊搜尋行為的心路歷程，受測者希望利用放聲思考法(Think Aloud)的方式將所遭遇到的各種情況表達出來。最後，當每個受測者完成三次的檢索任務之後，會請受測者接受一個簡短的問題及最終的訪談；而訪談可分為每個檢索任

<sup>29</sup> “Interaction lab : University of Saskatchewan”, Retrieved December 4, 2003, from <http://hci.usask.ca/index.xml>.

務完成之後的問題(見表 2-5), 和完成三次檢索任務之後的最終訪談(見表 2-6), 其目的就是要了解使用者對兩系統間的接受度。其訪談內容分別如下(表 2-5)、(表 2-6)。<sup>30</sup>

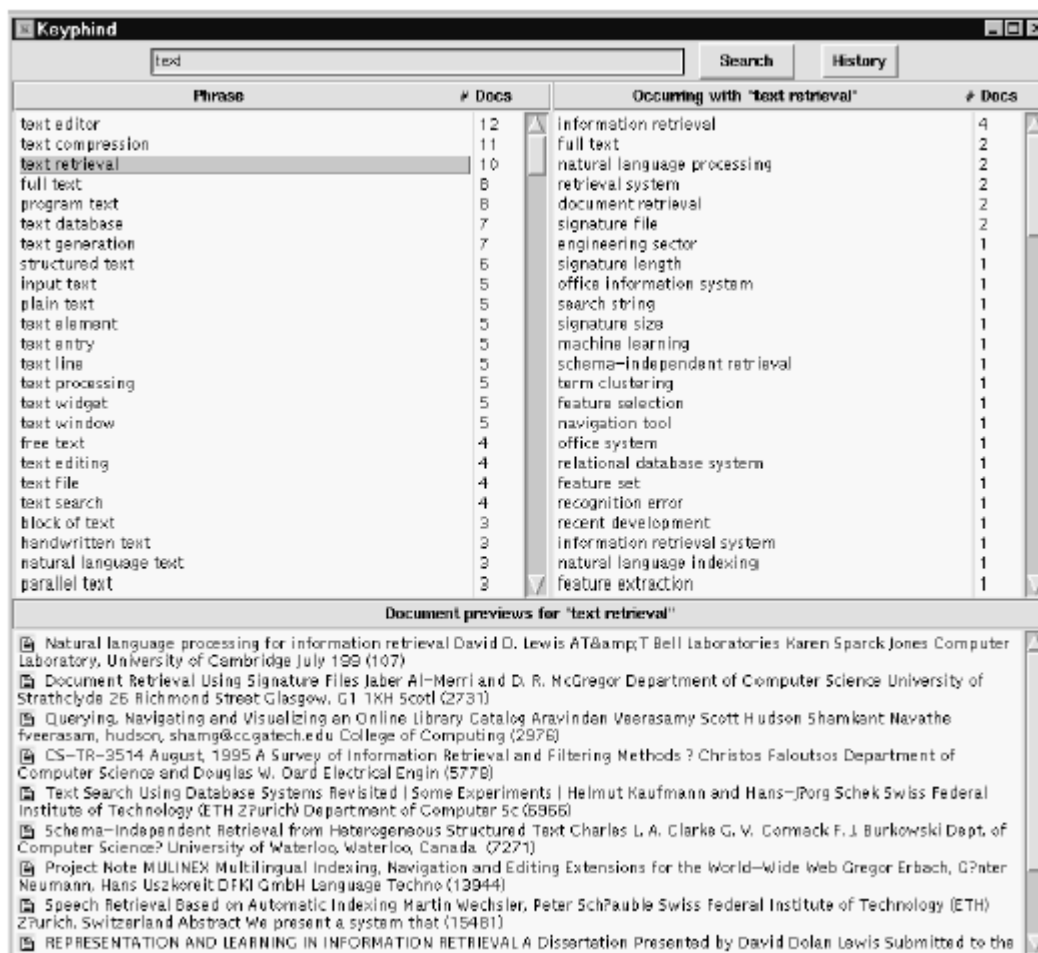


圖 2-7 : Keyphind user interface

資料來源：Gutwin C., Paynter G., Witten I.H., Nevill-Manning C., Frank, E.: Improving Browsing in Digital Libraries with Keyphrase Indexes, Journal of Decision Support Systems 27(1-2)(1999) 81-104  
<<http://hci.usask.ca/publications/1999/keyphind-jdss/keyphind-jdss.pdf>>

<sup>30</sup> Gutwin C., Paynter G., Witten I.H., Nevill-Manning C., Frank, E., "Improving Browsing in Digital Libraries with Keyphrase Indexes," *Journal of Decision Support Systems* 27(1-2) (1999) 81-104, Retrieved December 4, 2003, from <http://hci.usask.ca/publications/1999/keyphind-jdss/keyphind-jdss.pdf>



表 2-5：每項檢索任務完成之後的問題

1. Was it easier to carry out the task with the order of these systems?
2. If yes, which one?
3. If yes, was the task: slightly easier, somewhat easier or much easier?

表 2-6：最終的訪談問題

1. What process did you use to find out about a topic with each system?
2. What were the major differences between the two system?
3. Would you use a system like this one in your work?
4. How would you use it?

## (二) 分析結果：

分析方式是利用統計學上的假設驗證來評估其實驗之系統，而驗證的假設為：Keyphind 系統的自動索引及歸類的功能優於一般傳統的全文檢索系統；驗證的方法則利用可用性評估 (Usability)，實際建構出檢索任務模擬檢索的環境；此外，在檢索過程中，利用觀察法來觀察使用者在使用兩系統間的異同，藉由使用者實際的反應得知此系統設計是否符合使用者的心理認知。再者，共有三個檢索任務，實際的檢索任務可分為下列兩大原則：

### 1. 涵蓋面評估(Coverage Evaluation Task):

涵蓋面評估是指測試議題之涵蓋面是否影響使用者的輔助程度及差異；所以在涵蓋面上，作者共設計二個檢索任務，其主要目的是在使用者對檢索議題不熟悉的狀況下，也能夠輕易的找出相關文件。

### 2. 探索性評估(Exploration Task):

在探索性評估上，則是讓使用者在不限類別的情況下，依照使用者最感興趣的議題檢索出 2 篇相關文件，共一個檢索任務。此設計主要是驗證此功能是否容易找出自己感興趣的網路文件，並有模擬一般搜尋引擎的用意。

最後，其結果可以從下列兩項原則來分析：

### 1. 哪一個系統是較容易執行檢索任務：

評估方式是單純統計使用者勾選哪一種系統是較容易使用，選項如下

圖 2-8。結果顯示，在“檢索任務一 (Task1)”中有 7/8 的使用者覺得 Keyphind 較容易使用；而“檢索任務二 (Task2)”中有一半的人覺得沒差異，有三個人覺得 Keyphind 較容易使用。從上述的統計數字中，我們可以發現一個共同的項目，無論是在檢索任務一或檢索任務二中，使用者都未曾勾選傳統的檢索系統是較容易使用的。

Task	The task was easier with:			
	Keyphind	Traditional	No difference	No answer
1 (planning)	7	0	1	0
2 (computer graphics)	3	0	4	1

圖 2-8：難易度比較性問題的統計數字列表

## 2. 系統的功能如何影響使用者的搜尋策略：

至於在此項分析中，經由觀察法得知，大致上可以從“剛開始檢索”及“系統回應的檢索結果”，兩方面分析之。在剛開始檢索時，傳統的檢索系統和 Keyphind 並無顯著的差異，也就是說，都是由檢索者心目中的檢索替代物來下檢索詞彙。在系統回應的檢索結果方面，傳統的全文檢索系統，檢索者必須經由瀏覽檢索結果的文件內容，才會再修正自己的檢索詞彙；而 Keyphind 系統則可以透過所提示的片語再修正自己的檢索詞彙，可以有效的節省時間。

值得注意的地方是，研究結果顯示在“檢索任務二”中，Keyphind 所花費的時間比傳統的全文檢索系統還多，作者指出可能是因為“檢索任務二”較具專業性，一般的檢索者無法直接透過標題就可以判定是否為相關的文章，因此，檢索者必須多經過一道判斷是否為相關的查詢片語，最後再判斷其文章的正確性。是故，所花費的時間當然會多於傳統的全文檢索系統。而分析結果也強調，雖然花費的時間較多，但對於輔助檢索者的搜尋策略的確有很大的幫助。由上述的案例可以發現，評估資訊檢索系統的效能，不能從單一面向來判斷，必須結合多項面向分析來評估，才能得到公平且客觀的評估結果。

## 二、 University of Waikato Usability Laboratory(The Phind Interface)：

### (一) 互動式檢索功能的設計：

這是一個由紐西蘭 Waikato 大學電腦科學系的一個數位圖書館計畫所做

的研究，其研究是針對階層式片語的瀏覽方式，期望能夠達到自動建構之索引典的機制，且打破傳統紙本索引典的顯示格式。

一般圖書館的檢索系統大都提供主題瀏覽的形式，此階層式的瀏覽方式雖然符合大部分的使用者需求，但缺點是需要大量的人力來維護，且擴充性較低及不能應付各學科的成長，因此，文中嘗試利用自動建構片語及索引典的方式來降低人力負擔及提高效能。

在文中提及替代人工分類耗時、耗力的缺點，其解決方法就是利用文件中片語及關鍵字共現性來自動建構索引典。而此一學者將此方法稱為「Hierarchical Phrase Browsing」，因為它允許檢索者自行瀏覽自動產生的片語及關鍵字，且允許檢索者檢閱此一片語及關鍵字在文件中的位置，並在介面上設計利用 Java Applet 將檢索結果的片語及關鍵字以分割畫面來呈現(見圖 2-9)，有別於一般人工建構的主題瀏覽之索引典。

文中也提及 Phrase Browsing Interface 此互動式輔助檢索之功能於許多系統中都有類似的設計，但是在使用者評估方面仍然是付之闕如，且大部分的文章都是強調如何提高索引詞的品質，就算是有評估也只是做到效率的評估而已。因此，文章中特別強調其貢獻是在描述及評估人與系統互動的過程；也就是說做到以往資訊檢索系統缺乏之「質」的評估。

文中也提到 Gutwin 曾在使用者互動評估中，發現有提供片語瀏覽介面之檢索系統的使用情形優於只提供查詢功能的檢索系統，分析之後的原因是它提供檢索者另一層次更適合不同檢索任務的管道，如：當檢索者的資訊需求是較一般化且無法明確指出，他會需要盡可能找尋到相關的文件；此時有提供片語或關鍵詞瀏覽的功能便可幫助檢索者擴展詞彙。因此，自動建構的片語或關鍵詞的品質便會影響到此功能的成效。

## (二) 評估實驗設計：

### 1. 實驗系統：

利用設計的 Phind Interface 片語式瀏覽介面為實驗系統(見圖 2-9)，其展示系統的網址為：

<http://www.sadl.uleth.ca/nz/cgi-bin/library?a=p&p=about&c=fi1998>

對照系統即為 Greenstone Keyword Searching 的關鍵字搜尋引擎。

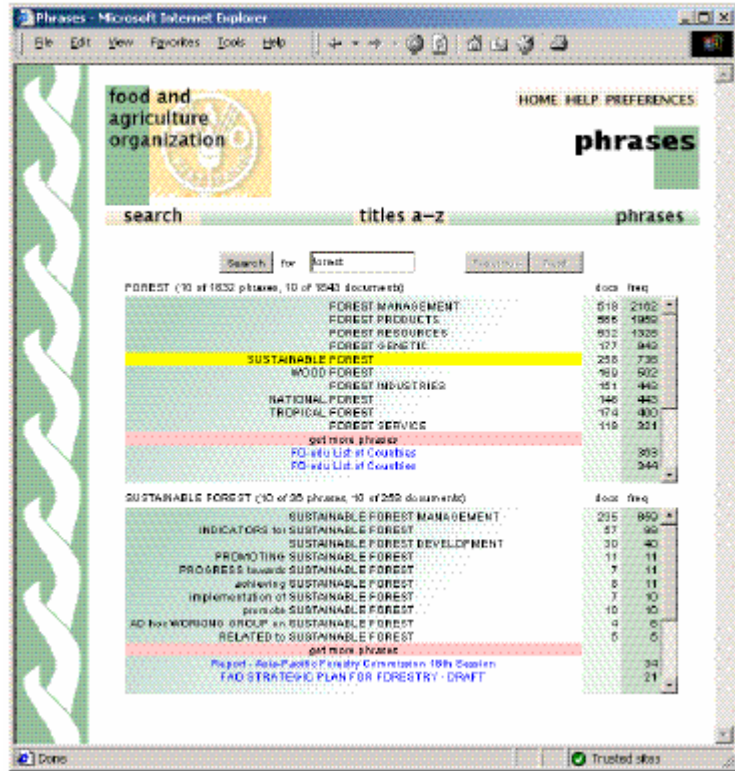


圖 2-9 : Phind 的瀏覽介面

## 2. 研究方法：

利用一些檢索任務的設計，配合實驗室觀察法和錄影機將受測者的一舉一動記錄下來，並以問卷調查法來收集各項檢索任務互動結果後的資料，最後再輔以訪談法將受測者的心理感受表達出來。

## 3. 實驗對象：

受測者共有 12 個學生，7 個是研究生；三分之一的受測者為男生，其學科背景為資訊科學及管理學的相關領域。

## 4. 檢索(電腦)經驗：

### 1) 使用電腦的經驗：

所有受測者每天皆有超過一小時使用電腦的經驗，只有一位受測者沒有。

### 2) 使用檢索的經驗：

所有受測者皆有使用關鍵字搜尋的經驗，但是只有 9 位受測者曾經接觸過數位圖書館的相關事務。

## 5. 檢索任務：

共有七個檢索任務，其檢索任務可分為：探索性的問題(Exploration Question)和專指性(Specific Retrieval Tasks)的問題。檢索任務範例如下：

### 1) 探索性的問題(Exploration Question)：

Find out more about national forest programmes in different countries.

### 2) 專指性的問題(Specific Retrieval Tasks)：

Where can golden apple snails be found  
Where was the locust numbers situation during May in Kuwait  
What does APFSOS stand for

## 6. 資料集：

實驗的資料較具專門性且以聯合國的食物暨農業組織(The Web site of the Food and Agriculture Organization (FAO) of the United Nations)的網頁資料為主，資料格式以網頁為主，共包含 21,700 個網頁及相關的檔案(如圖檔及 PDF 的文件格式)，約 140 萬個字。

## (三) 分析方式：

實驗結果從檢索任務的成效評估(Task Performance)、Phind 介面的評估(The Phind Interface)以及 Phind 可用性評估(Phind Usability Issues)三項分析之，其評估方式不單僅是做到單一檢索的評估，更是兼具全面的系統評估。

### 1. 檢索任務的成效(Task Performance)：

利用特定的檢索任務來評估單一檢索的檢索成效，且在檢索任務又可細分為下列幾個面向來評估此介面：

### **探索性的問題(Exploratory Questions)**

此部分以七個檢索任務的設計來進行分析。

### **專指性(Specific Retrieval Tasks)**

此部分由於檢索任務較具專指性，而此專指性可以測試系統之功能是否可以達到幫助使用者的目的，也就是利用檢索者是否有在特定的時間完成檢索任務的程度來分析之，將這些結果利用檢索後的訪談收集無法完成此檢索任務的原因及所遭遇到的困難。此實驗結果發現大部分的受測者都無法利用 The Phind Interface 順利完成較具專業性的檢索任務。筆者分析可能有下列幾點因素：

- 1) 受測者的專業背景皆不是屬於此一領域。
- 2) 介面的設計在可用性(Usability)還不夠成熟以及不熟悉此一介面的操作。

### 2. Phind 介面的評估(The Phind Interface)：

利用檢索後的問卷收集受測者對此一介面之設計的觀感，且經過整理摘要之後，歸納出幾個具代表性的問題，由於這些都是受測者在使用過後的感想，所以其建議對系統的設計有很大的幫助。

而相關的建議如下：

- 搜尋的詞彙應該不受限於單一詞彙或字。
- 當我要查詢兩個不同的主題時，其介面的設計可能會造成視覺的混淆。

### 3. Phind 可用性評估(Phind Usability Issues)：

在可用性評估方面，大都是以受測者使用系統之後的互動表現來分析，可用性評估中只要對使用者實施測試工作，如：檢索任務的測試及回答一些實驗性問題(Exploratory Questions)，即可對於網頁或系統的設計是否合乎一些重要的可用性原則，而利用這些可用性原則來進行評估；在測試中所顯示出的問題，也能立即被應用於重新設計，或是用作進一步的可用性測試與以使用者為中心的系統設計研究之中。經由可用性評估作者所歸納出下列幾個面向來探討：

1) Two Window Display

由於 Phind 的介面分為上下兩個視窗呈現，也就是子階層的片語會在點選之後於另外一個視窗開啟。而文件的連結則是另外開一個新視窗。

上述的介面在剛開始實驗時許多受測者都反應不知道系統已經回應了，尤其是以文件的連結視窗最容易被人所忽視。

2) Result Panels

有些受測者會一直認為下面的視窗永遠是上面視窗的子階層。這可以說是受測者的習慣所影響。

3) Title Display

標題展現的方式及在頁面的位置，是否會影響到受測者在介面的可用性程度。

4) User Feedback from Request

檢索結果的呈現是利用滑鼠在標題點兩下的連結方式，受測者往往不知道檢索結果的文章已經在另一個頁面開啟。所以在實驗過程中，有受測者建議應該利用語音的方式提示受測者文章結果已經開啟的訊息，這樣可以避免受測者重覆開啟頁面。

5) Navigation

在網頁的可用性評估方面來說，航行(Navigation)在資訊檢索系統而不迷失方向，一直是廣受討論的問題。因此，在這方面，實驗者特別針對其系統的各項按鈕及提示功能進行分析與評估，讓使用者來決定及設計模式。

三、 小結：

從上述的案例中可以發現，無論是 TREC Interactive Track 的評估實驗，或是自行設計的評估方法在互動式檢索系統的評估方式可以說慢慢兼具量化與質化的設計，且愈來愈注重與人互動之後的使用者評估。因此，無論是自動摘要、自動索引典或自動關聯提示的高階主動提示式檢索功能，其設計的原理都是在滿足使用者的資訊需求。是故，評估的方法只要是朝著上述的原則去執行，整體而言，就可以彌補「系統的設計者常常不是系統使用者，系統的使用挫折和不便常常無法為系統設計者所洞悉」的缺點<sup>31</sup>。下列將分析前面幾節介紹的互動式檢索系統之評估方法與設計，以及最後輔助檢索功能在檢索成效的分析，利用表格的方式來呈現，以利於後續相關研究之參考，分述如下：

在表 2-7 中，主要是說明與檢索者互動的檢索功能之設計，可以粗略分為兩部分，一是以介面的設計為主；另外一種則是偏重在動態詞彙的提示及摘要。介面的設計主要是利用資訊視覺化(Information Visualization)，打破傳統以條列式呈現文字檢索系統的檢索結果，如在表 2-7 中的 Huddle Search 就是利用資料夾(Folder)階層式的形式，將檢索結果利用動態的歸類將文件分類；而 The Phind Interface 的設計則是將傳統紙本的索引典瀏覽方式，利用 Java 程式語言將檢索結果分為上下視窗以方便使用者瀏覽(見圖 2-7)。至於在動態的辭彙提示及摘要方面，詞彙提示主要以動態分類及關聯詞提示為主，如在表 2-7 中的 HuddleSearch 和 Keyphind 都是利用自動歸類的方法動態產生分類目錄，有別於以往的人工分類目錄(如雅虎)，且其最大的優點就是可以節省人力及時間。

表 2-7：互動式輔助檢索功能的設計

互動式系統		互動式功能
TREC Interactive Track	Glasgow-HuddleSearch	文件階層式歸類、自動摘要。
	Rutgers University	檢索詞彙的長短、檢索結果文件的呈現方式。
Keyphind		自動歸類、視覺化的呈現方式。
The Phind Interface		階層式片語瀏覽(Hierarchical Phrase Browsing)。

資料來源：本研究整理。

表 2-8、2-9 內主要將文獻中提及互動式檢索系統評估實驗設計，將其整理成表格以方便分析。在表 2-8 主要可以分為兩大部分，第一部分就是 TREC

<sup>31</sup> 同註 1。



Interactive Track 的評估實驗設計之研究方法；另一部分則是非參與 TREC 互動式評估的檢索系統。其實，無論是否參與 TREC 評估實驗的互動式檢索系統，其主要的實驗設計都是在比較有無互動式檢索輔助功能之系統設計的異同，因此，大部分都是利用實驗室控制法，將系統分為實驗組和對照組，再依實驗的環境輔以各種問卷、放聲思考法、觀察法及訪談法互相搭配。

表 2-8：互動式檢索評估方式-評估實驗設計之研究方法

互動式系統		研究方法
TREC Interactive Track	Glasgow-HuddleSearch	利用問卷調查法、觀察法及實驗室控制法來收集資料，且實驗環境皆有兩個系統做比較。
	Rutgers University	
Keyphind		觀察法、放聲思考法、問卷調查法及訪談法，且實驗環境皆有兩個系統做比較。
The Phind Interface		觀察法、問卷調查法及訪談法，其實驗環境皆有兩個系統做比較。

資料來源：本研究整理。

表 2-9 中主要是將非參與 TREC 互動式評估實驗的檢索系統，就「實驗對象」及「檢索任務」的差異性列出來。由於 TREC 的實驗對象並無嚴格限制，但是至少需 16 人以上，推測其原因可能是這樣較符合現實環境中的使用者；還有 TREC 檢索任務的制定，受限於實驗的資料集規定為.GOV 網站的資料，所以檢索任務都是相同的。是故，在此只做非參與 TREC 評估實驗的比較。從表 2-9 可以得知「實驗對象」大都沒達到統計學上至少 30 人的要求，筆者分析可能是其評估實驗，主要的面向為使用的評估，用意在於得知使用者對系統的直覺反應，此反應並不適合利用數字來代替一切，因此，收集的資料較具質化的傾向，所以實驗的人數不受限於統計學上的限制；此外，實驗對象還有一個較特別的設計，就是利用使用者對資料集特性不了解的情況下直接進行評估，如表 2-9 中 The phind Interface 資料內容是以聯合國的農業暨食物網站的資料為主，而實驗對象學科背景皆非具備此相關知識的人，探究其原因，實驗設計者可能為了讓自行設計之檢索系統的功能更具說服力，即當使用者對不熟悉的領域進行檢索時，還能夠找出所需的資訊。

在「檢索任務」方面，參與 TREC 評估實驗的檢索系統，大都遵循其訂定的原則，而檢索任務則可以分為三大類：政府規章及法條、健康資訊、專案企劃及旅遊資訊等；各參賽系統之檢索任務的差異性較低，原因是為了讓參與此評估的系統，能夠在相同條件之下比較其系統功能的差異性。至於在非參與 TREC 的評估實驗之系統，檢索任務制定的原則，大都有探索性任務(Exploration Task)的設

計，分析其原因，可能是此種設計較符合現實環境中，使用者有時候並無特定目的，只是想找自己有興趣的主題瀏覽。此外，在 Keyphind 的檢索任務其中一個原則-「涵蓋面評估」，可能是為了配合資料集涵蓋面廣的特性，所制定的主題性任務，藉此利用不同主題的範圍來降低檢索者對各主題熟悉的程度不同而產生的判斷偏差。最後 The Phind Interface 中專指性任務設計之目的，主要是因為資料集較具專業性，因此，為了模擬現實環境中，當檢索者不具備此一專業知識的資訊需求者，而對此檢索系統有資訊需求的動機，用以看出檢索功能設計是否有達到輔助不具備專業知識檢索者的目的。

表 2-9：非參與 TREC Interactive Track 的評估實驗設計

互動式系統	實驗對象	檢索任務
Keyphind	10 位自願參與此檢索實驗之電腦科學系的大學生	其檢索任務設計的原則如下： 涵蓋面評估任務(Coverage Evaluation Task) 探索性任務(Exploration Task)。
The Phind Interface	受測者共有 12 個學生，7 個是研究生，三分之一的受測者為男生，其學科背景為資訊科學及管理學的領域。	探索性的問題(Exploration Question) 專指性的任務(Specific Retrieval Tasks) 共七個檢索任務。

資料來源：本研究整理。

至於在表 2-10 互動式檢索評估的分析方式，只要是參與 TREC 互動式評估實驗的系統，其分析方式會將問卷所收集的資料，依據下列三大原則進行分析：檢索任務完成之效能、檢索任務完成效率、及使用者滿意度分析。最後利用統計學上的方式，將兩系統間做一個多面向的比較。在非參與 TREC 互動式評估的 Keyphind 則利用涵蓋面評估來分析檢索任務會不會因為主題的不同而影響其檢索功能；探索性評估則讓使用者不限類別下，依照自己感興趣的主題去檢索。至於 The Phind Interface 除了有檢索任務的成效評估之外，還從介面上的評估及可用性評估方面去分析，與其他評估方式最大的不同是兼具面的系統評估。

表 2-10：互動式檢索評估方式-分析方式

互動式系統		分析面向
TREC Interactive Track	Glasgow-HuddleSearch	檢索任務完成之效能(Task Completion) 檢索任務完成之效率(Task Times) 使用者滿意度(User Satisfaction)
	Rutgers University	
Keyphind		涵蓋面評估(Coverage Evaluation Task) 探索性評估(Exploration Task) 可用性評估(Usability)
The Phind Interface		檢索任務的成效評估 (Task Performance) 介面的評估(The Phind Interface) 可用性評估(The Phind Usability Issues)

資料來源：本研究整理。

最後，就其互動式檢索功能在輔助檢索者的檢索成效進行分析。整體而言，系統設計者所研發的設計功能，經由互動式評估之後，都可以有效的幫助使用者縮短查詢的時間，且對系統輔助檢索功能的滿意度也都比傳統的全文檢索系統高；簡而言之，就是達到輔助檢索者的目的。但是，系統設計者不能因此而滿意，因為，就單一檢索效能的分析來說，不是所有項目的檢索效能評估皆以有輔助檢索功能的系統功能較優越，如：The Phind Interface 在其介面評估中就指出，使用者因為需先適應新系統的介面設計，必需花更多的時間去學習，縱使能夠有效的提示使用者相關詞彙，但是因為習慣性的問題，最後還是會選擇繼續使用原來的系統。分析其原因，個別的檢索成效會因為實驗的環境、實驗的資料、個人的檢索背景、習慣性及檢索任務的專指度等因素，而無法達到輔助檢索者的目的，因此，所呈現的使用者滿意度方面，反而是沒有提供輔助檢索功能的檢索系統之滿意程度較高。以下僅列出各系統輔助檢索功能的檢索成效之綜合分析（見表 2-11）：

