國立臺灣師範大學圖書資訊學研究所

碩士學位論文

指導教授：曾 元 顯 博士

會議與期刊文獻對預測主題趨勢之比較研究 ──

以「資訊檢索」領域為例

A comparison study on conference papers and journal articles for predicting topic trends – using「Information Retrieval」as an example

研究生：許 育 聞 撰

中 華 民 國 九 十 八 年 六 月

誌謝

在即將畢業的此刻，心中五味雜陳，求學的階段在此暫告一段落，心中終於可以放下心中一塊大石頭。未來即將邁向工作階段，有期待也有不捨，校園生活的點點滴滴讓我感動在心頭，即使一路走來跌跌撞撞，但不可否認地這確實是一顆甜美又豐碩的果實，很高興自己可以在台灣師範大學取得學位，校園裡的人、事、物和一切都將讓我終生難忘。

感謝我的師傅兼指導教授元顯老師，總是不厭其煩地指導育聞，也在我最無助與困惑的時候給予許多寶貴的建議。謝謝口委兼導師的小蝶老師，從剛進研究所到畢業，總是帶給育聞許多不同面向的觀點與思考模式，豐富了我對人生看法，謝謝口委明月老師巨細靡遺地協助育聞修正論文，謝謝建成老師在馬來西亞對育聞的照顧以及給予畢業班最大的精神支持 。

謝謝同屬畢業班兼玩樂團隊的怡君、老王、佩佩瑛、湯姆，有你們的歡笑才能讓我熬過寫論文的辛苦；謝謝從凌網變麻吉的慧婷、育云、宜臻，和你們一起分享生活瑣碎的大小事，讓我備感溫馨；謝謝一起奔向馬來西亞的好夥伴戴寧及同窗好友琳禎、郁秀、輝宏等，能和你們在師大相遇、相識是育聞很珍惜的福分；謝謝曾幫助過我的同學兼學姊惠萱與佳臻，及相識超久的死黨們，小瑾、家需、菊、蝦、陳政、牛、娟、A 里等，有你們的關心與問候讓育聞加速寫論文的動力。謝謝工作上遇到的每位主管與夥伴，玲莉姊、秀珍姊、姜文、凱真、梅均、語涵等，有你們的鼓勵，才能讓育聞順利完成學業。

最後，感謝在背後默默支持我的家人，謝謝父母親多年來的養育及栽培，讓育聞幸福又無後顧之憂的完成學業，也謝謝喜歡和我拌嘴的弟弟，願意和我分享生活上的大小事並善盡後勤支援的責任，及總是調皮搗蛋的巧克力，願意每天守在門口等我回家，並時常給我一個超可愛又無辜的眼神。要感謝的人太多，但能打的字太少，千言萬語道不盡育聞對每個人的感謝 ，謝謝大家、有你們真好！

育聞 謹誌

摘要

多數進行主題趨勢探測的學者，皆採用期刊文獻作為主要的分析素材，鮮少針對其他類型的文獻進行研究，然而在學術傳播中會議論文的重要性不可小覷，因此本研究以資訊檢索（Information Retrieval）領域為範圍，針對會議文獻與期刊文獻分別進行主題趨勢預測，以觀察不同類型的文獻進行主題趨勢預測時的差異性。

本研究收集 1990 年至 2007 年資訊檢索領域具代表性的 SIGIR 會議文獻及五本核心期刊中收錄主題為「資訊檢索」的期刊文獻，五本核心期刊分別是：Information Processing & management、JASIST&JASIS、Journal of Information science、Journal of Documentation、Information Retrieval，主題歸類的部份是以主題整併和自動化歸類兩種方式進行。為了確保預測的準確性，本研究以相同文獻類型和相異文獻類型分別進行預測，以比較其預測上的準確性，最後分別改變預測集和驗證集之年代範圍以比較其差異性。

研究發現會議文獻和期刊文獻在主題詞彙的用法上有所差異，且各自有較關注探討的主題。會議文獻大部分比期刊文獻較早出現，然而在主題預測上，會議文獻並未佔有優勢，當預測的主題範圍較廣時，期刊文獻預測之效果較佳，且相同類型文獻預測效果優於交叉預測之效果。

最後提出之建議為：期刊文獻之控制詞彙尚未完善，許多單複數詞彙和縮寫詞彙尚需統整；主題預測的部份，若要瞭解較廣泛的領域趨勢，以期刊文獻預測的效果較佳，而要了解細部領域的趨勢則是以相同的文獻類型進行預測較佳；在後續研究的部份，可以針對像是專利或部落格等其他類型的灰色文獻進行研究，或是以文獻之作者群進行社會網絡分析也是一個可行的研究方向。

關鍵詞：主題趨勢預測、會議文獻、共字分析、自動化歸類、資訊檢索

Abstract

Many scholars who study topic trends use journal articles as primary texts for analysis and hardly pay attention to other types of documents. However, the importance of conference papers cannot be neglected in the academic field of Scholarly Communication. Hence, the research focusing on Informal Retrieval puts topic trends into practice in two kinds of literature, conference papers and journal articles, and observes the discrepant results of those in different types of documents.

The research collects representative researches on "Information Retrieval" in SIGIR conference papers and five core journals: *Information Processing & Management*, *JASIST&JASIS*, *Journal of Information Science*, *Journal of Documentation*, and *Information Retrieval*. The methods of categorizing documents rely on topics of journal articles given in databases, session titles of conference papers, and then the previous articles and papers in automatic categorization. In order to ensure the accuracy of prediction, and prediction is experimented in two groups, the same and the different types of literature. Then, the research changes periods of prediction and validation set to compare the results.

The research finds that conference papers and journal articles differ not only in the uses of topic vocabulary but also in the topics of their concerns. Although most conference papers publish earlier than journal articles, the latter possesses more advantages in topic prediction. When the scope of the predicted topic is wider, the predicted results of journal articles are better. The predicted results of documents from the same type also generate superior outcomes than those from the different type.

Suggestions are proposed in the end of the research. Control terms of publication papers are defective because plenty of singular/plural vocabulary and abbreviations need arranging. In the part of topic prediction, if understanding trends in wilder fields is needed, the prediction of topic trends in journal articles leads to better effects. The result of using journal articles to predict topic trends is better. To understand trends in detailed field, the prediction of topic trends in same type of documents is more effective. Finally, further studies on Information Retrieval is recommended to study other types of gray literature, such as patents or articles on blogs, or make an social network analysis on authors of documents.

Keywords：topic trends predict, conference paper, co-word analysis, automatic categorization, information retrieval

目次

摘要 ii

目次 iv

表次 vi

圖次 xii

附錄表次 xiv

第一章 緒 論 1

第一節 研究動機與背景 1
第二節 研究目的與研究問題 4
第三節 研究範圍與限制 5
第四節 名詞解釋 6

第二章 文獻探討 7

第一節 會議文獻與期刊文獻 7
第二節 主題趨勢之相關研究 17

第三章 研究方法 27

第一節 研究概念 27
第二節 研究假設 28
第三節 研究方法 28
第四節 研究流程與實驗規劃 36
第五節 研究對象與工具 47

第四章 研究結果與分析 53

第一節 文獻主題判斷與歸類 53
第二節 文獻主題與年代分佈 65
第三節 會議與期刊文獻主題趨勢預測差異 100
第四節 小結 145

第五章 結論與建議 153

第一節 結論 153

第二節 建議.....157
參考文獻.....161
附錄.....167
附錄 一 第一種和第二種主題整併下主題年代序列表.....167
附錄 二 主題萃取系統歸類結果.....182
附錄 三 系統自動萃取主題之文獻數量年代排序.....188

表次

表 2- 1 會議與期刊文獻比較表 14
表 3- 1 文獻集熱門主題計算與排列方式範例表 30
表 3- 2 SLP 範例計算之原始資料 31
表 3- 3 SLP 範例計算之結果與排序 31
表 3- 4 評鑑排序方式之範例 32
表 3- 5 兩種評鑑排序指標之計算過程 33
表 3- 6 研究目的、研究問題與研究方法對照表 35
表 3- 7 預測集年代範圍大小之分組詳細年代表 39
表 3- 8 驗證集年代範圍大小之分組詳細年代表 40
表 3- 9 兩種趨勢預測指標之實驗各組詳細起始年代 41
表 3- 10 期刊文獻樣本中各期刊所佔之篇數與比例 48
表 4- 1 各年代缺少關鍵字詞彙與 DE 詞彙的文獻分佈比率 54
表 4- 2 ID 主題詞彙次數分佈比率 54
表 4- 3 資訊檢索領域中核心期刊文獻每年主題分佈情況 55
表 4- 4 ID 主題詞彙出現 10 次以上之數量分佈 55
表 4- 5 會議文獻各年度文獻主題平均數量分佈 57
表 4- 6 SIGIR 會議文獻主題數量分佈 58
表 4- 7 期刊文獻無法對應至會議文獻之主題 60
表 4- 8 會議文獻無法對應到期刊文獻之主題表 61
表 4- 9 會議文獻與期刊文獻主題相似但詞彙不同之主題 63
表 4- 10 第一種主題整併後會議文獻與期刊文獻統整比較表 80
表 4- 11 第二種主題整併後會議文獻與期刊文獻統整比較表 89
表 4- 12 會議文獻第一種與第二種主題整併結果比較表 90
表 4- 13 期刊文獻第一種與第二種主題整併結果比較表 91
表 4- 14 自動化歸類下系統萃取 38 個類別主題文獻統整表 93
表 4- 15 本研究與先前研究之比較 96
表 4- 16 自動化歸類與先前研究相同年代比較表 97
表 4- 17 自動化歸類接續先前研究之主題對應表 98
表 4- 18 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 102
表 4- 19 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之 Levene's Test 102
表 4- 20 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之二因子變異數分析摘要表 103
表 4- 21 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相

