

# 行政院國家科學委員會專題研究計畫 成果報告

## 資訊檢索評估中檢索問題特性之探究 研究成果報告(精簡版)

計畫類別：個別型  
計畫編號：NSC 95-2413-H-002-012-  
執行期間：95年08月01日至96年07月31日  
執行單位：國立臺灣大學圖書資訊學系暨研究所

計畫主持人：陳光華

計畫參與人員：碩士班研究生-兼任助理：吳恬安  
大學生-兼任助理：王俐涵  
工讀生：林純如

處理方式：本計畫可公開查詢

中華民國 96年10月01日

# 行政院國家科學委員會專題研究計畫報告

## 資訊檢索評估中檢索問題特性之探究

計畫編號：95-2413-H-002-012

執行期限：95年8月1日至96年7月31日

主持人：陳光華

國立臺灣大學圖書資訊學系 副教授

研究助理：吳恬安

國立臺灣大學圖書資訊學系 研究生

### 1. 摘要

自從美國的 TREC，歐洲的 CLEF，東亞的 NTCIR 的資訊檢索評估計畫，讓從事資訊檢索的學者專家，有一個公正、公開、公平的評量機制，各種資訊檢索研究可以在一致的比較基礎下，確認各種檢索技術的優劣，並進一步發展探討各深入的資訊檢索研究。目前 TREC、CLEF、NTCIR 三大資訊檢索評估會議舉辦的各項資訊檢索評估項目，已經受到資訊檢索學界的廣泛接受，而這些評估會議的評估結果已成為資訊檢索的研究者發表學術論文時，必須使用的數據，以作為研究成果與其他研究比較的基礎。資訊檢索評估必須使用測試集以進行各種項目的評估，因此測試集的適用性就成為必須深入探討的課題。本計畫擬藉由分析 NTCIR 評估會議使用之檢索問題與各研究團隊的檢索結果，探討檢索問題各項特性對檢索結果的影響。研究結果顯示，查詢主題若為「全球層次」、「時空皆不明確」、主題為「科技資訊」、「不明確的條件敘述」多於「明確的條件敘述」等特質，則可預測有較佳的檢索表現，而擁有這些特性的查詢主題可推論為一個較「容易」的查詢主題。

### 2. Abstract

The TREC, CLEF, and NTCIR have built open and fair mechanism of information retrieval (IR) evaluation. Therefore, researchers of IR could compare different retrieval techniques based on the concrete procedures provided by these IR evaluation forums. Furthermore, researchers could investigate the pros and cons of different techniques and models. Now, TREC, CLEF, and NTCIR have been broadly recognized as the must-be evaluation forums for researchers of IR. The test collections used for IR evaluation should be investigated in depth and in detail, since they have become the golden standards in some senses for the IR researches. This project will investigate the test collections used in NTCIR and the corresponding search results using these test collections. The investigated results showed global and technical topics with broad in temporal and spatial coverage will be much easier than others.

### 3. 緒論

由於資訊檢索技術被運用於網際網路，使得目前廣為眾人所知的「搜尋引擎」服務，已經成為熟悉網路的使用者獲取資訊的主要途徑。事實上，搜尋引擎的競爭極為激烈，雖然，Google 被許多人認為已經是網路

搜尋引擎的第一把交椅，然而，Yahoo 與微軟急起直追，而中國大陸的百度與臺灣的 Openfind 各自在中文的世界也有一定的影響力。既然有如此多的資訊檢索系統，且在商業環境下的競爭如此激烈，資訊檢索系統的績效自然受到業界與學界的重視，而運用於網際網路的資訊檢索系統，其實際的績效更是眾所矚目，但是經常卻以商業機密的理由，無法分析或評量搜索引擎的真正績效，而流於各說各話。

自從美國的 TREC，歐洲的 CLEF，東亞的 NTCIR 的資訊檢索評估計畫，讓從事資訊檢索的學者專家，有一個公正、公開、公平的評量機制，各種資訊檢索研究可以在一致的比較基礎下，確認各種檢索技術的優劣，並進一步發展探討各深入的資訊檢索研究。在這種研究的大環境下，資訊檢索評估的技術就益形重要，資訊檢索評估的項目、技術、方法，會引導資訊檢索研究的發展方向。因此，如何設計更好的資訊檢索評估方法，如何分析資訊檢索評估的結果，如何確定資訊檢索測試集可以真正評估資訊檢索技術的相對優劣，是非常重要的課題。

資訊檢索系統的評估非常倚賴測試集，測試集是由評估會議的舉辦人建構的，其中測試集的文件與問題是用以測試資訊檢索績效的重要資源，文件的適用性，問題的適用性與困難度，則是是否能夠真正評估資訊檢索系統績效的核心問題。本計畫探討用於 NTCIR 資訊檢索評估會議的檢索問題的困難度對於檢索績效的影響，作為後續建構資訊檢索測試集的參考，與公平進行資訊檢索評估的基礎。本計畫針對 NTCIR 使用之測試問題與評估結果，分析測試問題與評估結果之關係，以及問題的困難度對於資訊檢索績效評估的適用性的影響。

### 4. 研究方法與實施

本研究以 NTCIR-6 中 CLIR 的 50 個查詢主題(Topic)為研究對象，CLIR 文件集的特性為使用新聞文件，主要原因在於新聞的主題分佈非常廣泛，且能即時反映目前語言文字的使用情形與特性，可測試出資訊檢索系統是否能適應時代的走向及需求，也能切合一般資訊檢索系統或搜尋引擎的設計目的與應用對象。(江玉婷，民 88) 因此測試集的設計亦為新聞性資訊需求，此設計也可減少相關判斷時，因判斷人員的背景知識因素不足，而導致的判斷不正確。

#### (1) 查詢主題屬性分析

查詢主題的屬性類別參考 Saracevic & Baxter 提出

的範疇、專指度、複雜度與概念上的前提假定五種問題屬性，發展出(a)涵蓋地域、(b)時空明確性、(c)主題、(d)條件敘述明確性及(e)條件敘述個數等五個特性，用以測試檢索結果是否受到問題特性之影響。以下分別敘述各屬性的內涵及界定標準。

本研究	Saracevic & Baxter
涵蓋地域	範疇
時空明確性	專指度
主題	範疇
條件敘述明確性	專指度、前提假定
條件敘述個數	複雜度、前提假定

(a) 涵蓋地域：分為全球、區域及單一國家三個層次。因新聞討論的主題分佈廣泛，需求的資訊可能包含全球性論點，洲際及區域性問題，或是只發生於單一國家的議題，因此將地域屬性分為全球、區域及單一國家三個層次。

- 全球層次：若在查詢主題中(1)沒有明確指出涵蓋地域，是一個世界性的趨勢及議題，如 Topic #017，問題描述為「**尋找有關克服後天性免疫不全症候群（愛滋病，AIDS）的努力與研究之文章**」，沒有限定國家或地區。(2)牽涉範圍廣泛，事件發生地點超過一個洲或區域，如 Topic #019，問題描述為「**尋找有關涉及超過一個國家的國際海上意外事件的文章**」，資訊需求為國際性的海上意外事件，亦無指定特定區域。符合上述兩種條件則屬於全球層次。
- 區域層次：若在查詢主題中(1)明確指出屬於洲際性事件，如 Topic #079，問題描述為「**有關於歐洲貨幣組織的經濟影響之文章**」，明確指出涉及區域為歐洲。(2)牽涉範圍超過一個國家，但不至於構成全球性議題，如 Topic #048，問題描述為「**查詢建設國際太空站計畫相關的文章**」，且在背景陳述部份指出「**基於和平目的，日本、歐洲、俄羅斯和美國進行國際太空站國際合作計畫。...**」，可知涉及超過一個國家，但也僅限於此四國進行的國際太空站計畫，地域範圍不至構成全球性議題。符合上述兩種條件則屬於區域層次。
- 單一國家層次：若在查詢主題中(1)明確指出屬於單一國家事件，如 Topic #016，問題描述為「**尋找有關北韓長程飛彈試射的文章**」，明確指定尋找單一國家-北韓的相關文件。(2)提及的人物或事件可界定歸屬於單一國家，如 Topic #077，問題描述為「**有關於北野武導演的電影之文章**」，涉及人物-北野武可明確歸屬為單一國家。或如 Topic #095，問題描述為「**查詢在日本的二次世界大戰戰罪訴訟**」，涉及事件發生於單一國家-日本。符合上述兩種條件則屬於單一國家層次。