表 2-11：各輔助檢索功能的檢索成效綜合分析

輔助檢索功能系統	分析面向	檢索成效分析結果
Glasgow-HuddleSearch	效能	輔助檢索功能的檢索失敗率較低。
	效率	影響檢索的效率不僅受制於系統功能的差異，還會受檢索任務的不同而產生差異。但就檢索時間點上來分析，有明顯的降低。
	使用者滿意度	有輔助檢索功能系統的滿意度沒有明顯的提高。
Rutgers University	各項分析皆顯示有提供輔助檢索功能的效益較高。	
Keyphind	方便使用	由於採二元化的比較方式，所以分析結果僅能主觀的反應出使用者對系統的喜惡。但是，就使用的難易度方面，沒有使用者認為一般的全文檢索系統較容易使用。
	搜尋策略	使用者會因為新增的輔助檢索功能，使得搜尋策略需多增加判斷的步驟，因此，所花費的時間較一般的全文檢索系統較長。
The Phind Interface	檢索任務的成效評估	發覺大部分的使用者皆無法利用此功能完成專指性的檢索任務。
	介面的評估	當使用者不只是查詢一個主題時，上下分割的畫面會容易造成視覺上的混淆。
	可用性評估	使用者在可用性評估中，反應習慣性問題會影響使用者使用輔助檢索功能的最大動機。

資料來源：本研究整理。

### 第三章 研究方法與設計

有關互動式檢索評估的研究方法，根據 TREC Interactive Track 及相關文獻分析，大都採用下列幾種方法：1.問卷調查法 2.觀察法 3.放聲思考法 4.訪談法 5.實驗室控制法。本研究則採取實驗室控制法配合問卷調查法、放聲思考法及訪談法收集檢索者對互動式檢索功能設計的反應與觀感。至於在關聯提示詞的成效評估方面，則以資訊檢索較常用的「相關判斷」來檢視其自動提示之關聯提示詞的成效，而相關的判斷則採用主題相關的概念來評估；因此，這裡的相關判斷是可以行諸於文字。最後，利用 TREC 的評估程式(trec\_eval)計算不同大小的文件集間之輸出結果的數據。

本章共分以下章節來討論：第一節研究流程與架構；第二節說明研究對象，第三節介紹研究之設計；第四節說明研究中運用之研究方法；第五節則介紹使用之研究工具；第六節則說明資料之分析與處理。茲分述如下：

## 第一節 研究流程與架構

本研究流程與架構如(圖 3-1)所示：

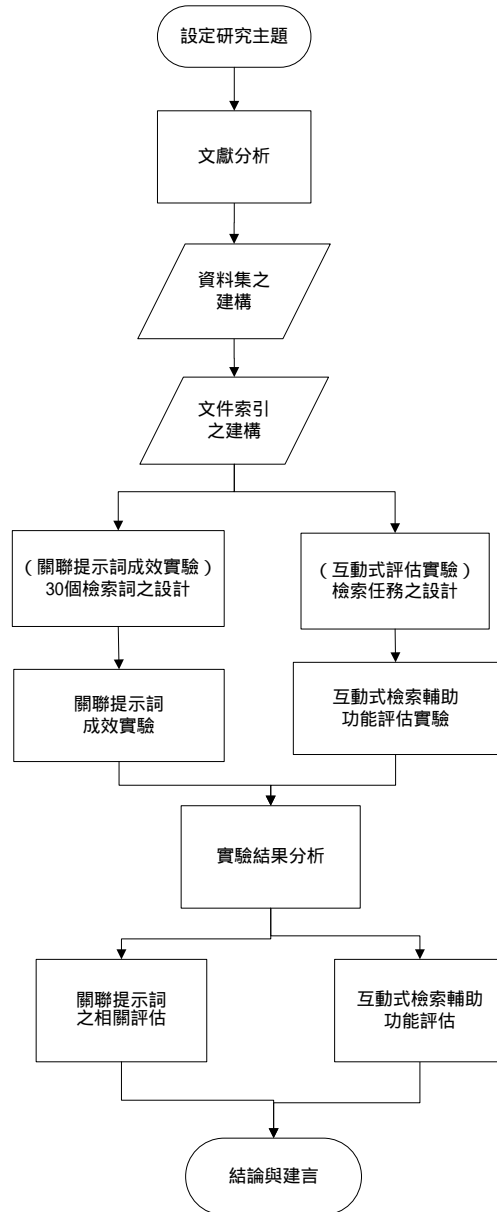


圖 3-1：研究流程與架構

各步驟詳細內容如下：  
選定研究主題與方向。

**文獻分析：**首先，蒐集現有有關互動式檢索輔助功能之設計與評估方式，並對國內外相關之系統進行探討與分析，選擇適合本研究之環境來作為實驗之依據。

**資料集之建構：**由於資訊檢索系統的評估必須具有一定程度的客觀性，因此，需模擬現實中的環境才能對系統進行有效的評估。所以本研究在進行效能評估實驗前，必須選擇適合之資料集來作為測試的環境。是故，本實驗選擇 NTCIR 的新聞資料集作為實驗的測試資料，資料集之細節及內容會在研究設計做詳細的闡述。

**文件索引之建構：**選擇適合之資料集後，利用本實驗系統 (WebGenie) 將上述的新聞資料分別在不同的電腦上建構大、中、小的索引資料，當索引建製完成之後就是實驗之研究環境。

**檢索詞與檢索任務之設計：**本研究之兩大實驗為「關聯詞成效評估」與「互動式檢索輔助功能評估」，前者必須先選擇系統中可以檢索到的檢索詞彙來作為檢索詞，再利用相關判斷讓使用者檢視系統所提供的關聯詞之相關度。至於在互動式輔助檢索評估方面，必須模擬現實環境中使用者的資訊需求，也就是說需建構一些檢索任務，讓實驗可以在相同的環境做比較與分析，因此，實驗之檢索任務依新聞資料的分類，選擇 NTCIR 中跨語言檢索的正式檢索問題。

**關聯提示詞成效與功能評估實驗：**設計完上述的檢索環境、檢索詞及檢索任務之後，分別執行「關聯詞成效評估」與「互動式檢索輔助功能評估」的實驗，前者選擇 5 位具有圖書資訊學背景的研究生作為相關判斷之研究對象。至於後者之實驗，依據 TREC 及相關文獻之分析，另外選擇 16 位作為互動式功能評估的實驗對象，實驗中配合各項研究方法來蒐集資料。

### **實驗結果分析：**

#### **(一) 關聯提示詞之相關評估：**

在此項實驗中，利用相關評估法收集檢索者對檢索詞與關聯提示詞之關聯程度，且利用 TREC 所發展的一套工具軟體—trec\_eval 計算出多種查準率與查全率的測量值，最後，再利用統計學上的 T 檢定來檢定是否因為資料集的大小而影響系統提供之關聯詞的品質。

#### **(二) 互動式檢索輔助功能評估：**

在互動式評估中，則利用問卷調查法、放聲思考法所收集到的資料，分別從效能、效率及使用者滿意度三方面來分析，再利用訪談法所收集到的訪談內容，依可用性評估的原則歸納出使用者對系統的回饋及建議。

**結論與建義：**根據這些結果歸納出影響自動關聯詞成效的品質與使用者對此互動式檢索輔助功能的使用情形。以作為未來互動式檢索輔助功能開發者在設計或改良系統時的參考和依據。並且在考量系統的效能之餘還能兼顧使用者之需求的檢索系統。

## 第二節 研究對象

本實驗其中一個目的就是對檢索詞做相關的判斷，而研究對象則徵求五位修習圖書資訊學的研究生，作為相關判斷的判斷者，為了讓相關判斷的過程能夠有一致性，這五位相關判斷者都是輔仁大學圖書資訊學系碩士班的研究生。

另一實驗則是評估互動式檢索輔助功能的成效。為了讓實驗能夠看出一致性及主要使用者(Heavy User)的差異，所以本研究的對象可分為兩群：一群是具備資訊檢索經驗且受過相關資訊檢索訓練的教育，如：曾經接受過「資訊檢索」、「資料蒐集」等相關課程的訓練，這些使用者我們以輔仁大學圖書資訊學系碩士班的研究生為主，共 8 位；另一群則是未曾接受過資訊檢索相關課程訓練且檢索經驗在 4 年以下者，鎖定在非圖書資訊學系背景且大學部的學生為主，共 8 位。參與此次實驗的人數共 16 位，並採用便利抽樣法，徵求參與意願高的受試者，以觀察他們在使用此互動式檢索結果的功能時，所遇到的問題和困難，及檢索時需求行為的質化分析。

評估互動式輔助檢索之功能的研究對象大略可分為兩群，如下：

### 圖書資訊學領域背景：

- 一、以接受過專業資訊檢索訓練之圖書資訊學領域的背景為主。
- 二、需具備 5 年以上使用過任何檢索系統的經驗(包含各類搜尋引擎及資料庫)。
- 三、以輔仁大學圖書資訊學系碩士班的研究生為主，共 8 位。

### 非圖書資訊學領域背景：

- 一、未曾接受過相關資訊檢索課程且非圖書資訊領域的檢索者。
- 二、以大學部的學生為主，共 8 位。
- 三、學科背景包含，經濟系、廣告系、法律系、新聞傳播學系、體育系及日文系，皆未曾接受過相關圖書資訊學課程的訓練。

## 第三節 研究設計

本研究設計分別從量化及互動式方法來評估本系統之關聯提示詞的成效，量化方面利用資訊檢索上的「相關」判斷，評估系統自動產生的詞彙和檢索詞的相



關度(Term Relevance Evaluation)；互動式評估方面則以 TREC Interactive Track 上的問卷設計作為依據，輔於放聲思考法及深度談法來觀察，記錄研究對象所使用之檢索敘述、檢索過程中遭遇之困難；而本研究之實驗過程也以驗證下列兩個研究假設作為設計之準則：

- (一) 大文件集所產生之關聯提示詞的關聯程度會優於小文件集。
- (二) 使用互動式檢索功能(關聯提示詞)的檢索效能會優於沒有使用互動式檢索功能的效能。

分述如下：

一、收集與研究主題相關的文獻，以了解應用於評估關聯提示詞之方法及互動式檢索過程的成效評估。之後，將文獻歸納整理，以作為本研究的理論架構與進行時之參考。

二、資料集之建構：

本次實驗主要採用的資料集為日本國立情報學研究所(National Institute of Informatics, 簡稱 NII)舉辦之 NTCIR 評估會議，其中的中文檢索(CHTR)競賽所使用的中文資訊檢索測試集 1.1 版(CIRB011, Chinese Information Retrieval Benchmark version 1.1)和 2.0 版(CIRB020, Chinese Information Retrieval Benchmark version 2.0)；而 CIRB011(見表 3-1)皆下載自五個新聞網站於 1998 年 5 月至 1999 年 5 月間的報導，這些新聞包括：中國時報、工商時報、中時晚報、中央日報以及中華日報。CIRB020 則下載自聯合新聞網站於 1998 年 1 月至 1999 年 12 月底的報導。選擇新聞文件作為評估的資料集，主要是取決於新聞文件內容主題豐富且符合現實生活的環境，其評估出來的結果可信度較高。至於另外一個 pub 的資料集，其資料內容為中國時報和中央日報的新聞資料，收錄的年代為 2000 年和 2001 年。

NTCIR 文件格式，包括：文件來源識別碼、新聞報導日期、新聞標題、新聞內容等，為使系統易於辨識，每篇文件都具有相同的格式與資料項目。文件採用 BIG5 編碼與 XML 型態的標記，將標記加入文件之後便可以辨識文件的特殊區域。文件範例如圖 3-2 所示。

表 3-1 及表 3-2 詳列文件集的數量與來源，CIRB011 共 132,173 篇文章，大小約 200MB 左右，其中以中國時報的資料佔了此資料集的 1/4，而此資料集大致上和 NTCIR Workshop 2 一樣，唯一不同的地方應該是在資料的格式。

表 3-2 中的 CIRB020 則收錄了 249508 篇的新聞資料。文件有各式各樣的主題，包括：政治、財金、社會綜合、生活、體育、娛樂、國際、資訊、科技等。<sup>3233</sup>

表 3-1 : CIRB011

新聞媒體	文件數	百分比
中國時報(Chinatimes)	38,163	28.8%
工商時報 (Chinatimes Commercial)	25,812	19.5%
中時晚報(Chinatimes Express)	5,747	4.4%
中央日報(Central Daily News)	27,770	21.0%
中華日報(China Daily News)	34,728	26.3%
Total	132,173	(200MB)

資料來源：陳光華，「資訊檢索系統的評估 - NTCIR 會議」，國立台灣大學圖書資訊學系四十週年系慶學術研討會論文集，台北（民 90 年），頁 73。

表 3-2 : NTCIR 中跨語言檢索的 CIRB011 &amp; CIRB020 文件集

CIRB011 (1998-1999): Chinese	132,173
CIRB020 (1998-1999): United Daily News (1998-1999): Chinese	249,508

資料來源：Kuang-hua Chen, "Introduction to Chinese Document Set for CLIR Task of NTCIR Workshop 3". 2003

<sup>32</sup> 陳光華，「資訊檢索系統的評估 - NTCIR 會議」，國立台灣大學圖書資訊學系四十週年系慶學術研討會論文集（民國 90 年 11 月），頁 67-86。

<sup>33</sup> Kuang-hua Chen, "Introduction to Chinese Document Set for CLIR Task of NTCIR Workshop 3". 2003

```

<DOC>
<DOCNO>cdn_chi_19980508_0002</DOCNO>
<LANG>CH</LANG>
<HEADLINE> 張榮恭強調江澤民所提不能視為統一時間表 </HEADLINE>
<DATE>1998-05-08</DATE>
<TEXT>
<P>
針對外傳中共將在全國對臺工作會議中以「促進中國完全統一」做為未來五年的重大政治任務，中國國民黨大陸工作會主任張榮恭昨日則表示，這項說法只不過是一種政治口號和目標而已。
</P>
<P>
據香港「星島日報」引述北京的消息來源報導，中共「國家主席」江澤民預定於即將召開的全國對臺工作會議中，在「江八點」的基礎上提出可操作性的「促進中國完全統一」的新對臺政策，並將把「完全統一」列為中共今後五年的重大政治任務。
</P>
<P>
張榮恭表示，中共已故元老鄧小平在八十年代會說希望實現中國統一，而中共領導人在一九九〇年的全國對臺工作會議中也講過類似「號召統一大業」的話，因此，不必解讀為中共對統一的具體時間表，而僅是一種政治號召。
</P>
<P>
不過，他認為，中共預定五月十一日起在北京舉行全國對臺工作會議，身兼中共中央對臺工作小組長的江澤民極有可能在這次會議上發表有關對臺問題的重要談話。
</P>
</TEXT>
</DOC>

```

圖 3-2：文件標記範例

### 三、關聯提示詞之建構：

利用自行研發的 WebGenie 知識檢索系統自動建構關聯提示詞，分別將不同大小的文件集建構索引，而文件集依大小可分為三種索引模式(檢索的資料集)：

- (一) 25,233 篇新聞文件集，檔案格式為 html，年代為 2000 年至 2001 年。
- (二) 原索引 25,233 篇的新聞文件加上 NTCIR 的跨語言檢索中的中文新聞文件集 CIRB011(見表 3-1)共 154,720 篇新聞文件，檔案格式為 xml，年代為 1998 至 2001 年。
- (三) 上述兩種索引的新聞文件集再加上 NTCIR 的跨語言檢索中 CIRB020(見表 3-2)聯合新聞網(United Daily News, UDN)的中文文件集，共有 406,953 篇新聞文件，年代從 1998 至 2001 年。

上述三種資料為了符合 WebGenie 建置索引的格式，我們利用 Perl 程式將原先資料格式為 html 和 xml 格式的資料轉為以逗點隔開的.csv 檔(見圖 3-3)的格式，其欄位可分為 ID 代表文件編號、Date 代表文件日期、Title 代表文件標題及 Text 代表文件內容，以方便系統建置索引；接著每隔一萬篇切為一個附檔名為.csv 的檔案格式，再將轉檔過的資料依上述三種索引格式，分別在三台個人電腦建立上述三個大中小的索引檔。

索引建置完成之後將檢索詞輸入，系統會自動產生 64 個關聯提示詞，再從中挑選前 50 個關聯提示詞作為檢索詞彙判斷相關的提示詞彙。

ID, Date, Title, Text
"cdn_chi_19980508_0001", "1998-05-08", " 港立法會選舉「中」英較勁搬上檯面 ", " 中共當局昨日對英國駐香港總領事館官員約談第一屆

圖 3-3：.csv 的檔案格式

#### 四、檢索任務之設計<sup>34</sup>

本研究之檢索任務綜合 NTCIR CLIRFormalRun Topic-CH98 和 TREC Interactive Track 上的設計作為本實驗之藍圖，且以加拿大 Saskatchewan 大學裡的人機互動實驗室<sup>35</sup>(HCI the interaction lab)所做的可用性評估方式為輔，再適時調整以符合本實驗之研究任務。由於 Interactive Track 上資料集的語言以英文為主，但是，本實驗之研究環境為中文的線上新聞資料，且又受限於年代的限制，所以必須將 Interactive Track 上的檢索任務重新編碼，再者，NTCIR 的正式檢索問題具有明確的指示和答案，是故，決定採用上述兩大評估會議的檢索任務之設計為原則。

而互動式評估主要是了解自動關聯詞能否有效輔助使用者做進一步的查詢；在此，因為資料集涵蓋面廣與具有時間性的限制，所以檢索任務的設計，是依據下列兩項原則來評估，可分為涵蓋面評估(Coverage Evaluation Task)及探索性評估(Exploration Task)。前者在於測試議題之涵蓋面是否影響「關聯提示詞」對使用者的輔助程度及差異，所以在涵蓋面上，會事先設計八個檢索任務，讓檢索者在檢索題目範圍不同的情況下使用本檢索系統，最後利用檢索後問卷(Post-search Questionnaire)，檢視使用者對系統所提示之關聯詞在滿意度及對問題信心指數是否有明顯的差異；在探索性評估上，則是讓使用者於不限類別的情況下，依照使用者最感興趣的議題檢索文件，共一個檢索任務(見附錄一)。<sup>36</sup>

為了避免使用關聯提示詞的先後而影響檢索者的心理認知，在涵蓋面評估中，先將 8 個檢索任務隨機分為兩群，以系統 A 來代表有關聯提示詞的功能；系統 B 則代表沒有關聯提示詞的功能，再將參與實驗之研究對象分為兩組，每組各 4 位非圖資背景及 4 位圖資背景的受試者，共 8 位；隨機抽取 4

<sup>34</sup> 邱立豐，「互動式概念查詢應用於網路文件自動摘要之效益」(雲林科技大學資訊管理系碩士班，碩士論文，民 91 年)，頁 41-49。

<sup>35</sup> "Interaction lab : University of Saskatchewan ," Retrieved December 4, 2003, from <http://hci.usask.ca/index.xml>.2003

<sup>36</sup> 同註 35。

個檢索任務先做系統 A 再做系統 B(見表 3-3), 另一組則先做系統 B 再做系統 A(見表 3-4), 此方法是為降低檢索者先使用某系統之後所造成的偏差。

至於在探索性評估方面, 由於檢索任務是由檢索者自行制定, 因此, 16 位實驗對象會產生出 16 個檢索任務, 無法用不同的題目互相比較, 所以只能設法降低使用系統的先後順序所產生的偏差進行設計。原則上, 實驗的對象分為兩群, 每群各 4 位非圖資背景及 4 位圖資背景的受試者, 共 8 位; 實驗進行的方式, 前 8 人先做系統 A 再做系統 B, 後 8 人則先做系統 B 再做系統 A(見表 3-5) 將上述產生出的分析結果, 經由敘述統計的方式對一個現象做一個簡單的描述。

無論是涵蓋面評估或探索性評估的檢索任務都沒有時間的限制, 其最大的目的就是模擬一般的檢索環境, 讓檢索者自己評定是否已經找到答案。

下列說明表格內各符號所代表的意義:

ABn1-1: 表示一般使用者先做系統 A 再做系統 B 來執行檢索任務 1 的代號。

A: 表示系統 A。

B: 表示系統 B

AB: 先做系統 A 再做系統 B

n1: 表示非圖資背景編號為 1; n 表示 novice 的意思。

h1: 表示具備圖資背景使用者編號為 1; h 表示 heavy user 的意思。

-1: 檢索任務 1 (共有八個檢索任務)

表 3-3: 涵蓋面評估先做系統 A 再做系統 B 的前 8 人檢索任務分配

系統	系統 A	系統 B
群一 (8 人)		
第 1 至 4 人	ABn1-1、ABn2-2 ABh3-3、ABh4-4	ABn1-5、ABn2-6 ABh3-7、ABh4-8
第 5 至 8 人	ABn5-5、ABn6-6 ABh7-7、ABh8-8	ABn5-1、ABn6-2 ABh7-3、ABh8-4

表 3-4：涵蓋面評估先做系統 B 再做系統 A 的後 8 人檢索任務分配

系統	系統 B	系統 A
群一 (8 人)		
第 9 至 12 人	BAn9-1、BAn10-2 BAh11-3、BAh12-4	BAn9-5、BAn10-6 BAh11-7、BAh12-8
第 13 至 16 人	BAn13-5、BAn14-6 BAh15-7、BAh16-8	BAn13-1、BAn14-2 BAh15-3、BAh16-4

表 3-5：探索性評估的檢索任務分配

系統	系統 A	系統 B
群一 (8 人)		
第 1 至 8 人	AB (4 位圖資, 4 位非圖資)	AB (4 位圖資, 4 位非圖資)
第 8 至 16 人	BA (4 位圖資, 4 位非圖資)	BA (4 位圖資, 4 位非圖資)

#### 五、詞彙和查詢語句的相關度評估設計(Term Relevance Evaluation)<sup>37</sup>：

本文所做的成效評估，主要是了解檢索詞與提示的關聯詞之間的關聯(Relatedness)情況是否受資料集的大小及時間所影響。本實驗承襲曾元顯之前的關聯詞相關性評估的做法。我們以兩種方式評估：

- (一)直接計數前 N 個關聯提示詞有哪幾個被每位受試者判定為相關，以統計關聯詞數的比例；
- (二)以資訊檢索常用的查準率與查全率評估哪一種文件數量的效果較好，再利用 TREC 所發展的一套工具軟體—trec\_eval，計算多種以查準率和查全率為基礎之指標的測量值，以對系統之檢索結果加以評估分析。

首先，我們邀請五位具有圖書資訊學背景的研究生，由於是為了探討資料集的大小影響關聯詞的情況，因此，在檢索詞方面將沿用曾元顯之前實驗由五位受試者從索引詞庫內挑選出的 30 個檢索詞彙(見表 3-6)，一併檢索

<sup>37</sup> 同註 3。

分別由大文件集共約 406,953 篇新聞文件，中文件集約 157,420 篇，小文件集約 25,233 篇即是曾元顯老師所實驗之資料集。為了避免同樣的檢索詞因為判斷的先後順序而影響關聯詞的成效，會在受試者進行完上述大中小的資料集實驗之後，再一併檢視由大中小產生的 50 個關聯詞，最後製作成一個 150 個關聯詞的詞彙列表，再請受試者判斷詞彙與查詢語句之間的關聯程度。最後將全部 30 個檢索詞的結果平均起來。

在這裡的成效評估是直接計數前 N (在此 N=50)個關聯提示詞有哪幾個被受試者判定為相關，以統計關聯詞數的比例。判斷方式說明如下：本實驗將相關的程度分為三個等級來評分，0 分為不相關(Irrelevant)；1 分為部份相關(Somewhat Relevant)；2 分為非常相關(Most Relevant)。再利用統計學上的配對樣本 T 檢定(Paired T-test)來檢定本實驗的假設一。

表 3-6：之前實驗的 30 個檢索詞彙

中東	地震	亞洲國家	通航	債券	環保署
中油	有線電視	邱義仁	博物館	奧運	職棒
主計處	朱鎔基	國安基金	晶圓代工	會計師	醫生
古蹟	李安	國科會	鄉鎮	調查局	顏慶章
生物科技	那斯達克	被害人	雅虎	選民	黨員

#### 第四節 研究方法

資訊檢檢索系統往往利用查準率與查全率來作為測量系統的準則，嚴格來說，它只是從某個角度來測量系統的效益，並沒有達到真正的「評估」層次，而真正的系統評估尚包括了許多層面，如：系統回應時間、親和力、顯示格式、使用者對檢索的主觀滿意度等。因此，本研究配合上述的研究設計利用量化及質化的方法來評估互動式檢索結果的成效；首先，利用系統實作法實際建構關聯詞，且採用「相關評估法」評估各關聯詞與檢索詞之關連程度，再輔以資訊檢索上查準率與查全率作為量化的評估依據。互動式評估方面則採取實證方式的質化研究取向，過程中運用「問卷調查」(Questionnaire)的方式，儘量完整地蒐集檢索互動前後的使用者記錄；利用「放聲思考法」(Thinking Aloud Analysis)彌補問卷不足的地方，以受試者在實驗過程中自行陳述其所遇到的問題及內心的觀感；最後，「深度訪談法」(Depth Interview)蒐集測試後受試者對互動式檢索輔助功能使用的想法。本研究將收集的資料包括：

##### 一、相關評估法：

評估系統的優劣本來就應該根據使用者主觀的感受進行評估，所以本研究之相關判斷採用「主題相關」(Subject Relevance)的概念。換言之，是一種可以形諸於文字的主題相關。而在本實驗中，各檢索詞即代表檢索者心目中對欲檢索之問題的替代物，在這樣的概念基礎上，判斷者應客觀的將檢索詞與關聯提示詞做相關性連結，如此可增進不同判斷者所做判斷的一致性與可靠性。在此我們所做的成效評估，主要是了解檢索詞與提示的關聯詞之間的關聯(Relatedness)情況。在測量尺度方面，為了使判斷者能客觀，相關程度必須有較明確的定義。本研究將相關分為不相關、部分相關及非常相關，分別給予 0 至 2 分的相關分數。

## 二、問卷調查資料：

本研究所收集的問卷資料主要分為四部份：(一)基本資料(Entry-Questionnaire)：檢索者的基本資料及背景，如：電腦技能和檢索經驗等；(二)每個檢索任務完成後的問卷(Post-search Questionnaires)，其目的主要是調查此檢索者對此檢索任務的熟悉程度；(三)完成所有檢索任務後的問卷(Post-system Questionnaires)：同一檢索系統完成各項檢索任務後的觀感；(四)檢索實驗後問卷(Exit-search Questionnaire)：目的在收集兩系統間的比較及使用者對系統所回饋的資料。最後將檢索後的問卷利用社會科學統計軟體進行分析與驗證。(請參見附錄二~五)由於此問卷所收集之資料偏向二元化的答案且較具主觀性，因此，為了深入探究使用者對系統的回應及意見，特別搭配訪談內容，利用可用性評估的原則來做質化的分析。

## 三、放聲思考法的資料：

本測試方法乃是主測者給予某項任務，讓受試者於使用系統的固定時間內，一邊想一邊說出自己對於操作程序上的想法、親身感覺、好處、壞處，並在規定時間內完成主測者所給予的任務，主測者將受試者的人聲全程記錄下。本研究除了利用問卷收集檢索實驗中的資料外，還利用放聲思考法將檢索者在執行檢索任務過程中的直覺反應記錄下來，以收集檢索者對系統最直接的反應。