對位差得分之各項細格平均數與邊緣平均數 103
表 4- 22 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之 Levene's Test 104
表 4- 23 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之二因子變異數分析摘要表 104
表 4- 24 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 105
表 4- 25 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之 Levene's Test 105
表 4- 26 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之二因子變異數分析摘要表 106
表 4- 27 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 106
表 4- 28 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之 Levene's Test 107
表 4- 29 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之二因子變異數分析摘要表 107
表 4- 30 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 108
表 4- 31 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之 Levene's Test 109
表 4- 32 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之二因子變異數分析摘要表 109
表 4- 33 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 110
表 4- 34 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之 Levene's Test 110
表 4- 35 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之二因子變異數分析摘要表 111
表 4- 36「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 113
表 4- 37「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 113
表 4- 38「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 114
表 4- 39「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 114
表 4- 40「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨

勢預測排序相對位差得分之 Levene's Test 115
表 4- 41「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 115
表 4- 42「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 116
表 4- 43「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 116
表 4- 44「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 117
表 4- 45「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 117
表 4- 46「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 117
表 4- 47「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 118
表 4- 48「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 119
表 4- 49「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 119
表 4- 50「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 120
表 4- 51「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 120
表 4- 52「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 121
表 4- 53「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 121
表 4- 54 不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 122
表 4- 55「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 122
表 4- 56「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 123
表 4- 57「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 123
表 4- 58「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 124
表 4- 59「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨

勢預測排序相對位差得分之各項細格平均數與邊緣平均數 126
表 4- 60「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 126
表 4- 61「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 127
表 4- 62「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 127
表 4- 63「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 128
表 4- 64「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 128
表 4- 65「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 129
表 4- 66「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 129
表 4- 67「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 130
表 4- 68「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 130
表 4- 69「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 131
表 4- 70「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 131
表 4- 71「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 132
表 4- 72「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 132
表 4- 73「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 133
表 4- 74「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 133
表 4- 75「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 133
表 4- 76「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 134
表 4- 77「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 135
表 4- 78「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨

勢預測排序相對位差得分之 Levene's Test 135
表 4- 79「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 136
表 4- 80「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 136
表 4- 81「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test 136
表 4- 82「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 137
表 4- 83「不同預測指標」與「各年度區間」在會議預測會議主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 139
表 4- 84「不同預測指標」與「各年度區間」在會議預測會議主題趨勢預測排序相對位差得分之 Levene's Test 139
表 4- 85「不同預測指標」與「各年度區間」在會議預測會議主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 140
表 4- 86「不同預測指標」與「各年度區間」在期刊預測會議主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 140
表 4- 87「不同預測指標」與「各年度區間」在期刊預測會議主題趨勢預測排序相對位差得分之 Levene's Test 141
表 4- 88「不同預測指標」與「各年度區間」在期刊預測會議主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 141
表 4- 89「不同預測指標」與「各年度區間」在會議預測期刊主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 142
表 4- 90「不同預測指標」與「各年度區間」在會議預測期刊主題趨勢預測排序相對位差得分之 Levene's Test 142
表 4- 91「不同預測指標」與「各年度區間」在會議預測期刊主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 143
表 4- 92「不同預測指標」與「各年度區間」在期刊預測期刊主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數 143
表 4- 93「不同預測指標」與「各年度區間」在期刊預測期刊主題趨勢預測排序相對位差得分之 Levene's Test 144
表 4- 94「不同預測指標」與「各年度區間」在期刊預測期刊主題趨勢預測排序相對位差得分之二因子變異數分析摘要表 144
表 4- 95 以 API 為預測指標之各項假設檢定之綜合整理 146
表 4- 96 以 SLP 為預測指標之各項假設檢定之綜合整理 148
表 4- 97 兩種不同預測指標之假設檢定統整 149
表 4- 98 採用兩種預測指標下的各次預測結果統計表 149
表 4- 99 交叉預測之情形依準確度大小排序 150

表 4- 100 依驗證集為區分對象之交叉預測情形 151

圖次

圖 2- 1 資訊循環圖 8
圖 2- 2 策略座標圖 23
圖 3- 1 研究架構圖 27
圖 3- 2 研究流程圖 36
圖 3- 3「預測集」與「驗證集」預測橫跨年代示意圖 37
圖 3- 4 實驗流程圖 42
圖 3- 5 第一種主題整併 API 為趨勢指標下 1995~2004 年 CC/JC 主題趨勢預測 42
圖 3- 6 期刊主題整理流程圖 43
圖 3- 7 判斷會議文獻重新分類流程圖 44
圖 3- 8 各階段歸類之原則與流程 46
圖 3- 9 樣本文獻中各本期刊所佔之比例圓餅圖 48
圖 3- 10 SIGIR 會議文獻 1990~2007 年各年度文獻篇數 49
圖 4- 1 主題系統萃取歸類期刊文獻之範例 64
圖 4- 2 會議文獻第一種主題整併後 18 個主題整體趨勢圖 Ⅰ 66
圖 4- 3 會議文獻第一種主題整併後 18 個主題整體趨勢圖 Ⅱ 66
圖 4- 4 期刊文獻第一種主題整併後 18 個主題整體趨勢圖 Ⅰ 67
圖 4- 5 期刊文獻第一種主題整併後 18 個主題整體趨勢圖 Ⅱ 67
圖 4- 6 以 Web 為主題的會議和期刊文獻各年度數量折線圖 68
圖 4- 7 以 Model(S)為主題的會議和期刊文獻各年度數量折線圖 69
圖 4- 8 以 Text Categoris(z)ation 為主題的會議和期刊文獻各年度數量折線圖 .. 69
圖 4- 9 以 User 為主題的會議和期刊文獻各年度數量折線圖 70
圖 4- 10 以 Interfaces 為主題的會議和期刊文獻各年度數量折線圖 71
圖 4- 11 以 Evaluation 為主題的會議和期刊文獻各年度數量折線圖 72
圖 4- 12 以 Efficiency 為主題的會議和期刊文獻各年度數量折線圖 72
圖 4- 13 以 Queries 為主題的會議和期刊文獻各年度數量折線圖 73
圖 4- 14 以 Query Expansion 為主題的會議和期刊文獻各年度數量折線圖 73
圖 4- 15 以 Language 為主題的會議和期刊文獻各年度數量折線圖 74
圖 4- 16 以 Relevance Feedback 為主題的會議和期刊文獻各年度數量折線圖 ... 74
圖 4- 17 以 Routing and filtering 為主題的會議和期刊文獻各年度數量折線圖 ... 75
圖 4- 18 以 Fusion 為主題的會議和期刊文獻各年度數量折線圖 75
圖 4- 19 以 Distributed 為主題的會議和期刊文獻各年度數量折線圖 76
圖 4- 20 以 Cross language 為主題的會議和期刊文獻各年度數量折線圖 77
圖 4- 21 以 Cluster 為主題的會議和期刊文獻各年度數量折線圖 77
圖 4- 22 以 Theory 為主題的會議和期刊文獻各年度數量折線圖 78
圖 4- 23 以 Multi-lingual IR 為主題的會議和期刊文獻各年度數量折線圖 78
圖 4- 24 會議文獻第二種主題整併後 11 個主題整體趨勢圖 82

圖 4- 25 期刊文獻第二種主題整併後 11 個主題整體趨勢圖 83
圖 4- 26 Web；Hypertext；Link analysis 的會議和期刊文獻各年度數量折線圖 83
圖 4- 27 Semantics；LSI；Probabilistic Models；(IR) Theory；Logic；Inference Networks 的會議和期刊文獻各年度數量折線圖 84
圖 4- 28 Classification&Categorisation(Classifiers) ；Clustering；Filtering；Fusion 的會議和期刊文獻各年度數量折線圖 84
圖 4- 29 Performance Evaluation；Test collections 的會議和期刊文獻各年度數量折線圖 85
圖 4- 30 Efficiency；Scalability 的會議和期刊文獻各年度數量折線圖 85
圖 4- 31 Indexing(structures)；Term Relationships；Statistics；Question and answering 的會議和期刊文獻各年度數量折線圖 86
圖 4- 32 Natural Language Processing；Cross language；MIR 的會議和期刊文獻各年度數量折線圖 86
圖 4- 33 User behavior；Human ；Interface Design and Display；Visualization 的會議和期刊文獻各年度數量折線圖 87
圖 4- 34 Relevance Feedback；Query；Query Expansion 的會議和期刊文獻各年度數量折線圖 87
圖 4- 35 Distributed(IR)(Systems) 的會議和期刊文獻各年度數量折線圖 88
圖 4- 36 DB&IR(Integration)(Structures) ；Systems；Knowledge Based Information Retrieval 的會議和期刊文獻各年度數量折線圖 88
圖 4- 37 主題萃取系統歸類後結果判讀示意圖 92

附錄表次

表 1 全部可對應之主題類別年代序列 167
表 2 補足對應之主題類別 174
表 3 第一種主題整併之文獻年代序列 176
表 4 第二種主題整併年代序列 180

第一章　緒 論

第一節　研究動機與背景

在學術傳播的過程中，通常可將資料分為一次資料(primary sources)、二次資料(secondary sources)、三次資料(tertiary sources)，科技研究者通常最重視一次資料，因為一次資料可以傳遞最新的研究消息與研究方向，並且常以研究日誌、書信、電話、簡訊、專利、會議、研究報告或期刊等方式進行學術交流(蔡明月，1997)。會議是學術傳播中重要的管道之一，且會議兼具正式與非正式的傳播形式，研究者往往得以從會議中先行發表自己的研究成果，並可於會議中和與會者進行交談，一起分享和討論研究成果，因此會議的舉辦對研究領域而言相當重要。

從學術會議中所產生的會議文獻，可以看出該領域的最新研究方向，會議文獻之所以可以備受科學家的重視是因具有下列特色 (Subramanyam, 1981)：

一、快速深入新研究和新發展的領域；

二、回顧過去的研究、現在的技術、展望新技術的研究方向；

三、透視科學理論，將其應用至新領域；

四、經由邀請的專家，可快速得到問題的現況和解決的趨勢；

五、由新的研究發現和創意的想法，融會貫通到新的技術領域；

六、蒐集目前工業所存在的應用技術，提出具體的研究；

七、將專家和設備重新組合，對進行中的研究實施模擬和修正；

八、整合技術、設備和理論的應用；

九、可瞭解目前研究情形，避免重複的研究和錯誤。

對於研究者而言，想要了解該領域的研究趨勢與動向，會議文獻將會是一個很好的研究素材。然而會議文獻具有灰色文獻的特性，蔡佳玲(1990)將灰色文獻定義為：「灰色文獻是指發行數量特定且缺乏書目控制，無法經由一般管道取得的公開文獻或電子資訊」，會議文獻因會議的舉行而印製，且大多缺乏書目控制，因