(b) 時空明確性：將時間、空間各區分為明確及不明確兩種，交叉配對後共產生時不明確空不明確、時不明確空明確、時明確空不明確、時明確空明確四種。

新聞文件談論的主題可能涉及事件發生的時間與空間，在陳述資訊需求時可能限制尋找特定時間或空間的事件，或者廣泛找出所有相關的文獻，因此將時空

特性各區分為明確及不明確兩種。

- 時不明確空不明確：在查詢主題中沒有明確指定事件發生的時間及空間，如 Topic #003，問題描述為「**查詢科學家成功培養人類胚胎幹細胞的相關介紹**」，此題沒有限制特定的時間或空間，屬於時不明確空不明確的命題。
  - 時不明確空明確：在查詢主題中沒有明確指定事件發生的時間，但指定事件發生的空間，如 Topic #102，問題描述為「**查詢歐盟與亞洲關係的相關報導**」，沒有指定確切的時間，但空間為歐盟與亞洲，因此屬於時不明確空明確的命題。
  - 時明確空不明確：在查詢主題中明確指定事件發生的時間，但沒有指定事件發生的空間，如 Topic #020，問題描述為「**尋找描述有關企業千禧年（Y2K）問題的文件**」，尋找文件之時間限定為千禧年，但空間不限定，屬於時明確空不明確的命題。
  - 時明確空明確：在查詢主題中沒有明確指定事件發生的時間，但指定事件發生的空間，如 Topic #065，問題描述為「**查詢朱鎔基擔任中國總理後所提出的經濟改革計畫**」，且在背景陳述部份指出「**朱鎔基於 1998 年就任中國國務院總理。朱鎔基的經濟改革計畫內容為何？...**」，可知指定尋找朱鎔基於 1998 年提出的經濟改革計畫，屬於時間、空間皆明確的命題。
- (c) 主題：將查詢主題分為國際政治、政治、財經、社會綜合(含教育、環保)、生活(含醫藥)、科技資訊、藝文、體育、娛樂(含旅遊)等九個類別。主題的分類標準參考江玉婷(民 88)的實證研究，將資訊需求類別分為 9 大類別。該研究是針對新聞文件，蒐集使用者的資訊需求統整產生的類別，因此也可用來區分本研究的查詢主題，且每個查詢主題只歸屬於一個主題類別。
- 國際政治：查詢主題涉及國際性的政治議題，如 Topic #044，問題描述為「**查詢探討聯合國至伊拉克檢查有關毀滅性武器後，美國與伊拉克關係發展情形的文章**」，為一國際政治類別的查詢主題。
  - 政治：查詢主題涉及單一國家的政治議題，在此次分析的查詢主題中，並未出現單一國家政治類別的查詢主題。
  - 財經：查詢主題涉及財經議題，不限於國際性或單一國家事件，如 Topic #015，問題描述為「**查找專家對國際貨幣基金組織（IMF）對亞洲國家政策的批評意見**」，為一財經類別的查詢主題。
  - 社會綜合(含教育、環保)：查詢主題涉及社會議題，教育與環保議題也隸屬於社會綜合類別，不限於國際性或單一國家事件，如 Topic #039，問題描述為「**查詢有關驅逐外勞的行動，或其人權問題之文章**」，為一社會綜合類別的查詢主題。
  - 生活(含醫藥)：查詢主題論及生活議題，醫藥相關文件也隸屬於生活類別，不限於國際性或單一國家事件，如 Topic #033，問題描述為「**尋找有關研究以蛋白質消除疾病的文章**」，為一生活類別的查詢主題。
  - 科技資訊：查詢主題論及科技及資訊議題，不限於國際性或單一國家事件，如 Topic #042，問題描述為「**查詢提到蘋果電腦成果或新產品的文章**」，為

一科技資訊類別的查詢主題。

- 藝文：查詢主題論及藝文活動，不限於國際性或單一國家的藝文資訊，在此次分析的查詢主題中，並未出現藝文類別的查詢主題。
- 體育：查詢主題論及體育競賽、活動或人物，不限於國際性或單一國家事件，在此次分析的查詢主題中，並未出現體育類別的查詢主題。
- 娛樂(含旅遊)：查詢主題論及娛樂設施、娛樂活動、影藝新聞及旅遊資訊，不限於國際性或單一國家事件，如 Topic #066，問題描述為「**查詢有關刻正興建中的主題樂園之相關文獻**」，為一娛樂類別的查詢主題。

(d) 條件敘述明確性：CLIR 結構化的查詢主題之中有一<REL>欄位，內容為對該查詢主題的條件敘述與需求限制，又可分為相關、部份相關及不相關三種需求限制的層次。因為皆以自然語言的方式陳述，是提供判斷員決定一個文件與查詢主題相關程度高低的條件敘述，可能包含人事物專有名詞、範圍限制、形容詞及一般名詞。

在這裡區分條件敘述明確性的主要原則為：對於該條件敘述的回覆是否具有完備的條件和固定答案，也就是問題的開放性與封閉性，若為封閉性條件敘述，則歸納為明確；若答案包含多種可能，為開放性條件敘述，則歸納為不明確。界定條件敘述為明確(封閉性)的情形可歸納為以下幾種，一個條件敘述可能符合一種或多種情形：

- 明確的人事物專有名詞：在條件敘述中，限制尋找與某專有名詞相關、部份相關或不相關的文件，如 Topic #016，條件敘述為「**相關文章包括北韓長程地對地飛彈試射之計畫**」，這裡所使用的「北韓長程地對地飛彈試射」為專有名詞。又如 Topic #041，條件敘述為「**提到利用手機查詢網頁或使用電子郵件(e-mail)的文章視為相關**」，這裡明確指出「手機」、「網頁」與「電子郵件」等專有名詞。Topic #046，條件敘述為「**討論基因治療用於癌症以外疾病的文章視為不相關**」，使用「基因治療」、「癌症」等專有名詞。Topic #083，條件敘述為「**若一文章描述金大中總統對亞洲的政策則為相關**」，使用「金大中」、「對亞洲的政策」。都可歸納為封閉性的條件敘述。
- 有範圍限制的敘述句：在條件敘述中，限制尋找某特定範圍的文件，如 Topic #046，條件敘述為「**討論基因治療用於癌症以外疾病的文章視為不相關**」，範圍限制為「癌症以外疾病」。Topic #053，條件敘述為「**描述透過網路傳遞音樂的著作權事件文章視為相關**」，以及「**文章如果針對著作權保護，但沒有特別提到音樂傳遞問題者，視為不相關**」，範圍限制為「透過網路傳遞音樂的著作權事件」。Topic #059，條件敘述為「**有關接收數位傳播的設備或電視機的文章視為部分相關**」，範圍限制為「接收數位傳播的設備或電視機」。Topic #064，條件敘述為「**電子商務之應用範圍如虛擬銀行、網路下單交易及網路購物等視為相關**」，範圍限制為「虛擬銀行、網路下單交易及網路購物等應用」。以上皆歸納為封閉性的條件敘述。
- 其它：歸納一個明確的條件敘述，可能還包括使用

某些特定詞彙，例如當條件敘述中使用「定義」、「案例」、「政策」比起使用「影響」、「介紹」、「未來發展方向」、「評論」、「意義」、「意見」、「反應」、「方法」要來得明確，具備固定答案的機率較高，而偏向於封閉性的條件敘述，但還要檢查前後文的敘述，不能單憑這些詞彙，就判定該條件敘述為明確或不明確。

- 因為使用詞彙而判定為明確，如 Topic #019，條件敘述為「**解釋處理海上意外方法的文章為部分相關**」，雖然要尋找關於「方法」的文章，但因加上「處理海上意外」的前文限定，因此仍判定為明確的條件敘述。Topic #027，條件敘述為「**包含台灣政府陳述中國仍不放棄武力犯台事實的官方言論或反應的文章為相關**」，但強調是「台灣政府陳述中國仍不放棄武力犯台事實」的「言論」及「反應」，因此判定為明確的條件敘述。Topic #046，條件敘述為「**討論癌症基因治療案例視為相關**」，因「案例」有特定限制，屬於明確的條件敘述。
- 因為使用詞彙而判定為不明確，如 Topic #016，條件敘述為「**討論外國對此飛彈試射之反應的文章為部分相關**」，只提及「反應」的文章，雖加上「外國對此飛彈試射」的限制，但範圍太廣泛，因此歸類為不明確的條件敘述。Topic #037，條件敘述為「**討論各國或國際性組織針對北韓飢荒或糧食短缺問題的反應之文章視為相關**」，限制範圍廣泛，可能答案不固定，因此亦為不明確的條件敘述。Topic #044，談論關於伊拉克製造毀滅性武器的主題，條件敘述為「**提到專家評論此類武器檢查的文章視為不相關**」，敘述中使用「評論」是開放性的敘述，判定為不明確的條件敘述。Topic #079，條件敘述為「**若一文章關於歐洲貨幣組織的經濟影響其為相關**」，敘述中使用「影響」，可能找出開放性的答案，判斷為不明確的條件敘述。Topic #097，條件敘述為「**對於人體複製倫理的意見視為不相關**」，「意見」也是開放性用語，判斷為不明確的條件敘述。
- (e) 條件敘述個數：依據上述判斷條件敘述明確性的方式，計算每一個查詢主題在<REL>欄位中，相關、部份相關及不相關三種需求限制的個數，交叉配對後形成六個數據。