## 四、深度訪談資料：

利用深入訪談法來收集資料是為了避免量化研究所得之概約性數據，且因為數據無法完全表示資訊搜尋行為中人的因素，對於各種情境因素解釋力



不足的問題。訪談進行是採用和受訪者面對面相互交談的方式，在受試者完成上述所有任務之後，利用非結構性的訪談法，引導受試者回答一些可用性評估的問題，再配合錄音將訪問的內容記錄下來。本研究利用深入訪談來瞭解使用者對於提供關聯提示詞圖示之功能及介面的使用情形及遭遇的困難，而訪談之問題則根據 Jakob Nielsen 的可用性原則來設計。(訪談內容請見附錄六)

## 第五節 研究工具

本研究主要是探討關聯詞對檢索者的輔助程度去進行評估，而下列則是針對實驗之系統的建制過程及各項所需的工具，簡述如下：

### 一、WebGenie 知識檢索系統的建制簡述：

以 WebGenie 知識檢索系統中的關聯詞提示功能作為實驗之系統，而關聯提示詞的建構方法，主要可分為三個步驟：先進行文件的關鍵詞擷取，再進行關聯詞的分析，最後累積所有的關聯詞建成關聯詞庫，而上述三種方法即是所謂的共現索引典。我們將建制的過程簡述如下：

從以往的研究指出，各種自動文獻處理的技術都必須進行關鍵詞自動擷取的步驟，因此，我們根據一個假設先對文件進行關鍵詞的擷取，這個假設就是：如果文件探討某個主題，那麼應該會提到某些特定的字串好幾次；而電腦就是依照這種重複的特性來擷取關鍵詞，擷取的關鍵詞也沒有長度的限制，在融入部分的自然語言處理方法之後其查準率高達 96%。

進行完關鍵詞擷取之後，就開始進行關聯分析。其建構的方法就是計算單一文件中任意兩個詞彙『個別出現』與『共同出現』的句子數，再套用 Dice、Cosine、Jaccard 或 Mutual Information 等資訊檢索常用的相似度計算公式，而求得兩詞彙間的關聯度。計算出每一篇文件重要詞彙間的關聯度後，累積關聯強度超過某個門檻值的關聯詞，即可完成整個文件資料庫的關聯詞庫。圖 3-4 則是展示一篇文件的內容，利用上述方法自動擷取出來的關聯詞，其中關聯詞的則以二維圖示方式顯示詞彙之間彼此的關聯。

文件內容	關聯詞
<p>記者葛大維/綜合報導】北野武執導新片「菊次郎的夏天」，代表日本角逐今年坎城影展，這是北野武導演的第八部電影，他將此片獻給父親。曾主演「俘虜」的北野武，去年執導「花火」獲得威尼斯影展最佳影片金獅大獎，所以今年進軍坎城影展，備受矚目，據知「菊次郎的夏天」是今年競賽項目的大熱門。北野武不想太早將「菊次郎的夏天」劇情曝光，不過坦言片中那位不成熟、愛作夢的男人翁，是自己父親的塑像，而在「菊次郎的夏天」裏，北野武改變過去作品的暴力風格，趨於寫實，人物對白很多，刻劃人性細膩感人，被認為是北野武重要的突破。北野武認為執導「菊次郎的夏天」，並非想挑戰傳統，只是想從畫抽象畫，改變成畫寫實畫，目的希望更親近大眾。在日本橫跨電影、電視、廣告多棲的北野武，有過人的精力與細密的心思，所以在密密麻麻的工作中導演「菊次郎的夏天」，有條不紊，不過動用大批工作人員數量驚人，因為北野武車禍臉受過傷，光是幫他化妝、梳頭、服裝和打光的人將近二十人。拍片時喜歡穿日本夾腳木屐的北野武，表示在拍「恣在少年」快殺青時，就構思了「花火」和「菊次郎的夏天」兩個劇本，最後決定先拍「花火」，再拍「菊次郎的夏天」，因為希望花更多時間整理對父親的記憶。而今年北野武是否會成為連續兩年奪得威尼斯金獅獎和坎城金棕櫚獎的日本導演，極受關注。</p>	

圖 3-4：一篇文件及其關聯詞

本研究主要是針對有關聯提示詞圖示和無關聯提示詞圖示功能的檢索系統來進行測試研究，而有關聯提示詞圖示功能的檢索系統可以讓檢索者利用系統所提供的關聯詞圖示搭配布林邏輯的運算來檢索，其餘在系統上所列的輔助檢索工具都與預設值為主；至於在無關聯提示詞圖示功能的系統方面，為了模擬一般網路的搜尋引擎，系統沒有提供關聯提示詞圖示的功能，其餘的檢索方式皆和有關聯提示詞圖示功能的系統相同，如圖 3-5 所列。而本系統之互動式檢索功能可大略分為「動態分類目錄提示詞」及「智慧型關聯詞圖示」。本研究只針對關聯提示詞圖示之功能作為評估的對象。最後利用社會科學統計軟體 (SPSS) 作為分析之工具。

有關聯提示詞圖示功能	無關聯提示詞圖示功能
	
實驗網址： <a href="http://140.136.85.107/kmap/kmap.asp">http://140.136.85.107/kmap/kmap.asp</a>	實驗網址： <a href="http://140.136.85.107/kmap/nokmap.asp">http://140.136.85.107/kmap/nokmap.asp</a>

圖 3-5：WebGenie 知識檢索系統介面

為配合研究方法中涉及的問題與訪談，因此設計下列幾項工具：

## 二、問卷之設計：

根據檢索任務及 TREC Interactive Track 之經驗，將問卷內容分為四大部分：(如附錄二~五)

- (一) 基本資料問卷(Entry-Questionnaire)：檢索者的基本資料及背景，如：在電腦技能和檢索經驗等方面；
- (二) 每個檢索任務結束後的問卷(Post-search Questionnaire)：旨在瞭解填答者對此次檢索的特性及了解程度；
- (三) 完成所有檢索任務後的問卷(Post-system Questionnaire)：旨在蒐集同一檢索系統內完成各項檢索任務後的觀感；
- (四) 檢索實驗後問卷(Exit-search Questionnaire)：目的在收集兩系統間的比較及使用者對系統所回饋的資料。

## 三、訪談內容表：

為記錄使用者在檢索實驗結束後所反應的意見及針對關聯提示詞圖示之功能的可用性評估，特別設計非結構性訪談的問題，引導使用者回答一些問題，再配合錄音筆將訪談的內容記錄下來。而介面的可用性評估的訪談大綱請參見附錄六。

## 第六節 資料之分析與處理

本實驗結果配合上述的研究方法所收集的資料，可以從兩個面向來分析：一是針對第一個實驗假設來檢定其資料集的大小是否會影響關聯提示詞的分佈及成效；另一個面向則是從使用的角度來評估此一互動式檢索功能對輔助檢索者的實際成效，評估方式分別從效能(Effectiveness)、效率(Efficiency)、使用者的滿意度(Satisfaction)及可用性評估(Usability Evaluation)來進行分析。

### (一) 關聯提示詞之相關評估結果分析：

第一個面向是關連提示詞之相關度的判斷；由於本實驗系統曾經做過小資料集(PUB)的25,233筆資料之關聯詞成效評估，且本實驗其中一個的目的就是要探討資料集的大小是否會影響關聯提示詞的成效，所以實驗結果將比較系統內三個索引之關聯提示詞的成效。將上述實驗設計的三等級之評分標準經過整理及分析，再利用統計學中的配對樣本T檢定，來檢定關聯提示詞是否會因為資料集的大小而影響其成效。

### (二) 互動式輔助檢索評估結果分析：

第二個面向就是利用問卷調查法、放聲思考法及訪談法所收集的資料，根據TREC Interactive Track上的使用者評估及Jakob Nielsen介面可用性評估方法之設計為準則，分析方式分述如下：

#### 1 檢索完成之程度(Task Completion)：

利用每個檢索任務結束後的問卷(Post-search Questionnaire)由檢索者自行評定是否成功的完成任務；因為檢索者如果認為自己在設定的時間內已經完成任務，表示檢索者滿意自己找出所需的文件，這也較符合現實網路環境中的搜尋狀況。然而本實驗的系統為有提供關聯提示詞之功能為實驗組，和無提供關聯提示詞之檢索功能作為對照組，利用上述的實驗設計來探討有無關聯提示詞對檢索成效的影響，最後在輔以開放式問題的答案及訪談等質性資料來解釋概約的數據。

#### 2 檢索所需之時間(Task Time)：

此步驟主要是針對兩個部份來分析：一是分析有無提供關聯提示詞之功能的系統完成任務所需時間(不同功能間的差異)；二是分析各任務完成所需之時間(不同任務的差異)。上述分析之數據主要是從每個檢索任務結束後的

問卷 (Post-search Questionnaire) 及完成所有檢索任務後的問卷 (Post-system Questionnaire) 得知，再經由列聯表輔以質性資料來分析不同任務和不同功能間效率的差異。

### 3 使用者滿意度調查 (User Satisfaction) :

進行完上述的步驟及任務，最後利用檢索實驗後問卷 (Exit-search Questionnaire) 收集使用者對此關聯提示詞之功能的各項滿意度調查，與在檢索過程中利用放聲思考法記錄任何檢索者的言行舉止，希望檢索者能夠大聲的表達出來，最後再搭配面對面訪談以引導式之非結構式的方法將檢索者不敢表達出的意見記錄下來，訪問受試者對關聯提示詞圖示的可行性及建議，瞭解使用者對於互動式檢索功能應用在關聯提示詞之功能的認知情況和一些檢索需求。最後，在利用統計學上的無母數相關檢定，來檢定各項因素影響檢索滿意度的相關性，以作為進一步推論的依據。

### 4 介面的可用性評估 (Usability Evaluation):

本實驗可用性評估主要是採用 Jakob Nielsen 提出的十個評估規則為準則，在根據實驗之環境適時調整。

Keith Instone 認為，對使用者實施測試工作，是讓網站及介面更容易使用的最重要方法。在各種測試方法之中，經驗評估只需少數幾位受試者的情形之下，即可對於系統是否合乎一些重要的可用性原則做出評估。在測試中所顯示出的問題，也能立即被應用於重新設計，或是用作進一步的系統可用性測試與以使用者為中心的系統設計研究之中。

經驗評估不但簡單、快速，而且十分經濟。只需要受過約一小時可用性相關訓練的測試者，就可以進行。如果系統已經建置，測試的範圍更可隨網路而遍及全球。基本上，經驗評估牽涉到判斷使用經驗、獲取一些關於可用性的意見，以及找出問題及其解決方法。

經驗評估是 Jakob Nielsen 在 1990 年代初期創始的，他整合了大部分的探索方法，但他並未一一列示出來，而從約 200 個可用性問題中做出統計，找出十個最重要的評估規則。<sup>38</sup>這十個規則如下<sup>39</sup>：

---

38 Keith Instone, "Site Usability Evaluation," Retrieved May 18, 2004, [http://www.webreview.com/1997/10\\_10/strategists/10\\_10\\_97\\_1.shtml](http://www.webreview.com/1997/10_10/strategists/10_10_97_1.shtml).

39 Keith Instone, "Site Usability Heuristics for the Web," Retrieved May 18, 2004, [http://www.webreview.com/1997/10\\_10/strategists/10\\_10\\_97\\_2.shtml](http://www.webreview.com/1997/10_10/strategists/10_10_97_2.shtml).

- 1) Visibility of system status ( 視覺設計 )<sup>40</sup>
- 2) Match between system and the real world ( 與現實環境的關聯程度 )
- 3) User control and freedom ( 操作系統的方便性 )
- 4) Consistency and standards ( 一致性 )
- 5) Error prevention ( 錯誤偵測 )
- 6) Recognition rather than recall ( 以認知取代記憶 )
- 7) Flexibility and efficiency of use ( 彈性與效率 )
- 8) Aesthetic and minimalist design ( 美學及最小努力原則之設計 )
- 9) Help users recognize, diagnose, and recover from errors ( 協助認知、  
錯誤之避免與修正 )
- 10) Help and documentation ( 系統輔助 )

根據 Jakob Nielsen 的 10 個可用性評估原則中，本實驗挑選出 9 個適合本實驗之研究環境來作為介面可用性評估的訪談大綱（請參見附錄六 可用性訪談大綱及問題），根據這些大綱研擬出一些訪談問題。在完成所有檢索任務後，利用面對面的訪談方式，請受試者回答上述的訪談問題，最後，利用訪談所收集的質化資料及放聲思考法所回饋的意見經過歸納及分析，以作為人與系統互動之後改善的依據。

---

40 Jakob Nielsen, "Ten Usability Heuristics," Retrieved May 18, 2004, [http://www.useit.com/papers/heuristic/heuristic\\_list.html](http://www.useit.com/papers/heuristic/heuristic_list.html).

## 第四章 實驗結果分析

本研究依照第三章的研究方法與設計來進行資料的收集與分析，本章將所得到的實驗結果加以整理及分析，並開始進行論文撰寫。本章共分為三節，第一節關聯提示詞之相關評估結果分析；第二節 互動式輔助檢索功能評估實驗分析；第三節 綜合分析與討論。

### 第一節 關聯提示詞之相關評估結果分析

本節主要是從系統的角度評估關聯詞的成效，而評估的面向可以分為：一、關聯詞的相關比例分析，簡而言之就是探討關聯詞的查準率；二、利用 trec\_eval 輸出結果的平均查準率和查全率來評估關聯詞的分佈情形。各分析詳述如下：

#### 一、 關聯詞的相關比例分析：

我們邀請 5 位圖書資訊學系碩士班的研究生作為受試者，延襲曾元顯在共現索引典之自動建構 評估與應用中的關聯詞相關性評估實驗的 30 個檢索詞彙(參見表 3-6)，各自挑選六個查詢詞彙，並就每個查詢詞彙，檢視前 N(在此 N=50)個提示詞，判斷其與查詢詞的相關比例，且為了提高檢索詞與相關詞的相關比例，將本實驗相關判斷分數中的部分相關和非常相關合併為相關；最後在經過相關詞的一致性修正步驟，修正同一位判斷者在判斷過程中的不一致性問題。從表 4-1 中得知 30 個檢索詞彙的相關比例分別為：小文件集的相關比例為 0.6987，中文件集的相關比例為 0.7833，大文件集的相關比例為 0.7913。

從上述的相關比例數據來說，文件集的大小的確會影響關聯詞的關聯程度，例如：小文件集至中文件集的關聯性增高約一成左右的關聯程度。但是，接下來中文件集至大文件集增加幅度不大的原因，可以推測年代的因素影響其關聯詞的程度，因為同一時期的新聞文件用語及詞彙較集中，例如：小文件集的資料年代為 2000 至 2001 年，而中文件集的資料則是小文件集的資料加上年代為 1998 至 1999 年的 CIRB011，大文件集的資料則是前兩個資料集再加上年代同為 1998 至 1999 年的 CIRB020，從時間分佈狀況可以得知小文件集至中文件集多增加了 1998 和 1999 年的新聞資料，所以相關比例增加的數據很明顯(小文件 0.6987 小於中文件 0.7833)，至於從大文件集至中文集資料雖然增加快一倍的資料量，但是年代同為 1998 至 1999 年的資料，所以縱使文件量增加，但是新聞事件的報導還是相同，故沒有增加其他相關的詞彙。

表 4-1：前 50 個提示出的詞彙被判斷為相關的相關比例

	個數	最小值	最大值	平均值	標準差
小文件集的相關比例	30	0.44	1	0.6987	0.1315
中文件集的相關比例	30	0.52	1	0.7833	0.1239
大文件集的相關比例	30	0.60	1	0.7913	0.1127

## 二、 利用 trec\_eval 之輸出結果分析：

### (一) 查詢詞在各文件集的 trec\_eval 輸出分析：

單就相關比例而言，即可知大文件集所產生的關聯詞效果的確有提昇，若需探討單一檢索詞的關聯詞在每個文件集的分佈情形，可以利用 trec\_eval 程式計算其查準率與查全率以顯示不同文件集的差異。例如：在三個實驗的資料集中，前 50 個關聯詞中，呈現的關聯提示詞也許不盡相同，透過 trec\_eval 的查準率與查全率計算，將三種資料集找到的關聯提示詞全部集合起來計算查全率，也可以大略得知，哪個資料集能夠找到其他資料集不能找到的關聯提示詞。或者是哪個檢索詞的關聯詞在每個資料集的分佈情形。首先，在表 4-2 中則是將 trec\_eval 程式輸出的 30 個查詢詞之平均查準率加以總計平均，其中中文件集是最佳的，平均查準率為 0.4770，其次是大文件集：0.4475，再其次是小文件集：0.3459。

表 4-2：各文件集之查詢詞的平均查準率統計表

	個數	最小值	最大值	平均值	標準差
小文件集查詢詞的查準率	30	0.1409	0.5618	0.3459	0.1159
中文件集查詢詞的查準率	30	0.2194	0.8102	0.4770	0.1321
大文件集查詢詞的查準率	30	0.2263	0.6631	0.4475	0.1151

為了驗證其數據的可信度，我們利用統計學上的配對樣本 T 檢定(Paired-T Test)來檢定關聯詞在不同大小的文件集是否有顯著的差異，以做為進一步推論的依據。可分下列兩個步驟來檢定：

#### 1. 小文集與中文件集的 T 檢定：

虛無假設  $H_0 : u = 0$

對立假設  $H_1 : u > 0$

虛無假設為關聯詞沒有因為文件的增加而有改善，對立假設為中文件集的關聯程度優於小文件集。從下表 4-5 可以得知  $P\text{-valued} = 0.000 < 0.05$ ，所以無



法接受虛無假設，也就是說關聯詞會因為文件集的增加而有顯著的改善，且顯著性非常的明顯，也可以證明中文件集關聯詞的查準率確優於小文件集的查準率。

表 4-3：小文件與中文件集的敘述統計量表

成對樣本	平均數	個數	標準差
小文件查詢詞的平均查準率	0.3458	30	0.1158
中文件查詢詞的平均查準率	0.4769	30	0.1320

表4-4：小文件與中文件的觀察值

	個數	相關	顯著性
成對 1 小文件 和 中文件	30	.465	.010

表4-5：小文件與中文件集的配對樣本T檢定表

	成對變數差異					t	自由度	顯著性 (雙尾)
	平均數	標準差	平均數的 標準誤	差異的 95% 信賴區間				
				下界	上界			
成對 1 小文件 - 中文件	-.131100	.129005	2.355E-02	-.179271	-8.29E-02	-5.566	29	.000

## 2. 中文件集與大文件集的配對 T 檢定

虛無假設  $H_0: u = 0$

對立假設  $H_1: u > 0$

虛無假設為關聯詞沒有因為文件的增加而有改善，對立假設為大文件集的關聯程度優於中文件集。從下表 4-8 可以得知  $P\text{-valued} = 0.091 > 0.05$ ，所以接受虛無假設，也就是說關聯詞沒有因為大文件集的增加而有顯著的改善。

表4-6：中文件與大文件的敘述統計量表

成對樣本	平均數	個數	標準差
中文件查詢詞的平均查準率	0.4769	30	0.1320
大文件查詢詞的平均查準率	0.4474	30	0.1150

表4-7：中文件集與大文件集的觀察值

	個數	相關	顯著性
成對 1 中文件 和 大文件	30	.729	.000

表4-8：中文件集與大文件集的配對樣本T檢定

	成對變數差異					t	自由度	顯著性 (雙尾)
	平均數	標準差	平均數的 標準誤	差異的 95% 信賴區間				
				下界	上界			
成對 1 中文件 - 大文件	2.95E-02	9.225E-02	1.68E-02	-4.96E-03	6.394E-02	1.751	29	.091

經由統計學上的 T 檢定，文件集的大小在某些情況之下會影響到關聯詞的分佈情形，就如同實驗得出的數據中，小文件集的平均查準率至中文件集的平均查準率確在統計數字上呈現出顯著的提昇，但是相對的中文件集至大文件集雖然增加了更多的新聞文件，查準率卻呈現出降低的現象。初步推論，中文件集與大文件集的查準率不顯著的原因，可能是因為各資料集收錄的年代不同而產生的差異。因此，接下來則是針對個別詞彙在每個文件集的關聯詞分佈情況，做更進一步的分析，以解釋表 4-2 數據背後所隱含的意義，且進而探討中文件集和大文件在平均查準率上呈現降低的原因。

## (二) 個別詞彙的平均查準率分佈情況分析：

為了探討中文件集至大文件集平均查準率降低的原因，我們嘗試分析個別詞彙在不同文件集的平均查準率數值的分佈情形。由統計軟體的分析結果顯示，大部分查詢詞的平均查準率在小文件集至中文集的分佈都呈現出逐漸增加的趨勢，而在中文件集至大文件集則呈現出不均勻的分佈情形，我們試圖歸納出幾個面向來分析，分述如下：

### 1. 文件集的增加反而降低詞彙的平均查準率：

在 30 個查詢詞中有 2 個查詢詞都呈現出因文件量的增加而平均查準率降低的趨勢，分別為：第 23 個詞彙『會計師』及第 24 個詞彙『調查局』(見表 4-9)，從表 4-10 中可以看出『會計師』查詢詞在各文件集的實際分佈情形，被判定為相關的詞彙，越是集中在前面，其平均的查準率越高。我們可以發現在中文件集與大文集的相關詞彙同為 33 個，但是因為大文件集其分佈情形沒有中文件集來的集中(見表 4-10 在前 50% 的分佈情形)，故其查準率才會呈現出逐漸降低的趨勢。

而『調查局』的查詢詞也約略呈現出這種分佈情形。初步分析，可能是因為文件量的增加反而降低各詞彙的集中分佈，讓原本出現較集中的關聯詞，因文件量的增加而降低查詢詞的關聯度，故才會呈現出分佈不均勻的情形。

表 4-9：平均查準率呈現降低的查詢詞彙列表

查詢詞	小文件查準率	中文件查準率	大文件查準率
第 23 個 會計師	0.3898	0.3095	0.2406
第 24 個 調查局	0.3999	0.3667	0.2263

表 4-10：查詢詞『會計師』在各文件集的 trec\_eval 輸出

Queryid ( Num ): 23 ( 及查詢詞 : 『會計師』 )			
Total number of documents over all queries			
	小文件集 ( small )	中文件集 ( medium )	大文件集 ( large )
Retrieved :	50	50	50
Relevant :	65	65	65
Rel_ret :	34	33	33
	Interpolated Recall - Precision	Interpolated Recall - Precision	Interpolated Recall - Precision
	Averages:	Averages:	Averages:
	at 0.00 1.0000	at 0.00 1.0000	at 0.00 0.6600
	at 0.10 0.7857	at 0.10 0.6600	at 0.10 0.6600
	at 0.20 0.7857	at 0.20 0.6600	at 0.20 0.6600
	at 0.30 0.7857	at 0.30 0.6600	at 0.30 0.6600
	at 0.40 0.7222	at 0.40 0.6600	at 0.40 0.6600
	at 0.50 0.6800	at 0.50 0.6600	at 0.50 0.6600
	at 0.60 0.0000	at 0.60 0.0000	at 0.60 0.0000
	at 0.70 0.0000	at 0.70 0.0000	at 0.70 0.0000
	at 0.80 0.0000	at 0.80 0.0000	at 0.80 0.0000
	at 0.90 0.0000	at 0.90 0.0000	at 0.90 0.0000
	at 1.00 0.0000	at 1.00 0.0000	at 1.00 0.0000
	Average precision (non-interpolated) for all rel docs(averaged over queries)	Average precision (non-interpolated) for all rel docs(averaged over queries)	Average precision (non-interpolated) for all rel docs(averaged over queries)
	0.3898	0.3095	0.2406
	Precision:	Precision:	Precision:
	At 5 docs: 0.8000	At 5 docs: 0.8000	At 5 docs: 0.0000
	At 10 docs: 0.7000	At 10 docs: 0.5000	At 10 docs: 0.1000
	At 15 docs: 0.6000	At 15 docs: 0.4667	At 15 docs: 0.2667
	At 20 docs: 0.7000	At 20 docs: 0.4500	At 20 docs: 0.3500
	At 30 docs: 0.7333	At 30 docs: 0.5333	At 30 docs: 0.4333
	At 100 docs: 0.3400	At 100 docs: 0.3300	At 100 docs: 0.3300
	At 200 docs: 0.1700	At 200 docs: 0.1650	At 200 docs: 0.1650
	At 500 docs: 0.0680	At 500 docs: 0.0660	At 500 docs: 0.0660

	At 1000 docs: 0.0340 R-Precision (precision after R (= num_rel for a query) docs retrieved): <b>Exact: 0.5231</b>	At 1000 docs: 0.0330 R-Precision (precision after R (= num_rel for a query) docs retrieved): <b>Exact: 0.5077</b>	At 1000 docs: 0.0330 R-Precision (precision after R (= num_rel for a query) docs retrieved): <b>Exact: 0.5077</b>
--	---	---	---

**表格欄位說明：**

Retrieved：即提示出的 n 個詞彙（本實驗 n=50）

Relevant：從三種文件集提示出來的所有詞彙中，被判斷為相關的詞彙個數。

Rel\_ret：單一文件集提示出來的 50 個詞彙中，被判斷為相關的詞彙個數。

**2. 單一查詢詞彙的個別分析：**

由於 30 個查詢詞的平均精確最高為 0.4770，但是，我們觀察到『國安基金』的查詢詞在中文件集的查準率高達 0.8102，且在大文件集時的查準率則降為 0.6631 的分佈情形，因此我們特別針對此一詞彙做個簡單的分析。由表 4-11 的 trec\_eval 輸出得知，小文件集與中文件集的相關詞彙雖然同為 36 個，但是這些相關詞彙相對於小文件集則顯得較集中（前 80 % 的分佈情形）；然而當文件量增加至大文件集時，分佈情形約與中文件集相同，但是其相關詞彙則明顯的降為 31 個，所以查準率則呈現出明顯的降低。

就『國安基金』查詢詞彙來說，文件量的增加雖然沒有相對的增加關聯詞彙，但是，只要增加的文件是有效的，卻可以提高相關詞彙的集中分佈情形，例如：小文件集與中文件集的平均查準率。然而，如果增加的文件量很大卻是無效的，只會產生出反效果的影響，例如：中文件集與大文件集的分佈情形。