此許多國家皆把會議文獻列為灰色文獻的範疇，像是英國、美國、日本等，再加上會議文獻並沒有特定的出版形式，有的刊載在學會、協會的期刊上，作為專刊、特輯或增刊，有些則發表在專門刊載會議紀錄或會議論文摘要的期刊上(中國大百科全書智慧藏)。由此可知，會議文獻的多樣性更是讓使用者難以取得 。

期刊文獻係指以定期或不定期連續出版的出版品之一，在我國則習慣將雜誌(magazine)或期刊(journal)分別定位在不同屬性的定期刊物，雜誌泛指一般大眾通俗性定期刊物；而期刊之內容則多屬於學術與專業性論文為主體的定期刊物(邱炯友，2006)。相較之下期刊文獻有特定的出版管道與形式，使用者可以透過期刊出版或資料庫查詢等方式取得期刊資源，像是引文索引資料庫中即收錄大量的期刊文獻與相關的引用索引資料，可以讓研究者自行下載所需文獻並進行分析，對專門研究文獻分析的學者而言是一個相當方便的研究工具，因此許多專門研究文獻的學者多會以該資料庫作為研究工具下載期刊文獻。蔡明月與劉瓊芳(2007)針對 1992 至 2005 年資訊計量學領域的發展與演變研究中指出，資訊計量學領域的研究對象多集中在期刊文獻，採用會議文獻作為研究樣本所進行的研究數量就明顯少很多。然而書目計量學的引用研究卻提出，學術界引用或參考灰色文獻的情形有明顯增加的趨勢(Perry, 1996)，且在獸醫學領域引用灰色文獻的書目計量學研究中亦發現，該領域所引用的灰色文獻中有 50%即是屬於會議文獻(Pelzer, 2003)。也就是說，會議文獻的重要性已不可同日而語，然而書目計量學的相關研究卻少以會議文獻作為研究對象，因而引發本研究更進一步思考，在學術傳播的過程中，尚有許多其他學術產物可瞭解該學科領域的知識架構，如會議文獻、技術報告、專利文獻等，並非只有期刊文獻可以做為進行研究趨勢的探測。

此外，會議所發表的文獻平均而言大多比期刊早約一年至兩年，因此本研究推測會議文獻在預測該領域的研究趨勢時，與期刊文獻所得出的結果相較之下應更具有說服力，是故本研究採實證之方法，以資訊檢索領域為研究範疇，提出相關數據展現會議文獻與期刊文獻在預測該領域的研究趨勢時，兩者之間的相對性效果。

本研究以資訊檢索做為研究範疇，主要因當前資訊科技的快速發展與網際網路的普及下，資訊檢索領域直到今日仍有大量的文獻產出，也有許多資訊檢索系統成功發展完成，由此可見該領域已成為一門成熟的學科領域，值得我們進一步去探究。此外，資訊檢索領域也有許多學者已針對相關文獻進行書目計量學分析，如新加坡的三位學者已於 1999 年找出該領域 1987 至 1997 年最高被引用的 39 位作者，並以作者雙被引的分析方式繪製資訊檢索領域的知識架構，並將資訊檢索研究主題劃分成六個主題，分別是「資訊檢索基本理論與技術」、「資訊檢索模式」、「資訊檢索系統技術」,「資訊尋求與檢索行為」、「電腦化資訊檢索系統及數學模式」和「資訊檢索通則」，從中也發現在資訊檢索主題下，隨著科技技術的進步與學科領域的衍生，形成更多樣的研究主題(Ding, Chowdhury & Foo, 1999a, 1999b)。由此可知，選用資訊檢索領域為研究範疇，不但可以對照許多先前的相關研究，也可以更進一步了解資訊檢索領域近年來的發展與研究趨勢，因此本研究將選擇資訊檢索領域作為研究範疇。

綜合上述所言，本研究以實證之方式比較資訊檢索領域中，會議文獻與期刊文獻應用於趨勢研究時的相對性效果。首先分別收集資訊檢索領域裡，關注於資訊檢索議題之計算機組織社群(Association for Computing Machinery's Special Interest Group on Information Retrieval，簡稱 ACM SIGIR)的會議論文以及下載自 Web of Science 資料庫的五本核心期刊文獻，分別是：Journal of Information Science、Information Processing & Management、Journal of the American society for Information Science and Technology，簡稱 JASIST(其前身為 Journal of the American Society for Information Science)、Journal of Documentation、Information Retrieval。探討該領域會議文獻與期刊文獻主題趨勢，並分別比較兩者對於預測研究趨勢上相對性的效果，以數據的方式檢驗會議文獻是否在預測研究趨勢時有較好的表現。此外本研究也針對預測主題趨勢時，採用的文獻集年代範圍與預測指標間是否會有差異存在加以探討，作為未來圖書館及資料庫廠商收錄，以為文獻後續進行文獻主題趨勢預測研究的參考依據。

第二節　研究目的與研究問題

根據本研究之宗旨，具體而言本研究的研究目的有以下四點：

一、瞭解資訊檢索領域，核心會議文獻與期刊文獻的主題分佈與研究趨勢變化。

二、比較會議與期刊文獻進行預測研究趨勢的差異。

三、瞭解會議與期刊文獻在預測研究趨勢上的穩定性。

四、瞭解趨勢預測指標對預測研究趨勢的影響情形。

五、綜合研究結果，提供圖書館與其他資料庫廠商未來收集主題資源、後續進行文獻主題趨勢預測時的參考方向。

根據上述研究目的，本研究所欲了解的問題如下：

一、資訊檢索領域的主要研究議題有哪些？各年代之間的主題變化為何？

二、預測研究趨勢時，採用會議論文和期刊文獻作為研究樣本之相對性效果為何？會議文獻是否比期刊文獻較為準確？

三、會議文獻與期刊文獻對研究趨勢預測的情形是否穩定？是否隨著選定預測範圍的年代不同而有差異？

四、不同的趨勢預測指標應用於預測研究趨勢時有何變化？

第三節　研究範圍與限制

本研究旨在比較會議與期刊文獻對預測主題趨勢上的相對性效果，由於主題趨勢預測上多以期刊文獻為主，故本研究探討以會議文獻進行主題趨勢預測之結果，其準確度是否優於期刊文獻預測之結果，故以「資訊檢索」領域作為本研究之主題範圍。

在年代範圍的限制上，因期刊文獻受限於人工主題整併時需採用資料庫中的關鍵字詞彙(KeyWords Plus®)[1]，而關鍵字詞彙受限於 Web of Science 資料庫由 1991 年起開始收錄，故人工主題整併之研究樣本年代受限於 1991 年至 2007 年；而自動化歸類則是採用 1990 年至 2007 年的期刊文獻。會議文獻的年代範圍則是與期刊文獻同步，故人工主題整併之年代範圍同樣為 1991 年至 2007 年；自動化歸類則為 1990 年至 2007 年。

在資料收集的限制上，會議文獻僅包括正式收錄於會議論文集中的論文，且發表於 SIGIR 會議上之會議文獻，不包括會議進行中的其他相關資料，亦不包含以海報(poster)發表等其他論文；期刊文獻的收集因受限於資料庫的取樣便利性，故研究樣本僅止於收錄在 Web of Science 資料庫中，且由於期刊文獻的研究樣本數量過大，故進一步將期刊文獻的範圍限制在資訊檢索領域的核心期刊作為本研究期刊文獻樣本的範圍。

本研究所採用的會議與期刊文獻因本身性質上即有差異，且各自有偏重探討之議題，在進行主題對應上會產生部份無法對應之情形，故本研究在探討資訊檢索領域之主題時，僅以可對應之主題進行探討。在主題趨勢預測的部份，除了盡可能將主題詞彙進行統整，也另搭配主題萃取系統進行自動化萃取，以排除主題在對應上需靠人工統整的主觀性差異，然因人工主題整併後之主題數量仍偏多，故僅針對文獻數量多於 10 篇之主題進行後續實驗，而自動化歸類之主題則不再以文獻數量為門檻。

[1] KeyWords Plus® 是由 Thomson Reuters 創建的索引詞，是取自文獻所引用的參考文獻標題中，頻繁出現的重要詞語。

第四節　名詞解釋

一、會議文獻(Conference Literature)

本研究所指的會議文獻為發表於 ACM SIGIR 研討會上，正式收錄於會議論文集中的文獻資料，不包含研討會中以海報或其他方式發表之論文，或論文集中的前言、序論等文獻。

二、熱門主題(Hot topic)

本研究所稱之熱門主題是以驗證集的文獻數量加總後，再依文獻總數多寡排序作為本研究的熱門主題排序。

三、趨勢預測指標(Trend Indices)

本研究所採用之趨勢預測指標共有兩種，其一為平均文獻成長率(average percentage of increase，簡稱 API)，其二是年代序列上線性迴歸的最佳曲線(slope of the linear regression line，簡稱 SLP)。其作用皆為計算預測集熱門排序之依據，亦即以這兩種指標分別計算針對預測集主題文獻之趨勢預測值，再依照計算值的大小排序，作為預測集的熱門主題排序。

四、趨勢預測評鑑(Trend Detect Evaluation)

本研究所指的趨勢預測評鑑是用於評估預測之準確性，所採用的評鑑方式為 AR(All Ranking)，亦即將所有排列之主題皆做為主題比對的依據，也就是說，本研究計算驗證集與預測集兩種主題排序之間的相對位差，作為評估預測準確性之得分。

第二章　文獻探討

本研究旨在探討會議與期刊文獻對於主題趨勢預測的相對性效果。本章將分為三節，第一節簡述比較會議與期刊文獻的特性；第二節則進一步探討主題趨勢預測相關方法與研究。

第一節　會議文獻與期刊文獻

會議文獻與期刊文獻是本節所探討的重點，以下先簡介兩者在學術傳播上的角色，再分別探討兩種文獻的特性、起源與相關研究，最後比較兩者之間的差異。

一、　會議文獻與期刊文獻在學術傳播中所扮演的角色

Borgman and Furner(1990)將學術傳播定義為所有學門領域的學者透過正式與非正式管道來使用與傳播資訊的過程。在傳播過程中所產生的資料，依照其資料類型又可區分為一次資料、二次資料及三次資料。Subramanyam(1981)指出透過追蹤科學循環的過程可以瞭解科學文獻的演變，於是提出科學資訊進展的概要圖，透過此概要圖可以看清楚所有文獻在資訊傳播中所扮演的角色 。