相關層次		部份相關層次		不相關層次	
不明確	明確	不明確	明確	不明確	明確

陳明君(民88)的研究提及，判斷一個查詢主題複雜度的方式為問題中包含概念的多寡，問題中包含的概念越多，則複雜程度越高。因此本研究計算條件敘述個數，做為判定題目複雜程度的依據。判斷條件敘述個數時需要注意的原則如下，但仍要搭配前後文：

- 條件敘述中使用「包含...」、「...等」、「如...」，則計算為一個敘述，如 Topic #070，條件敘述為「**反聖嬰現象形成的原因、特徵、循環性等基本介紹視為相關**」，這裡強調的是反聖嬰現象的「基本介紹」，而「原因、特徵、循環性」只是提供參考的敘述，讓相關判斷者在判斷一篇文件時有所憑藉，因此計算為一個相關層次的條件敘述。Topic

#079, 條件敘述為「若一篇文章有關其他地區貨幣組織的實踐, 如亞洲國家, 則為部分相關」, 由此例可知「如」是一個說明的用法, 不需視為單一獨立的條件。Topic #095, 條件敘述為「相關內容應描述有關日本在二次大戰期間戰罪的民事訴訟案件, 包括新的訴訟案件、過程、判決結果、對於判決的輿論」, 此處強調的是「日本在二次大戰期間戰罪的民事訴訟案件」, 而「包括...」則是對於民事訴訟案件的說明, 計算為一個相關層次的條件敘述。Topic #096, 條件敘述為「相關內容應描述在核能議題上特定的負面反映案例, 如請願、抗議或是示威運動」, 敘述中「如...」是對「負面反映案例」的說明, 計算為一個相關層次的條件敘述。

- 條件敘述中使用「或」、「與」、「及」、「、」、「區隔前後兩敘述, 則前與後兩個敘述皆計數一次, 如 Topic #074, 條件敘述為「非關特定事件或病毒的文章而為事件的一般性統計之文章為不相關」, 敘述中「非關特定事件」及「病毒」都可視為單獨的條件, 故此敘述句包含兩個不相關層次的敘述。Topic #075, 條件敘述為「若一篇文章描述複製小牛的出生與其使用的技術則為相關」, 敘述中「複製小牛的出生」及「其使用的技術」都可視為單獨的條件, 故包含兩個相關層次的敘述。Topic #100, 主題為「查詢電視台在除夕及新年期間所撥放的特別節目表」, 而條件敘述「對於平時的電視節目或只有節目表的報導則視為不相關」, 其中「平時的電視節目」及「只有節目表的報導」可視為兩個單獨的條件, 計算為兩個不相關層次的敘述。Topic #103 查詢全球性自然災害的文章, 條件敘述為「相關內容應包括受災地區的確切名稱、對於自然災害的詳細描述、傷亡人數及財物的損失」, 計算為四個相關層次的敘述, 「受災地區的確切名稱」、「對於自然災害的詳細描述」、「傷亡人數」與「財物的損失」。

將各查詢主題依據(1)涵蓋地域、(2)時空明確性、(3)主題、(4)條件敘述明確性及(5)條件敘述個數等五個查詢主題屬性予以編碼, 以利後續統計分析。

## (2) 查詢主題屬性與查詢結果分析

資料來源為 NTCIR-6 之中 CLIR 之相關判斷結果, 共有 20 個團隊參與 CLIR 競賽, 每個團隊可選擇使用特定的欄位, 以各自研發的機制進行檢索, 但指定每個參賽團隊都必須執行(1)從 Title 擷取詞彙及(2)從 Description 擷取詞彙的檢索方式(Mandatory Runs), 各團隊也可自行選擇從其他單一欄位或從多個欄位擷取詞彙。因此每個團隊交出的檢索結果至少要有兩組(T-run & D-run)。

另外, CLIR 不只是檢索機制的競賽, 還比較各團隊翻譯文獻、執行跨語言檢索的功效, CLIR 的文件集與測試集皆包括中文(繁)、日文及韓文三種語言, 參與競賽的隊伍必需選擇測試集以及文件集的語種, 因此共發展出 12 種跨語言檢索項目, C-CJK、C-C、C-J、E-C、E-J、E-K、J-C、J-J、J-K、K-C、K-J、K-K, 其中 C=中文(繁), J=日文, K=韓文, E=英文。

20 個參賽團隊選擇欄位、測試集與文件集的語種之後, 總計產生 152 組檢索結果, 其標記方式為「參賽隊伍代號-測試集語言-文件集語言-選擇欄位-檢索序

號」, 例 BKY-C-C-D-03 = 參賽隊伍 BKY, 選擇中文的測試集, 中文的文件集, 從 Description 欄位篩選詞彙, 第 03 號檢索。

相關判斷的層次又可分为四種: 高度相關(S)、相關(A)、部份相關(B)與不相關(C), 經過判斷者人工比對檢索出的文件與查詢主題的相關程度後, 採用 AP<sub>g</sub> 計算(又稱為 MAP), 針對每一題算出參賽隊伍在該題的表現, 計算相關程度的方式又可分为 Rigid relevant (S+A)及 Relaxed relevant (S+A+B)兩種, 因此 Rigid relevant 會低於 Relaxed relevant 的分數, 共計產生 152 組 \* 2 種計算方式 \* 50 個查詢主題的檢索結果, 檢索結果的表現介於 0~1 之間, 越接近 1 表現越佳。

## 5. 研究結果

(1) 查詢主題分析: 根據上述(a)涵蓋地域、(b)時空明確性、(c)主題、(d)條件敘述明確性及(e)條件敘述個數等五個特性, 分析 50 個查詢主題的檢索表現。

### (a) 涵蓋地域分析

- 全球層次: 沒有明確指出涵蓋地域, 為一世界性的趨勢及議題。或是牽涉範圍廣泛, 事件發生地點超過一個洲或區域, 則為全球層次。
- 區域層次: 明確指出屬於洲際性事件。或是牽涉範圍超過一個國家, 但不至於構成全球性議題。
- 單一國家層次: 明確指出屬於單一國家事件。或是提及的人物或事件可界定歸屬於單一國家

在 50 個查詢主題中, 分析結果如下:

- 37 個(74%)全球層次的查詢主題, 平均表現為 0.2561 (表現最佳!)
- 8 個(16%)區域層次的查詢主題, 平均表現為 0.2359
- 5 個(5%)單一國家層次的查詢主題, 平均表現為 0.2103 (表現最差)

平均分數

涵蓋地域	平均數	個數	標準差	最小值	最大值
全球	.256141	37	.140014	.0409	.5458
區域	.235888	8	.117522	.1078	.3958
單一國家	.210300	5	.103016	.1583	.3943
總和	.248316	50	.132141	.0409	.5458

### (b) 時空明確性分析

- 時不明確空不明確: 在查詢主題中沒有明確指定事件發生的時間及空間。
- 時不明確空明確: 在查詢主題中沒有明確指定事件發生的時間, 但指定事件發生的空間。
- 時明確空不明確: 在查詢主題中明確指定事件發生的時間, 但沒有指定事件發生的空間。
- 時明確空明確: 在查詢主題中沒有明確指定事件發生的時間, 但指定事件發生的空間。

在 50 個查詢主題中, 分析結果如下:

- 34 個時不明確空不明確的查詢主題, 平均表現為 0.2787 (表現最佳!)
- 6 個時不明確空明確的查詢主題, 平均表現為 0.1351
- 2 個時明確空不明確的查詢主題, 平均表現為 0.1176 (表現最差)
- 8 個時明確空明確的查詢主題, 平均表現為 0.2369

平均分數

時空明確性	平均數	個數	標準差	最小值	最大值
時不明確空不明確	.278682	34	.135203	.0546	.5458
時不明確空明確	.135067	6	5.727E-02	.0409	.2066
時明確空不明確	.117550	2	3.345E-02	.0939	.1412
時明確空明確	.236888	8	.114203	.1203	.3958
總和	.248316	50	.132141	.0409	.5458

## (c) 主題分析

- 國際政治：查詢主題涉及國際性的政治議題
- 政治：查詢主題涉及單一國家的政治議題
- 財經：查詢主題涉及財經議題，不限於國際或單一國家事件
- 社會綜合(含教育、環保)：查詢主題涉及社會議題，教育與環保議題也隸屬於社會綜合類別，不限於國際或單一國家事件
- 生活(含醫藥)：查詢主題論及生活議題，醫藥相關文件也隸屬於生活類別，不限於國際或單一國家事件
- 科技資訊：查詢主題論及科技及資訊議題，不限於國際或單一國家事件
- 藝文：查詢主題論及藝文活動，不限於國際或單一國家事件
- 體育：查詢主題論及體育競賽、活動或人物，不限於國際或單一國家事件
- 娛樂(含旅遊)：查詢主題論及娛樂設施、娛樂活動、影藝新聞及旅遊資訊，不限於國際或單一國家事件