表 4-11：查詢詞『國安基金』在各文件集的 trec\_eval 輸出

Queryid ( Num ): 13 ( 及查詢詞：『國安基金』)			
Total number of documents over all queries			
	小文件集 ( small )	中文件集 ( medium )	大文件集 ( large )
Retrieved :	50	50	50
Relevant :	44	44	44
Rel_ret :	36	36	31
	Interpolated Recall - Precision Averages:	Interpolated Recall - Precision Averages:	Interpolated Recall - Precision Averages:
	at 0.00 0.7200	at 0.00 1.0000	at 0.00 1.0000
	at 0.10 0.7200	at 0.10 1.0000	at 0.10 1.0000
	at 0.20 0.7200	at 0.20 1.0000	at 0.20 1.0000
	at 0.30 0.7200	at 0.30 1.0000	at 0.30 1.0000
	at 0.40 0.7200	at 0.40 1.0000	at 0.40 1.0000
	at 0.50 0.7200	at 0.50 1.0000	at 0.50 1.0000
	at 0.60 0.7200	at 0.60 1.0000	at 0.60 0.6750
	at 0.70 0.7200	at 0.70 1.0000	at 0.70 0.6200
	at 0.80 0.7200	at 0.80 0.7347	at 0.80 0.0000
	at 0.90 0.0000	at 0.90 0.0000	at 0.90 0.0000
	at 1.00 0.0000	at 1.00 0.0000	at 1.00 0.0000
	Average precision (non-interpolated) for all rel docs(averaged over queries)	Average precision (non-interpolated) for all rel docs(averaged over queries)	Average precision (non-interpolated) for all rel docs(averaged over queries)
	0.5004	0.8102	0.6631
	Precision:	Precision:	Precision:
	At 5 docs: 0.6000	At 5 docs: 1.0000	At 5 docs: 1.0000
	At 10 docs: 0.6000	At 10 docs: 1.0000	At 10 docs: 1.0000
	At 15 docs: 0.5333	At 15 docs: 1.0000	At 15 docs: 1.0000
	At 20 docs: 0.5500	At 20 docs: 1.0000	At 20 docs: 1.0000
	At 30 docs: 0.6000	At 30 docs: 1.0000	At 30 docs: 0.8667
	At 100 docs: 0.3600	At 100 docs: 0.3600	At 100 docs: 0.3100
	At 200 docs: 0.1800	At 200 docs: 0.1800	At 200 docs: 0.1550
	At 500 docs: 0.0720	At 500 docs: 0.0720	At 500 docs: 0.0620
	At 1000 docs: 0.0360	At 1000 docs: 0.0360	At 1000 docs: 0.0310
	R-Precision (precision after R (= num_rel for a query) docs retrieved):	R-Precision (precision after R (= num_rel for a query) docs retrieved):	R-Precision (precision after R (= num_rel for a query) docs retrieved):
	Exact: 0.6818	Exact: 0.7955	Exact: 0.6364

**表格欄位說明：**

Retrieved：即提示出的 n 個詞彙（本實驗 n=50）

Relevant：從三種文件集提示出來的所有詞彙中，被判斷為相關的詞彙個數。

Rel\_ret：單一文件集提示出來的 50 個詞彙中，被判斷為相關的詞彙個數。

**3. 中文件集的平均查準率高於大文件集的平均查準率：**

表 4-12 顯示出 30 個檢索詞彙中共有 15 個查詢詞的平均查準率分佈情形，呈現出中文件集高於大文件集的分佈，可見下列幾個詞彙沒有因為文件量的增加，而關聯詞的平均查準率就相對的提高，例如：以表 4-13 『國科會』這個查詢詞來說，三個文件集共被判斷出 60 個關聯詞，分別在小文件集中有 27 個被判斷是關聯詞，中文件集則增加至 37 個關聯詞，大文件集則降為 32 個關聯詞；由此可見，中文件集至大文件集之平均查準率降低的原因是增加了一些雜訊，使得原本增加的關聯詞反而因文件集的增加而降低它的關聯詞數。

再者，由資料集收錄的內容也可以推論出：1) 中文件集的資料內容包含小文件集，且增加了中時報系體下的工商時報、中時晚報及中華日報，故會多增加一些相關詞彙，例如：小文件集原本只收錄中國時報和中央日報的新聞資料，經過中文件集再增加 CIRB011 底下五大報紙 13 萬篇的新聞資料其平均查準率均優於小文件集；2) 大文件集雖多收錄了約 24 萬多篇的新聞資料，但是由於其收錄的新聞資料內容，主要是以 UDN 聯合新聞網的資料集為主，且其報導的新聞內容取向可能異於中時體系報紙的新聞資料（新聞報導有集中性的趨勢），所以文件量雖然增加了，但是雜訊也有可能相對的增加，而降低了大文件集的查準率；3) 由於在大文件集增加的資料內容年代同屬於 1998 至 1999 年的新聞資料，所以同時期的新聞事件往往是重複的，整體而言文件平均查準率才會呈現出增加不大的趨勢。

**表 4-12：查詢詞的平均查準率之中文件集優於大文件集**

查詢詞	小文件查準率	中文件查準率	大文件查準率
第 2 個 中油	0.2624	<b>0.5738</b>	<b>0.5728</b>
第 5 個 生物科技	0.2431	<b>0.4675</b>	<b>0.3269</b>
第 6 個 地震	0.1409	<b>0.5106</b>	<b>0.3587</b>
第 7 個 有線電視	0.2941	<b>0.5515</b>	<b>0.4277</b>
第 12 個 邱義仁	0.2792	<b>0.4871</b>	<b>0.4793</b>
第 13 個 國安基金	0.5004	<b>0.8102</b>	<b>0.6631</b>
第 14 個 國科會	0.2656	<b>0.5652</b>	<b>0.3936</b>
第 16 個 通航	0.4692	<b>0.6237</b>	<b>0.5844</b>
第 17 個 博物館	0.2548	<b>0.5338</b>	<b>0.3123</b>

第 20 個 雅虎	0.4642	<b>0.6494</b>	<b>0.5628</b>
第 21 個 債券	0.5139	<b>0.5460</b>	<b>0.4252</b>
第 23 個 會計師	0.3898	<b>0.3095</b>	<b>0.2406</b>
第 24 個 調查局	0.3999	<b>0.3667</b>	<b>0.2263</b>
第 25 個 選民	0.2167	<b>0.5268</b>	<b>0.4735</b>
第 27 個 職棒	0.4468	<b>0.5454</b>	<b>0.4775</b>

表 4-13 : 查詢詞『國科會』在三個文件集的 trec\_eval 輸出

Queryid ( Num ): 14 ( 及查詢詞 : 『國科會』)			
Total number of documents over all queries			
	小文件集 ( small )	中文件集 ( medium )	大文件集 ( large )
Retrieved :	<b>50</b>	<b>50</b>	<b>50</b>
Relevant :	<b>60</b>	<b>60</b>	<b>60</b>
Rel_ret :	<b>27</b>	<b>37</b>	<b>32</b>
	Interpolated Recall - Precision Averages:	Interpolated Recall - Precision Averages:	Interpolated Recall - Precision Averages:
	at 0.00 <b>0.6667</b>	at 0.00 <b>1.0000</b>	at 0.00 <b>1.0000</b>
	at 0.10 <b>0.6471</b>	at 0.10 <b>1.0000</b>	at 0.10 <b>1.0000</b>
	at 0.20 <b>0.6471</b>	at 0.20 <b>1.0000</b>	at 0.20 <b>0.9286</b>
	at 0.30 <b>0.6471</b>	at 0.30 <b>0.9231</b>	at 0.30 <b>0.6400</b>
	at 0.40 <b>0.6154</b>	at 0.40 <b>0.9231</b>	at 0.40 <b>0.6400</b>
	at 0.50    0.0000	at 0.50 <b>0.8378</b>	at 0.50 <b>0.6400</b>
	at 0.60    0.0000	at 0.60 <b>0.7400</b>	at 0.60    0.0000
	at 0.70    0.0000	at 0.70    0.0000	at 0.70    0.0000
	at 0.80    0.0000	at 0.80    0.0000	at 0.80    0.0000
	at 0.90    0.0000	at 0.90    0.0000	at 0.90    0.0000
	at 1.00    0.0000	at 1.00    0.0000	at 1.00    0.0000
	Average precision (non-interpolated) for all rel docs(averaged over queries) <b>0.2656</b>	Average precision (non-interpolated) for all rel docs(averaged over queries) <b>0.5652</b>	Average precision (non-interpolated) for all rel docs(averaged over queries) <b>0.3936</b>
	<b>Precision:</b> At 5 docs: 0.4000 At 10 docs: 0.5000 At 15 docs: 0.6000 At 20 docs: 0.5500 At 30 docs: 0.6000 At 100 docs: 0.2700	<b>Precision:</b> At 5 docs: 1.0000 At 10 docs: 1.0000 At 15 docs: 1.0000 At 20 docs: 0.9000 At 30 docs: 0.9000 At 100 docs: 0.3700	<b>Precision:</b> At 5 docs: 1.0000 At 10 docs: 0.9000 At 15 docs: 0.8667 At 20 docs: 0.7000 At 30 docs: 0.5000 At 100 docs: 0.3200

At 200 docs: 0.1350	At 200 docs: 0.1850	At 200 docs: 0.1600
At 500 docs: 0.0540	At 500 docs: 0.0740	At 500 docs: 0.0640
At 1000 docs: 0.0270	At 1000 docs: 0.0370	At 1000 docs: 0.0320
R-Precision (precision after R (= num_rel for a query) docs retrieved):	R-Precision (precision after R (= num_rel for a query) docs retrieved):	R-Precision (precision after R (= num_rel for a query) docs retrieved):
<b>Exact: 0.4500</b>	<b>Exact: 0.6167</b>	<b>Exact: 0.5333</b>

**表格欄位說明：**

Retrieved：即提示出的 n 個詞彙（本實驗 n=50）

Relevant：從三種文件集提示出來的所有詞彙中，被判斷為相關的詞彙個數。

Rel\_ret：單一文件集提示出來的 50 個詞彙中，被判斷為相關的詞彙個數。

Precision：Rel\_ret / Relevant = 27 / 60 = 0.4500

## 第二節 互動式檢索輔助功能評估實驗分析

第二節則是從使用者的角度評估受試者使用關聯詞提示詞圖示之後的使用分析（Use Study），而分析的面向為：一、受試者的基本資料分析，包含：電腦技能及檢索經驗等；二、各面向的分析，包含：檢索完成的程度、檢索所需的時間、使用者的滿意度及介面的可用性評估。各分析如下：

### 一、受試者基本資料分析：

#### （一）實驗對象：

本實驗為了取樣的方便，抽樣的方式採取非隨機抽樣中的立意抽樣及配額抽樣，且徵求參與意願高的受試者，並將受試者分為兩群：一群是具有圖書資訊學系背景的研究生 8 位（見表 4-15 和 4-17），另一群則是非圖書資訊學系背景的大學生 8 位（見表 4-14 和 4-16），共 16 位受試者；受試者中有 10 位男生 6 位女生，非圖資系的系別有經濟系、廣告系、法律系、新聞傳播學系、體育系及日文系共 8 位。受試者的年齡範圍為 18 至 37 歲。

表 4-14：非圖書資訊學領域之背景的受試者先從 A 系統進行實驗

受試者編號	檢索任務分配
ABn1	A-system 1-4, B-system 5-8
ABn5	A-system 5-8, B-system 1-4
ABn2	A-system 1-4, B-system 5-8
ABn6	A-system 5-8, B-system 1-4



表 4-15：圖書資訊學領域之背景的受試者先從 A 系統進行實驗

受試者編號	檢索任務分配
ABh3	A-system 1-4,B-system 5-8
ABh7	A-system 5-8,B-system 1-4
ABh4	A-system 1-4,B-system 5-8
ABh8	A-system 5-8,B-system 1-4

表 4-16：非圖書資訊學領域之背景的受試者先從 B 系統進行實驗

受試者編號	檢索任務分配
BAn9	b-system 1-4,a-system 5-8
BAn13	b-system 5-8,a-system 1-4
BAn10	b-system 1-4,a-system 5-8
BAn14	b-system 5-8,a-system 1-4

表 4-17：圖書資訊學領域之背景的受試者先從 B 系統進行實驗

受試者編號	檢索任務分配
BAh16	b-system 5-8,a-system 1-4
Bah12	b-system 1-4,a-system 5-8
Bah15	b-system 5-8,a-system 1-4
Bah11	b-system 1-4,a-system 5-8

## (二) 電腦技能：

在電腦技能中，我們利用基本資料問卷(Entry-Questionnaire)，收集檢索者使用電腦的頻率以及相關技能，而問卷可以從幾個面向來代表受試者的電腦技能，分別為：使用電腦的頻率、瀏覽器的頻率以及利用電腦從事工作、學術、以及個人用途的頻率，最後再請受試者對自己在電腦方面的技能評定一個分數；而上述的評量方式皆採七等式量表來區分。下表 4-18 與 4-19 則是利用統計軟體初步分析的結果：

具備圖書資訊學背景領域的受試者在各項的平均數分別為：電腦經驗 7、使用瀏覽器的頻率為 7，以及利用電腦從事工作的頻率為 5、學術用為 6.125、個人用為 6.75，最後這些受試者評定自己在電腦技能的分數也高達 4.75，由統計數據可以初步得知這些受試者對自己在電腦方面的技能皆有一定的程度（見表 4-18）。而非圖書資訊背景領域之受試者的各項平均數則分別為：電腦經驗 5.75、瀏覽器 5.75 以及工作（頻率）4.62、學術（頻率）4.75、個人用（頻率）為 6.5，最後，在評定自己在電腦的技能平均分數為 3.5（見表 4-19）。

從上述的數據可以得知具備圖書資訊背景領域的受試者在各項分數均高於非圖書資訊學領域的受試者，且在電腦技能的自我評分中，也高於非圖書資訊學領域的受試者。因此，可以初步得知具有圖書資訊學背景的受試者在電腦技能是優於非圖書資訊學背景的受試者。

表 4-18：具備圖書資訊學背景的電腦技能統計量表

	個數	最小值	最大值	平均數	標準差
電腦經驗	8	7.00	7.00	7.0000	.0000
瀏覽器IE	8	7.00	7.00	7.0000	.0000
工作(頻率)	8	1.00	7.00	5.0000	2.5635
學術(頻率)	8	5.00	7.00	6.1250	.6409
個人(頻率)	8	6.00	7.00	6.7500	.4629
電腦技能評分	8	4.00	7.00	4.7500	1.0351
有效的 N (完全排除)	8				

表 4-19：非圖書資訊學背景的電腦技能統計量表

	個數	最小值	最大值	平均數	標準差
電腦經驗	8	3.00	7.00	5.7500	1.3887
瀏覽器IE	8	3.00	7.00	5.7500	1.8323
工作(頻率)	8	1.00	7.00	4.6250	2.2638
學術(頻率)	8	2.00	7.00	4.7500	1.4880
個人(頻率)	8	4.00	7.00	6.5000	1.0690
電腦技能評分	8	1.00	6.00	3.5000	1.5119
有效的 N (完全排除)	8				

### (三) 檢索經驗：

在檢索經驗中，也是經由基本資料問卷(Entry-Questionnaire)收集檢索者在檢索的相關經驗，包含：使用搜尋引擎、圖書館的線上書目資料查詢系統(OPAC)，索引或摘要檢索系統的頻率，還詢問對於個人經驗中利用檢索系統都可以找到資訊的滿意度調查，以及利用搜尋引擎從事作業或工作、購物、旅遊、醫療或健康、政治及娛樂等的頻率，最後，再詢問每位受試者利用線上搜尋引擎的時間有多久，以及最常用的搜尋引擎。綜合上述的資料，我們可以初步判斷每位受試者在檢索經驗的程度，以作為本實驗分析及推論之依據。而上述的評量方式也是採取七等式量表由檢索者自行評定。

從下表4-20數據可以得知，具備圖書資訊學領域背景的受試者平均使用搜尋引擎來檢索的經驗為7（經常使用）、查詢圖書館公用目錄的經驗為5.875以及利用索引或摘要檢索系統的經驗為5.75，且他們對於上述的檢索系統都可以找到資訊的評分也高達5.50；而在查詢有關下列資訊的頻率分別為：作業或工作有關為5.875、購物為4、旅遊及醫療皆為3、政治為2以及娛樂為5.62，最後在評定自己在資訊檢索的平均能力為5.375且利用線上搜尋引擎的時間平均為6年。至於表

4-21則是顯示非圖書資訊學背景領域的受試者，平均使用搜尋引擎來檢索的經驗為5.62、查詢圖書館公用目錄的經驗為4.125以及利用索引或摘要檢索系統的經驗為3.5，且他們對於上述的檢索系統都可以找到資訊的評分為5.25；而在查詢有關下列資訊的頻率分別為：作業或工作有關為5、購物為4.375、旅遊為3.25、醫療為3.37、政治為4以及娛樂為5.37，最後在評定自己於資訊檢索的平均能力為4.12且利用線上搜尋引擎的時間平均約為3.875年。

從上述的各項統計數字得知，在使用各種資料庫或檢索系統的頻率中，具有圖書資訊學背景的受試者優於非圖書資訊背景領域的受試者，且對於上述檢索系統都能夠找到資訊的滿意程度也略優於非圖書資訊學領域背景的受試者；至於所有受試者均以查詢有關作業或工作以及娛樂上的頻率為最高，其餘的頻率皆差不多，由此可顯示本實驗對象在利用檢索系統查詢的資訊類型皆大同小異，是故，可以初步排除因為受試者對本實驗設計之檢索任務的熟悉程度不同，而產生的個別差異進而影響實驗的結果。最後，在資訊檢索能力的自我評分中，也是以具有圖書資訊學背景領域的受試者高於非圖書背景領域之受試者，且利用搜尋引擎的頻率也大於非圖書領域之受試者（6年 > 3.875年），至於在使用搜尋引擎的系統調查中，受試者大都偏好使用 Google 和 Yahoo 為主，且所有受試者均反應出未曾使用過類似本實驗系統之聯提示詞圖示功能，因此，大家皆在相同的條件之下參與實驗，只是個人檢索經驗差異的不同而已。上述各項敘述統計的分析，皆可作為本實驗推論統計分析之依據。

表4-20：具備圖書資訊學背景的檢索技能統計量表

	個數	最小值	最大值	平均數	標準差
搜尋引擎	8	7.00	7.00	7.0000	.0000
OPAC	8	2.00	7.00	5.8750	1.6421
索引或摘要檢索系統	8	5.00	7.00	5.7500	.7071
當我利用檢索系統搜尋時都可以找到資訊	8	5.00	7.00	5.5000	.7559
作業或工作(頻率)	8	5.00	7.00	5.8750	.8345
購物(頻率)	8	2.00	6.00	4.0000	1.4142
旅遊(頻率)	8	2.00	4.00	3.6250	.7440
醫療或健康(頻率)	8	1.00	4.00	3.0000	1.0690
政治(頻率)	8	1.00	3.00	2.1250	.8345
娛樂(頻率)	8	4.00	7.00	5.6250	.9161
資訊檢索能力	8	4.00	7.00	5.3750	.9161
利用線上搜尋引擎時間有多久	8	3.00	8.00	6.0000	1.5119
有效的 N (完全排除)	8				

表4-21：非圖書資訊學背景的檢索技能統計量表

	個數	最小值	最大值	平均數	標準差
搜尋引擎	8	4.00	7.00	5.6250	1.4079
OPAC	8	1.00	7.00	4.1250	1.6421
索引或摘要檢索系統	8	1.00	7.00	3.5000	2.2039
當我利用檢索系統搜尋時都可以找到資訊	8	4.00	6.00	5.2500	.7071
作業或工作(頻率)	8	3.00	7.00	5.0000	1.6903
購物(頻率)	8	1.00	6.00	4.3750	1.5980
旅遊(頻率)	8	1.00	5.00	3.2500	1.4880
醫療或健康(頻率)	8	1.00	5.00	3.3750	1.3025
政治(頻率)	8	1.00	7.00	4.1250	1.8077
娛樂(頻率)	8	4.00	7.00	5.3750	1.0607
資訊檢索能力	8	2.00	6.00	4.1250	1.1260
利用線上搜尋引擎時間有多久	8	.50	7.00	3.8750	2.1171
有效的 N (完全排除)	8				

## 二、各面向之分析：

此部分的分析主要是根據第三章第六節資料分析與處理的方式來進行，將利用問卷調查法、放聲思考法及介面可用性訪談訪所收集的資料，經由四個面向來分析：分別是（1）檢索完成之程度（Task Completion），也就是檢索的效能（Effectiveness）；（2）檢索所需的時間（Task Time），也就是檢索的效率（Efficiency）；（3）使用者的滿意度調查（User Satisfaction）；（4）介面的可用性評估（Usability Evaluation），各項分析如下：

### （一） 檢索完成之程度（Task Completion）：

檢索完成之程度，主要是針對單一檢索任務在有無提示關聯詞下，是否能夠有效的幫助檢索者提高檢索效果，而檢索效果的評量主要是針對每個檢索者在執行檢索任務後，利用「單一檢索任務結束後問卷」收集下列資訊，且利用七等式量表讓受試者勾選滿意程度。經由統計軟體利用敘述統計的方式，各項數據如下：

#### 1 在對檢索任務的熟悉程度方面：

##### 1) 具備圖書資訊學背景的受試者：

表 4-22 顯示 8 位具備圖書資訊學背景的受試者，對各檢索任務之熟悉程度的平均分數分別為：Task1 是 3.625、Task2 是 4、Task3 是 5.375、Task4 是 5、Task5 是 4、Task6 是 4.375、Task7 是 5.250、Task8 是 5.250。從數據可以初步分析，除了“檢索任務 1”之外，具備圖書資訊學背景之受試者對各項檢索任務的熟悉程度皆超過 4 分以上。

表 4-22：圖資背景在檢索任務的熟悉程度

	最小值	最大值	總和	平均數
Task1	1	5	29	3.625
Task2	1	6	32	4.
Task3	1	6	43	5.375
Task4	1	7	40	5
Task5	2	7	42	5.25
Task6	1	6	35	4.375
Task7	3	7	42	5.250
Task8	2	7	42	5.250

## 2) 非圖書資訊學背景的受試者：

表 4-23 顯示 8 位非圖書資訊學背景的受試者，對各檢索任務的熟悉程度平均分數分別為：Task1 是 4.625、Task2 是 4.125、Task3 是 5、Task4 是 5.125、Task5 是 4.25、Task6 是 4.875、Task7 是 5、Task8 是 5.375。由此可知，非圖資背景在各項檢索任務皆有一定的熟悉程度(分數皆大於 4)。

表 4-23：非圖資背景在檢索任務的熟悉程度

	最小值	最大值	總和	平均數
Task1	3	7	37	4.625
Task2	3	5	33	4.125
Task3	3	7	40	5
Task4	3	7	41	5.125
Task5	3	5	34	4.25
Task6	4	6	39	4.875
Task7	3	7	40	5
Task8	4	7	43	5.375

## 2 對檢索任務的專業程度方面：

## 1) 具備圖書資訊學背景的受試者：

表 4-24 顯示 8 位圖書資訊學背景的受試者，在檢索任務的專業程度平均分數分別為：Task1 是 4.125、Task2 是 4.625、Task3 是 5.625、Task4 是 5.625、Task5 是 5.375、Task6 是 5.75、Task7 是 5.75、Task8 是 5.25。因此在專業程

度方面，具備圖資背景的受試者皆達到一定的程度(4分以上)，而政治類(Task1和 Task2)的檢索任務略低於其他類型的檢索任務，這與基本資料問卷中，搜尋資料的目的大都是非政治類新聞為主相符。

表 4-24：圖資背景在檢索任務的專業程度

	最小值	最大值	總和	平均數
Task1	1	5	33	4.125
Task2	3	6	37	4.625
Task3	5	6	45	5.625
Task4	4	7	45	5.625
Task5	4	6	43	5.375
Task6	4	6	46	5.750
Task7	5	6	46	5.750
Task8	4	6	42	5.250

## 2) 非圖書資訊學背景的受試者：

表 4-25 顯示 8 位非圖書資訊學背景的受試者，在檢索任務的專業程度平均分數分別為：Task1 是 4.125、Task2 是 4.625、Task3 是 5.625、Task4 是 5.625、Task5 是 5.375、Task6 是 5.75、Task7 是 5.75、Task8 是 5.25。因此在專業程度方面，非圖資背景的受試者也達到一定的程度(4 分以上)，而政治類(Task1 和 Task2)的檢索任務也略低於其他類型的檢索任務，上述結果與具備圖資背景的受試者相同，唯一不同的地方是整體的平均分數略低於具備圖資背景的受試者。

表 4-25：非圖資背景在檢索任務的專業程度

	最小值	最大值	總和	平均數
Task1	2	7	33	4.125
Task2	2	5	33	4.125
Task3	3	7	38	4.75
Task4	3	5	38	4.75
Task5	3	6	35	4.375
Task6	2	7	36	4.5
Task7	3	6	37	4.625
Task8	4	6	40	5

### 3 對自行找出的答案有多少把握是正確：

在此方面的評估，我們是利用受試者對自己查出的檢索結果，自行評定一個信心分數，評估的方式也是利用七等級分數來評分，且分別探討有無圖書資訊學背景的受試者，分別使用 A 系統及 B 系統後，所檢索出的結果之信心分數的差異，最後再配合受試者在實驗結束後，所記錄下的檢索結果，與 NTCIR 檢索任務的答案做個比對與分析。各項數據及分析如下：

#### 1) 具備圖書資訊學背景的受試者：

表 4-26 顯示具備圖資背景受試者對自己利用系統所檢索出的答案，皆有一定程度的信心分數(4 分以上)，從表格的數據可以初步分析，在探索性任務的信心分數 A 系統略高於 B 系統；政治類(Task1 和 Task2)的檢索任務之信心分數 B 系統略高於 A 系統；運動類的檢索任務(Task3 和 Task4)之信心分數 A 系統與 B 系統無明顯的差異；生活類的檢索任務(Task5 和 Task6)也無明顯的差異；娛樂類的檢索任務則是 A 系統的信心程度優於 B 系統(Task7 和 Task8)。

表 4-26：圖資背景分別在兩系統之檢索結果的信心分數

	有關聯提示詞圖示之系統 (A-system)				無關聯提示詞圖示之系統 (B-system)			
	最小值	最大值	總和	平均數	最小值	最大值	總和	平均數
Explore	4	7	49	6.125	4	7	48	6
Task1	1	6	16	4	4	6	19	4.75
Task2	5	5	20	5	4	6	21	5.25
Task3	5	7	24	6	5	7	24	6
Task4	5	7	26	6	6	7	26	6.5
Task5	6	7	25	6.25	6	7	25	6.25
Task6	3	7	22	5.5	2	7	20	5
Task7	5	7	26	6.5	4	6	21	5.25
Task8	6	7	26	6.5	1	7	19	4.75

#### 2) 非圖書資訊學背景：

表 4-27 數據可以初步分析，非圖資背景的受試者在探索性任務也是 A 系統略高於 B 系統；政治類(Task1 和 Task2)與運動類(Task3 和 Task4)的檢索任務之信心分數 B 系統略高於 A 系統；生活類(Task5 和 Task6)與娛樂類(Task7 和 Task8)的檢索任務則是 A 系統的信心程度優於 B 系統。與具備圖資背景領域的受試者比