由圖 2-1 可知，研究發展幾乎都由同儕間非正式的溝通開始，而會議文獻則是所有正式的論文、技術報告、學會、專書等不同媒體之前的預印本，由會議所衍生出的三種出版品分別是：1.由作者或主辦單位所印製的預印本；2.會議議程中所包含的文件與討論的摘要及相關資訊；3.會議結束後發行的重印本。同樣地期刊文獻也會衍生出類似於會議的三種出版品。會議文獻與期刊文獻都同屬於一次資料，也同屬於正式傳播的管道，兩者之間在性質上十分相似，但仍有些許差異之處，將於後續內容進行探討。

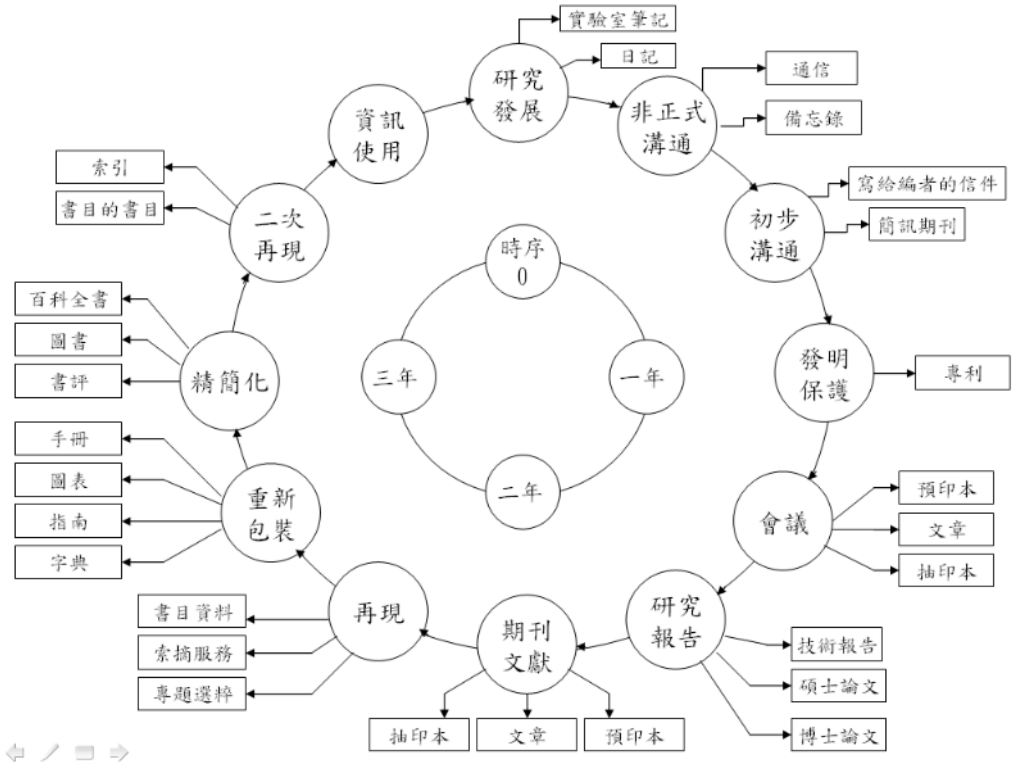

圖 2- 1 資訊循環圖

資料來源：K. Subramamyam, *Scientific and Technical Information Resources* (New York: Marcel Dekker, 1981)p.5

科學傳播是資訊科學和知識組織的研究範疇之一(Hjerland, 2003)，且資訊科學家比較偏好出版於學術圈的期刊文獻(Drott, 1995)，因此有學者認為會議文獻所具備的角色應該是期刊文獻的補充說明，然而後續並沒有很多的研究者認同此論點，反而是比較多研究者接著探討會議文獻和期刊文獻的相對性研究(Montesi & Owen, 2008)。由於網路科技的發達，使得學術傳播的模式也已經改變，在 XML 領域的學者已習慣先將研究成果發佈於網路上，再與同領域的研究人員分享成果，最後才是發表於期刊或會議上(Zhao & Strotmann, 2007)。由此可知，在科學傳播中，人們的習慣早已隨著科技的發達而有所改變，且會議文獻的重要性已不僅是期刊文獻的補充而已。

(一) 會議文獻本質與特性

會議文獻的產生主要是起源於會議的舉辦，會議除了提供研究者公開發表的

場所，更能藉此機會讓學者與其他研究者彼此交流、交換意見，以迅速取得直接的回饋。根據英國國家標準 BS 4446 解釋「會議」(Conference)是指「任何一型的集會，於集會中發表論文，討論某一專題，或針對單一主題進行廣泛的討論」(轉引自蔡明月、鄭琚媛，2004)。會議的大小規模各有不同，小型會議通常是針對部分章節或特殊興趣的團體集會，彼此討論新的知識，具有教學性質；大型會議則有數千名來自世界各國的代表出席，為了回顧過去的發展，獲得現階段的知識，及指出未來的研究目標。會議文獻的主要特色為專深、新穎、簡化，學術會議主要針對單一主題或多個議題進行探討，參與者大多為該領域的專家或是正在從事研究的人員，在科技會議上，會議論文所闡述的觀點是最新的研究成果，且會議文獻的發行時間一般比期刊早約一年(郭華，2006)。

會議文獻在內容形式上大致可分為兩種：會議論文集和會議實錄。預定於會議中發表的文章摘要，通常會預先發佈給可能的參與者或刊出在期刊當中，除能告知參與者研討會大概的內容，也能讓無法前來參加的人對活動有一些大致的了解。有時作者也會先將文章的複本散佈出去，希望參與者能夠事先閱讀，以便能有更多的時間在會議上進行討論。會議進行時，參與者通常可以取得文章的摘要或預印的文章；在會議結束後，出版的文獻通常包括完整的會議紀錄及研討會中個人發表的文章(Subramanyam, 1981)。

(二)　期刊文獻本質與特性

期刊的概念源自於英國皇家學會主席 Robert Moray 爵士，他認為以個人通信作為科學資訊交流媒介有諸多缺點，因而提出此概念，其後法國在 1665 年 1 月 5 日正式創立了第一本期刊－學者雜誌(scavans)。期刊的興起隨著高速印刷技術的進步、科學與技術的專門化、科學人員與技術人員積極的研究與出版等發展而來，期刊的暴增也帶來許多困擾，並減緩了資訊流通的速度，詳盡的編審程序導致研究文獻從投稿至出版經過一年至兩年左右的延遲，因此迫使某些需要快速新資訊的研究人員轉向其他傳播管道，例如研討會等。

Subramanyam(1981)指出當科學家或圖書館員選擇以期刊作為學術傳播的管道時，會面臨到的缺點有五項：1.出版上的延遲。2.期刊的暴增且相關主題被分散至各種期刊上，選擇的相關期刊變多，科學人員很難再去掌握最新發展；圖書館或從事二次資料收集的人員僅能針對小範圍的資料收錄。3.連續出版品的增加讓科學人員感到緊張，且導致期刊可以提供給研究人員的幫助越來越小。4.期刊成本的增加，影響到出版商、圖書館、讀者，作者也受到影響，因其必須負擔一部分出版成本。5.時間上的耗費，消耗科學人員的時間和精力，導致研究人員寧願轉往其他管道。由此可知，期刊的暴增帶給出版商、讀者、圖書館員、科學人員等許多困擾，圖書館在徵選期刊時，面對眾多期刊應如何選擇成了當前主要的課題，核心期刊即是能解決此種困擾的方式之一。

(三) 核心期刊(Core journal)

核心期刊的概念起源於布萊德福定律(Bradford's Law)，是布萊德福於 1934 年觀察 300 種索引與摘要的文獻分佈所提出。他為特定領域或主題中大部分的文獻會集中於少部分的期刊，而少部分的文獻會集中於大部分的期刊。也就是說，如果將期刊按照文獻數量多寡進行排列時，可以找出該學科或主題中最有貢獻的核心區，以及含有數量與之相等的其他幾個區域，這時核心區與其他相鄰的區域，所屬的期刊數量會呈現 $1：n：n^2$ 的情況(昌炎新，2006)。核心期刊的認定可以作為圖書館在評選期刊時的參考依據，同時也可以做為讀者在查找期刊時的一個便利的管道。觀察特定領域的文獻分佈情形，進而界定核心期刊相關的實證研究，在國內有蔡明月(1989)針對「索引摘要」及「電腦在化學分析」兩大主題分別進行實證研究，其後也陸續有許文華(2001)針對基因工程學等領域；劉瓊芳(2006)針對資訊計量學；林秀惠(2006)針對資訊科學，驗證布萊德福定律並找出核心期刊與文獻分佈情形。

Ding et al.(1999)針對資訊檢索領域的期刊文獻進行期刊共被引分析(co-citation analysis)，利用 SSCI資料庫下載1987-1997年間資訊檢索領域的期刊

文獻，找出在該領域中被引用最多次的期刊。其研究發現在資訊檢索領域中的期刊引用與被引用情形很穩定，並提出六種在資訊檢索領域中核心期刊的核心，分別是：Journal of the American Society for Information Science(JASIS) 、Information Processing & Management 、 Journal of Documentation 、 Annual Review of Information Science and Technology、Journal of Information Science、Proceedings of the ASIS Annual Meeting。

二、 會議文獻相關研究

會議文獻是具有快速發表且深具學術內涵的文獻。會議文獻所報告的是目前最流行的資訊，其相關出版品可以成為發展中科學的權威性調查報告及參考資料，也是受研究人員或學者重視的資料來源。但會議文獻因書目控制不易、書目不完整等因素，使其成為灰色文獻。Subramanyam(1981)提出會議論文集有五項缺點分別是：1.出版上的延遲。2.文獻重複性高，通常會出版在很多地方讓與會者方便取得。3.參考書目不完整：許多會議文獻沒有提供完整的參考書目 。4.審查過程不嚴謹，許多會議文獻或投稿並未經過合格的審查 。5.資料來源難以認定，很難認定所檢索到資訊是出自於會議論文集中 。