在 50 個查詢主題中，分析結果如下：

- 11 個(22%)國際政治的查詢主題，平均表現為 0.1689 (表現最差)
- 6 個(12%)財經的查詢主題，平均表現為 0.2087
- 14 個(28%)社會綜合的查詢主題，平均表現為 0.2363
- 5 個(10%)生活的查詢主題，平均表現為 0.2206
- 11 個(22%)科技資訊的查詢主題，平均表現為 0.3740 (表現最佳！)
- 3 個(6%)娛樂的查詢主題，平均表現為 0.2603
- 政治、藝文及體育的查詢主題皆未出現

平均分數

主題	平均數	個數	標準差	最小值	最大值
國際政治	.168918	11	.102323	.0409	.3761
財經	.208700	6	9.372E-02	.1500	.3958
社會綜合	.236264	14	.115318	.0581	.4695
生活	.220620	5	.139492	.0546	.3910
科技資訊	.373991	11	.122796	.1412	.5458
娛樂	.260267	3	.152788	.0939	.3943
總和	.248316	50	.132141	.0409	.5458

## (d) 條件敘述明確性與個數分析

判定一條件敘述為明確的情形包括提及明確的專有名詞、有範圍限制的敘述句，以及使用某些特定詞彙，例如「定義」、「案例」及「政策」。又條件敘述可分為相關、部份相關及不相關三個層次。

將三個層次明確的條件敘述個數加總，可反映該查詢主題的複雜度，條件敘述個數越少，則複雜度越低；條件敘述個數越多，則複雜度越高。

## ◎ 條件敘述明確

- 6 個查詢主題明確敘述為 0 個，平均表現為 0.2801
- 13 個查詢主題明確敘述為 1 個，平均表現為 0.2437
- 14 個查詢主題明確敘述為 2 個，平均表現為

0.2777

- 12 個查詢主題明確敘述為 3 個，平均表現為 0.2254
- 3 個查詢主題明確敘述為 4 個，平均表現為 0.2815 (表現最佳！)
- 1 個查詢主題明確敘述為 5 個，平均表現為 0.0409 (表現最差)
- 1 個查詢主題明確敘述為 6 個，平均表現為 0.0890

平均分數

條件敘述明確	平均數	個數	標準差	最小值	最大值
0	.280133	6	.100967	.1662	.3958
1	.243738	13	.137752	.0581	.5458
2	.277650	14	.120309	.1203	.4448
3	.225408	12	.132864	.0939	.4695
4	.281500	3	.221559	.0546	.4973
5	4.090E-02	1	.	.0409	.0409
6	8.900E-02	1	.	.0890	.0890
總和	.248316	50	.132141	.0409	.5458

## ◎ 條件敘述不明確

- 7 個查詢主題不明確敘述為 0 個，平均表現為 0.1939 (表現最差)
- 17 個查詢主題不明確敘述為 1 個，平均表現為 0.2204
- 15 個查詢主題不明確敘述為 2 個，平均表現為 0.2782
- 6 個查詢主題不明確敘述為 3 個，平均表現為 0.2524
- 2 個查詢主題不明確敘述為 4 個，平均表現為 0.3531 (表現最佳！)
- 3 個查詢主題不明確敘述為 5 個，平均表現為 0.3063

平均分數

條件敘述不明確	平均數	個數	標準差	最小值	最大值
0	.193929	7	.177740	.0409	.4973
1	.220353	17	.116994	.0783	.4695
2	.278187	15	.119751	.0890	.4448
3	.252433	6	.163523	.0581	.5458
4	.353050	2	6.046E-02	.3103	.3958
5	.306267	3	.121301	.1662	.3765
總和	.248316	50	.132141	.0409	.5458

## ◎ 條件敘述總個數(明確+不明確)

- 8 個查詢主題包含 2 個條件敘述，平均表現為 0.2096
- 12 個查詢主題包含 3 個條件敘述，平均表現為 0.2385
- 22 個查詢主題包含 4 個條件敘述，平均表現為 0.2764
- 4 個查詢主題包含 5 個條件敘述，平均表現為 0.1659
- 1 個查詢主題包含 6 個條件敘述，平均表現為 0.3761 (表現最佳！)
- 2 個查詢主題包含 7 個條件敘述，平均表現為 0.3346
- 1 個查詢主題包含 8 個條件敘述，平均表現為 0.0890 (表現最差)

平均分數

條件敘述個數	平均數	個數	標準差	最小值	最大值
2	.209550	8	.105871	.0783	.3943
3	.238475	12	.104777	.0939	.3910
4	.276359	22	.155779	.0546	.5458
5	.165900	4	9.833E-02	.0409	.2812
6	.376100	1	.	.3761	.3761
7	.334550	2	5.933E-02	.2926	.3765
8	8.900E-02	1	.	.0890	.0890
總和	.248316	50	.132141	.0409	.5458

## ◎ 條件敘述差異個數

- 34 個查詢主題的明確條件敘述多於不明確條件

敘述，平均表現為 0.2339 (表現較差)

- 16 個查詢主題的明確條件敘述多於不明確條件敘述，平均表現為 0.2790 (表現較佳！)

平均分數

條件敘述個數	不明確	明確	平均數	個數	標準差
0			.233868	34	.134514
1			.279019	16	.125515
總和			.248316	50	.132141

## (2) 檢索結果分析

### (a) 各語言類別的參賽組數

根據「測試集-文件集」12 個單一語言或跨語言類別：C-CJK、C-C、C-J、E-C、E-J、E-K、J-C、J-J、J-K、K-C、K-J、K-K (其中 C=中文(繁)，J=日文，K=韓文，E=英文)，區分 148 組參賽隊伍的檢索結果。

※以下分析檢索結果皆採用 MAP Scores (rigid relevant)

由下表可知單一語種(X→X)參賽組數最多為 J-J 的 43 組，其次為 C-C 的 32 組，之後為 K-K 的 20 組。雙語種(X→Y)參數組數最多為 C-J 的 17 組，其次為 E-J 的 9 組。多語種(X→XYZ)只有 2 個參數組。

語言別	參數組數	%
C-CJK	2	1.3
C-C	32	21.1
C-J	17	11.2
E-C	8	5.3
E-J	9	5.9
E-K	5	3.3
J-C	4	2.6
J-J	43	28.3
J-K	5	3.3
K-C	2	1.3
K-J	5	3.3
K-K	20	13.2

### (b) 各語言類別的表現

C-CJK：共 2 組檢索結果，在 50 個查詢主題的平均表現分別為：

組別	CYUT-T-01	CYUT-D-02
平均分數	0.0704	0.0584

C-C：共 32 組檢索結果，在 50 個查詢主題中平均表現最佳的三組為：

組別	I2R-D-01	I2R-T-01	UniNE-D-05
平均分數	0.3136	0.3097	0.2891

C-J：共 17 組檢索結果，在 50 個查詢主題中平均表現最佳的三組為：

組別	TSB-T-04	TSB-T-03	TSB-D-02
平均分數	0.3233	0.3219	0.3119

E-C：共 8 組檢索結果，在 50 個查詢主題中平均表現最佳的三組為：

組別	I2R-T01	I2R-D-01	pircs-T-03
平均分數	0.2013	0.1911	0.1686

E-J：共 9 組檢索結果，在 50 個查詢主題中平均表現最佳的三組為：

組別	TSB-T-03	TSB-D-01	TSB-D-02
平均分數	0.3072	0.3043	0.3012

E-K：共 5 組檢索結果，在 50 個查詢主題中平均表現最佳的三組為：

組別	NICT-TDNC-02	NICT-D-05	NICT-T-04
平均分數	0.3191	0.2915	0.2813

J-C：共 4 組檢索結果，在 50 個查詢主題中平均表現最佳的三組為：

組別	BKY-TDNC-01	BKY-DN-02	BKY-D-03
平均分數	0.1748	0.1660	0.0770

J-J：共 43 組檢索結果，在 50 個查詢主題中平均表現最佳的三組為：

組別	TSB-T-04	TSB-T-03	TSB-D-02
平均分數	0.3600	0.3598	0.3255

J-K：共 5 組檢索結果，在 50 個查詢主題中平均表現最佳的三組為：

組別	NICT-TDNC-04	NICT-DN-03	NICT-D-05
平均分數	0.3401	0.3343	0.2866

K-C：共 2 組檢索結果，在 50 個查詢主題的平均表現分別為：

組別	IASL-T-01	IASL-D-01
平均分數	0.0704	0.0584

K-J：共 5 組檢索結果，在 50 個查詢主題中平均表現最佳的三組為：

組別	NICT-D-02	NICT-TDNC-04	NICT-T-01
平均分數	0.2671	0.2596	0.2452

K-K：共 20 組檢索結果，在 50 個查詢主題中平均表現最佳的三組為：

組別	KLE-TDNC-05	NICT-TDNC-02	KLE-N-03
平均分數	0.4789	0.4710	0.4631

由 12 種語言別表現，可看出 K-K 表現最佳，J-K 次之，其餘 10 種語言別表現依序為：E-K、E-J、J-J、K-J、C-C、C-J、E-C、J-C、K-C、C-CJK。