較來說，對利用 A 系統查詢探索性及娛樂類檢索任務的檢索結果，較有把握是正確的，而在政治類則認為利用 B 系統檢索出的檢索結果較有把握是正確的。

表 4-27：非圖資背景分別在兩系統之檢索結果的信心分數

	有關聯提示詞圖示之系統 (A-system)				無關聯提示詞圖示之系統 (B-system)			
	最小值	最大值	總和	平均數	最小值	最大值	總和	平均數
Explore	1	7	38	4.75	1	7	37	4.625
Task1	2	7	16	4	4	7	20	5
Task2	1	6	15	3.75	4	6	20	5
Task3	3	5	17	4.25	6	7	27	6.75
Task4	4	6	19	4.75	6	7	26	6.5
Task5	6	7	26	6.5	3	7	17	4.25
Task6	6	7	26	6.5	3	6	19	4.75
Task7	6	7	26	6.5	4	7	22	5.5
Task8	6	7	27	6.75	3	7	18	4.5

#### 4 能夠在此系統找到多少資訊：

針對每個檢索任務詢問其在檢索過程中，對單一檢索任務利用兩系統後，對所收錄之資訊量的調查，其目的是想探知檢索者利用系統的功能之後，是否能夠有效的了解檢索任務在此系統所收錄的資料範圍，而在問卷中的題目則是「你認為此檢索任務能夠在此系統找到多少資訊」及「詢問你回答此一答案的信心程度」（見表 4-28），各數據如下：

##### 1) 具備圖資訊學背景的受試者：

表 4-28 顯示 8 位具備圖書資訊學背景的受試者，對各項檢索任務能夠找到的資訊量的平均分數分別為：探索性任務 A 系統高於 B 系統(5.375>5.25)；Task1 之平均分數 B 系統高於 A 系統(5.75>3.5)；Task2 之平均分數 A 系統高於 B 系統(5>4.75)；Task3 之平均分數 B 系統與 A 系統相同(6=6)；Task4 之平均分數 B 系統高於 A 系統(5.5>5.25)；Task5 之平均分數也是 B 系統高於 A 系統(7>6)；Task6 之平均分數 B 系統高於 A 系統(3.75>3)；Task7 之平均分數 A 系統高於 B 系統(6.25>5.5)；Task8 之平均分數 B 系統高於 A 系統(5.75>5.5)。歸納出受試者認為 A 系統所檢索出的資訊量優於 B 系統的檢索任務分別為：探索性任務、Task2、Task7；反之 B 系統優於 A 系統的則為：Task1、Task4、Task5、Task6、Task8；最後在 Task3 方面，受試者則認為兩系統所檢索出的資訊量無明顯的差



異。

表 4-28：具備圖資背景的受試者在資訊量與回答之答案的信心統計表

	有關聯提示詞圖示之系統 (A-system)				無關聯提示詞圖示之系統 (B-system)			
	找到多少資訊		信心		找到多少資訊		信心	
	總和	平均數	總和	平均數	總和	平均數	總和	平均數
Explore	43	5.375	46	5.75	42	5.25	45	5.625
Task1	14	3.5	15	3.75	23	5.75	18	4.5
Task2	20	5	19	4.75	19	4.75	20	5
Task3	24	6	24	6	27	6	25	6.25
Task4	21	5.25	23	5.75	22	5.5	24	6
Task5	24	6	25	6.25	28	7	26	6.5
Task6	12	3	25	6.25	15	3.75	21	5.25
Task7	25	6.25	23	5.75	22	5.5	21	5.25
Task8	22	5.5	24	6	23	5.75	23	5.75

## 2) 非圖書資訊學背景：

在表 4-29 中 8 位非圖書資訊學背景的受試者，對各項檢索任務能夠找到的資訊量之平均分數分別為：探索性任務 A 系統高於 B 系統(5.125>5)；Task1 之平均分數 B 系統高於 A 系統(5.25>5)；Task2 之平均分數 A 系統高於 B 系統(4.5>4.25)；Task3 之平均分數 B 系統高於 A 系統(6.25>4.75)；Task4 之平均分數 A 系統高於 B 系統(5>4)；Task5 之平均分數 A 系統高於 B 系統(6.25>4.75)；Task6 之平均分數 A 系統高於 B 系統(5.5>4.25)；Task7 之平均分數 A 系統和 B 系統相同(5.25>5.25)；Task8 之平均分數 A 系統高於 B 系統(6>5)。歸納出受試者認為 A 系統所檢索出的資訊量優於 B 系統的檢索任務分別為：探索性任務、Task2、Task4、Task5、Task6、Task8；反之 B 系統優於 A 系統的則為：Task1 和 Task3；最後在 Task7 受試者則認為兩系統所檢索出的資訊量無明顯的差異，但是從表 4-29 的信心平均分數可以約略看出受試者對此檢索任務，使用 A 系統時的信心程度高於在 B 系統使用。

表 4-29：非圖資背景的受試者在資訊量與回答之答案的信心統計表

	有關聯提示詞圖示之系統 (A-system)				無關聯提示詞圖示之系統 (B-system)			
	找到多少資訊		信心		找到多少資訊		信心	
	總和	平均數	總和	平均數	總和	平均數	總和	平均數
Explore	41	5.125	40	5	40	5	39	4.875
Task1	20	5	18	4.5	21	5.25	25	6.25
Task2	18	4.5	18	4.5	17	4.25	23	5.75
Task3	19	4.75	17	4.25	25	6.25	27	6.75
Task4	20	5	19	4.75	16	4	25	6.25
Task5	25	6.25	27	6.75	19	4.75	21	5.25
Task6	22	5.5	25	6.25	17	4.25	17	4.25
Task7	21	5.25	25	6.25	21	5.25	20	5
Task8	24	6	27	6.75	20	5	18	4.5

#### 5 關聯提示詞圖示之功能對檢索的輔助程度：

在此項分析中，受限於實驗樣本的數量以及實驗設計的關係，無法針對有無圖書資訊學背景之受試者進行個別分析，因此，只針對各檢索任務利用 A 系統執行檢索時，分析當受試者「第一次輸入檢索詞時的滿意度」，以及當受試者「利用系統所提示之關聯詞的輔助程度」，利用此兩項數據前後的差異分析受試者對關聯提示詞圖示對檢索的幫助程度，各項數據如下：

表 4-30 顯示，各項檢索任務的數據在第一次輸入檢索詞的滿意程度皆達到 4 分以上，而當系統主動提示關聯詞時，對在提高檢索效果的程度上也高於 4 分以上，但是整體的分數略低於受試者對第一次所輸入檢索詞之檢索結果的滿意程度，尤其是在檢索任務 6 方面，其平均分數幾乎顯示是無法提高檢索效果。故初步得知，受試者皆滿意自己輸入之檢索詞，且當系統提示之關聯詞無法滿足自己心目中的認知以及達到某種程度的要求時，其結果會呈現出關聯提示詞圖示是有效果的，但是效果需再提昇至受試者可以接受的程度。

表 4-30：關聯提示詞圖示之功能對檢索的幫助程度比較表

	輸入之檢索詞能夠幫助你				提示之功能夠有效提高檢索效果			
	最小值	最大值	總和	平均數	最小值	最大值	總和	平均數
Explore-a	2	7	84	5.25	1	7	78	4.875
Task1	1	6	36	4.5	1	6	34	4.25
Task2	2	6	35	4.375	1	6	34	4.25
Task3	4	7	46	5.75	1	7	35	4.375
Task4	4	6	42	5.25	3	7	41	5.125
Task5	5	7	45	5.625	1	7	43	5.375
Task6	3	7	44	5.5	1	6	13	1.625
Task7	4	7	47	5.875	1	7	35	4.375
Task8	5	7	46	5.375	5	7	52	6.5

## (二) 檢索所需之時間 (Task Time) :

經由每個檢索任務結束後的問卷(Post-search Questionnaire), 收集每位檢索者在不同功能及不同任務間的時間分佈情形, 此步驟主要是針對兩個部份來分析: 一是分析有無提供關聯提示詞之功能的系統完成任務所需時間(不同功能間的差異); 二是分析各任務完成所需之時間(不同任務的差異)。各項分析分述如下:

### 1. 有無提示關聯提示詞之系統完成任務所需的時間 (不同功能間的差異):

由於本實驗設計為了避免相同的檢索任務因先使用某系統之後所造成的影響, 是故, 在不同功能間的差異可以從兩個角度去分析: 一是檢索任務 1 至檢索任務 4 分別在系統 A 及系統 B 的檢索時間分佈情形; 二是檢索任務 5 至檢索任務 8 分別在系統 A 及系統 B 的檢索時間分佈情形, 從上述兩個角度來代表不同功能間的時間效率差異。

檢索任務 1 至檢索任務 4 可以從表 4-31 及表 4-32 中得知, 在 A 系統的各檢索任務所花的檢索時間均高於在 B 系統檢索時的檢索時間, 且在 A 系統執行檢索任務 1 至檢索任務 4 總共所花的時間也高於 B 系統 (158 分鐘>135 分鐘); 至於檢索任務 5 至檢索任務 8 的分佈情形從下表 4-33 及表 4-34 得知, 也是呈現出 A 系統所花的檢索時間略高於 B 系統。最後, 就 16 位受試者較難的政治類檢索議題來比較, 也是呈現出 A 系統所花的檢索時間高於 B 系統; 因此, 綜合上述的數據可以初步分析, 有關聯提示詞圖示之功能在時間的效率上似乎反而花費較多的時間。

然而本實驗綜合開放式問題及訪談的質化資料可以初步解釋上述數據所代表的意義，一是有許多受試者在訪談中皆強調在 A 系統所花費的時間，往往是因為他提供了更多的資訊去連結其他資訊，且圖示之功能會激發檢索者去探詢更多的資料，因此如果所花的時間可以找到更精確的答案，也是檢索者願意樂見的情況；另一個原因則是系統提示之關聯詞往往也會造成檢索者需多一道步驟來判斷相關詞之關聯程度，因此，無形之中檢索者便會增加有別於一般關鍵字搜尋引擎之檢索路徑；然而，也有可能是因為受限於實驗之設計的關係，因為每個檢索任務在本實驗中只是單純記錄其檢索時所花費的時間，沒有嚴格限定在一定的時間內完成，且其中幾位受試者也強調如果實驗有規定時間的話，應該 A 系統會花費較少的時間。

表 4-31：利用 A 系統執行檢索任務 1 至 4 所需的時間

	ABn1	ABn2	ABh3	ABh4	BAh16	BAh15	BAn13	BAn14	平均	加總
Task1	10	23	4	11	14	6	6	2	9.5	76
Task2	7	7	1	4	11	3	2	3	4.8	38
Task3	3	3	5	1	2	3	2	2	2.6	21
Task4	4	6	1	4	1	4	1	2	2.9	23
										158

表 4-32：利用 B 系統執行檢索任務 1 至 4 所需的時間

	ABn5	ABn6	ABh7	ABh8	BAh12	BAh11	BAn9	BAn10	平均	加總
Task1	2	9	12	10	15	7	5	5	8.1	65
Task2	4	2	2	10	3	7	6	3	4.6	37
Task3	1	1	2	3	5	3	2	2	2.3	19
Task4	1	1	1	1	4	2	3	1	1.8	14
										135

表 4-33：利用 A 系統執行檢索任務 5 至 8 所需的時間

	ABn5	ABn6	ABh7	ABh8	BAh12	BAh11	BAn9	BAn10	平均	加總
Task5	3	3	2	9	4	1	2	2	3.3	26
Task6	1	3	3	6	7	1	2	2	3.1	25
Task7	3	4	3	10	3	3	8	2	4.5	36
Task8	1	3	1	6	5	2	3	1	2.8	22
										109

表 4-34：利用 B 系統執行檢索任務 5 至 8 所需的時間

	ABn1	ABn2	ABh3	ABh4	BAh16	BAh15	BAn13	BAn14	平均	加總
Task5	2	3	3	1	2	6	6	3	3.3	26
Task6	2	3	2	1	1	6	1	5	2.6	21
Task7	1	3	2	2	1	2	1	1	1.6	13
Task8	4	3	2	1	4	5	2	4	3.1	25
										85

## 2. 各任務完成所需的時間（不同任務間的差異）：

而下表4-35則是紀錄每個檢索任務所花費之時間的總和及平均數，從表中可以得知在探索性任務評估方面，由於每位受試者所訂定的檢索任務都不同，因此只能針對不同功能來做比較，從表中可以得知A系統所花費的時間略高於B系統（平均數A系統為2.625大於B系統2.312），也就是說探索性任務在有關聯提示詞圖示的功能會花費較多的檢索時間。至於在涵蓋面評估方面，為了模擬現實的檢索環境且測試議題之涵蓋面是否影響『關聯提示詞圖示』對使用者的輔助程度及差異，從時間點上來分析，檢索任務1（約8分鐘）及檢索任務2（約5分鐘）平均所花費的時間皆多於其他檢索任務（約2-3分鐘）。探究其原因有可能是基本資料問卷中，得知檢索者在檢索經驗中利用搜尋引擎檢索政治類的資訊頻率較低，因此，在此項數據中，也顯示政治類的檢索任務所花的時間均異於其他類型的檢索任務，故可以初步推論，檢索任務的涵蓋面和檢索者之搜尋經驗中的取向的確會影響到檢索的時間。是故，在設計檢索任務時，需盡可能考量檢索者的背景（如檢索經驗等），以便能夠真正呈現出系統真實面進而達到評估的效果。

表4-35：各項任務完成所需之時間敘述統計量表

	個數	最小值	對大值	總和	平均數	標準誤	標準差	變異數
Explore-a	16	1.00	6.00	42.00	2.6250	0.3966	1.5864	2.517
Explore-b	16	1.00	5.00	37.00	2.3125	0.3619	1.4477	5.096
Task1	16	2.00	23.00	141.00	8.8125	1.3669	5.4677	29.896
Task2	16	2.00	11.00	85.00	5.3125	0.7344	2.9375	8.629
Task3	16	1.00	5.00	41.00	2.5625	0.2882	1.1529	1.329
Task4	16	1.00	6.00	37.00	2.3125	0.4054	1.6215	2.629
Task5	16	1.00	9.00	52.00	3.2500	0.5284	2.1134	4.467
Task6	16	1.00	7.00	46.00	2.8750	0.5072	2.0290	4.117
Task7	16	1.00	10.00	49.00	3.0625	0.6289	2.5158	6.329
Task8	16	1.00	6.00	47.00	2.9375	0.4028	1.6612	2.596

說明：Explore-a代表探索性任務在A系統執行。

### (三) 使用者滿意度調查 (User Satisfaction) :

主要是利用完成單一檢索系統的問卷(Post-system Questionnaire)和檢索實驗後的問卷(Exit-search Questionnaire)來收集使用者的滿意度，可分為：

1. 『單一檢索系統的滿意度分析』，利用完成單一檢索系統的問卷(Post-system Questionnaire)收集具備圖書資訊學背景及非圖書資訊學背景的受試者，分別在A系統及B系統對各項功能的使用狀況分析，而分析之面向，可以從對『系統操作上的容易程度』、『對檢索任務的容易程度』、『對檢索結果的整體滿意程度』、『先前的知識背景對此次檢索任務的影響程度』以及『詢問此系統(檢索功能)能夠幫助你找到所預期不到的結果』等面向，構成使用者對單一檢索系統的整體滿意度之分析；2. 『完成所有檢索實驗後的整體滿意度分析』，此部分的分析主要是調查受試者，在進行完所有檢索任務的實驗之後，利用二元式的答題方式搭配開放性的問題，請受試者根據自己的使用經驗，直接判斷兩系統間的優劣；最後，再利用交叉分析的方式，以統計學上的無母數檢定資訊檢索能力是否會影響到檢索結果的滿意程度，細節分述如下：

#### 1. 單一檢索系統的滿意度分析：

此部分分析則是針對圖資背景的受試者與非圖資背景的受試者，分別利用有關聯提示詞圖示功能之系統或無關聯提示詞圖示功能之系統的整體滿意度分析，再搭配一些經由問卷及訪談所得出的質性資料做個綜合的比較，而在問卷的評量方式則是採用七等式量表，各項分析如下：

#### 1) 具備圖書資訊學背景的受試者：

##### (1) 有關聯提示詞圖示的系統(A系統)：

在表4-36中8位具有圖書資訊學背景之受試者在A系統的各項滿意度分數分別為：『上手(容易操作)』為6、『檢索任務容易程度』是5.375、『對在A系統執行之檢索任務的滿意程度』是4.875、『先被知識及背景影響此檢索實驗的程度』為5.25及『利用A系統的功能能夠找到你所預期不到的結果之程度』為5.25。從數據可以得知，整體滿意度的分數約接近為5分，故具備圖書資訊學背景的受試者，對於利用關聯提示詞圖示之功能執行的檢索任務，大致上來說是滿意的。

表4-36：具備圖書資訊學背景的受試者在A系統各項滿意度

	個數	最小值	最大值	總和	平均數	標準差	變異數
(上手)容易操作	8	5.00	7.00	48.00	6.0000	.7559	.571
檢索任務容易嗎	8	4.00	6.00	43.00	5.3750	.7440	.554
檢索結果滿意嗎	8	2.00	6.00	39.00	4.8750	1.2464	1.554
先被知識及背景的幫助	8	2.00	7.00	42.00	5.2500	1.6690	2.786
能夠找到您所預期不到的結果嗎	8	3.00	7.00	42.00	5.2500	1.2817	1.643
有效的 N (完全排除)	8						

## (2) 無關聯提示詞圖示的系統 (B 系統)：

至於在B系統的各項滿意度分數從表4-37可得知，分別為：『上手(容易操作)』是6.125、『檢索任務容易程度』是5.5、『對在B系統執行之檢索任務的滿意程度』是5、『先被知識及背景影響此檢索實驗的程度』為5.875及『利用B系統的功能能夠找到你所預期不到的結果』之程度為2.875；與在A系統的數據比較，除了在利用B系統的功能能夠『找到你所預期不到結果』分數只有2.875之外，其餘各項分數皆略高於在A系統的分數，故具備圖書資訊學背景的受試者，對於在B系統執行的檢索任務，除了不能達到『找到你所預期不到結果』之功用外，大致上來說，B系統的滿意度是高於在A系統的，也就是說沒有關聯提示詞圖示的功能的滿意度是高於有關聯提示詞圖示之功能的系統。

表 4-37：具備圖書資訊學背景的受試者在 B 系統的各項滿意度

	個數	最小值	最大值	總和	平均數	標準差	變異數
(上手)容易操作	8	5.00	7.00	49.00	6.1250	.6409	.411
檢索任務容易嗎	8	5.00	7.00	44.00	5.5000	.7559	.571
檢索結果滿意嗎	8	4.00	6.00	40.00	5.0000	.9258	.857
先備知識及背景的幫助	8	4.00	7.00	47.00	5.8750	1.1260	1.268
能夠找到您所預期不到的結果嗎	8	1.00	6.00	23.00	2.8750	1.7269	2.982
有效的 N (完全排除)	8						

## (3) 小結：

加上質性的開放式問題答案及訪談資料來說，發現具備圖書資訊學背景領域的受試者，雖然 A 系統的整體滿意度略低於 B 系統，但是可以初步得知有關聯提示詞圖示之功能，的確能夠幫助受試者『找到所預期不到的結果』。以下是幾位具備圖書資訊學背景之受試者的開放式問題答案：

『我很喜歡 A 系統中的關聯提示詞圖示，一目了然，而且提供了許多

我原本沒想到的相關詞彙。』 ABh7

『A系統有很多的自動關聯詞可選，可幫助我沒想到的』 BAh11

## 2) 非圖書資訊背景的受試者：

### (1) 有關聯提示詞圖示的系統（A系統）：

表4-38中8位非圖書資訊學背景的受試者各項平均分數分別為：『上手（容易操作）』是5.375、『檢索任務容易程度』是5.125、『對在A系統執行之檢索任務的滿意程度』是5.125、『先被知識及背景影響此檢索實驗的程度』為4.625及『利用A系統的功能能夠找到你所預期不到的結果之程度』為5.25。整體滿意度的分數也接近為5分，故非圖書資訊學背景的受試者，對於在A系統執行的檢索任務大致上來說也是滿意的。

表4-38：非圖書資訊學背景的受試者在A系統各項滿意度

	個數	最小值	最大值	總和	平均數	標準差	變異數
(上手)容易操作	8	4.00	7.00	43.00	5.3750	1.1877	1.411
檢索任務容易嗎	8	4.00	7.00	41.00	5.1250	1.3562	1.839
檢索結果滿意嗎	8	3.00	7.00	41.00	5.1250	1.3562	1.839
先被知識及背景的幫助	8	1.00	7.00	37.00	4.6250	1.7678	3.125
能夠找到您所預期不到的結果嗎	8	4.00	7.00	42.00	5.2500	1.2817	1.643
有效的 N (完全排除)	8						

### (2) 無關聯提示詞圖示的系統（B系統）：

表 4-39 顯示在 B 系統的各項平均分數分別為：『上手（容易操作）』5.375、『檢索任務容易程度』是 5、『對在 B 系統執行之檢索任務的滿意程度』是 5、『先被知識及背景影響此檢索實驗的程度』為 4.25 及『利用 B 系統的功能能夠找到你所預期不到的結果之程度』為 4.25。對於在 B 系統執行的檢索任務來說，整體滿意度的分數也接近為 5 分。



表4-39：非圖書資訊學背景的受試者在B系統各項滿意度

	個數	最小值	最大值	總和	平均數	標準差	變異數
(上手)容易操作	8	4.00	7.00	43.00	5.3750	1.0607	1.125
檢索任務容易嗎	8	3.00	7.00	40.00	5.0000	1.3093	1.714
檢索結果滿意嗎	8	4.00	7.00	40.00	5.0000	1.0690	1.143
先備知識及背景的幫助	8	1.00	5.00	34.00	4.2500	1.3887	1.929
能夠找到您所預期不到的結果嗎	8	2.00	7.00	34.00	4.2500	1.6690	2.786
有效的 N (完全排除)	8						

### (3) 小結：

加上質性的開放式問題答案及訪談資料來說，可以發現在 A 系統的各項分數皆略高於在 B 系統執行時，也就是說，非圖書資訊學背景領域之受試者認為，利用關聯提示詞圖示之功能的使用情況之滿意度高於無關聯提示詞圖示之功能的系統，且這些受試者還反應出 A 系統的關聯提示詞圖示功能有增加使用者掌握題目方向之用處，也認為此介面有別於一般文字式介面，其介面較生動活潑。以下是幾位非圖書資訊學系背景之受試者的開放式問題答案：

『系統 A 的關聯詞部分連結很多，且分關聯性遠近。系統 B 的邏輯符號功能很清楚，不會記錯。』 ABn5

『A 系統的圖示關聯詞系統的確有助增加使用者掌握題目方向，耕精確地找到資料，使用起來也很有趣(拉來拉去的..)，為關聯詞的簡寫和縮寫無法兼顧是一個小缺點。』 ABn6

### 3) 圖資與非圖資領域之受試者的綜合分析：

綜合圖資與非圖資之領域的受試者分析，可以發現除了『能夠找到你所預期不到的結果之程度』的分數是兩類受試者，皆一致認為 A 系統的滿意度高於 B 系統之外，其餘則呈現出相對的結果，換句話說，具備圖資領域之受試者，認為在系統 B 執行的各項面向都高於在 A 系統執行；反之，非圖資領域之受試者則認為 A 系統的各項滿意度則高於在 B 系統執行，雖然上述的數據差距都很小，但是，約略可以反應出一些現象，例如：具備圖資領域之受試者對自身的檢索技巧深具信心，因此，大都習慣於自己去嘗試擴展搜尋的關鍵詞，且也相對的要求系統提示之關聯詞的品質要達到一定的程度，才會採納系統建議之關聯提示詞，故呈現出沒有關聯提示詞圖示功能之整體滿意度略高於有關聯提示詞圖示之功能。然而，非圖資領域之受試者由於對自身的檢索技能較缺乏信心，因此，對有關聯提示詞圖示之功能接受的程度也相對的較高，所以才會呈現出與具備圖資領域之受

試者相反的結果。

## 2. 完成所有檢索實驗後的整體滿意度分析：

在進行完所有檢索任務指派的實驗之後，利用檢索實驗後問卷(Exit-search Questionnaire)收集受試者對使用之狀況的意見，因此，分別詢問受試者對『關聯詞』的認知狀況、以及是否使用過類似如系統 A 的檢索方式，以及詢問受試者是否有發覺兩系統之間的異同處，最後，再請受試者選擇心目中較容易學習的系統、哪個系統較方便使用以及選擇自己心目中最好的系統，利用受試者直覺的喜好來評定系統的使用情形，且搭配開放性問題請受試者將喜好的原因及對系統的建議將也記錄下來，各項數據及分析如下：

### 1) 具備圖書資訊學背景的受試者：

從表 4-41 顯示 8 位具備圖書資訊學背景的受試者，在五等式量表中各項平均數據分別為：對關聯詞的熟悉程度為 4.5、大都未使用過如關聯提示詞圖示之功能是 2.25、在使用的過程中慢慢發現兩系統間的異同處為 4.75。且從表 4-42 可以得知在詢問哪個系統較容易學習方面，有 3 個人選擇系統 A，5 個人認為差不多；在哪個系統較方便使用方面，有 6 個人認為系統 A 較方便使用，2 個人認為差不多；最後，在選擇心目中最好的系統方面，有 7 個人選擇系統 A，1 個人認為差不多，沒有人選擇系統 B。搭配上上述數據與下表 4-40 開放式問題答案的分析來說，在大部分的受試者都未曾使用過如關聯提示詞圖示之功能的系統下，還是認為此系統較容易學習及方便使用，且幾乎一致認為有關聯提示詞圖示功能之系統（選擇 A 系統的有 7 人）是心目中理想的系統，因為從表 4-40 與表 4-42 中可以得知都沒有人選擇 B 系統。

表 4-40：8 位具備圖書資訊學背景之受試者的開放式問題答案

受試者	開放式問題答案
ABh3	關聯詞的部份可以提示出想不到的關鍵字。
ABn7	我很喜歡 A 系統中的關聯提示詞圖示，一目了然，而且提供了許多我原本沒想到的相關詞彙。
ABn4	A 系統提供關聯詞圖示，讓使用者可以在不了解問題時，可進一步選擇其他相關詞彙檢索。
ABh8	A 系統的圖示有助於知悉相關詞彙。
BAh16	關聯詞可以幫助對主題不熟悉的檢索者對問題更能掌握。即使對主題熟悉，關聯詞可以擴大獲得答案的範圍。介面簡潔容易上手。
BAh12	圖示很有趣。

BAh11	A 系統有很多的自動關聯詞可選，可幫助我沒想到的。
BAh15	A 系統有關聯詞圖示，可將主題做分類，且擴充與主題相關的新聞。

表 4-41: 具備圖書資訊學背景之受試者的檢索實驗後問卷統計表

	您知道何謂關聯詞嗎	曾經使用過如系統a的檢索方式	您有發現兩系統有不同的地方嗎
個數	8	8	8
有效的	0		0
遺漏值			
平均數	4.5000	2.2500	4.7500
眾數	4.00 <sup>a</sup>	1.00	5.00
標準差	.5345	1.4880	.4629
最小值	4.00	1.00	4.00
最大值	5.00	4.00	5.00
總和	36.00	18.00	38.00

a. 存在多個眾數，顯示的為最小值。

表4-42：具備圖書資訊學背景之受試者的次數分配表

	較容易學習		較方便使用		心目中最好的	
	次數	百分比	次數	百分比	次數	百分比
系統 A	3	37.5	6	75	7	87.5
系統 B	0	0	0	0	0	0
差不多	5	62.5	2	25	1	12.5
總和	8	100	8	100	8	100