會議文獻具有灰色文獻的特質，造成研究人員、圖書館員、讀者的取用不便，也因而產生許多值得探討的議題，如多語言問題、索引和摘要品質不均和延續出版等。會議文獻的取用不易，再加上資料庫的快速發展，且資料庫又多以收錄期刊文獻為主，讓主題趨勢預測的研究多以期刊文獻做為主要的分析對象，而忽略會議文獻也可以作為主題趨勢研究的樣本，因此本研究延續這樣的想法，進行會議與期刊文獻對預測主題趨勢的比較研究。

(一) 會議文獻的延續出版研究

針對會議文獻進行研究的議題，早年多集中於其特質上的探討，進而找出會議文獻潛藏的問題所在，其後的相關學者則針對會議文獻的延續出版等議題多做

探討。1958 年 Felix Liebesny 的研究中發現會議論文集中約有 48.5%的論文沒有延續性出版，約有 51%的會議文獻進行延續出版，且其中的 1/3 以期刊型式出版，其他則僅以摘要方式出版，而延續出版中，多在一至兩年後才出版(轉引自黃淑娟、蔣嘉寧、黃擎天與宋雪芳，1998)。

Garvey(1972)調查 1966 年至 1971 年間九個科技和社會科學的會議，針對會議延續出版品研究發現參與會議之後有 50%的論文發表者，會在 1 至 2 年內進行會後延續性出版，而在 2 年或 2 年後以期刊型式出版(轉引自傅雅秀，1996)；50%則表示在會議中所獲得的回饋對其會後出版相當有幫助 。次年，Garvey 等在接續研究中，將參與會議者分成論文發表者、出席討論會與索取論文者，以問卷的方式進行會議文獻在同儕間資訊同化的調查。結果顯示，會議上所發表的論文有半數是剛完成於六個月內的研究，另有半數是在兩年內開始的研究工作，此結果印證會議文獻是研究人員收集最新研究趨勢的重要管道 。

Subramanyam and Schatter(1981)的研究報告則指出，隨著不同學科，會議文獻延續出版的比例約在 30%至 70%之間。Allen(1995)以普渡大學圖書館的物理館藏進行研究，結果發現 32%的期刊包括會議文獻，總頁數佔 28%。Drott(1995)以 1987 年美國資訊科學學會(American Society for Information Science，簡稱 ASIS)的年會論文和同年 ASIS 的期刊為研究對象，結果發現會議文獻只有 13%延續出版於期刊中發表，但是期刊文獻卻有 33%為會議文獻的延續出版。

而針對簡訊或摘要類型的延續出版研究則有 Athreton(1961)以「物理評論簡訊」(Physical Review Letters)的作者進行問卷調查，了解其延續出版的情況，結果有 39%會在兩年內發表於簡訊的論文，又以正式論文型式出版者中，有 4%以會議文獻型式出現。Keane and Ronayne(1972)則以「化學期刊」(Chemical Abstract)的兩種期刊簡訊 2 至 3 年後被 SCI 索引的情形進行研究，其基本假設是以期刊簡訊上的第一作者於延續出版後，會引用自己在簡訊上的文獻。研究結果發現，再出版的比率分別是 20%和 29%。

(二) SIGIR 會議

在資訊檢索領域中，SIGIR 會議為資訊檢索領域中具有代表性的社群之一，ACM SIGIR 全名為 Association for Computing Machinery's Special Interest Group on Information Retrieval，是一個專門以資訊檢索為研究興趣的協會組織。SIGIR 關注於所有關於資訊儲存、檢索和散佈的面向，包括研究策略、產出架構和系統評鑑等。

SIGIR 會議起源於 1971 年，Jack Minker 和 Sam Rosenfeld 組織 ACM SIGIR 資訊儲存與檢索研討會，第一個正式的 SIGIR 會議由 James Iverson 主辦，於 1978 年在美國紐約羅徹斯特(Rochester, New York)舉行[2]。針對 SIGIR 會議進行研究的學者有 Smeaton, Keogh, Gurrin, McDonald, and Sødring (2002)利用 25 年來發表於 SIGIR 會議的 853 篇文獻進行內容分析。

三、 會議與期刊文獻的比較研究

會議文獻與期刊文獻在資訊循環中所扮演的角色有所不同，因此兩者所重視的面向、文獻特性、出版形式等也會有所差異，會議重視最新的研究以及學術交流，主題通常比期刊更為新穎。會議大部分以口頭報告為主，期刊則注重嚴謹的文獻結構；期刊的同儕審查(peer-review)制度亦較會議文獻嚴謹。Gardner(1980)和 Montesi & Owen(2008)分別比較期刊文獻與會議文獻之間的差異，歸納如表 2-1。Montesi & Owen 以會議文獻的延續出版至期刊的過程進行研究，該研究訪問 22 位編輯與助理編輯，並利用網路發放問卷給 122 位將會議文獻出版成期刊文獻的作者填寫，由表中可知，除了可以很清楚的看出會議文獻與期刊文獻在時效性、完整性上有較大的不同以外，其閱讀的對象和所扮演的角色也有一些差異。

[2]上網日期：2007/06/03，檢索自 http://www.sigir2007.org/history.html

表 2- 1 會議與期刊文獻比較表

文獻類型 / 特性比較	會議文獻	期刊文獻
目的	以傳遞進行中或最新研究的成果為主，重點在於發佈新知。快速更新的作品，供作者展現其想法和概念。	具有完整的研究成果，通常是會議文獻的延續出版品，注重推翻舊資料、修改理論、提供外延的文獻。
出版速度與對象	出版速度很快。 為特定或專門人士所準備。	出版過程通常很久。 為大眾所準備，可作為學生的作業基礎。
表現方式與完整度	通常只刊載發現、結論和影響，有些甚至只出版問題的詢答和有價值的答覆。精簡且具有投稿的本質。	具有完整的理論、實驗資料、結論和引文，表現方式較為嚴謹。較長且主題完整。
內容與特性	題材比較新穎和創新。 協助領域中其他研究者可以跟上腳步。	承接先前研究者的想法並完整瞭解先前研究的過程與成果。 類似檔案形式，且想要成為某時點的最終作品
同儕審查制度	審查制度較為鬆散，甚至以邀稿的方式進行特殊主題的探討。	純粹是個人科學發現優先權的主動動機，甚至發表時須符合該期刊所預定的研究主題，並接受審查制度。
引文特性	引用文獻較為新穎，而且會有特別的引用文獻出現。	出版需經一至兩年的延誤，引用文獻較老，但屬於較正式的資訊載體。
延續發展	作品有可能會進一步發展成後續文獻	作品注重完整的敘述，且文獻具有自我完備的特性(self-contained)

資料來源：本研究整理

會議文獻與期刊文獻雖然在許多面向上都有所不同，但並非所有領域皆是如此。在不同的學門中，有不同的特殊發表管道，像是自然科學界經常將研究直接發表於期刊上；應用科學界則將研究成果發表於會議上；醫學、食物製造、農業和能源科學，並以國際型式的會議方式進行(黃秀琴，1998)。社會科學則傾向全文式的發表；科學與工程類常以非描述性的文字表達，如使用大量的圖表或口語式的敘述，因此，對非與會的資料使用者而言，會議文獻對其研究幫助較有限(蔡

明月、鄭琚媛，2004)。

會議文獻對某些特定領域的研究者來說，則是很重要性的。傅雅秀曾於 1996 年針對生命科學專家進行研究調查指出，該領域 33%的研究員每年參加國內二個以上的會議，51.4%的研究員經常由會議報告中獲得資訊者，且 44%的研究人員認為在會議中間休息的交談中可以得到幫助。蔡明月等於 2000 年觀察半導體文獻進行研究時亦提出，會議文獻尤在新興領域中越來越受到重視。此外該研究亦發現在 INSPEC 資料庫裡，會議文獻與期刊文獻的數量總和已達全部文獻的 99.67%，其中期刊文獻佔 66.85%；會議文獻佔 32.82%，顯示出學術傳播過程中，會議文獻已日趨受重視(Tsay, Jou, & Ma, 2000)。

在會議與期刊文獻的比較研究中可發現，兩者在引用文獻的比較上，會議的引用文獻數量均少於期刊文獻。Zhao and Strotmann 於 2007 年的研究發現 XML 領域中發現，從 CiteSeer 下載的會議文獻，其參考文獻明顯少於 SCI 所下載的文獻 20%。Cai and Card 於 2008 年發現軟體工程(software engineering)領域會議文獻的引用文獻數量明顯少於期刊文獻約三分之一。而會議的引用文獻數量少於期刊文獻其原因的探究上，Zhao and Strotmann(2007)推測是因為期刊文獻通常會仔細地建構論點，因此需要大量的參考文獻，此外會議文獻通常在長度上也會受到會議主辦單位的限制。

在會議與期刊文獻的被引用情形可看出，大多的學者引用文獻時會以期刊文獻為主。Funk(1988)於研究中發現，以心臟血管為主題的會議和期刊文獻的引用情形發現，會議文獻和期刊文獻的影響係數分別為 0.98 及 3.86，期刊被引用次數約是會議文獻被引次數的四倍之多。Drott(1995)則指出資訊科學家在引用文獻的題材選擇上偏好期刊文獻，其結果顯示會議文獻只有 10%被引用，而該領域的期刊文獻卻高達 67%被引用。究其原因，Funk 曾以五種影響因素的差異顯著性考驗下，該研究仍無法找出確切的原因可解釋會議比期刊文獻較少被引用的情形。

由以上的文獻可知，若以會議與期刊引用或被引用的文獻數量來比較其重要

性或實用性，則會忽略新穎性、取用的便利性、時間上的落差等其他因素對兩者之間的影響。此外，圖書館對於灰色文獻的重視也應日趨增加，期刊文獻並非學術傳播上的唯一管道，專利、會議文獻、技術報告等對研究人員等皆是重要的文獻，因此針對這些文獻加以保存的重要性應受到重視。