### (c) 各語言類別的表現 (測試集)

將 12 種檢索語言類別依據測試集語言種類分為四組，找出表現最佳的測試集語言種類。

- C：包括 C-CJK、C-C、C-J，共有 47 組參賽，平均表現為 0.1673
- E：包括 E-C、E-K、E-J，共有 22 組參賽，平均表現為 0.2310
- J：包括 J-C、J-J、J-K，共有 52 組參賽，平均表現為 0.2274
- K：包括 K-C、K-J、K-K，共有 27 組參賽，平均表現為 0.2460

統計結果：四組測試集中以 K 組的表現最佳，其次為 E 組，而後為 J 組，C 組表現最差。

### (d) 各語言類別的表現 (文件集)

將 12 種檢索語言類別依據文件集語言種類分為四組，找出表現最佳的文件集語言種類。

- CJK：包括 C-CJK，共有 2 組參賽，平均表現為 0.0644
- C：包括 C-C、E-C、J-C、K-C，共有 42 組參賽，平均表現為 0.1510
- J：包括 C-J、E-J、J-J、K-J，共有 74 組參賽，平均表現為 0.2447
- K：包括 E-K、J-K、K-K，共有 30 組參賽，平均表現為 0.3227

統計結果：四組文件集中以 K 組的表現最佳，其次為 J 組，而後為 C 組，CJK 組表現最差。

(e) 不同語言類別中查詢主題的表現

找出在 12 種不同「測試集-文件集」語言類別中，表現最佳的查詢主題。各題之表現以 MAP (rigid) 為衡量準則。

- C-CJK：在 50 個查詢主題中表現最好的三題為：Topic #003(0.2051)、Topic #070(0.1987)、Topic #042(0.1924)
- C-C：在 50 個查詢主題中表現最好的三題為：Topic #075(0.5105)、Topic #070(0.4999)、Topic #077(0.4868)
- C-J：在 50 個查詢主題中表現最好的三題為：Topic #021(0.6008)、Topic #074(0.5866)、Topic #053(0.4964)
- E-C：在 50 個查詢主題中表現最好的三題為：Topic #077(0.5069)、Topic #070(0.3526)、Topic #080(0.3475)
- E-J：在 50 個查詢主題中表現最好的三題為：Topic #021(0.8538)、Topic #074(0.7200)、Topic #050(0.7101)
- E-K：在 50 個查詢主題中表現最好的三題為：Topic #074(0.8545)、Topic #021(0.8474)、Topic #014(0.6900)
- J-C：在 50 個查詢主題中表現最好的三題為：Topic #075(0.6542)、Topic #044(0.3973)、Topic #074(0.3308)
- J-J：在 50 個查詢主題中表現最好的三題為：Topic #021(0.7248)、Topic #059(0.6224)、Topic #058(0.6027)
- J-K：在 50 個查詢主題中表現最好的三題為：Topic #021(0.9104)、Topic #053(0.8142)、Topic #080(0.6377)
- K-C：在 50 個查詢主題中表現最好的三題為：Topic #077(0.4615)、Topic #097(0.3788)、Topic #044(0.3609)
- K-J：在 50 個查詢主題中表現最好的三題為：Topic #074(0.7675)、Topic #050(0.6863)、Topic #053(0.6832)
- K-K：在 50 個查詢主題中表現最好的三題為：Topic #074(0.7885)、Topic #053(0.7553)、Topic #077(0.6499)

統計上述資料可得：

- 在 6 種語言類別中表現佳：Topic #074，問題描述為「有關於感染電腦病毒引起的問題之文章」。
- 在 5 種語言類別中表現佳：Topic #021，問題描述為「尋找有關電子商務交易所需的電子簽章

之文件」。

- 在 4 種語言類別中表現佳：Topic #053，問題描述為「查詢有關因網際網路傳遞音樂所引發的版權保護議題文章」及 Topic #077，問題描述為「有關於北野武導演的電影之文章」。
- 在 3 種語言類別中表現佳：Topic #070，問題描述為「查詢何謂反聖嬰現象及其與聖嬰現象的比較與影響」。
- 在 2 種語言類別中表現佳：Topic #044，問題描述為「查詢探討聯合國至伊拉克檢查有關毀滅性武器後，美國與伊拉克關係發展情形的文章」、Topic #050，問題描述為「查詢反對地底核武試爆的文章」、Topic #075，問題描述為「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」及 Topic #080，問題描述為「有關於日產與雷諾汽車公司資本結合的文章」。
- 在 1 種語言類別中表現佳：Topic #003，問題描述為「查詢科學家成功培養人類胚胎幹細胞的相關介紹」、Topic #014，問題描述為「查詢「環境荷爾蒙」對人體造成的影響及威脅」、Topic #042，問題描述為「查詢提到蘋果電腦成果或新產品的文章」、Topic #058，問題描述為「查詢非接觸式智慧卡技術或使用的相關文章」、Topic #059，問題描述為「查詢電視傳播數位化相關文章」及 Topic #097，問題描述為「查詢政府或國際上對於人體複製禁令所做的努力」。

TOPIC

	次數	百分比	有效百分比	累積百分比
有效的	3	2.8	2.8	2.8
	14	2.8	2.8	5.6
	21	13.9	13.9	19.4
	42	2.8	2.8	22.2
	44	5.6	5.6	27.8
	50	5.6	5.6	33.3
	53	11.1	11.1	44.4
	58	2.8	2.8	47.2
	59	2.8	2.8	50.0
	70	8.3	8.3	58.3
	74	16.7	16.7	75.0
	75	5.6	5.6	80.6
	77	11.1	11.1	91.7
	80	5.6	5.6	97.2
	97	2.8	2.8	100.0
總和	36	100.0	100.0	

- 表現最佳的 Topic #074，問題描述為「有關於感染電腦病毒引起的問題之文章」。查詢主題屬性為全球層次、時不明確空不明確、有 2 個明確的相關層次條件敘述及 2 個明確的不相關層次的條件敘述，主題為科技資訊。
- 表現次佳的 Topic #021，問題描述為「尋找有關電子商務交易所需的電子簽章之文件」。查詢主題屬性為全球層次、時不明確空不明確、有 2 個不明確及 1 個明確的相關層次條件敘述，1 個不明確的不相關層次條件敘述，主題為科技資訊。

由上述兩個例子，再次驗證表現較佳的查詢主題之可能特性：(1)涵蓋地域=全球層次，(2)時空明確性=時不明確空不明確，(3)主題=科技資訊，(4)問題複雜度=條件敘述總個數(明確+不明確)=4。

與查詢主題分析的結果相似，但在問題複雜度的部



份出現歧異，上述統計結果為：當一查詢主題包含 6 個條件敘述時表現最佳。

(f) 前十名表現最佳查詢主題之特性

依據 148 組檢索表現的平均分數，列出前十名表現最佳的查詢主題，依序為 Topic #021(0.5458)、Topic #074(0.4973)、Topic #053(0.4695)、Topic #075(0.4448)、Topic #058(0.4395)、Topic #059(0.4328)、Topic #042(0.4246)、Topic #080(0.3958)、Topic #077(0.3943) 以及 Topic #036(0.3910)。

- 涵蓋區域：全球 x 7 題、區域 x 1 題、單一國家 x 1 題
- 時空明確性：時不明確空不明確 x 9 題、時明確空明確 x 1 題
- 主題：科技資訊 x 6 題、生活 x 1 題、社會綜合 x 1 題、娛樂 x 1 題、財經 x 1 題
- 不明確的條件敘述：2 個 x 4 題、1 個 x 2 題、0 個 x 2 題、3 個 x 1 題、4 個 x 1 題
- 明確的條件敘述：2 個 x 3 題、3 個 x 3 題、0 個 x 2 題、1 個 x 1 題、4 個 x 1 題
- 在表現最佳的前十名查詢主題中，最常見的特色為全球、時不明確空不明確的科技資訊議題，且包含 2 個不明確的條件敘述、2~3 個明確的條件敘述。

(g) 前十名表現最差查詢主題之特性

依據 148 組檢索表現的平均分數，列出前十名表現最差的查詢主題，依序為 Topic #019(0.0409)、Topic #110(0.0546)、Topic #018(0.0581)、Topic #105(0.0783)、Topic #103(0.0890)、Topic #100(0.0939)、Topic #027(0.1078)、Topic #037(0.1203)、Topic #017(0.1224) 以及 Topic #026(0.1386)

- 涵蓋區域：全球 x 8 題、區域 x 2 題
- 時空明確性：時不明確空不明確 x 5 題、時不明確空明確 x 3 題、時明確空不明確 x 1 題、時明確空明確 x 1 題
- 主題：國際政治 x 5 題、社會綜合 x 2 題、生活 x 2 題、娛樂 x 1 題
- 不明確的條件敘述：0 個 x 4 題、1 個 x 4 題、2 個 x 1 題、3 個 x 1 題
- 明確的條件敘述：1 個 x 2 題、2 個 x 2 題、3 個 x 3 題、4 個 x 1 題、5 個 x 1 題、6 個 x 1 題
- 在表現最差的前十名查詢主題中，最常見的特色為全球、時不明確空不明確的國際政治議題，且包含 0~4 個不明確的條件敘述、3 個明確的條件敘述。