## 2) 非圖書資訊學背景的受試者：

表 4-44 則顯示 8 位非圖書資訊學背景的受試者，在五等式量表中各項平均數據分別為：對關聯詞的熟悉程度為 4.125、大都未使用過如關聯提示詞圖示之功能是 1.375、在使用的過程中慢慢發現兩系統間的異同處為 3.75。且從表 4-45 可以得知在詢問哪個系統較容易學習方面，有 4 個人選擇系統 A，2 個人選擇系統 B，2 個人認為差不多；在哪個系統較方便使用方面，有 5 個人認為系統 A 較方便使用，1 個人選擇系統 B，2 個人認為差不多；最後，在選擇心目中最好的系統方面，有 5 個人選擇系統 A，3 個人認為差不多，沒有人選擇系統 B。搭配上述數據與下表開放式問題答案（見表 4-43）可以得知，大部分的非圖書資訊學背景的受試者，也都未曾使用過如關聯提示詞圖示之功能的系統，且也是認為此系統較容易學習及方便使用，但是可以從表 4-45 中發現與具備圖書背景的受試者比較來說，三項數據中還是有人選擇傳統的無關聯提示詞圖示之功能的系統較容易學習及方便使用，且表 4-43 中也顯示出非圖書資訊背景的受試者，會選擇 B 系統的原因與數據相符。例如：ABn6 受試者的開放式問題答案。

表 4-43：8 位非圖書資訊學背景之受試者的開放式問題答案

受試者	開放式問題答案
ABn1	都差不多，所以也沒有比較喜好的部份，但 A 系統旁邊的 2D 有時候也會用到且比較引人注意，畫面也比較生動
ABn5	系統 A 的關聯詞部分連結很多，且分關聯性遠近。系統 B 的邏輯符號功能很清楚，不會記錯。
ABn2	A 系統的關聯圖。
ABn6	A：有關聯詞圖示系統，節省鍵入關聯詞的時間。 B：因為可以控制關聯詞的長短，有時反而會出現更多的資料。
BAn9	關聯詞圖示。
BAn13	A：關聯詞有時可提示有幫助的詞彙。 B：簡便，易上手。
BAn10	A：有關聯提示之功能。 B：較熟悉的介面。
BAn14	關聯詞圖示。

表 4-44：非圖書資訊學背景之受試者的檢索實驗後問卷統計表

	您知道何謂關聯詞嗎	曾經使用過如系統a的檢索方式	您有發現兩系統有不同的地方嗎
個數	8	8	8
有效的遺漏值	0	0	0
平均數	4.1250	1.3750	3.7500
眾數	4.00	1.00	4.00
標準差	.3536	.7440	.7071
最小值	4.00	1.00	3.00
最大值	5.00	3.00	5.00
總和	33.00	11.00	30.00

表4-45：非圖書資訊學背景之受試者的次數分配表

	較容易學習		較方便使用		心目中最好的	
	次數	百分比	次數	百分比	次數	百分比
系統 A	4	50	5	62.5	5	62.5
系統 B	2	25	1	12.5	0	0
差不多	2	25	2	25	3	37.6
總和	8	100	8	100	0	100

### 3) 圖資與非圖資領域之受試者的綜合分析：

為了明顯的比較有無圖書資訊學背景之受試者的差異，我們再將 16 位受試者的數據利用表格來顯示（見表 4-46），在五等式量表中各項平均數據分別為：

對關聯詞的熟悉程度為 4.3125、所有受試者幾乎都是第一次使用關聯提示詞圖示之功能、在使用的過程中慢慢發現兩系統間的異同處，分數為 4.25；而在詢問哪個系統較容易學習方面（見表 4-47），有 7 個人選擇系統 A，有 2 個人選擇系統 B，另外有七個人認為差不多；在哪個系統較方便使用方面，有 11 個人認為系統 A 較方便使用，4 個人認為差不多，只有 1 個人選擇系統 B，最後，在選擇心目中最好的系統方面，有 12 個人選擇系統 A，4 個人認為差不多，沒有人選擇系統 B。

如果是單純從實驗的數據，可以發現就使用的狀況而言，所有受試者大都認為有關聯提示詞圖示之功能的系統（系統 A）是較容易學習且方便使用，只有少數非圖資領域之受試者選擇 B 系統，且所有受試者在心目中認為有關聯提示詞圖示之功能（A 系統）比沒有關聯提示詞圖示之功能（B 系統）好，因為沒有人選擇 B 系統是心目中較好的系統。

表 4-46：16 位受試者的檢索實驗後問卷統計表

	對於關聯詞的熟悉程度	曾經使用過如系統 A 的檢索方式	您有發覺兩系統兼有不同的地方
個數	16	16	16
平均數	4.3125	1.8125	4.25
平均數的標準誤	0.1179	0.3058	0.1936
眾數	4	1	5
標準差	0.4787	1.2230	0.7746
變異數	0.2292	1.4958	0.6000
最小值	4	1	3
最大值	5	4	5

表 4-47：16 位受試者的次數分配表

	較容易學習		較方便使用		心目中最好的	
	次數	百分比	次數	百分比	次數	百分比
系統 A	7	43.8	11	68.8	12	75
系統 B	2	12.5	1	6.3	0	0
差不多	7	43.8	4	25	4	25
總和	16	100	16	100	16	100

### 3. 各項變數與檢索結果的交叉分析：

此部分分析的目的，主要是因為在評估互動式檢索系統的文獻中，其實驗結

果大都顯示受試者的資訊檢索能力會影響到檢索結果的滿意度，因此，特別針對此一因素是否適用在本實驗的檢索環境中來做一個探討與比較；再者，進一步檢定『資訊檢索能力』與『系統操作上的容易程度』是否相關，以作為進一步分析與推論的依據。

由於本實驗的樣本數在統計學上是屬於小樣本的資料量，且為了不受母體分配形狀的限制，我們採用無母數的檢定方法來檢定上述的分析。細節分述如下：

### 1) 資訊檢索能力與檢索結果的相關檢定：

#### (1) 有關聯提示詞圖示的系統（A 系統）：

$$H_0 : u = 0$$

$$H_1 : u \neq 0$$

虛無假設為資訊檢索能力與 A 系統檢索結果的滿意度不相關

從表4-48中可以得看出，資訊檢索能力與檢索結果滿意度的相關係數雖然呈現出正相關0.293，這表示資訊檢索能力愈高的受試者，對A系統檢索結果的滿意度也愈高。但是在資訊檢索能力與檢索結果之滿意度互不相關的假設檢定值為  $P = 0.270 > 0.05$ ，故無法接受虛無假設，即得到資訊檢索能力的高低對於在A系統檢索結果的滿意度是不相關的。

表4-48：A系統資訊檢索能力與檢索結果滿意度相關表

			資訊檢索能力	檢索結果滿意嗎
Spearman ' s rho係數	資訊檢索能力	相關係數	1.000	0.293
		顯著性（雙尾）	.	0.270
		個數	16	16
	檢索結果滿意	相關係數	0.293	1.000
		顯著性（雙尾）	0.270	.
		個數	16	16

#### (2) 無關聯提示詞圖示的系統（B 系統）：

$$H_0 : u = 0$$

$$H_1 : u \neq 0$$

虛無假設為資訊檢索能力與 B 系統檢索結果的滿意度不相關

從表4-49中可以得看出，資訊檢索能力與檢索結果滿意度的相關係數雖然呈現出正相關0.137，但是在資訊檢索能力與檢索結果之滿意度互不相關的假設檢定值為  $P = 0.270 > 0.05$ ，故無法接受虛無假設，即得到資訊檢索能力

的高低對於在B系統檢索結果的滿意度是不相關的。

表4-49：B系統資訊檢索能力與檢索結果滿意度相關表

			資訊檢索能力	檢索結果滿意嗎
Spearman ' s rho係數	資訊檢索能力	相關係數	1.000	0.137
		顯著性 ( 雙尾 )	.	0.614
個數		16	16	
	檢索結果滿意	相關係數	0.137	1.000
		顯著性 ( 雙尾 )	0.614	.
		個數	16	16

## 2) 資訊檢索能力與系統操作上的相關檢定：

### (1) 有關聯提示詞圖示的系統 ( A 系統 )：

$$H_0 : u = 0$$

$$H_1 : u \neq 0$$

虛無假設為資訊檢索能力與 A 系統操作的容易程度不相關

從表4-50中可以得看出，資訊檢索能力與A系統操作的容易程度之相關係數雖然呈現出正相關0.395，但是在資訊檢索能力與A系統操作的容易程度互不相關之假設檢定值為  $P = 0.130 > 0.05$ ，故無法接受虛無假設，即得到資訊檢索能力的高低對於在A系統操作的容易程度是不相關的。

表4-50：A系統資訊檢索能力與容易操作相關檢定表

			資訊檢索能力	檢索結果滿意嗎
Spearman ' s rho係數	資訊檢索能力	相關係數	1.000	0.395
		顯著性 ( 雙尾 )	.	0.130
個數		16	16	
	( 上手 ) 容易操作	相關係數	0.395	1.000
		顯著性 ( 雙尾 )	0.130	.
		個數	16	16

### (1) 無關聯提示詞圖示的系統 ( B 系統 )：

$$H_0 : u = 0$$

$$H_1 : u \neq 0$$

虛無假設為資訊檢索能力與 B 系統操作的容易程度不相關。

從表4-51中可以得看出，資訊檢索能力與B系統操作的容易程度之相關係數雖然呈現出正相關0.223，但是在資訊檢索能力與A系統操作的容易程度互不

相關之假設檢定值為  $P = 0.407 > 0.05$ ，故無法接受虛無假設，即得到資訊檢索能力的高低對於在B系統操作的容易程度是不相關的。

表4-51：B系統資訊檢索能力與容易操作相關檢定表

			資訊檢索能力	檢索結果滿意嗎
Spearman ' s rho係數	資訊檢索能力	相關係數	1.000	0.223
		顯著性 ( 雙尾 )	.	0.407
		個數	16	16
	( 上手 ) 容易操作	相關係數	0.223	1.000
		顯著性 ( 雙尾 )	0.407	.
		個數	16	16

## (2) 小結：

從上述A系統與B系統的統計數據可以得知，在本實驗環境中，受試者的『資訊檢索能力』與『檢索結果的滿意程度』和『操作上的容易程度』，沒有呈現出顯著的相關。初步分析，可能是受制於實驗的設計，每個人對於自身的資訊檢索能力檢定有不同的標準，且再仔細探究其相關係數的數值，可以發現大都是呈現出正相關的分配，因此雖然在統計學上的角度是不相關的，但是至少可以看出出資訊檢索能力愈好的人其對檢索結果的滿意度也會愈高。

## 4. 綜合分析：

綜合單一檢索系統及檢索實驗後的問卷調查結果可以初步歸納出，無論是圖資與非圖資領域之受試者，皆一致認為有關聯提示詞圖示之功能的系統皆能夠幫檢索者找到『所預期不到的檢索結果』，就系統的整體滿意度來說，具備圖資背景的受試者對有關聯提示詞圖示功能之系統，雖然也是滿意的，但是其滿意程度略低於無關聯提示詞圖示功能之系統，這或許與具備圖資背景之受試者在檢索技巧較熟悉，且相對的對提示之關聯詞的品質要求較高，因此會習慣的相信自己所下的關聯詞；然而，對非圖資領域的人來說，由於本身在檢索技巧較缺乏，故沒有先入為主的觀念，對任何系統只要在操作上沒有太大的難處，縱使對檢索結果只有一點的幫助，其接受的程度自然會比有圖資背景領域之受試者來的高且滿意程度也會呈現出較明顯的提升。以下分別為圖資與非圖資領域之受試者的訪談答案：

『當我使用 A 系統時，若關聯提示詞中沒有出現任何相關詞時，我會覺得 A 系統的相關資料很少；但若再以布林邏輯的方式檢索，卻又會得到許多



資料。』 ABh7(圖資領域)

『有關聯詞圖示系統，節省鍵入關聯詞的時間』 ABn6(非圖資領域)

#### (四) 介面的可用性評估 (Usability Evaluation):

在可用性評估方面的分析，利用附錄六 Jakob Nielsen 的 10 個可用性評估準則所設計出的可用行訪談問題，將 16 位受試者經由錄音所收集的資料，以質化分析的方式 (Top-Down) 的方式，重新歸納及整理出下列五項原則來評估關聯提示詞圖示之功能的可用性：分別是 1. Visibility of system status (視覺上的設計)；2. Match between system and the real world (與現實環境的關聯程度)；3. User control and freedom (使用系統的自由程度及操作系統的方便性)；4. Recognition rather than recall (以認知取代記憶)；5. Aesthetic and minimalist design (美學及最小努力原則之設計)。

##### 1. Visibility of system status (視覺上的設計):

綜合所有受試者的意見，均一致表示關聯詞上的名次顯示功能沒有太大的意義，因為就視覺上的直覺反應，當系統提示出關聯詞時，受試者會很直接從螢幕上去瀏覽他心中目中認為相關的詞彙；至於系統利用遠近的方式來展現詞彙的關聯度，許多受試者均反應，距離檢索詞愈近的關聯詞愈容易造成視覺上的混淆，希望能夠利用顏色來區別前幾個相關的詞彙；而在關聯詞的顯示數目上，有些受試者認為顯示關聯詞的數目多少是不重要的，最重要的是顯示出來是否相關(即關聯程度)，有些受試者則認為顯示出的關聯詞數目約介於 5 至 10 個左右即可，因為太多關聯詞在狹小的區塊裡反而會造成畫面的混淆；另外就關聯詞展現的方式，均認為此 2D 的放射狀圖示有助於提高探索新知的欲望，且有別於一般的文字展現模式。就介面上的創新性是值得嘉許的。

值得注意的是有一位受試者建議此關聯詞圖示介面可以增加讓使用者**直接在顯示出的關聯詞上輸入或減少相關字詞的功能**，以便直接進行檢索，而不需要額外在檢索欄位裡重新輸入檢索詞彙，因為當檢索者想要重新修正第一次下的檢索詞彙時，往往需要在回檢索欄位裡重新輸入，感覺上使用的方式過於複雜且繁瑣。以下是幾位受試者的訪談內容及答案：

##### 訪談問題：

- (一)本系統利用 2D 的圖示呈現各個關聯詞的方式是否能夠協助您了解所收集資料的範圍？ (請說明原因)
- (二)關聯詞的關聯程度利用距離檢索詞的遠近來表示是否能夠增進您

對欲查詢主題的初步瞭解？

(三)關聯詞的關聯名次是否能夠幫助您縮小查詢範圍？

**受試者的回答：**

➤ **圖資領域：**

『有幫助的，但是沒有去點選那個名次，比較容易瞭解資料的範圍，但是有些詞彙不太相關。』 ABh3

『一開始會不太適應，因為在使用上的經驗較缺乏。數字上的提示反而會混淆視覺。關聯提示詞的功能在介面上擺放的位置，感覺上不太方便，因為還要往下拖曳才能看到提示的詞彙。』 ABh7

『圖示的功能滿清楚的，可以大概了解一下主題的範圍。但是沒有使用到名次這個功能。有時覺得詞彙還滿擠的。』 ABh4

『有幫助，比較奇怪的是鐵達尼號哪題的檢索任務，竟然沒有提示出奧斯卡此一主題。關聯詞的遠近沒有感覺，如奧斯卡提示出的關連詞彙的第一名是電影，比自己心目中的關聯詞彙還廣泛。』 ABh8

『很明顯不會去點太遠的關聯詞彙。』 BAh16

『第一個畫面的顏色可以區分，如：檢索出的檢索結果區塊可以和其他區塊區分，版面上有一點擁擠。』 BAh12

➤ **非圖資領域：**

『習慣上不太會注意到名次的顯示，因為很直覺的就會直接去尋找關聯詞。』 ABn1

『有縮小範圍，以自己的資訊需求來說，如在查詢有關廣告企劃案之類的文章，它可以進一步做到幾個面向分析的功能，不會像一般搜尋引擎出來一大堆資料。沒有利用到名次顯示這個功能，是因為直接從關聯詞的遠近就可以判斷了，但是，較近的關聯詞不一定是心中所要的提示詞且就算是最相關不一定是相關的。』 ABn5

『名次上的功能沒有注意到，因為心目中有相關的詞彙。圖示展現出的程度跟自己心目中的認知不一樣。』 ABn2

『可以再加一些額外的功能，如用拖曳的方式讓詞彙集中，且可以直接在圖示上輸入心目中的檢索詞彙，可以限定主題的範圍。名次的功能沒有幫助的原因是因為直接從遠近就可以得知關聯詞程度。』 ABn6

『有幫助，但是乍看之下會有點眼花撩亂，會亂彈。關聯詞和檢索詞的距離如果很近的話，容易擠在一起。』 BAn13

『放射狀的展現方式，需避免去混淆到視覺上的觀感。』 BAn10

『可能對不熟悉的題目較有幫助，』 *BAn14*

## 2. Match between system and the real world (與現實環境的關聯程度)

由於此系統的關聯詞功能在現行的資訊檢索環境中，應用的對象大都以公司行號企業內部的知識檢索為主，而本實驗為了探討是否能夠應用在一般的網路環境中，也就是網頁的全文檢索，特別利用開放式訪談問題，詢問受試者對此關聯詞提示詞圖示功能應用在一般搜尋引擎之意見。

經過分析整理之後，所有受試者均一致認為可以嘗試應用在搜尋引擎上，且認為應可以有效過濾網路上龐大的資料量，因為檢索者在進行一道檢索議題時，心中的檢索詞彙往往以一個檢索詞來代表整個問題，這與國內學者卜小蝶曾經分析台灣網路搜尋引擎的檢索詞彙具有檢索詞彙短、存在核心詞彙的研究相似，所以當使用者只輸入一個檢索詞時系統會自動提示其他關聯詞，無疑是幫助使用者增加檢索的詞彙，進而提升檢索效果，且還認為此一功能還可以意外的發現意想不到的相關資訊。然而，有許多受試者也反應此一功能也有可能會在找資料的過程中偏離原先的主題，而迷失在龐大的資料量中，因此必須有相關機制的設計，來提示使用者目前所在的路徑或檢索歷程等。以下是幾位受試者的訪談內容及答案：

### 訪談問題：

- (一)您覺得本系統的關聯詞圖示功能是否適用於一般的搜尋引擎？
- (二)對於系統的關聯詞圖示而言，您是否有任何的意見？

### 受試者的回答：

#### ➤ 圖資領域：

『合適，因為可以提示出沒有想到的關鍵字。』 *ABh3*

『合適，可以幫助使用者，尤其是一般使用者，且還可以和一般的布林邏輯一起使用。』 *ABh7*

『合適，因為可以依照檢索詞彙的相關主題來縮小範圍。』 *ABh4*

『可以嘗試配合，因為網路的資料太多，且感覺上它是從資料庫的內容去篩選，會比較貼切檢索環境。』 *ABh8*

『合適，因為網路的資料量太多，且如果只是漫無目的在瀏覽網路資料，此一功能可以意外的發現許多沒想到的相關資訊，縱使在搜尋特定的主題或目的也可以因為關聯詞的提示而找到更精確的答案。』 *BAh16*

『需要多一些宣導及配套措施,因為一般大眾可能會排斥再多學習新的搜尋功能。』 *BAh12*

『沒什麼特別的感覺。』 *BAh15*

➤ **非圖資領域：**

『圖形介面較有趣不會太呆版,會吸引人使用。』 *ABn1*

『可以試試看,因為可以幫助過濾較多的資料。』 *ABn5*

『因為可以稍微提示使用者其它詞彙。』 *ABn2*

『合適,因為網路的資料太多,且可以提示出幾個較具體的方向。』 *ABn6*

『滿合適的,打一個字可以提示出很多沒想到的關鍵字,』 *BAn9*

『此功能適合較龐大的資料內容,如果是用在資料量小的會得到反效果。』 *BAn13*

『合適的原因是因為可以稍微縮小資料範圍,不合適的原因是因為花太多時間在找資料,覺得有可能會偏離原先要尋找的主題。』 *BAn10*

『只要有關聯詞的這個功能就可以,至於呈現的方式影響比較不大。』 *BAn14*

3. **User control and freedom (使用系統的自由程度及操作系統的方便性)：**

無論是具備圖書資訊學領域或非圖書資訊學領域之背景的受試者皆認為在轉換系統的過程中,由於習慣性的關係,縱使 A 系統有提供關聯提示詞的功能,還是會傾向於利用傳統的布林邏輯檢索方式,且大部分的使用者都把關聯提示詞圖示的功能當作輔助布林邏輯的檢索工具,因此,必須經過幾次的檢索經驗之後,受試者才會直接在關聯詞圖示上面直接操作,進而取代文字輸入的方式。其原因有可能是因為檢索者在未熟悉的狀況下往往懼於使用新的資訊科技,且以目前的檢索環境而言,文字式檢索系統大都以文字的輸入方式最為普遍,如將來需導入此視覺化的關聯提示詞圖示功能,必須花相當多的時間在宣導上面。

至於在操作上,有許多受試者反應利用點選滑鼠左鍵的一下和兩下來區別布林邏輯中的 And 和 Or 的關係,在操作上易造成混淆,因為受試者常常點錯次數,且也建議系統應該有提供重新檢索的功能,因為當檢索詞需回至上一層時,往往需要按下回上一頁的按鈕。以下是幾位受試者的訪談內容及答案：

**訪談問題：**

(一)您覺得本系統的關聯詞圖示是否容易操作？

(二)當您操作上述的功能,是否有遇到任何挫折？

- (三) (接上題) 如果有, 請您說說看遇到挫折的情形如何?
- (四) (接上題) 您描述的挫折對您使用系統的便利性造成多大的影響?

**受試者的回答：**

➤ **圖資領域：**

『有一些關聯詞提示不太相關, 最後會放棄提示的功能然後在直接去下檢索詞彙。』 ABh3

『操作上只要習慣了應該還可以。』 ABh7

『操作上如果是確定是要用上述的方式, 應該有一個操作的使用說明。』 ABh4

『覺的操作上, 還可以再友善一點, 如點選之後要移除, 好像沒有此一功能, 需點選上一頁的按鈕才會回到前一層。』 ABh8

『此操作方式只要習慣之後應該可以接受, 也可以考慮利用顏色來區分前幾名的關聯詞。』 BAh16

『此一構想還不錯, 但是不符合一般的使用情況, 需額外說明使用方式。』 BAh12

『此一構想是不錯, 但是如果沒有說明可能會不知所措, 且不太習慣。勾選的名次可以一開始就設定好了。』 BAh15

➤ **非圖資領域：**

『操作方法沒有什麼差異。在平常使用的習慣很少運用到布林邏輯。習慣直接使用單一詞彙檢索。』 ABn1

『B 系統已經習慣了。A 系統由於我自己心目中已經有預設的關鍵字及關聯詞, 所以覺得操作上可能在圖示上有點混淆, 因為必須尋找相關的詞彙去做比對。』 ABn5

『操作上和一般檢索系統沒有什麼不同。』 ABn2

『點一點和點兩點容易點錯, 希望利用區別性較大的方式這樣才不會混淆。』 ABn6

『按錯的時候無法回上一步, 或重新檢索。』 BAn9

『點選的方式會混淆, 希望區別性較大。』 BAn13

『點選的方式易混淆, 希望區別更大一點。』 BAn10

**4. Recognition rather than recall (以認知取代記憶):**

在減輕使用者對未知資訊認知的負擔方面, 受試者大部份都覺得由關聯詞圖示可以大略知道欲查詢之主題的範圍, 因為一圖勝千言。至於在已知或熟悉的主

題上，受試者皆反應提示之關聯詞和自己心目中的關聯詞不太吻合，且受試者往往傾向利用自己心目中的檢索詞彙來檢索。這或許跟檢索者對系統認知的落差所產生的反應，因為檢索者往往會忽略檢索的資料庫內容不是涵蓋所有的資訊，而從自身的角度去衡量此系統所產生的關聯提示詞，因此，此情況的發生可能是受陷於實驗環境的限制，所產生出特有的實驗結果。最後，關聯提示詞圖示的功能的確有指引使用者在未知的領域去尋求相關資料的方向，且降低使用者記憶上的負擔，因為受試者不需要記住每個與檢索詞相關的關聯詞。但是，必須事先告訴使用者此系統提示的關聯詞，是根據所涵蓋的資料內容本身而自動產生的。以下是幾位受試者的訪談內容及答案：

#### 訪談問題：

與一般搜索引擎相比，本系統利用 2D 的關聯詞圖示介面，提示相關之檢索概念，是否可以有效減輕您認知上的負擔？

#### 受試者的回答：

##### ➤ 圖資領域：

『在不熟悉的主題上，可以從關聯題詞的功能提示幾個方向。可以稍微降低心理認知上的負擔。』 ABh7

『畫面可以再大一點，因為詞彙之間的關係感覺上有點混亂。』 ABh8

『如果抓的關聯詞程度能夠再提高的話，是很有幫助，顯示的方式讓人家一目了然的知道整體資料的範圍，可以當機立斷決定所要的資訊。因為如果沒有下到我想到的檢索詞，系統可以主動提示，』 BAh16

『所謂一圖勝千言，有圖可以激發使用的興趣，關聯詞的遠近沒注意到，一開始還是會先用眼睛瀏覽一次自己心目中的詞彙，有一些專業性的題目，的確可以幫助。』 BAh12

##### ➤ 非圖資領域：

『A 系統至少知道資料收集的範圍。』 ABn1

『在不熟悉的主題上 B 系統較有幫助，因為可以直接控制檢索詞彙，將自己心中所想的東西一直增加下去，B 系統沒有幫助的原因是提示之關聯詞我還是要一個一個點選下去。放射狀的圖示會有渾淆的感覺，可以利用較具規則性的樹狀排列，一層一層展開。』 ABn5

『本身對議題不清楚有幫助，議題較不熟悉的會比較依賴關聯詞圖示。但是就剛剛的檢索經驗來說，好像沒有什麼幫助，因為提示出的詞彙覺得不太相關。』 ABn2

『此一方式還滿清楚。』 ABn6

『可以稍微了解一下主題，不熟悉的會使用系統 A，且可以稍微縮小範圍。但是熟悉的主題還是會以自己心目中的詞彙為主。』 BAn9

『對不熟悉的議題較有幫助，發覺提示出的詞彙關聯程度，和心中所想的不一樣。和習慣上的操作也不太一樣。』 BAn13

『是有幫助，對不熟悉的領域可以靠周邊的關聯詞來了解資料的範圍。』 BAn10

## 5. Aesthetic and minimalist design (美學及最小努力原則之設計):

受試者一致覺得本系統之關聯提示詞圖示功能能夠一目了然的瞭解欲查詢之主題的資料分佈情形，且可以透過放射狀圖示迅速的瞭解詞彙間的關聯程度。但是就時間上來說，有些受試者認為此功能如果在限定時間內需完成某些問題，有提供關聯提示詞圖示之功能的確能夠在短時間之內瞭解不熟悉的專業知識，因為其呈現的方式能夠讓檢索者預先瀏覽資訊；然而某些受試者卻呈現出相反的意見，主要原因是關聯提示詞圖示功能的確有幫助，但是很容易迷失在層層的資訊量中，且必須增加額外的判斷時間來過濾系統提示之關聯詞是否為相關的詞彙，再加上如果系統提示出意想不到的關聯詞時，檢索者往往會因為好奇心而去點選與查詢主題不相關的資訊，進而迷失在茫茫的資訊量中。