而在會議與期刊文獻的主題比較研究上，Cai and Card(2007)發現期刊文獻有73%被集中歸類在 ACM 所擬定的前三大類主題；而會議文獻中則有 89%被歸類於前三大之相同主題，顯示出此研究中會議文獻的主題較期刊文獻集中。Funk(1988)則發現影響係數的高低並不會造成會議文獻的熱門主題分佈不均，亦即熱門主題不論在影響係數高或低的會議文獻中均可被看見。

第二節　主題趨勢之相關研究

要深入瞭解特定領域的研究發展，其方法之一就是對該領域所發表的文獻進行分析與分類。透過對發表文獻的分析與分類，可以歸納出該研究領域的研究範圍、議題以及在不同階段的發展重心，而文獻與資料的主題分析向來是圖書資訊學的研究重心之一。羅思嘉、陳光華與林純如(2001)指出當研究者在進行主題分析研究時，其主題類目的擬定大致有兩種方式：一是針對所研究的學科範圍，訂定一個包含整個學科領域的完整類表；另一則是事先檢視已蒐集文獻之內容，根據在文獻中出現的主題擬定主題類目。前者如 Cai and Card(2007)利用 ACM 的分類架構將文獻進行主題分類，後者即是書目計量學的相關研究所常進行之方法，本研究即是以後者做為主題分析之方式，因此會經過以下幾項步驟：

(1) 決定該領域欲進行分析的文獻或資料範圍

(2) 收集研究樣本

(3) 利用相關研究方法找出文獻探究之主題並加以歸類

(4) 以時間順序呈現主題脈絡或趨勢圖像

(5) 設法進行解讀，以驗證研究結果

資訊計量學(informetrics)即是利用量化的統計分析，描述各領域出版文獻的特性及發展模式，藉以檢視各學科的起源與進展，進而對未來的研究趨勢加以預測，這是資訊計量學在出版文獻、學科評價及學術傳播方面的重要應用(蔡明月、劉瓊芳，2007)。針對特定領域的主題研究趨勢進行探測時，找出該領域的研究主題是一個很重要的課題。因此本章節先針對文獻主題探測的書目計量學相關研究進行瞭解，最後將目前已針對資訊檢索領域文獻進行主題探測的研究加以統整。

一、　書目計量學應用於文獻主題探測相關研究

引用文獻分析(citation analysis)是指有系統的組織文獻及其所引用的資料，其目的希望藉由兩者關係圖的發展，了解文獻與文獻之間的關係，進一步探索學

術間的知識傳播模式(羅思嘉，2001)。引用文獻分析方法大致上可分為兩種模式：數量上的計算及文獻連結關係的分析。前者指的是單純以統計的方式來分析期刊、作者或文獻等被引用的次數；後者則是透過書目連結等方式分析文獻之間的關係(Baker, 1990)。

(一)文獻主題探測之方法

書目計量學中可應用於探討研究文獻之間主題的關係，則有三種方法：共被引分析(co-citation analysis)、書目對分析(bibliographic coupling)和共字分析(co-word analysis)。共被引分析的基本概念乃是指一群引用文獻依據他們共同出現於期刊文獻的參考書目中的數量所群聚，根據其分析對象又可細分為文獻共被引、期刊共被引、作者共被引；書目對分析的基本概念是指依據共同分享的參考文獻數量所群聚的引用文獻(Persson, 1994)；共字分析是指出現在兩篇文獻同一欄位，相同的字或詞來測量兩篇文獻的聯結關係(蔡明月，2003)。

書目對與共被引分析都同樣是利用第三文獻與兩篇存在的文獻之間的引用與被引用關係，建立兩篇文獻之間的聯結關係。而這兩者之間最主要的差別在於書目對是以引用文獻(citing document)為叢集的主體，共被引則是以被引用文獻為叢集的組成分子。因此兩種方法所得的結果會有一些偏差。利用書目對方法僅能測出科學發展趨勢的成長點，屬於科學計量學的預測作用；共被引方法則有較廣的應用，除了可以瞭解學科分支的進化情形，亦可用於科學的歷史研究。

蔡明月(2003)指出共被引分析比書目對分析更可有效的指出相似的主題，亦能反應出被引用文獻之間的語意關係，進而對科學結構，科學或文獻之間的聯結掌握其歷史性的動態研究。而曾元顯(2007)則認為依照書目對歸類的好處是論文 A 和 B 的用語可以差異極大，只要兩者共同引用足夠多相同的文獻，但若兩者之間沒有足夠的相同參考文獻時，卻又使用相同的主題詞彙，則兩者之間就可以藉由共字分析對應起來，因此，書目對分析與共字分析可以互補，也可以彼此強化對方的結果。

共字分析是計算和分析文獻間經常一起出現的關鍵字詞，能夠直接的將文獻的主題展現出來，且共字分析是分析自然語言，因此很重要的是這種方法可以代出文獻中的科學概念、構想和知識(van Raan & Tijssen, 1993)。此外，蔡明月(2003)也提到共字分析不同於共被引分析只限於引用的文獻，共字分析的字詞來源相當普遍，許多書目記錄的欄位都可以是共字分析的對象。在不同的研究中，研究者所採用的欄位皆有所不同，像是 Ding et al.(2001)利用文章的關鍵字、標題、摘要等三個欄位進行分析；Chaomei Chen(2005)是以標題、摘要、 描述、引用文獻進行分析；曾元顯、林瑜一(2006)則利用文章標題和摘要進行分析。

傳統的書目計量學分析像是期刊或作者共被引分析，都是藉由分析科學文獻的引用文獻，但卻無法直接將文獻的主題顯示在文獻中，共字分析可以解決此傳統書目計量學無法直接於文獻中顯示出文獻主題的缺失 (Ding et al., 2001)。從大量文獻中歸類出重要的研究主題，除了可以應用上述的書目計量學等方法以外，也有其它方式可以進行主題判斷。Funk(1988)判斷熱門或冷門主題的方法是先以 1973 年為基礎年度，再計算每年度的詞彙增長比例作為熱門主題的判斷依據，計算辭彙的增長斜率以藉此找出熱門或冷門的主題 。Smeaton et al.(2002)則是利用 Clustan Graphics 軟體輔助分類後再以人工進行檢視並給予主題；Cai and Card(2007)是先利用 ACM 的分類架構再以人工之方式將文獻進行主題分類，而涉及以人工方式進行分類的缺點除了耗時耗工以外，也很難排除掉個人的主觀意識。賴芊卉與劉瓊芳(2007)則是採取較客觀的方式，從索引摘要資料庫中取出各篇文獻的關鍵字欄位進行分析，此種方法雖可避免人為的主觀因素，但會受制於資料庫的限制。而黃惠美(2007)也提出書目計量學的研究需要相關資料庫做為基礎，假使資料庫沒有收錄研究者所需分析的資料時，即無法採用此種方式進行分析，像是會議文獻顯少被收錄於資料庫中，若需要將會議文獻進行主題分析與歸類時就會遇到此問題。因此欲瞭解會議文獻的研究主題，若無法採用資料庫中所給予的主題詞彙，又欲以節省時間的方式進行歸類，就需使用其他相關分析軟體進行輔助歸類。

最後，在主題呈現的方式上，書目對和共被引分析皆是以引用文獻或被引用文獻所存在的連結進行分析量化，都可以得出文獻叢集，計算叢集之間的關聯性強度則可以繪製出科學之間的關係圖，常見的有網路圖、時序圖、樹狀圖、圓形圖及二維空間圖。但是根據研究結果所繪製出的圖形或叢聚具有高度複雜的分析，需要有專業知識才有辦法解讀(Chaomei Chen, 2005)，因此在專家驗證上會是趨勢研究最終的挑戰。

(二)應用於文獻主題探測之相關研究

Persson(1994)利用書目對和共被引分析 JASIA 中 209 篇期刊文獻，發現資訊科學領域的核心文獻大多集中於 1970 年以後，且該領域的兩大研究主流分別是資訊計量學和資訊檢索。Åström(2007)針對圖書資訊學領域的文獻進行共被引分析，該研究印證資訊計量學和資訊檢索為兩大研究主流外，亦發現在資訊計量學的文獻集叢中，網路計量學形成了新的群聚；而在資訊檢索領域中，研究議題由系統導向的資訊檢索議題轉變為注重使用者互動的檢索系統以及資訊檢索行為研究，也印證 Persson(1994)提到資訊檢索領域中關於使用者等相關議題的軟性(soft)研究逐漸增加並影響關於技術方面的硬性(hard)研究。

Zhao and Strotmann(2007)利用作者共被引分析 XML 的相關文獻並進行該領域的研究趨勢預測，以比較兩種不同的學術傳播管道對預測研究趨勢上的表現。他們分別下載代表期刊的 ISI 資料庫與代表網路文獻的 CiteSeer 兩種管道的 XML 相關文獻，並以 SPSS 軟體進行作者共被引分析該領域的作者群，透過年代看出作者所發表的文獻其所屬領域變化，再利用因素分析繪製該領域的知識架構，以獲得更清楚的圖像。其研究結果發現從 1996 至 2001 年間於 CiteSeer 所下載的文獻作者群所屬的主題領域中，有部份與從 2001 至 2006 年間於 SCI 所下載的文獻作者群所屬領域有相同之處。也就是說，他們發現 CiteSeer 所下載較早期的文獻已有部分出現於 SCI 所下載的文獻中，因此作者認為 CiteSeer 比 SCI 更能夠預測該領域的研究趨勢。

在國內則有蔡明月與吳嘉雯於 2000 年針對半導體文獻進行期刊與作者雙被引分析，以 SCI 資料庫下載 30 種期刊為研究樣本，共得 435 組數據進行分析；以 46 位作者，共得 1035 組數據，以相關係數分別進行分析，並繪製二維圖像進行探討，該研究發現半導體文獻具有強烈跨學科性質，且以物理學為主要領域，但無法判斷核心主題。