(3) 一致性檢定-Kendall's W 檢定

Kendall's W 檢定：若有 b 種特性，以 k 個受測者來衡量其對每一特性的能力或看法，並依每一特性，排列此 k 種能力的順序。再衡量此 k 個受測者對於 b 種特性的能力是否均具一致性，其一致性的強度有多大？或者此 b 種特性具有關聯性，其關聯性有多高。

從上段敘述，將 Kendall's W 檢定套用至本研究則為：有 50 個查詢主題，以 148 個組別來衡量其對每一查詢主題的表現效益，以及 148 個組別對於檢索 50 個查詢主題的表現效益是否具一致性，一致性強度有

多大？或者 50 個查詢主題之間具有關聯性，則其關聯性有多高。也就是將每一組別對 50 個查詢主題的表現 (MAP score-rigid) 相加，若 148 個組別對 50 個查詢主題的檢索表現具一致性，則 148 個組別的加總應相近。當此假設成立時，可推論 148 個組別之間的檢索效益且一致性，表現分數的高低起因於查詢主題的個別差異。

Kendall's W 檢定的雙尾假設( $\alpha=0.05$ )為：

$H_0$ : 148 個組別對 50 個查詢主題的表現不具一致性

$H_1$ : 148 個組別對 50 個查詢主題的表現具一致性

因 Kendall's W 檢定為無母數分析的一種，故需先將 148 組的分數等級化，將每一組的評分轉換為四個等級(1、2、3、4)，再進行 Kendall's W 檢定。

檢定結果如下表：顯著性 p 值=0.200 大於顯著水準  $\alpha=0.05$ ，所以不拒絕  $H_0$ 。即 148 個組別對 50 個查詢主題的表現不具一致性，表示組別之間也存在個別差異。

※ 個數=14，因為 Kendall's W 檢定需要成對的數據才能執行，但每個組別不一定交出 50 個查詢主題的檢索結果，亦即一個題目可能只有 140 個組別交出結果，最後只有 14 個查詢主題是 148 個組別皆有分數，故 Kendall's W 檢定個數為 14。

檢定統計量

個數		14
Kendall's W 檢定	a	.078
卡方		161.227
自由度		147
漸近顯著性		.200

a. Kendall 和諧係數

(4) 交叉分析

交叉分析的主要功能是用以探討多個離散變數的關聯分佈。因此需先將 148 組對 50 個查詢主題的檢索表現編碼，在 7400 (=148\*50) 筆結果中，有 135 筆為遺漏值。最後依據有效的 7265 筆檢索表現之四分位數 (Q1=0.0582, Q2=0.1987, Q3=0.4067) 將檢索表現分為 4 組。

範圍	編碼	次數
0~0.0582	1	1820
0.0583~0.1987	2	1813
0.1988~0.4067	3	1816
0.4067~1	4	1816

※ 7265 筆結果中包含 6 次 0.0582，故第 1 組比其他 3 組次數多。

以下將進行五項交叉表的卡方檢定：(a)涵蓋地域 x 檢索表現，(b)時空明確性 x 檢索表現，(c)主題 x 檢索表現，(d)語言種類 x 檢索表現，(e)檢索欄位 x 檢索表現，以瞭解行變數與列變數是否統計獨立。若是，則表示行變數與列變數無統計相關。應用卡方檢定時，期望次數低於 5 的方格數不能多於 5%。若多於 5%，可能導致卡方檢定不準確，則須考慮合併期望次數低於 5 的方格，使其能適合條件。

(a) 涵蓋地域 x 檢索表現

涵蓋地域包含：全球、區域及單一國家三個層次。

交叉表

			分數分4組				總和
			1	2	3	4	
地域因素	全球	個數	1383	1268	1297	1435	5383
		期望個數	1348.5	1343.3	1345.6	1345.6	5383.0
	區域	個數	210	358	349	248	1165
		期望個數	291.9	290.7	291.2	291.2	1165.0
	單一國家	個數	227	187	170	133	717
		期望個數	179.6	178.9	179.2	179.2	717.0
總和		個數	1820	1813	1816	1816	7265
		期望個數	1820.0	1813.0	1816.0	1816.0	7265.0

卡方檢定

	數值	自由度	漸近顯著性(雙尾)
Pearson卡方	94.464 <sup>a</sup>	6	.000
概似比	95.726	6	.000
線性對線性的關連	15.446	1	.000
有效觀察值的個數	7265		

a. 0格(.0%)的預期個數少於5。最小的預期個數為178.93。

P值為0.000小於顯著水準 $\alpha=0.05$ ，故判定地域因素與檢索表現有關。

## (b) 時空明確性 x 檢索表現

時空明確性包含：時不明確空不明確、時不明確空明確、時明確空不明確、時明確空明確。

交叉表

			分數分4組				總和
			1	2	3	4	
時空因素	時不明確空不明確	個數	1079	1137	1254	1477	4947
		期望個數	1239.3	1234.5	1236.6	1236.6	4947.0
	時不明確空明確	個數	291	350	186	42	869
		期望個數	217.7	216.9	217.2	217.2	869.0
	時明確空不明確	個數	147	78	43	19	287
		期望個數	71.9	71.6	71.7	71.7	287.0
	時明確空明確	個數	303	248	333	278	1162
		期望個數	291.1	290.0	290.5	290.5	1162.0
總和		個數	1820	1813	1816	1816	7265
		期望個數	1820.0	1813.0	1816.0	1816.0	7265.0

卡方檢定

	數值	自由度	漸近顯著性(雙尾)
Pearson卡方	470.309 <sup>a</sup>	9	.000
概似比	524.483	9	.000
線性對線性的關連	64.207	1	.000
有效觀察值的個數	7265		

a. 0格(.0%)的預期個數少於5。最小的預期個數為71.62。

P值為0.000小於顯著水準 $\alpha=0.05$ ，故判定時空明確性與檢索表現有關。

## (c) 主題 x 檢索表現

查詢主題包含：國際政治、政治、財經、社會綜合(含教育、環保)、生活(含醫藥)、科技資訊、藝文、體育、娛樂(含旅遊)等九個類別，其中政治、藝文及體育並未出現在此次分析的50個查詢主題中。

交叉表

			分數分4組				總和
			1	2	3	4	
主題因素	國際政治	個數	564	517	322	197	1600
		期望個數	400.8	399.3	399.9	399.9	1600.0
	財經	個數	177	287	268	133	865
		期望個數	216.7	215.9	216.2	216.2	865.0
	社會綜合	個數	519	567	454	488	2028
		期望個數	508.0	506.1	506.9	506.9	2028.0
	生活	個數	199	190	172	166	727
		期望個數	182.1	181.4	181.7	181.7	727.0
	科技資訊	個數	215	203	473	720	1611
		期望個數	403.6	402.0	402.7	402.7	1611.0
	娛樂	個數	146	49	127	112	434
		期望個數	108.7	108.3	108.5	108.5	434.0
總和		個數	1820	1813	1816	1816	7265
		期望個數	1820.0	1813.0	1816.0	1816.0	7265.0

卡方檢定

	數值	自由度	漸近顯著性(雙尾)
Pearson卡方	809.574 <sup>a</sup>	15	.000
概似比	825.326	15	.000
線性對線性的關連	326.574	1	.000
有效觀察值的個數	7265		

a. 0格(.0%)的預期個數少於5。最小的預期個數為108.31。

P值為0.000小於顯著水準 $\alpha=0.05$ ，故判定主題類別與檢索表現有關。

## (d) 語言種類 x 檢索表現

12個單一語言或跨語言類別：C-CJK、C-C、C-J、E-C、E-J、E-K、J-C、J-J、J-K、K-C、K-J、K-K (C=中文(繁)，J=日文，K=韓文，E=英文)

交叉表

			分數分4組				總和
			1	2	3	4	
語言(1-12)	c-cjk	個數	59	37	4	0	100
		期望個數	25.1	25.0	25.0	25.0	100.0
	c-c	個數	260	431	447	222	1360
		期望個數	340.7	339.4	340.0	340.0	1360.0
	c-j	個數	301	183	163	175	822
		期望個數	205.9	205.1	205.5	205.5	822.0
	e-c	個數	153	122	92	33	400
		期望個數	100.2	99.8	100.0	100.0	400.0
	e-j	個數	99	125	99	127	450
		期望個數	112.7	112.3	112.5	112.5	450.0
	e-k	個數	69	47	57	76	249
		期望個數	62.4	62.1	62.2	62.2	249.0
	j-c	個數	108	28	45	10	191
		期望個數	47.8	47.7	47.7	47.7	191.0
	j-j	個數	486	545	533	538	2102
		期望個數	526.6	524.6	525.4	525.4	2102.0
	j-k	個數	54	49	68	79	250
		期望個數	62.6	62.4	62.5	62.5	250.0
	k-c	個數	50	31	16	3	100
		期望個數	25.1	25.0	25.0	25.0	100.0
	k-j	個數	58	71	51	63	243
		期望個數	60.9	60.6	60.7	60.7	243.0
	k-k	個數	123	144	241	490	998
		期望個數	250.0	249.1	249.5	249.5	998.0
總和		個數	1820	1813	1816	1816	7265
		期望個數	1820.0	1813.0	1816.0	1816.0	7265.0