還有受試者反應可能受到顏色、區塊大小、數字、文字等因素的干擾，反而更不易找到所需的主題，這與資訊視覺化的研究中常發生的視覺混淆 (Visual cluttering) 有很大的關係。但在美學觀點上，部份受試者表示較一般分類式目錄更佳。以下是幾位受試者的訪談內容及答案：

### 訪談問題：

### 受試者的回答：

#### ➤ 圖資領域：

『A 系統應該會比較快。』 ABh7

『沒有差別。』 ABh8

『因為關聯詞可以擴充搜尋的好奇心，系統 A 有多提示所以會花較多的時間。不管主題熟不熟悉，對熟的題目如果有感興趣的話，願意花較多時間去看，題目是會有影響，如華航空難這題系統 A 會提示關聯詞，還是會想知道更多的相關資料 對議題不熟悉關聯詞幫助較大，可以收斂查詢題的範圍。』 BAh16

『系統 A 可以縮短檢索的時間。』 BAh12

『覺得 A 系統必須花較多的時間去思考此一詞彙是否相關，所以反而會花較多

的時間。』 BAh15

➤ 非圖資領域：

『會縮短時間，不用打字且會縮短範圍。』 ABn1

『時間上沒有大的差異。』 ABn5

『在 A 系統的題目好像較專業，且因為有提示出其它相關的詞彙，所以會花較多的時間去過濾資料，如果是在 B 系統查詢時，當檢索結果出現時因為沒有額外的資訊，所以會傾向直接去尋找答案。』 ABn2

『題目影響較大，相較起來 A 系統應該會比較快。』 ABn6

『有時候會花較多的時間去過濾一些相關的資料，但是有時間壓力的話應該 A 系統可以較快找到答案。』 BAn10

『A 系統因為有提示可以一目了然。』 BAn14

### 第三節 綜合分析與討論

本研究主要是嘗試建立出評估互動式檢索系統的方式，且進而探討互動式檢索系統中關聯提示詞圖示之功能的成效與品質，期望藉由量化與質化的方式來評估，以作為未來互動式檢索系統在考量功能成效之餘，還能夠兼顧使用者的意見及回應。

因此在本章的最後一節，將針對目前文獻與本研究探討議題相關的研究發現中，來進行綜合的分析與比較，以期能夠得到更深入的認識，本節主要分成兩大議題來進行探討，分別為：(1) 關聯詞的品質與文件集之大小的關係；(2) 互動式檢索輔助功能的使用評估。利用兩者的分析與比較，進而達到兼具量化與質化的評估方式。

#### 一、 關聯詞的品質與文件集之大小的關係：

綜合關聯詞的比例與計算不同文件集相關詞彙分佈之平均查準率的分析，可以明顯的發現有一個共通的現象，就是從小文件集至中文件集的各项數據都是呈現出明顯的增加，且在嚴格的統計學 ( Paired T-test ) 檢定下，也證明在此實驗環境中 ( 小文件集與中文件集的比較 )，數據上有一定的效用，故我們初步的推論，無論是在利用關聯詞的比例或是計算不同文件集間相關詞分佈的查準率數據下，都可以明顯的發現其關聯詞的數量都有顯著的增加，而增加的最大原因，應該是受到文件集間年代分佈不同而產生出的影響，例如：小文件集的資料是 2000 年至 2001 年，增加至中文件集，多了



1998 年和 1999 年的新聞資料。且在新聞的資料特性中又有『集中報導』的一個趨勢，故相同年代的新聞資料其報導內『大同小異』，因此，縱使增加一倍以上新聞資料，無論是在關聯詞的比例或平均查準率的數據均不會受到太大的影響，例如：中文件集增加至大文件集的資料都是相同年代的資料，不同的地方是，大文件集多了一倍的新聞資料，故在『大同小異』之下，此『小異』的情形有時候還可能會造成關聯詞的降低。

而因為『小異』情形，所以我們進一步的檢定中文件集至大文件集的查準率是否在統計學上有顯著增加的趨勢，而結果顯示在統計學上的確不具代表任何意義，可見其關聯詞的平均查準率，不單只是受到文件集大小的影響，還有受到其它因素的影響。故我們進一步去探析單一詞彙在三個文件集的分佈情形，分析結果顯示，30 個查詢詞中有 15 個查詢詞的查準率是呈現出中文件集優於大文件集的分佈情形，且相關詞的詞數約與中文件集相同，也就是說關聯詞詞數沒有明顯的降低，但是其關聯詞的分佈情形就沒有中文件集來的集中，例如：50 個提示的詞彙中，分佈在前 80% 的差異。所以整體而言，平均查準率(Interpolated Recall - Precision Averages)當然降低。總而言之，當相同年代的文件量增加至一定的大小時，如果沒有增加新的關聯詞，反而會降低其關聯詞分佈的集中情形，換句話說，讓原本關聯詞程度較高的詞彙，會因為一些不相關的雜訊而降低其門檻，使其關聯詞的分佈呈現出較發散的情形。

再者，本實驗與曾元顯的研究比較發現，在不同的『判斷者』與『判斷尺度』下，本實驗小文件集的相關比例與曾元顯相關性評估實驗中依『強度』的相關比例數據同為 69%。是故，就本實驗的相關比例數據而言，相關判斷的結果沒有因為『判斷者』與『判斷尺度』不同而產生出明顯的差異，分析其影響的因素是因為相關判斷的『文件』(資料內容)影響之因素大於上述兩項因素，因為在文獻中指出，較特定及專指的內容，較容易產生出相同的判斷結果。由此可知，相關的評估結果是要『因地制宜』的，也就是說依照不同的實驗環境及系統來設計不同相關判斷的準則。

## 二、 互動式檢索輔助功能的使用評估：

『關聯提示詞』是一種行之有年的互動式檢索輔助功能，國內外不乏已有建置完成的系統且開發出更創新的呈現方式，例如：國外亞利桑納大學人工智慧實驗室的 ET-Map，利用一種自我組織的演算法 (Kohonen Self-Organizing Map, SOM)，以地圖的方式呈現出詞彙間的關聯度；而國內在關聯詞的展現方式大都還是停留在文字介面的形式。姑且不論其在關聯詞的展現形式如何，在評估此類的系統功能時，國外已經發展到利用量化與質

化的方式來評估關聯提示詞視覺化後的地步，但國內大都還停留在利用量化的方式來評估文字式的關聯提示詞介面。因此，本研究嘗試建立評估中文的互動式檢索輔助功能之評估方式，藉以達到讓檢索系統更趨完善且符合系統發展的使用者導向需求，以下針對本研究所歸納的幾個面向來進行探討。

### 1. 效能 (Task Completion):

『效能』就是指檢索完成之程度，其最大的目的就是針對本實驗的互動式檢索輔助功能是否能夠有效的提高檢索效果進行探討，根據研究調查發現，具備圖資背景的受試者，由於自身的檢索經驗和電腦經驗大都優於非圖資背景的受試者，因此造成對系統所提示的關聯詞產生出不信賴感，故會傾向於相信自己所下的檢索詞彙，相對的也會要求系統提高更精確的關聯詞；然而對非圖資領域背景來說，在第一次輸入完檢索詞彙時，對系統所提示的關聯詞，其依賴程度高於圖資領域背景的受試者，因為這些受試者本身對檢索系統的使用經驗較缺乏，且在資訊檢索技巧也比較生疏，因此，當系統只要能夠在他們的檢索歷程中，提供一點的小暗示，對他們來講都是一種很大的幫助。整體而言，除了上述因素是比較明顯的差異之外，所有受試者皆反應只要關聯詞的關聯程度高，對不熟悉的檢索主題的確是有很大的幫助。

在提示之關聯詞方面，本研究還發現，檢索者對系統提示的關聯詞認知是有差異的，無論是有無圖資背景領域的受試者，在檢索的過程中，都將關聯詞的關聯程度利用自身的知識背景及認知來衡量，因為檢索者不習慣這些關聯詞是由檢索系統本身的內容所產生，所以往往系統判定為相關的詞彙，但是檢索者卻不認為是相關的落差；再者，有些受試者還反應系統斷定為最相關的詞彙，往往比自己所下的檢索詞彙的概念還廣泛。而從實驗的數據還得知，當檢索者輸入的檢索詞是人名時，對系統所提示出關聯詞的滿意度為最高，例如：“檢索任務八”為有關於北野武導演的電影文章，而檢索者大都輸入「北野武」人名的檢索詞，對系統提示之關聯詞的程度，檢索者都給予很高的分數，且在“檢索任務四”中檢索者所輸入的檢索詞大都為「佐佐木主浩」，其評分也都高於 5 以上。上述這些調查結果，也正好反應出關聯詞技術發展的現況，因為，此一系統是經由單一文件中詞彙間共現 (Co-Occurrence) 的方式，求得兩詞彙間的關聯度，而新聞文件中又以人名出現的頻率為最高且最具代表性，故經由系統斷定出的關聯詞品質往往也會高於其他詞類。因此，實驗的結果才會出現此一情形。

最後，檢索任務的不同也和檢索的效能有極大的關聯，例如：在“檢索任務六”是詢問『與使用被稱為體細胞核移植的技術創造複製牛相關的文章』中，16 位受試者中只有 1 位受試者所輸入的檢索詞，系統有提示出關

聯詞，因為檢索者所輸入的檢索詞大都為「體細胞核」，只有 1 位檢索者輸入「細胞核」系統才提示出關聯詞，故造成此檢索任務在「提高檢索效果方面」是無效的。分析其原因，可以發現此檢索任務是非常專業性的題目，故所有參與實驗的受試者都是處於不熟悉且不知道如何使用檢索詞彙的狀況，再加上此類題目的報導量在本系統是較少的，因此，檢索者必須下更精確的檢索詞彙才能完全表達出此檢索任務。故也反應出系統需針對文件量不足且檢索任務過於專業性時，提高其關聯詞數及品質，這樣才不會讓檢索者誤以為此系統沒有收錄此相關文章。

## 2. 效率 (Task Time)

『效率』是指檢索所需的時間，分析面向可從不同功能及不同任務間的差異來比較，根據研究調查發現，有利用關聯提示詞圖示功能的受試者，平均花費的檢索時間比無關聯提示詞圖示功能的系統還多，如果是單純從檢索的時間觀點來看，似乎有關聯提示詞圖示的功能反而需花費較多的時間，但是，深入探討其原因，發現無論是有無圖書資訊學背景的受試者，皆強調利用有關聯提示詞圖示的功能在檢索的歷程中，需額外在多一道判斷關聯詞是否為相關的步驟，而受試者也反應如果多了此一步驟是可以更精確的找到答案，寧願花費較多的時間在有關聯提示詞圖示功能的系統上。此一觀點和加拿大 Saskatchewan 大學裡的人機互動實驗室(HCI the interaction lab)針對自行研發之檢索系統所做的評估結果相同，在其研究結果指出，檢索者往往需多增加一道有別於一般搜尋引擎的檢索歷程，判斷系統所提示的輔助功能是否符合自身的資訊需求，因此反而需花費更多時間。但是，研究中也強調，檢索者在意的不是時間上的多寡，而是找尋的資訊是否精確且詳細，故在功能面的效率分析，如果再結合檢索的效能一併探討，有關聯提示詞圖示功能的系統在效率上還是不錯的。

另一個受試者會覺得在有關聯提示詞圖示功能的系統會花費較多時間的原因是，如果在無時間的壓力下進行檢索時，檢索者會因為系統所提示的關聯詞產生好奇心，而無限制的搜尋下去，進而迷失在此檢索系統，所以反而需要回頭重新修正欲查詢的題目，因此這時候所花費的時間就比在一般檢索系統還多，故系統在設計此互動式功能，可能需要考量增加一個隨時提醒檢索者所在的檢索歷程之機制，而這個機制需能夠讓檢索者隨時修正自己的檢索詞彙。

最後，就不同檢索任務的差異來說，根據本研究調查顯示，無論是有無圖書資訊學背景的受試者，對於政治類的檢索議題，平常是較少觸及的，因此，在此類新聞議題的檢索時間上，明顯的多於其他類型的新聞檢索任務，

且利用有關聯提示詞圖示功能系統所花費的時間，也是明顯的多於無關聯提示詞圖示功能的檢索系統。綜合訪談資料得知，有受試者反應由於此類型的檢索任務是較乏味且平常也較少查詢此類型的資料，因此，必須額外花費較多的時間去瞭解議題的內容；但是，這些受試者也強調，利用有關聯提示詞圖示功能的檢索系統，可以讓原本較少觸及的資訊需求，引發出檢索者的興趣，進而提高檢索效果。

### 3. 使用者的滿意度(User Satisfaction)：

根據本研究調查結果顯示，在大多數的受試者都是第一次使用有關聯提示詞圖示功能的系統下，整體而言雖然還是滿意，也強調其最大的功能就是『能夠找到你所預期不到的結果』，但是，可以發現有無圖書資訊學背景的受試者還是有明顯的差異，因為有圖書資訊學背景的受試者，在無關聯提示詞圖示功能的滿意度高於有關聯提示詞圖示功能的系統，分析原因是具備圖資領域的受試者對於有關聯提示詞圖示功能系統的關聯詞品質有高標準的要求；相對於非圖書資訊學背景領域的受試者而言，只要檢索系統有一點點小小的幫助，縱使只是在介面上的改變而已，這些受試者都會覺得新奇有趣。然而從訪談資料還可以得知，無論是有無圖資背景的受試者均反應出出，除非系統所提示的功能是真的很有幫助，不然還是會選擇操作較『習慣』的檢索系統或檢索方式，而這點也與許多國外學者在評估資訊視覺化的技術結果雷同，因為畢竟文字式的檢索方式已經行之有年，要讓檢索者在短暫的時間習慣此一操作方式，不是所有人都能夠接受，所以在考量使用者的滿意度之餘，也應該深入去探討這些數據背後所代表的意義。

然而根據參與 TREC 2001 Interactive Track 的評估結果中指出，檢索者的『檢索經驗』會影響到利用輔助檢索系統功能在『檢索結果的滿意度』。這點與本研究的分析不相同，分析其影響的因素是因為 16 位受試者都是第一次使用此關聯提示詞圖示功能，所以雖然在檢索經驗方面，具備圖資背景領域的受試者優於非圖資背景領域的受試者，但是就此實驗環境而言，所有受試者都是在未熟悉的狀態下進行實驗，再加上實驗的樣本數可能不夠龐大，而造成檢索經驗和檢索結果的滿意度呈現出不相關的結果。然而相關性雖然在統計上不夠顯著，但是從訪談資料的分析得知，檢索經驗也不全然是和檢索結果不相關的，因有些檢索經驗較缺乏的受試者在訪談中回答，影響檢索結果的滿意度有時候會和一些外在因素有關，例如：檢索者的習慣性、對於新功能的接受度等都會影響檢索結果的滿意度。

### 4. 介面可用性評估方面 (Usability Evaluation)：

在介面的可用性評估方面，無論是有無圖書資訊學背景的受試者沒有太大的差異，故綜合本研究的五項分析指出，在視覺設計上，關聯詞的名次對檢索者來說沒有太大的意義，因為檢索者認為系統斷定為最相關的詞彙，有時候檢索者並不認同，且檢索者者還反應當系統提示出幾個關聯詞時，檢索者會很直覺的找尋自己認為最相關的詞彙，因此，系統提示的關聯詞數對檢索者來說不是很重要，重要的是提示出關聯詞的關聯程度。但是，可以肯定的事，所有受試者皆認為此 2D 的展現形式，的確能夠提高檢索者的使用興趣。至於如果將此功能應用在一般的搜尋引擎上，受試者也都給予肯定的態度，但是，必須搭配一些相關機制的輔助，例如：增加提醒檢索者所執行之檢索歷程的機制，以降低檢索者因為好奇心而點選不是很相關的關聯詞，進而迷失在龐大的資料量中。

最後在減輕心理認知負擔方面，關聯提示詞圖示的功能的確有指引使用者在未知的領域去尋求相關資料的方向，且降低使用者記憶上的負擔，因為受試者不需要記住每個與檢索詞相關的關聯詞。但是，必須事先告訴使用者此系統提示的關聯詞，是根據涵蓋的資料內容本身而自動產生的。因此，如果能夠克服檢索者在使用新的檢索功能，因為不『習慣』的狀況下而懼於嘗試的難題，相信此檢索功能在輔助檢索者方面還是優於一般的檢索系統。

## 第五章 結論與建議

本研究旨在探討文件量的大小對於關聯詞的品質與成效，以及嘗試建立出評估中文互動式檢索輔助系統中關聯提示詞圖示功能的方式，期望除了系統設計者常關注的關聯提示詞之成效外，更希望藉由與檢索者的互動中所回饋的使用資訊，達到兼具量化與質化的評估方式。在關聯詞的評估方面，藉由相關評估法輔以 trec\_eval 輸出的平均查準率和平均查全率作為評估關聯提示詞的成效；在互動式檢索輔助的評估則是利用量化分析統計配合質化的訪談內容，進行過程中則採用問卷調查法、訪談法與放聲思考法，來進行資料的收集，最後，綜合上述的內容進行資料的分析與歸納。綜合研究的發現與文獻的探討，提出下列的結論與建議。

### 第一節 結論

根據本研究發現，關聯提示詞圖示功能應用在互動式檢索輔助系統的成效結論如下：

**結論一：文件量的確是影響關聯詞的關聯程度，但不是和文件量呈現出絕對正比的關係。**

根據本研究實驗分析顯示，當相同年代的新聞文件集增加至某種程度之後，關聯詞的詞數便不會再增加，且這些關聯詞的品質反而會因為文件量的增加而降低它的品質（強度），例如：本實驗比較中文件集和大文件集的查準率，便發現 30 個查詢詞中，就有 15 個查詢詞的平均查準率呈現出中文件集優於大文件集的現象。因此，如果欲使用文件量來提高關聯詞的品質時，必須依照文件的特性（內容）、年代等相關因素，一併來探討文件量需增加至多少數量，關聯詞的品質才會呈現出顯著的增加。

**結論二：影響關聯詞品質的因素，除了文件量的大小之外，還需考量文件的時間、文件的特性等相關因素。**

由於本實驗的設計，是將文件分為大、中、小三個不同資料量的文件集，小文件集至中文件集的資料，除了增加文件量之外，還增加了不同年代的新聞資料，從實驗的數據可以得知無論是關聯比例或平均查準率均呈現出明顯的增加；然而，中文件集至大文件集雖增加了一倍的新聞資料量，但是文件的年代則是相同的，故得出關聯詞的平均查準率出現不增反降的情形。因此，就新聞資料的特

性，在相同時期所報導事件往往是大同小異的，因此，沒有新的新聞事件進來，就無法產生出新的關聯詞，且原本被系統判斷為相關的關聯詞，會因為一些雜訊而降低關聯詞的關聯強度，所以大文件集的平均查準率才會出現不增反降的情形。

**結論三：在本實驗影響相關判斷的因素中，以『文件內容』為最主要因素。**

與國內學者曾元顯在中文索引典評估的研究結果比較，發覺利用不同的相關判斷尺度及不同的判斷者，在同樣的文件集裡，被判斷為相關詞比例的數據是同為 69%，這與 Saracevic 所提出的影響相關判斷五大因素中的『文件』因素非常吻合，因為在本實驗相關判斷的文件內容是以新聞資料為主，其用語是屬於較特定及專指的內容，所以縱使利用了不同的判斷尺度及判斷者，最後所判斷出的相關比例數據不會受到太大的影響。故可以得出影響相關判斷的因素，會隨著實驗環境的不同，而產生出不同的因素。

**結論四：檢索者對關聯詞的認知，往往是從自身的知識背景去斷定關聯詞的關聯程度，不會從系統的角度去考量，關聯詞的產生是隨著系統的資料內容所建構出來的。**

在評估關聯提示詞的效能中，有許多受試者都會不自覺的反應此關聯詞似乎和自己心理的認知有一段的落差，所以會傾向於不相信系統所提示的關聯詞，因而造成受試者需重新建構自己的檢索詞彙，這樣便有違於系統設計者在設計此功能的原意。所以筆者認為必須在檢索者利用此系統前，有適當的教育訓練，教導檢索者此關聯詞的建構概況，且必須隨時提醒檢索者此輔助功能，最大的目的是從檢索系統的內容幫助檢索者修正及擴展自己的檢索詞彙，雖然檢索系統設計的導向，已經慢慢轉變為使用者的角度，但是畢竟一般的檢索系統所涵蓋的資料內容往往是不同的。最後，根據本研究的結果推測，此輔助功能如果是應用在較具專業領域的檢索方面，例如：專利資料的檢索及生物資訊的檢索等，相信必能提供非專業人士在檢索上的一種指引。

**結論五：檢索經驗較豐富的檢索者，在利用系統所提供的輔助檢索功能時，往往較相信自己的檢索技巧，且對系統所提供的關聯詞品質採取較高的標準。**

在本研究中具備圖資背景領域的受試者，是屬於檢索經驗較豐富的檢索者，而這些受試者在訪談中皆表示，系統所提供的關聯詞涵蓋的概念有時候比自己所下的檢索詞彙還廣且關聯程度不高，故最後還是會重新利用自己的檢索技巧來修正及擴展詞彙；然而對非圖資領域的受試者，其反應出只要系統所提示出的關聯詞的程度相差不會太遠，且又可以省去自己去建構檢索詞彙的時間，其接受的程

度都很高。所以根據研究結果可以推測，此關聯提示詞在幫助檢索經驗較缺乏的檢索者，及對檢索議題較不熟悉的檢索者，才能發揮出最大的效用。

**結論六：在檢索的效率上，有關聯提示詞圖示功能的檢索系統，比一般的搜尋引擎需額外再增加一道判斷關聯詞之關聯程度的步驟。**

一般來說，檢索者在使用網路上的搜尋引擎進行檢索時，花費最多的時間往往是在判斷列出的檢索結果是否符合自己的資訊需求。而關聯提示詞的最大功用，就是節省檢索者直接去過濾龐大的資料量；然而，本系統增加了輔助檢索的功能，卻也相對的增加了檢索者在檢索歷程中的步驟，因為畢竟檢索者還是需要對檢索結果做一次判斷步驟。因此，就效率上來說，關聯詞沒有明顯的降低檢索時所需的時間，但是對於檢索者在瞭解此檢索系統所涵蓋的內容的確有很大的幫助。而國外之相關系統的評估研究指出，為了改善檢索的效能，特別利用當檢索者移動到系統所提示的關聯詞時，自動顯示出有此關聯詞文章的自動摘要，而檢索者可以直接利用所顯示出的摘要，再來判斷是否繼續點選此關聯詞。此方式也是本系統持續改進的方向。

**結論七：關聯提示詞最大的功能，就是提供檢索者意想不到的檢索提示或檢索結果。**

綜合本研究的調查結果及分析指出，無論是有無圖書資訊學背景領域的受試者，認為在進行檢索時，發覺系統有時候會提示出心理都沒想到過的檢索詞彙，而這些詞彙往往是一般人不會想到的檢索詞彙，而這些詞彙對於一些較艱深的檢索議題，通常都有很大的幫助。所以此觀點就符合系統設計者，利用關聯提示詞圖示功能在檢索系統想要達到的效用。

**結論八：受試者認為『習慣』會影響關聯提示詞圖示的可用性。**

根據本實驗的可用性訪談之分析結果指出，在有關聯提示詞圖示功能的系統中，有許多受試者在剛開始進行檢索時，還是會利用文字式的輸入方式來擴展自己的檢索詞彙，且必須經過多次的檢索歷程之後，才會直接去點選系統所提示的2D 關聯提示詞圖示功能；再者，受試者也反應利用點選滑鼠左鍵一下和二下，來區別詞彙間的布林邏輯關係，在操作上有時候會造成混淆。上述的情形，其實都是『習慣性』而產生出的實驗結果，這與許多國外學者在利用資訊視覺化的技術應用在文字式檢索系統上，所產生出的研究結果相符，因為文字本身就是一個一維的資訊載體，且文字式輸入的檢索系統到現在還是最為普遍的檢索方式，要讓檢索者將已經行之有年的習慣改變，必須有相關的配套措施且要循序漸進，故利用2D的圖示展現關聯詞不是不可行的，但是，一些更符合使用者導向的設計



必須盡可能的考量進去，當此功能融入一般檢索者的『習慣』中就會不自覺的使用此輔助檢索功能。

## 第二節 建議

以下針對本研究未臻完善之處，以及後續值得研究之議題，提出一些建議。

### 一、 本研究之建議：

由研究發現的結論，提出了幾點建議，作為未來評估互動式檢索輔助功能成效的參考與依據。

#### (一) 對影響關聯提示詞的成效建議：

**建議一：文件集的大小的確會影響到新聞文件之關聯詞的相關比例及其查準率，但應該更進一步探討會影響關聯詞品質的文件量門檻。**

從相關研究報告可以得知影響關聯詞成效的因素，會隨著資料內容的不同而有不同的因素，故資料內容的型態如果是屬於特定的領域，應該針對其特定的領域去研究。而在本實驗的新聞資料集中，文件量在某些情況下的確會影響到關聯詞的品質，可以再進一步去探討當文件增加至何種程度之後，關聯詞的品質便不會再增加了。

#### (二) 對互動式檢索輔助功能之建議：

**建議一：利用滑鼠移動至關聯詞上，顯示出有此關聯詞之文章的自動摘要，以增進檢索上的效率（時間）。**

為了降低檢索者需多一道判斷相關詞所需的時間，建議可以參考國外相關設計，例如：利用當檢索者移動到系統所提示的關聯詞時，自動顯示出有此關聯詞文章的自動摘要，這或許可以降低檢索者點選關聯詞後，才發現沒有符合自己心理認知的檢索結果，所產生出的偏差。

**建議二：允許檢索者直接在關聯提示詞圖示上，增加或修改檢索詞彙。**

本系統的關聯提示詞圖示功能，雖然允許檢索者利用點選滑鼠左鍵的方式，來取代傳統的文字式輸入方式，但是，卻無法讓檢索者在關聯提示詞圖

示上直接修正原來的檢索詞彙，而必須在回到傳統的文字輸入方塊欄位裡修正檢索詞彙，這點讓檢索者在操作上感覺複雜且不一致，故將來的設計方向，應該可以嘗試兩者的輸入方式並行，由檢索者自行評定何者為最適當的檢索模式。