曾元顯、林瑜一(2006)針對教育評鑑領域的318篇期刊文獻進行趨勢偵測，該研究利用SCI資料庫篩選Education Evaluation Policy Analysis(EAAP)期刊做為分析對象，採用自動化的方式偵測，以書目對分析和共現字分析兩種方法，找出EAAP期刊中歷時不衰的研究主題與研究前沿，該研究主要採用自動化的方式針對非結構化的資料進行探測，不透過人工的方式進行處理。該研究進行分析時，因資料集內欄位的關鍵字、文獻來源等控制詞彙不足，因此採用曾元顯開發的關鍵詞及關聯詞擷取技術，自標題、摘要的欄位內容中，自動擷取重要的辭彙，並將主題相近的詞彙歸類成「概念」，以消除單一詞彙主題模擬兩可之情形，並以書目對分析和共字分析兩種方法進行該領域研究趨勢的探測，最後對照兩種方法所共同出現的主題詞彙(曾元顯、林瑜一，2006)。

黃惠美(2007)以臺灣地區圖書資訊學文獻進行作者雙被引分析。該研究以因素分析、多維尺度法分析、群聚分析以及作者間的相互引用情形，探討 1955 年至 2002 年間，九種臺灣地區圖書資訊學領域期刊，該研究從中挑選文獻高生產的 40 位作者分析所屬領域，並找出分屬六區文獻的主題。

(三) 共字分析相關研究

共字分析是以文字或詞彙之間經常同時出現的文字對，計算其頻率以定義出呈現在文字間主題的關連圖像(He, 1999)。共字分析的技術首見於 1980 年位於法國的 CNRS(Centre National de la Re-cherche Scientifique)機構，他們將其系統命名為 LEXIMAPPE；而應用該技術進行研究的則有 Serge Bauin(1986)分析 1979 至 1981 年關於水產養殖相關文獻(Callon, Courtial, & Laville, 1991)。

He 於 1999 年針對共字分析的發展歷程作出完整的回顧，提出共字分析的優勢是可量化與具有彈性，而其限制則是在索引者的影響(indexer effect)。其中又有許多因素會影響至共字分析的結果，像關鍵字的品質、索引詞彙、資料庫範圍、統計方法的適用性等，此外，他也依據文獻探討提出三種共字分析的侷限：1.共字分析展現的結果難以解讀。2.資料庫涵蓋的範圍並不完整，像是專利與灰色文獻皆未被納入探討。3.出版時間與索引時間的落差，導致其結果無法即時反應出最新的研究趨勢。

隨著時代的進步，共字分析研究的素材，已從取自索引的文字到取自文獻標題、摘要，直至目前的全文分析，索引者的影響程度已降低(Callon, Courtial, & Laville, 1991)，但共字分析所進行的研究仍多侷限於期刊文獻的探討(Courtial, 1994; Ding et al., 2001)，直至近年來才陸續有學者將擴增至專利文獻(Tseng, Lin, & Lin, 2007)。

共字分析的研究進行大致上可分為三個步驟：1.萃取出字彙。可從標題、摘要或關鍵字中擷取。2.資料標準化，建立共字矩陣。3.彙製結果對應圖(Rokaya, Atlam, Fuketa, Dorji, & Aoe, 2008)。在彙製研究結果對應圖時，早期如圖 2-2 繪製出策略座標圖(strategic diagram)，以字彙間的密度(Density)和向心性(Centrality)作為 X 軸和 Y 軸的兩個分析面向，將所的結果分為四個區塊(He, 1999)。此外，也有相關研究會以多維尺度法(Multidimensional Scaling, MDS)及階層群聚分析(Cluster Analysis)展現其結果(Ding et al., 2001)。

共字分析的相關研究，在國內有張正宏(2008)針對知識管理領域進行關鍵詞分析，歸納國內與「知識管理」相關的博碩士論文，並比較資訊系所與非資訊系所呈現出的關鍵詞分佈，最後該研究以群集結果樹狀圖和策略座標圖繪製出以共詞分析所呈現之系所間的差異。在大陸地區則是有周靜怡、孫坦與陳濤(2007)探討「人類基因組」領域，藉由共字分析的策略圖描繪該領域的熱門議題與發展趨勢。

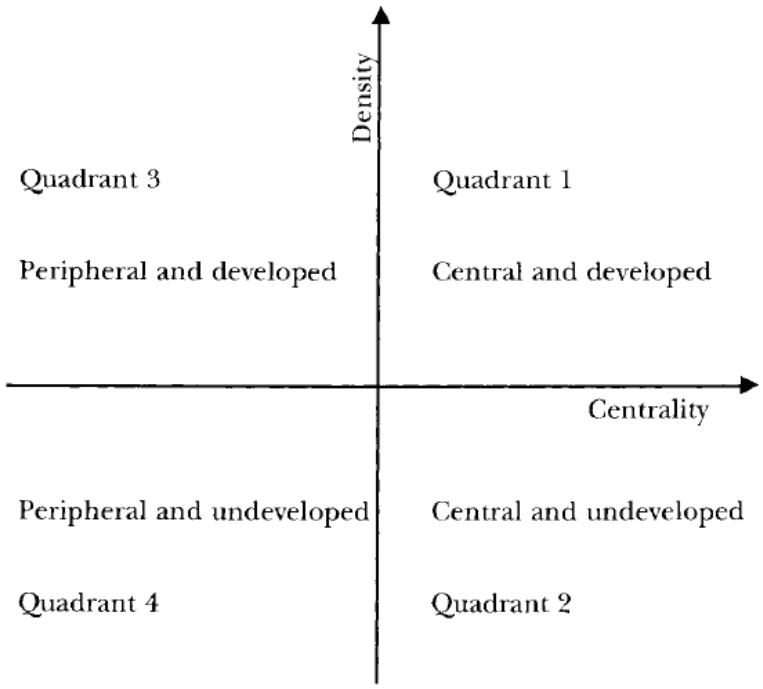

圖 2- 2 策略座標圖

資料來源：Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, *22*(1), 155-205.

二、 資訊檢索文獻相關研究

在資訊日新月異的時代中，資訊檢索(information retrieval)隨著資訊爆炸而更顯重要，資訊透過有效率的方式進行整理後，得以讓取用者能夠更有效率的取得需要的資訊，而資訊檢索亦是圖書資訊領域中所關切的議題之一 。

針對資訊檢索領域所進行文獻分析的學者有 新加坡的三位學者(Ding et al., 1999, 2001)分別利用三種不同的書目計量學方法針對資訊檢索十一年來的文獻進行相關研究。在 1999 年期刊共被引的研究中 ，他們找出資訊檢索領域被引用最多次的期刊，以多維尺度分析法(Multidimensional Scaling, MDS)將高被引的期刊，依照主題的相近度，繪製出該領域期刊的主題地圖，也進一步繪製出被資訊檢索領域的研究者高度引用的圖書資訊學期刊的主題地圖 。該研究將年代分為 1987-1991、1992-1997 前後兩階段，觀察兩階段之間的期刊主題地圖的變化，發

現在資訊檢索領域中的期刊引用與被引用情形很穩定，電腦科學、心理學、科學與自然等皆在資訊檢索領域中扮演很重要的角色。

另外，他們在同年度也以作者共被引分析方式找出 39 位在資訊檢索領域中高度被引的學者，並且以多維尺度分析法(Multidimensional Scaling, MDS)和群聚技術(Cluster techniques, CT)繪製研究圖像，再以因素分析(Factor analysis, FA)法找出每位學者的研究領域，其研究結果發現作者共被引分析確實可以用在研究學術領域的知識架構(Intellectual Structure)，此外他們也提出，經過長時間研究而累積的文獻，陸續被後續的研究者所引用，即可以作為一個正式的歷史回顧。

其後他們又於 2001 年以共字分析關鍵字數量上與群體鏈結的變化，他們以 LISA、LCSH、Thesaurus of Information Technology Terms 三種辭典同時進行詞彙控制，而文獻的關鍵字暴增代表了該領域正在快速發展，並發現有些主題是清楚明確的，像是使用者理論(user theory)；而某些主題是多元性的，像是由人工智慧擷取智慧型資訊檢索、神經網絡等；其他則是傳統的議題，像是網際網路、網路等(Ding et al., 2001)。最後他們將所取得的關鍵字以視覺化呈現出來，並且觀察 1987-1991、1992-1997 前後兩階段所取得的關鍵字數量上與群體間鏈結的變化，該研究找出某些字彙在兩階段中有明顯的正成長和負成長，代表了不同主題之間的興起與衰落，此外也發現有一些新興的字彙出現，像是網際網路(World Wide Web ,WWW)、網路(Internet)、資訊檢索行為(Information seeking behaviour)、線上資料庫(Online databases)、多媒體(Hypermedia)、生物學(Biology)、電子出版(Electronic publishing)、伺服器(Server)、決策支援系統(Decision support systems)、知識圖像(Knowledge representation)、視覺化(Visualization)等。

Smeaton et al.(2002)等利用25年來發表於SIGIR會議的853篇文獻進行內容分析，以找出1971~2002年SIGIR會議文獻的所探討的主題趨勢。此外，他們也利用文獻作者進行共同作者分析(Co-authorship analysis)，找出在SIGIR會議文獻中最具有影響力的作者。該研究取自853篇文獻的標題、作者名、摘要，形成853x853的相似矩陣圖，並且利用clustan graphics5.52版軟體進行分析，最後分成 29個群

體，最小的群體有5筆資料，最大的有126筆資料，再由人工給每個群體一個主題。其研究發現以Databases, NL Interfaces為主題的文獻於1980開始興盛，但在近年來開始衰落；以question answering為主題的文獻原先被視為是近年來才開始有發展的，但卻早在1971年開始就已經有相關文獻出現，Cross lingual的資訊檢索文獻從1996年開始出現並且持續發展中，而Evaluation則是於1993年開始出現，與TREC的發展有密切相關，因為TREC是從1991年開始舉辦的。

而在國內針對資訊檢索領域的文獻進行研究者甚少，僅有許雅婷於 2006 年針對資訊檢索領域進行文獻老化之研究，該研究以資訊檢索的兩個主題-系統研究和使用者研究進行分析，了解同領域中不同主題的文獻老化現象，而該研究指出在資訊檢索領域中兩個主題的文獻老化現象並沒有顯著差異，也就是說在資訊檢索領域中系統研究和使用者研究的文獻老化現象差異不大。