卡方檢定

	數值	自由度	漸近顯著性(雙尾)
Pearson卡方	885.716 <sup>a</sup>	33	.000
概似比	890.643	33	.000
線性對線性的關連	239.095	1	.000
有效觀察值的個數	7265		

a. 0格(.0%)的預期個數少於5。最小的預期個數為24.96。

P值為0.000小於顯著水準 $\alpha=0.05$ ，故判定語言種類與檢索表現有關。

## (e) 檢索欄位 x 檢索表現

148個參賽組別所使用的檢索欄位包含：C (concept)、D (description)、N (narrative)、T (title)、DN (D+N)、TDNC (T+D+N+C)五種。

交叉表

			分數分4組				總和
			1	2	3	4	
檢索詞選擇	c	個數	2	8	16	24	50
		期望個數	12.5	12.5	12.5	12.5	50.0
	d	個數	727	765	756	710	2958
		期望個數	741.0	738.2	739.4	739.4	2958.0
	n	個數	3	4	10	32	49
		期望個數	12.3	12.2	12.2	12.2	49.0
	t	個數	736	696	672	704	2808
		期望個數	703.4	700.7	701.9	701.9	2808.0
	dn	個數	106	185	193	166	650
		期望個數	162.8	162.2	162.5	162.5	650.0
	tdnc	個數	246	155	169	180	750
		期望個數	187.9	187.2	187.5	187.5	750.0
總和		個數	1820	1813	1816	1816	7265
		期望個數	1820.0	1813.0	1816.0	1816.0	7265.0

卡方檢定

	數值	自由度	漸近顯著性(雙尾)
Pearson卡方	126.898 <sup>a</sup>	15	.000
概似比	126.009	15	.000
線性對線性的關連	2.087	1	.149
有效觀察值的個數	7265		

a. 0格(.0%)的預期個數少於5。最小的預期個數為12.23。

P值為0.000小於顯著水準 $\alpha=0.05$ ，故判定檢索欄位與檢索表現有關。

上述五個交叉表的卡方檢定，p 值皆小於顯著水準，可知行變數與列變數並非統計獨立，亦即兩者存在統計相關，但從交叉分析無法判定統計相關的正負向。

相關

			檢索表現	涵蓋地域	時空明確性	主題	語言種類	檢索欄位	
Kendall's tau_b 統計量數	檢索表現	相關係數	1.000	-.031**	-.127**	.197**	.140**	-.011	
		顯著性(雙尾)	.	.003	.000	.000	.000	.268	
		個數	7265	7265	7265	7265	7265	7265	
	涵蓋地域	相關係數	-.031**	1.000	.721**	-.391**	.000	.000	
		顯著性(雙尾)	.003	.	.000	.000	1.000	1.000	
		個數	7265	7400	7400	7400	7400	7400	
	時空明確性	相關係數	-.127**	.721**	1.000	-.414**	.000	.000	
		顯著性(雙尾)	.000	.000	.	.000	1.000	1.000	
		個數	7265	7400	7400	7400	7400	7400	
	主題	相關係數	.197**	-.391**	-.414**	1.000	.000	.000	
		顯著性(雙尾)	.000	.000	.000	.	1.000	1.000	
		個數	7265	7400	7400	7400	7400	7400	
	語言種類	相關係數	.140**	.000	.000	.000	1.000	.006	
		顯著性(雙尾)	.000	1.000	1.000	1.000	.	.520	
		個數	7265	7400	7400	7400	7400	7400	
	檢索欄位	相關係數	-.011	.000	.000	.000	.006	1.000	
		顯著性(雙尾)	.268	1.000	1.000	1.000	.520	.	
		個數	7265	7400	7400	7400	7400	7400	
	Spearman's rho 係數	檢索表現	相關係數	1.000	-.035**	-.145**	.240**	.176**	-.013
			顯著性(雙尾)	.	.003	.000	.000	.000	.266
			個數	7265	7265	7265	7265	7265	7265
		涵蓋地域	相關係數	-.035**	1.000	.754**	-.443**	.000	.000
			顯著性(雙尾)	.003	.	.000	.000	1.000	.999
			個數	7265	7400	7400	7400	7400	7400
時空明確性		相關係數	-.145**	.754**	1.000	-.480**	.000	.000	
		顯著性(雙尾)	.000	.000	.	.000	1.000	.998	
		個數	7265	7400	7400	7400	7400	7400	
主題		相關係數	.240**	-.443**	-.480**	1.000	.000	.000	
		顯著性(雙尾)	.000	.000	.000	.	1.000	.998	
		個數	7265	7400	7400	7400	7400	7400	
語言種類		相關係數	.176**	.000	.000	.000	1.000	.007	
		顯著性(雙尾)	.000	1.000	1.000	1.000	.	.520	
		個數	7265	7400	7400	7400	7400	7400	
檢索欄位		相關係數	-.013	.000	.000	.000	.007	1.000	
		顯著性(雙尾)	.266	.999	.998	.998	.520	.	
		個數	7265	7400	7400	7400	7400	7400	

\*\* . 在 .01水準(雙尾)上的相關才會顯著。

(5) 相關

相關係數的主要功能是用以衡量兩配對隨機變數之關係，相關係數介於-1與+1之間。在此研究中，討論條件敘述明確性與條件敘述個數與檢索結果的表現是否相關。若存在相關，則為正相關或為負相關。

由統計報表的資訊可知：

- 檢索表現與「不明確相關敘述個數」、「不明確部份相關敘述個數」、「不明確不相關敘述個數」、「條件敘述不明確總個數」及「條件敘述總個數」為正相關。
  - 不明確的條件敘述個數越多，則檢索表現佳；條件敘述總個數越多，檢索表現亦佳。
  - 與檢索表現的正相關強度依序為：
    - (a) 條件敘述不明確總個數 (0.157)
    - (b) 不明確部份相關敘述個數 (0.116)

- (c) 不明確不相關敘述個數 (0.109)
- (d) 不明確相關敘述個數 (0.076)
- (e) 條件敘述總個數 (0.042)

- 檢索表現與「明確相關敘述個數」、「明確部份相關敘述個數」、「明確不相關敘述個數」及「條件敘述明確總個數」為負相關。
  - 明確的條件敘述個數越多，則檢索表現不佳。
  - 與檢索表現的負相關強度依序為：
    - (a) 明確相關敘述個數 (-0.114)
    - (b) 條件敘述明確總個數 (-0.113)
    - (c) 明確部份相關敘述個數 (-0.037)
    - (d) 明確不相關敘述個數 (-0.035)
- 檢索表現與「條件敘述不明確個數」的相關度顯著水準達0.05；而檢索表現與「不明確相關敘述個數」、「明確相關敘述個數」、「不明確部份相關

敘述個數」、「明確部份相關敘述個數」、「不明確不相關敘述個數」、「明確不相關敘述個數」、「條件敘述明確個數」及「條件敘述總個數」等八個屬性的相關度顯著水準達 0.01。

- 亦即明確的條件敘述個數越多，檢索表現越差。

#### (6) 無母數相關-Spearman's Rho 等級相關係數 Kendall's Tau 相關係數

如上述的 Pearson 相關係數適用於量尺變數，當變數為名義或次序尺度時，必須使用無母數相關係數分析。本研究應用「Spearman's Rho 等級相關係數」以及「Kendall's Tau 相關係數」，多重驗證分析檢索表現與「涵蓋地域」、「時空明確性」、「主題」、「語言種類」及「檢索欄位」之相關程度強弱。

由統計報表的資訊可知：

- 檢索表現與「涵蓋地域」、「時空明確性」、「主題」、「語言種類」的相關達 0.01 的顯著水準
- 檢索表現與「主題」(0.197)與「語言種類」(0.140)為正相關
- 檢索表現與「涵蓋地域」(-0.031)與「時空明確性」(-0.127)為負相關
- 檢索表現與「檢索欄位」在交叉分析檢定時，雖顯示兩者存在統計相關，但之間的相關度不顯著。
- 與檢索表現的相關強度依序為
  - ◆ 主題
  - ◆ 語言種類
  - ◆ 時空明確性
  - ◆ 涵蓋地域
  - ◆ 檢索欄位

#### (7) 迴歸模型

迴歸模型的假設為「依變數 Y 之期望值為自變數 X1、X2、... 之線性函數」。本研究中依變數為「檢索表現」，自變數為「不明確相關敘述個數」、「明確相關敘述個數」、「不明確部份相關敘述個數」、「明確部份相關敘述個數」、「不明確不相關敘述個數」、「明確不相關敘述個數」、「條件敘述不明確個數」、「條件敘述明確個數」及「條件敘述總個數」等九項連續變數。