**建議三：操作上盡可能朝向符合檢索者的『習慣』且不易混淆的方向來設計。**

無論是有無檢索經驗的檢索者，對於資訊檢索系統都依賴著行之有年的模式下來進行檢索，而這個模式說穿了就是檢索者的一種『習慣』，因此，除非新的檢索功能真的能夠達到非常顯著的幫助，不然對於資訊檢索者還是會選擇他們較常接觸的檢索系統。

(三)對評估互動式檢索輔助功能之方法的建議：

所謂互動式檢索輔助功能，就是經由系統主動提示檢索者相關資訊，再讓檢索者從提示出的建議回饋給系統，經過一連串與系統互動之後達到輔助檢索者表達其資訊需求，而這些功能除了本研究的關聯提示詞圖示功能之外，還包括：自動歸類、自動摘要等。上述這些功能的環境大都以英文的資料為主且評估的環境異於一般的關鍵字檢索系統，故評估環境是否直接可以應用在中文的檢索環境，也是值得相關研究者去思考的地方。因此，本研究試圖去建立一個以評估中文的互動式檢索輔助功能的方法，然而，從本實驗的過程中，發覺仍有許多不周之處，例如：檢索任務的設計、評估的面向分析等，都是評估者需重新思考是否需特別為中文檢索的評估環境建立一個準則。

## 二、 後續研究之建議：

而本研究也由於人力、時間，以及可取得資源等限制，仍有許多待改善與加強之處。

**建議一：針對文件集時間的不同來探討影響關聯詞的程度。**

根據本實驗的研究結果，可以初步推論出年代的不同會影響到關聯詞的成效，因此，後續研究者可以針對年代的因素，深入去分析時間的先後影響關聯詞的程度。例如：在本實驗的中文件集中，增加年代較久遠的資料，關聯詞的成效還是呈現出明顯的增加。

**建議二：**除了利用相關比例及平均查準率來評估關聯詞的品質之外，應該結合對此實驗資料內容熟悉的專家及語言學家，針對單一檢索詞彙的分佈情形做更深入的探討與分析。

由於本實驗的資料集是新聞資料，而新聞資料的內容及其用語都有一定的格式，故利用本實驗之相關比例及 trec\_eval 的輸出結果確能夠看出詞彙的分佈情形，但是，如要在更深入探討這些詞彙對此資料集的影響，應該結合相關新聞從業人員，從語言學的角度來分析提示出的關聯詞，應該會更具體的表達數據背後所代表的意義。

**建議三：**嘗試將實驗的環境擴及網路的資料，以較不具專指性的資料集來評估關聯詞的品質。（為了探討檢索者對關聯詞的關聯程度與系統認知的差異）

本實驗的環境受限於取得資料內容管道的限制，無法針對現有的網路資源來作為實驗之資料集，而也因為這個因素，使得參與實驗之受試者對系統所提示之關聯詞的認知，產生出極大的落差，因為受試者常常會忽略實驗的環境只是限於新聞資料，而不是擴及所有的網路資源，所以受試者會用自己的心理認知來評定系統所提示出的關聯詞程度。故如果實驗的環境能夠擴及網路資源，或許便能夠降低檢索者對關聯詞認知的落差，且更能夠看出系統提示出關聯詞在輔助檢索者的成效。

**建議四：**為了深入探討關聯詞對檢索者的輔助情形，應該針對每個檢索者選擇提示之關聯詞的檢索點差異。

在早期無關聯提示詞的資訊檢索系統，檢索者必須依賴自己與系統多次的互動之後，才能夠找到符合自己的資訊需求，因此，便有許多學者相繼投入研究檢索者在檢索歷程中，使用檢索詞彙的變化情形，而此模式無疑在研究檢索者的心理認知模式。而相對的關聯提示詞圖示的功能，只是主動幫助檢索者接下來所應該下的檢索詞彙，因此，如果在時間的允許之下，應該針對檢索者對每個檢索歷程中，選擇系統所提示之關聯詞的變化，相信更能夠完全的表達出此功能對檢索者的輔助情形。

**建議五：**評估的方式如能再輔與更深入的質化研究方法，來解釋量化後所產生出概約的數據，便能夠達到兼具量化與質化之評估檢索系統的最佳方式。

本實驗的量化數據雖然在統計學上是屬於小樣本，但是利用一些質性的資料來解釋數據所代表的意義，似乎比直接利用數據來推測一些表象的東西要來的具體。且因為互動式檢索系統就是在乎檢索者與系統之間的互動情形，而這些互動

的使用情形，往往是量化的數據無法表達出來的。故本研究建議，受試者的樣本數多寡不是絕對的因素，關鍵在於利用多深入的質化方式來分析。

## 參考書目

### 一、 中文文獻：

#### (一) 圖書：

1. 吳美美。「試論資訊檢索理論」,王振鵠教授七秩榮慶祝壽論文集編輯小組。當代圖書館事業論集：慶祝王振鵠教授七秩榮慶論文集。台北市：正中，民 83。
2. 黃慕萱，資訊檢索中「相關」概念之研究。台北市：台灣學生，民 85。
3. 曾元顯，數位文件之資訊擷取與檢索。台北市：全壘打文化，民 88 年。
4. 林傑斌、陳湘、劉明德合著，SPSS 11 統計分析實務設計寶典。台北縣：博碩文化，民 91。
5. David S. Moore 著；王念孫、鄭玉卿、王鴻龍譯，實用統計學。台北市：科大文化，民 91。
6. 陳燦珠，使用者與中文全文檢索系統互動關係研究：以中央研究院漢籍全文資料庫的使用為例。台北縣：淡大教資系，民 88。

#### (二) 期刊：

1. 卜小蝶，「網路使用者查詢語彙紀錄之分析與應用初探」，資訊傳播與圖書館學 6 卷 2 期（民 88），頁 49-62。
2. 曾元顯，「共現索引典之自動建構、評估與應用」，台灣大學圖書資訊學系四十週年系慶研討會（民 90 年 11 月），頁 87-105。
3. 江玉婷，陳光華，「TREC 現況及其對資訊檢索研究之影響」，政治大學圖書與資訊學刊 29 期（民國 88 年 5 月），頁 36-59。
4. 莊雅蓁，「資訊檢索之索引典研究」，中國圖書館學會會報 63 期（民 88 年 12 月），頁 77-89。
5. 曾元顯、林瑜一，「模糊搜尋、相關詞提示與相關詞回饋在 OPAC 系統中的成效評估」，中國圖書館學會會報 61 期，（民 87 年 12 月），頁 103-125。

6. 陳光華,「資訊檢索系統的評估 – NTCIR 會議」, 國立台灣大學圖書資訊學系四十週年系慶學術研討會論文集 (民國 90 年 11 月), 頁 67-86。
7. 曾元顯「關鍵詞自動擷取技術與相關詞回饋」, 中國圖書館學會會報 59 期 (民 86 年 12 月), 頁 59-64。
8. 邱立豐。「互動式概念查詢應用於網路文件自動摘要之效益」, 雲林科技大學資訊管理系碩士班, 碩士論文, 民 91 年。
9. 蔡孟竹。「中文 OCR 文件檢索測試集之製作與應用」, 輔仁大學圖書資訊學系碩士班, 碩士論文, 民 91 年。
10. 林瑜一。「模糊搜尋與相關詞提示整合在 OPAC 系統之評估研究」, 台灣師範大學工業科技教育研究所, 碩士論文, 民 87 年。

## 二、 英文文獻：

### (一) 圖書：

1. Ricardo Baeza-Yates, Berthier Ribiero-Neto, Berthier Ribeiro-Neto. Modern Information Retrieval. 1st edition. Addison-Wesley Pub Co, 1999.

### (二) 期刊：

1. N.J. Belkin, C. Cool\*, D. Kelly, G. Kim, J.-Y. Kim, H.-J. Lee, G. Muresan, M.-C. Tang, X.-J. Yuan (2002). "Rutgers Interactive Track at TREC 2002." Retrieved May 18, 2004, from <http://trec.nist.gov/pubs/trec11/papers/rutgers.belkin.pdf>
2. Yuen-Hsien Tseng, "Fast Co-occurrence Thesaurus Construction for Chinese News," Proceedings of the 2001 IEEE International Workshop on Natural Language processing and Knowledge Engineering (NLPKE 2001) in conjunction with the IEEE International Conference on Systems, Man, and Cybernetics SMC 2001 Tucson, Arizona, USA, October 7-10, 2001.
3. Kuang-hua Chen, 「Introduction to Chinese Document Set for CLIR Task of NTCIR Workshop 3」, 2003
4. "Text Retrieval Conference (TREC) Data – English Relevance Judgments Files List," Retrieved May 18, 2004, from [http://trec.nist.gov/data/qrels\\_eng/index.html](http://trec.nist.gov/data/qrels_eng/index.html).
5. Ellen M. Voorhees and Donna Harman. "Overview of the Ninth Text Retrieval Conference (TREC-9)," Retrieved May 18, 2004, [http://trec.nist.gov/pubs/trec9/papers/overview\\_9.pdf](http://trec.nist.gov/pubs/trec9/papers/overview_9.pdf).
6. Michael Eisenberg and Carol Barry. "Order Effects: A study of the Possible

Influence of Presentation Order on User Judgments of Document Relevance,” Journal of the American Society for Information Science 39:5 (1988):293-300.

7. Joseph W. Janes. “Other People’s Judgments: A comparison of User’s and Other’s Judgments of Document Relevance, Topicality, and Utility,” Journal of the American Society for Information Sciences 45:3 (April 1994):160-171.
8. William Hersh, Oregon Health & Science University, Portland, OR, USA.”TREC 2002 Interactive Track Report”. Retrieved May 18, 2004, from <http://trec.nist.gov/pubs/trec11/papers/INTERACTIVE.OHSU.pdf>
9. Richard Osdin, Iadh Ounis and Ryen W. White. “Using Hierarchical Clustering and Summarization Approaches for Web Retrieval :Glasgow at the TREC 2002 Interactive Track.”, Retrieved May 18, 2004, from <http://trec.nist.gov/pubs/trec11/papers/glasgow.int.pdf>
10. N.J. Belkin, C. Cool, D. Kelly, G. Kim, J.-Y. Kim, H.-J. Lee, G. Muresan, M.-C. Tang, X.-J. Yuan.“Rutgers Interactive Track at TREC 2002.”, Retrieved May 18, 2004, from <http://trec.nist.gov/pubs/trec11/papers/rutgers.belkin.pdf>
11. Hsinchun Chen, Bruce Schatz, Tobun Ng, Joanne Martinez, Amy Kirchhoff, Chienting Lin, “A Parallel Computing Approach to Creating Engineering Concept Spaces for Semantic Retrieval: The Illinois Digital Library Initiative Project,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, Retrieved May 18, 2004, from <http://ai.bpa.arizona.edu/papers/pami96/>
12. Gutwin C., Paynter G., Witten I.H., Nevill-Manning C., Frank, E, “Improving Browsing in Digital Libraries with Keyphrase Indexes, “ Journal of. Decision Support Systems 27(1-2)(1999) : 81-104 , Retrieved May 18, 2004, from <http://hci.usask.ca/publications/1999/keyphind-jdss/keyphind-jdss.pdf>

## (二) 網路資源：

1. Vivisimo Clustering Engine. <http://vivisimo.com/>
2. WiseNut. <http://www.wisenut.com/>
3. Openfind 台灣線上新聞搜尋。 <http://news.openfind.com.tw/>
4. “Draft Guidelines for TREC 2002 Interactive Track”, Retrieved May 18, 2004, from <http://www-nlpir.nist.gov/projects/t11i/guidelines.html>
5. “TREC Web Corpus : .GOV”, Retrieved May 18, 2004, from, <http://www.ted.cmis.csiro.au/TRECWeb/govinfo.html>.

6. “TREC-9 Interactive Track Guidelines”, Retrieved May 18, 2004, from, <http://www-nlpir.nist.gov/projects/t9i/spec.html>
7. TREC-9 Interactive Track Guidelines , Retrieved May 18, 2004, from, <http://www-nlpir.nist.gov/projects/t9i/spec.html>。
8. “Interaction lab : University of Saskatchewan”, Retrieved May 18, 2004, from <http://hci.usask.ca/index.xml>.
9. Keith Instone, “Site Usability Evaluation, ” Retrieved May 18, 2004, from [http://www.webreview.com/1997/10\\_10/strategists/10\\_10\\_97\\_1.shtml](http://www.webreview.com/1997/10_10/strategists/10_10_97_1.shtml)。
10. Keith Instone, “Site Usability Heuristics for the Web, ” Retrieved May 18, 2004, from [http://www.webreview.com/1997/10\\_10/strategists/10\\_10\\_97\\_2.shtml](http://www.webreview.com/1997/10_10/strategists/10_10_97_2.shtml)。
11. Jakob Nielsen, “Ten Usability Heuristics, ” Retrieved May 18, 2004, from [http://www.useit.com/papers/heuristic/heuristic\\_list.html](http://www.useit.com/papers/heuristic/heuristic_list.html)
12. “Notes on Trec Eval,” Retrieved May 18, 2004, from [http://ir.iit.edu/~dagr/cs529/files/project\\_files/trec\\_eval\\_desc.htm](http://ir.iit.edu/~dagr/cs529/files/project_files/trec_eval_desc.htm)



## 附錄一：檢索任務

受限於實驗資料集的特性，且新聞資料往往具備時間性，所以檢索任務根據聯合新聞網及中時電子報的分類目錄，初略可分為下列四大類，「政治」包含國際、財經、社會等，「運動」則包含各項體育賽事的報導，「生活」則包含醫藥保健、消費資訊等，而「娛樂」則包含各項影視新聞資料。上述的檢索任務還需根據資料集的年代限定符合其年代的各项新聞，如：在 CIRB020 的資料年代從 1998 至 1999 年，而 PUB 的資料則是從 2000 至 2001 年

下列檢索任務主要可分為涵蓋性任務(Coverage evaluation task)及探索性任務(Exploration task)，涵蓋性任務主要是針對主題的涵蓋面是否會影響此關聯提示詞的成效，因此依據 NTCIR 的正式檢索題目，選擇適合本實驗之環境的檢索任務，共八個檢索任務，依序分為政治、運動生活、娛樂。而探索性任務則是模擬現實環境中，使用者有時候只是隨意的瀏覽網頁，尋找自己有興趣的東西且沒有特別目的，共一個檢索任務。

### 涵蓋性(Coverage evaluation task)

#### 政治

檢索任務一：“查詢精省的法規內容有哪些，以及台灣省廢省後前省長宋楚瑜的態度”(根據NTCIR CLIRFormalRun Topic013 )

檢索任務二：有關金大中總統對亞洲的政策之文章

#### 運動

檢索任務三：查詢曼谷亞洲運動會的相關報導。(根據NTCIR CLIRFormalRun Topic012 )

檢索任務四：有關於決定加入西雅圖水手隊的投手佐佐木主浩的文章(根據NTCIR CLIRFormalRun Topic016 )

#### 生活

檢索任務五：查詢有關華航於桃園中正機場失事的相關報導。(根據NTCIR CLIRFormalRun Topic007 )

**檢索任務六：**與使用被稱為體細胞核移植的技術創造複製牛相關的文章  
(根據NTCIR CLIRFormalRun Topic015 )

**娛樂**

**檢索任務七：**查詢一九九八年電影「鐵達尼號」獲得奧斯卡獎之相關報導。(根據NTCIR CLIRFormalRun Topic008 )

**檢索任務八：**有關於北野武導演的電影之文章(根據NTCIR CLIRFormalRun Topic017 )

**探索性(Exploration task)**

電子新聞包含了「兩岸國際」、「地方新聞」、「財經產業」、「股市理財」、「資訊科技」、「運動賽事」、「影視娛樂」、「消費流行」、「生活健康」、「閱讀藝文」、「旅遊休閒」等，請您檢索 1998-2001 年中您最感興趣的新聞議題。

## 附錄二：基本資料問卷

### 基本資料

1. 請問您的身份是？

1. 館員 (請跳答至第三題)

2. 圖書資訊學系相關系所(級別：\_\_\_\_)請說明大學部碩士班或博士班  
(請跳答至第三題)

3. 非圖書資訊學系相關系所(科系：\_\_\_\_)請回答第二題。

2. 您曾經修習過圖書資訊領域的相關課程嗎？如資訊檢索、索引與摘要及資料收集等相關課程(請將您修過的課程列在下列空白處)

\_\_\_\_\_

3. 性別？

\_\_\_\_ 女

\_\_\_\_ 男

4. 年齡

\_\_\_\_ 18 - 27 歲

\_\_\_\_ 28 - 37 歲

\_\_\_\_ 38 - 47 歲

\_\_\_\_ 48 - 57 歲

\_\_\_\_ 58 - 67 歲

\_\_\_\_ 68+歲

### 電腦技能

請根據問題勾選最適當的數字以描述您在使用電腦方面的經驗。

您使用下列各項的經驗為何	None 未曾 使用			Some 偶爾 使用			A great deal 經常 使用
1. 電腦	1	2	3	4	5	6	7
2. 網際網路的瀏覽器	1	2	3	4	5	6	7

您使用電腦從事下列工作的頻率	Never 未曾 使用			Monthly 每月			Daily 每天
1. 工作用 work tasks?	1	2	3	4	5	6	7
2. 學術用 academic tasks?	1	2	3	4	5	6	7
3. 個人用 personal tasks?	1	2	3	4	5	6	7

請為自己在電腦方面的技能評分：

初學者 Novice			普通			專家級 Expert
1	2	3	4	5	6	7

請列出您經常使用的軟體、作業系統或任何您所會的程式語言。

**檢索經驗(Searching Experience)**

請根據問題勾選最適當的數字（分數）以描述您在檢索方面的經驗。

Please indicate the number that most closely describes your searching experience.

您使用下列各項的經驗為何	None 未曾 使用			Some 偶爾			A great deal 經常
1. 使用搜尋引擎來檢 索	1	2	3	4	5	6	7
2. 查詢圖書館的公用 目錄(書目系統)	1	2	3	4	5	6	7
3. 利用索引或摘要的 功能來檢索。如：博碩 士論文索引系統等。	1	2	3	4	5	6	7
4. 請列出曾經使用過 的檢索系統(資料庫)：							
a. _____	1	2	3	4	5	6	7
b. _____	1	2	3	4	5	6	7
c. _____	1	2	3	4	5	6	7
d. _____	1	2	3	4	5	6	7

5. 每當我利用檢索系統 搜尋時，都可以找到我 想要的資訊。	Rarel y 很少			Some- times 有時 候			Ofte n 常常
	1	2	3	4	5	6	7

您查詢有關下列資訊的 頻率為何	Never \未曾 使用			Monthly 每月			Daily 每天
1. 作業 / 與工作有關	1	2	3	4	5	6	7
2. 購物	1	2	3	4	5	6	7
3. 旅遊	1	2	3	4	5	6	7
4. 醫療 / 健康	1	2	3	4	5	6	7
5. 政治	1	2	3	4	5	6	7

2003 Interactive Study  
 受測者編號 Participant # \_\_\_\_\_  
 日期 Date \_\_\_\_\_

附錄二

8. 娛樂	1	2	3	4	5	6	7
9. 其他，請列出：							
a. _____	1	2	3	4	5	6	7
b. _____	1	2	3	4	5	6	7
c. _____	1	2	3	4	5	6	7
d. _____	1	2	3	4	5	6	7

請您評定自己在檢索資訊方面的能力：

Novice 初學者			普通			Expert 專家
1	2	3	4	5	6	7

最後，請問您利用線上搜尋引擎的時間有多久？ \_\_\_\_\_年  
 請列出您常用的檢索引擎： \_\_\_\_\_.

2003 Interactive Study  
 受測者編號(Participant) # \_\_\_\_\_  
 Condition \_\_\_\_\_  
 檢索任務(Task) # \_\_\_\_\_  
 Search # \_\_\_\_\_

**附錄三：單一檢索任務結束後問卷**  
**POST-SEARCH QUESTIONNAIRE**

檢索之任務：

檢索任務一：“查詢精省的法規內容有哪些，以及台灣省廢省後前省長宋楚瑜的態度”(根據NTCIR CLIRFormalRun Topic013)

1. 您一開始輸入之查詢語句能幫助您執行此檢索任務：

Not at all 非常不滿意			Somewhat 還可以			Extremely 非常滿意
1	2	3	4	5	6	7

2. 所提示之自動關聯提示詞能有效提高檢索效果。

Not at all 無效			Somewhat 還可以			Extremely 非常有效
1	2	3	還可以	5	6	7

3. 請指出您自己對此一檢索任務的熟悉程度

Not at all 不熟悉			Somewhat 還可以			Extremely 很熟悉
1	2	3	4	5	6	7

4. 請為自己在此檢索任務的專業程度評定一個分數

Novice 初學者			還可以			Expert 專家
1	2	3	4	5	6	7

2003 Interactive Study  
 受測者編號(Participant) # \_\_\_\_\_  
 Condition \_\_\_\_\_  
 檢索任務(Task) # \_\_\_\_\_  
 Search # \_\_\_\_\_

5. 如果您已經知道此一檢索任務之答案，請在下列空白處寫下您的答案。

對您上述回答的答案有多少把握是正確的，請勾選出一個分數。

Extremely Uncertain 很不確定			Neutral 還可以			Extremely Certain 很確定
1	2	3	4	5	6	7

請您指出此檢索任務能夠在此網站(資料集)找到多少資訊

None 少			Some 還可以			A great deal 很多
1	2	3	4	5	6	7

6. 最後，對您在此檢索任務所找出的答案有多少信心把握是正確的：

Not at all 不太有信心			Somewhat 還可以			Extremel y 很有信心
1	2	3	4	5	6	7



2003 Interactive Study  
 受測者(Participant) # \_\_\_\_\_  
 Condition \_\_\_\_\_  
 檢索任務(Task) # \_\_\_\_\_  
 Search # \_\_\_\_\_

**附錄四：完成單一檢索系統之問卷**  
**POST-SYSTEM QUESTIONNAIRE**

請根據下列問題對此次檢索實驗做個簡單的評量。

	Not at All 不同意			Some- what 還可以			Extremel y 非常同意
1. 此檢索引擎容易上手(操作)?	1	2	3	4	5	6	7
2. 此檢索任務容易檢索?	1	2	3	4	5	6	7
3. 您對此次檢索結果滿意嗎?	1	2	3	4	5	6	7

	None 沒有			Som 有一點			A great deal 很大的幫助
5. 您先前的知識及背景能夠幫助此次的檢索任務?	1	2	3	4	5	6	7
6. 此檢索功能能夠幫您找到任何您所預期不到的結果嗎?	1	2	3	4	5	6	7

**附錄五：檢索實驗後問卷**  
**EXIT-SEARCH QUESTIONNAIRE**

(To minimize the potential for ordering effects, it is suggested that experimental systems be assigned neutral names)

當您回答下列問題時，請仔細思考與您剛才所執行的檢索經驗之關聯度。

	Not at all 不知道		Somewhat 有一點		Completely 知道
1. 您知道何謂 " 關聯詞 " 嗎?	1	2	3	4	5
2. 在您的搜尋經驗中，曾經使用過類似的關聯詞圖示(系統 A)檢索方式嗎?	1	2	3	4	5
3. 您有發覺系統 A 和系統 B 有異同的地方嗎?	1	2	3	4	5

4. 哪一個系統較容易學習?

系統 A \_\_\_\_\_ 系統 B \_\_\_\_\_  
 差不多 \_\_\_\_\_

5. 哪一個系統較方便使用?

系統 A \_\_\_\_\_ 系統 B \_\_\_\_\_  
 差不多 \_\_\_\_\_

6. 請挑選出您心目中最好的系統?

系統 A \_\_\_\_\_ 系統 B \_\_\_\_\_  
 差不多 \_\_\_\_\_

2003 Interactive Study  
Searcher # \_\_\_\_\_  
Condition \_\_\_\_\_

附錄五

7. 請列出您對兩系統的設計或功能中，指出較喜好的部分。
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
8. 請列出您對兩系統的設計或功能中，指出較不喜好的部分。
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
9. 請列出您對此次檢索任務及檢索系統的看法與意見。THANK YOU!!!!

## 介面可用性訪談大綱及問題

本實驗可用性評估主要是採用 Jakob Nielsen 提出的十個評估規則為準則，在根據實驗之環境適時調整為 9 個可用性評估原則，作為訪談之依據。

**注意事項:**除了回答下列的問題之外，儘量引導受試者說明回答此一答案的原因。

### 一、Visibility of system status (視覺設計)

- (一)本系統利用 2D 的圖示呈現各個關聯詞的方式是否能夠協助您了解所收集資料的範圍？（請說明原因）
- (二)關聯詞的關聯程度利用距離檢索詞的遠近來表示是否能夠增進您對欲查詢主題的初步瞭解？
- (三)關聯詞的關聯名次是否能夠幫助您縮小查詢範圍？

### 二、Match between system and the real world (與現實環境的關聯程度)

- (一)您覺得本系統的關聯詞圖示功能是否適用於一般的搜尋引擎？
- (二)對於系統的關聯詞圖示而言，您是否有任何的意見？

### 三、User control and freedom (操作系統的方便性)

- (一)您覺得本系統的關聯詞圖示是否容易操作？
- (二)當您操作上述的功能，是否有遇到任何挫折？
- (三)（接上題）如果有，請您說說看遇到挫折的情形如何？
- (四)（接上題）您描述的挫折對您使用系統的便利性造成多大的影響？

### 四、Error prevention (錯誤偵測)

- (一)您在操作本系統時，是否有遇到任何困難？或者是希望系統能主動提示的機制？
- (二)（接上題）如果有，請您說說看您是如何發現這個問題的？
- (三)（接上題）您認為系統提供的訊息是否足？

### 五、Recognition rather than recall (以認知取代記憶)

- (一)與一般搜索引擎相比，本系統利用 2D 的關聯詞圖示介面，提示相關之檢索概念，是否可以有效減輕您認知上的負擔？

六、 Help users recognize, diagnose, and recover from errors ( 協助  
認知、錯誤之避免與修正 )

- (一) 本系統所提示的相關詞彙是否能夠幫助您找到所需的資料？
- (二) 您覺得本系統提供的的相關詞彙提示是否能協助您表達欲查詢的  
資訊需求？

七、 Flexibility and efficiency of use ( 彈性與效率 )

- (一) 在您不熟悉的主題下，此關聯提示詞的功能是否能夠幫您找到所需  
的資訊？
- (二) ( 接上題 ) 如果有，請您試著說說看它如何幫助您？
- (三) ( 接上題 ) 如果沒有，請說明原因？

八、 Aesthetic and minimalist design ( 美學及最小努力原則之設計 )

- (一) 與一般搜索引擎相比，您覺得本系統的關聯詞圖示是否能夠有效縮  
短您查詢的時間？
- (二) 您覺得本系統的關聯詞圖示介面需增加哪些設計及功能？

九、 Help and documentation ( 系統輔助 )

- (一) 您覺得是否需要提供使用系統的輔助說明？
- (二) ( 接上題 ) 如果需要，請說明需提供哪些輔助說明？
- (三) ( 接上題 ) 如果不需要，請說明原因？