經由統計分析的結果，發展出的線性函數解釋力只有 4.0%，可能原因包括九個自變數的代表性不足，不適合以線性迴歸模型預測「檢索表現」。

## 6. 結論

從上述統計分析的結果，最後雖然無法得到一個有效的線性迴歸模型，但仍可歸納出表現較佳的查詢主題之特性如下：

「一個查詢主題若為「全球層次」、「時空皆不明確」、主題為「科技資訊」、「不明確的條件敘述」多於「明確的條件敘述」等特質，則可預測有較佳的檢索表現，而擁有這些特性的查詢主題可推論為一個較「容易」的查詢主題。」

經檢驗 50 個查詢主題中，共有 5 題符合以上敘述，分別為 Topic #003、Topic #021、Topic #058、Topic #059 以及 Topic #075，且此 5 題的平均表現分數為 0.5189。」故上述對於一個表現較佳的查詢主題之特質敘述為有效的。

至於其他特性，如選擇檢索的欄位、三個相關層次（相關、部份相關、不相關）的條件敘述個數，雖與檢索表現有正或負向相關，但統計上的相關性皆不夠顯著，因此不納入影響檢索表現的屬性。

本研究分析 50 個查詢主題的特性與檢索表現之關係，在將特性編碼及界定範圍及明確性時，已儘量遵從既定的原則，但仍可能因個人主觀認知而產生偏誤；另一方面，雖本研究的結果顯示各特性之間並無交互影響，但是受限於研究的時間與使用的語料，無法直接推論於其他的語料。仍待更為全面與完整的語料，進一步的研究，才能有顯著的實驗結果，得到確切的結論。

## 7. 參考文獻

- Akiba, T., Itou, K. & Fujii, A. (2004). Question Answering using "Common Sense" and Utility Maximization Principle. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 297-303). Tokyo: NII.
- Buckley, C. & Voorhees, E. M. (2004). Retrieval Evaluation with Incomplete Information. In *Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval* (pp. 25-32). New York: ACM Press.
- Buckley, C. (1991). *Trec\_eval IR Evaluation Package* [Computer software and manual]. Retrieved from <ftp://ftp.cs.cornell.edu/pub/smart>
- Bush, V. (1945). As We Will Think. *Atlantic Monthly*, 176(1), 641-649.
- Chen, K.-H. et al. (2002). Overview of CLIR Task at the Third NTCIR Workshop. In *Proceedings of the Third NTCIR Workshop Part II: Cross-Lingual Information Retrieval Task* (pp. 1-38). Tokyo: NII.
- Cleverdon, C. (1967). The Cranfield Tests on Index Language Devices. *Aslib Proceedings*, 19(6), 173-194.
- Eguchi, K. et al. (2002). Overview of Web Retrieval Task at the Third NTCIR Workshop. In *Proceeding of the third NTCIR Workshop Part VI: Web Retrieval Task* (pp. 1-24). Tokyo: NII.
- Fox, E. (1983). *Characteristics of Two New Experimental Collections in Computer and Information Science Containing Textual and Bibliographic Concepts* (Tech. Rep. TR 83-561). Ithaca, NY: Cornell University, Computing Science Department.
- Fuchigami, M., Ohnuma, H. & Ikeno, A. (2004). Oki QA System for QAC-2. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 304-309). Tokyo: NII.
- Fukumoto, J. et al. (2004). Rits-QA: List answer detection and Context task with ellipses handling. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 310-314). Tokyo: NII.
- Fukumoto, J., Kato, T. & Masui, F. (2004). Question Answering Challenge for Five ranked answers and List answers- Overview of NTCIR4 QAC2 Subtask 1 and 2. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 283-290). Tokyo: NII.
- Fukushima, T. & Okumura, M. (2002). Text Summarization Challenge 2: Text Summarization Evaluation at NTCIR Workshop. In *Proceeding of the*

- third NTCIR Workshop Part V: Text Summarization Challenge 2 (pp. 1-6). Tokyo: NII.
- Harman, D. (1993). The First Text REtrieval Conference (TREC-1). *Information Processing and Management*, 29(4), 411-414.
- Harman, D. (1996). Panel: Building and Using Test Collections. In *Proceedings of the 19th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval* (pp. 335-337). New York: ACM Press.
- Hidaka, N., Masui, F. & Tosaki, K. (2004). MAIQA: Mie Univ. Participated System at NTCIR4 QAC2. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 315-319). Tokyo: NII.
- Ismail, I. & Yukawa, T. (2004). Question Answering System using Concept-based Vector Space Model. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 320-325). Tokyo: NII.
- Isozaki, H. (2004). NTT's Question Answering System for NTCIR QAC2. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 326-332). Tokyo: NII.
- Järvelin, K. & Kekäläinen, J. (2000). IR Evaluation Methods for Retrieving Highly Relevant Documents. In *Proceedings of the 23rd Annual International Conference on Research and Development in Information Retrieval* (pp. 41-48). New York: ACM Press.
- Kando, N. (2001). Overview of the Second NTCIR Workshop. In *Proceedings of the Second NTCIR Workshop on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization* (pp. 51-72). Tokyo: NII.
- Keyes, J. G. (1996). Using Conceptual Categories of Questions to Measure Differences in Retrieval Performance. *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, 33, 238-242.
- Kishida, K., Chen, K.-H., Lee, S., Kuriyama, K., Kando, N., & Chen, H.-H. (2007). Overview of CLIR Task at the Sixth NTCIR Workshop. *Proceedings of NTCIR-6 Workshop Meeting*. May 15-18, 2007, Tokyo, Japan. 1-19
- Kurata, G., Okazaki, N. & Ishizuka, M. (2004). GDQA: Graph Driven Question Answering System - NTCIR-4 QAC2 Experiments. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 338-344). Tokyo: NII.
- Mori, T. (2004). Japanese Q/A System using A\* Search and Its Improvement - Yokohama National University at QAC2. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 345-352). Tokyo: NII.
- Na, S.-H., Kang, I.-S. & Lee, J.-H. (2004). POSTECH Question-Answering Experiments at NTCIR-4 QAC. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 361-366). Tokyo: NII.
- Nelson, M. J. (1995). The Effect of Query Characteristics on Retrieval Results in the TREC Retrieval Tests. *Proceedings of the 23rd Annual Conference of the Canadian Association for Information Science (CAIS 95)*. H.A. Olson, and D.B. Ward. Edmonton, Alberta: CAIS and the School of Library and Information Studies, University of Alberta. 156-163.
- Ramamonjisoa, D. (2004). Finding Relevant Answers in Question Answering System Contest. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 381-386). Tokyo: NII.
- Salton, G. (1972). A New Comparison between Conventional Indexing (MEDLARS) and Automatic Text Processing (SMART). *Journal of the American Society for Information Science*, 23(1), 75-84.
- Saracevic, T. & Kantor, P. (1988). A Study of Information Seeking and Retrieving. II. Users, Questions, and Effectiveness. *Journal of the American Society for Information Science*, 39(3), 177-196.
- Saracevic, T. & Baxter, M. A. (1983). On a Method for Studying the Structure and Nature of Requests in Information Retrieval. *Proceedings of the 46th Annual Meeting of the American Society for Information Science*, 20, 22-25.
- Sparck Jones, K. & van Rijsbergen, C. J. (1976). Information Retrieval Test Collections. *Journal of Documentation*, 32, 63-73.
- Takaki, T. (2004). NTT DATA Question-Answering Experiment at the NTCIR-4 QAC2. In *Proceedings of the 4<sup>th</sup> NTCIR Workshop* (pp. 402-406). Tokyo: NII.
- van Rijsbergen, C. J. (1975). *Information Retrieval*. London: Butterworth & Co.
- 王怡人 (民 93)。國立台灣大學學生使用線上百科全書之資訊尋求行為 - 以 Grolier Multimedia Encyclopedia 為例。國立臺灣大學圖書資訊學研究所碩士論文，未出版，台北市。
- 江玉婷 (民 88)。中文資訊檢索測試集設計與製作之研究。國立臺灣大學圖書資訊學研究所碩士論文，未出版，台北市。
- 陳明君 (民 88)。檢索背景與檢索問題對檢索技巧及檢索結果之影響研究。國立臺灣大學圖書資訊學研究所碩士論文，未出版，台北市。
- 陳景堂 (民 94)。統計分析：SPSS for Windows 入門與應用 (第六版)。臺北市：儒林。
- 黃怡如 (民 88)。終端使用者與系統互動前後檢索問題、檢索概念與檢索詞彙變化之研究。國立臺灣大學圖書資訊學研究所碩士論文，未出版，台北市。
- 蔡育欽 (民 94)。查詢擴展之詞彙篩選應用於主題檢索之研究。輔仁大學圖書資訊學研考所碩士論文，未出版，台北市。