在資訊檢索領域中的文獻研究上，新加坡的三位學者已利用書目計量學的多種方式進行研究，然而資料類型都限定在期刊文獻，Smeaton et al.(2002)則是以會議文獻為分析對象，兩者皆並未針對不同的文獻類型進行比較分析，在國內針對資訊檢索領域的文獻進行研究者亦尚未針對主題趨勢方面進行探討，因此本研究除了比較兩種資料類型在預測研究趨勢上的差異外，亦針對資訊檢索領域的主題趨勢進行探討。

第三章　研究方法

為瞭解會議文獻與期刊文獻對預測主題趨勢之相對性的效果，本研究以實驗法進行實證比較研究。本章分為研究概念、研究假設、研究方法、研究流程與實驗設計、研究對象與工具等五部份加以闡述。

第一節 研究概念

本研究之基本構想為比較會議文獻與期刊文獻預測相同主題時的各自表現，再依據其表現進行統計檢定，以判別會議文獻是否比期刊文獻更能有效預測資訊檢索領域的主題趨勢，而改變預測集和驗證集的年代範圍及預測指標之目的在於觀察自變項是否會影響主題趨勢預測。其研究架構如圖 3-1，圖中上半部是以兩種不同的文獻進行交叉預測與三種歸類情況下的年代推演，觀察其進行主題趨勢預測後之得分情形；下半部則是以自動化歸類下的年代推演進行預測集範圍、驗證集範圍、預測指標為自變項，觀察其進行主題趨勢預測後之得分情形。

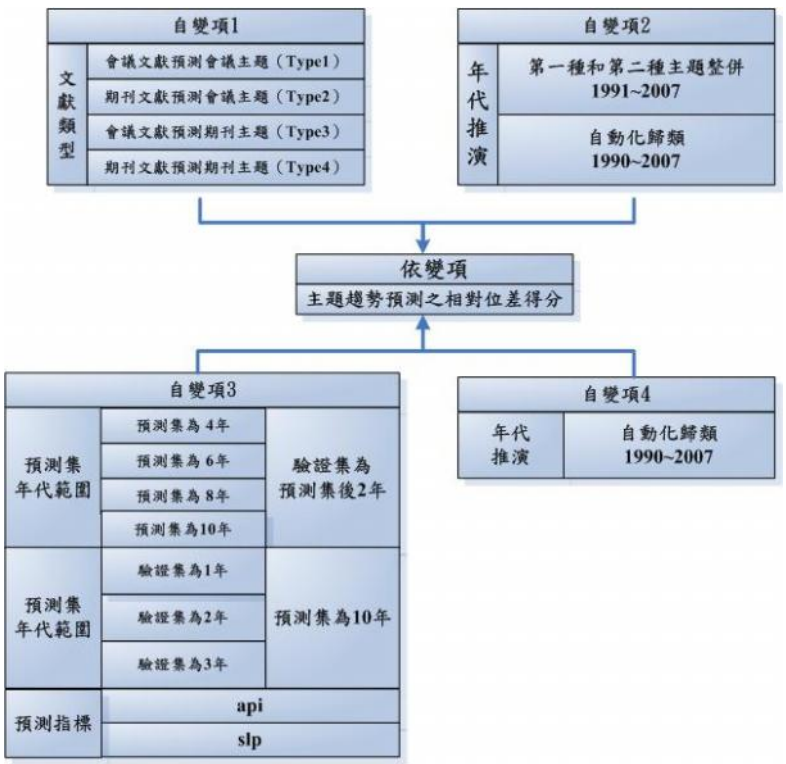

圖 3- 1 研究架構圖

第二節 研究假設

根據研究目的、研究問題及研究架構，提出本研究之研究假設如下：

假設一：兩種文獻類型交叉預測於不同年代序列，在主題趨勢預測的結果上有顯著差異。

假設二：不同預測集年代範圍於不同年代序列，在主題趨勢預測的結果上有顯著差異。

假設三：不同驗證集年代範圍於不同年代序列，在主題趨勢預測的結果上有顯著差異。

假設四：不同預測指標兩種於不同年代序列，在主題趨勢預測的結果上有顯著差異。

第三節 研究方法

本研究主要採用實驗法比較兩種不同的文獻，預測資訊檢索領域時之相對性效果，其作法為先將文獻分成「預測集」與「驗證集」，利用「預測集」所取出的主題趨勢預測「驗證集」所呈現的主題。首先必須萃取出預測集和驗證集的主題，再以趨勢預測指標的計算值將「預測集」依熱門程度依序排列主題，而驗證集則是以文獻總數量作為熱門程度之排序，最後比對「預測集」和「驗證集」中各主題的相對位差作為主題趨勢預測的結果，即完成主題趨勢預測的部份。

驗證的部份則是以統計的雙因子變異數分析，檢定各年度所計算的相對位差平均值，檢驗各組文獻集預測表現及其穩定度，以下針對主題萃取方式、主題趨勢預測方式、驗證集與預測集之取樣、資料分析的方法等加以詳述。

一、 主題萃取方式

本研究的主題萃取方式主要可分為以下三種：

(1) 期刊文獻：取出 WOS 資料庫所下載之關鍵字詞彙(KeyWords Plus®，下載後欄位名稱為 ID，故以下簡稱 ID 詞彙)搭配主題萃取系統。

(2) 會議文獻：採用會議主題(Session topic)，並進行人工檢驗。

(3)　期刊和會議文獻：合併後匯入主題萃取系統進行歸類。

期刊文獻因數量龐大且不易由人工逐一比對，故採用資料庫所下載取得的 ID 詞彙再搭配系統自動萃取之方式，如此較為省時省力。資料庫所提供的 ID 詞彙即可代表每篇文獻的主題，雖然每篇文獻的 ID 詞彙不只一個，但每個 ID 詞彙卻只會對應至一篇文獻中，因此本研究將每個 ID 詞彙視為一篇文獻之主題進行年代統計後，做為本研究的期刊主題依據的一部份，另一部分則是由於會議文獻與期刊文獻的主題會有不一致的情形產生，故本研究再針對會議文獻中的重要主題但卻無法對應至期刊文獻的部份，由主題萃取系統分析期刊文獻後取出各年代篇數，以補強無法對應的部分期刊主題。

會議文獻的主題萃取方法則採用專家分類法，因 SIGIR 會議每年度的文獻均已分類至各會期的主題(Session Topic)，本研究以 SIGIR 會議的分類主題為依據，再透過人工的方式給予權威控制，將 SIGIR 會議文獻更嚴謹地分類至各主題中，最後由一位具有資工相關背景的教授協助檢視歸類。

最後，因上述兩種主題萃取與歸類方式皆需透過部份人工進行，難以完全排除個人主觀之因素干擾，故本研究最後再將會議文獻和期刊文獻全部合併後，一併利用主題萃取系統進行文獻主題的歸類，再取出可看出明顯主題的文獻群組進行統計分析。而本研究透過主題萃取系統萃取主題之方式是以書目計量學中的共現字分析作為系統的萃取方法，將標題和摘要由文獻中取出並經由系統自動偵測主題。系統主題偵測的方式為先將文件中的重要詞彙萃取出來後歸類為概念；再將各種概念進一步歸類為主題；再以主題歸類為次領域和大領域，以逐步地方式將文獻主題歸納出來(曾元顯、林瑜一，2006)。

二、　主題趨勢預測方式

本研究主題趨勢的預測方式分為兩階段：第一階段為利用趨勢預測指標計算預測集中各主題之熱門程度，再按熱門程度排序主題。第二階段則是以趨勢預測評鑑之方式比對「預測集」與「驗證集」的相對位差，以觀察會議文獻和期刊文

獻的預測相對性效果，以下分別就趨勢預測指標和評鑑方式兩部份做詳細說明。

(一) 趨勢預測指標

本研究之預測集所採用排序熱門主題的方法，係採用曾元顯(2008)在網路計量與探勘學術研討會中所提出，常見的兩種趨勢評鑑指標作為主要依據。一為平均文獻成長率(API)，其主要之計算方式為將文獻年代區間分為前後兩階段，再計算兩階段文獻成長之斜率作為各主題的趨勢成長指標，其計算公式為 $api = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{d_{i+1}-d_i}{d_i}$，$n$ 值為文獻篇數，由表 3-1 為以 API 指標所計算之各主題文獻成長之範例，最後再依平均文獻成長率之大小做排序。

表 3- 1 文獻集熱門主題計算與排列方式範例表

主題 \ 年代	97-01	02-03	總計	計算方式	排序
Question answering	5	12	17	(12-5)/5=140%	3
Relevance feedback	8	17	25	(17-8)/8=112.5%	4
Signature files	5	4	9	(4-5)/5=-20%	5
Evaluation	0	34	34	(35-1)/1 =340%	1
Cross lingual	0	16	16	(17-1)/1=160%	2

資料來源*: Tseng, Y.-H., Lin, Y.-I., Kuo, C.-H., & Lee, Y.-Y. (2008).*

另一種趨勢預測的指標為在年代序列上線性迴歸的最佳曲線(slope of the linear regression line)，簡稱 *SLP*，此為統計學上較常使用於趨勢預測之方式，其詳細之計算方式如下：

$$slp = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}}, \text{ where } \quad x_i = i - \frac{1}{n}\sum_{i=1}^{n} i \quad \text{and} \quad y_i = d_i - \frac{1}{n}\sum_{i=1}^{n} d_i$$

下表 3-2 和 3-3 為此計算方式之範例，表 3-2 為五個主題依照年代排列的原始文獻篇數，表 3-3 即為透過公式計算後的 *SLP* 趨勢指標。

表 3- 2 SLP 範例計算之原始資料

主題 \ 年代	2000	2001	2002	2003	2004	總計
Web(IR,search,structure,www)	4	3	4	3	3	17
Retrieval Model(S)	1	9	0	4	10	24
Classification、Categorisation(Classifiers)	0	3	5	7	3	18
User Issues	0	3	3	0	0	6
Evaluation = PERFORMANCE	4	3	2	0	0	9

其計算方式為先將各年度的文獻數量和年代加總後再除以個數，以求得各年代和文獻數量的平均，再以各年度文獻數量減去此平均數，以求得 x_i 和 y_i，再將兩者相乘得出表 3-3，最後再除以 $\sqrt{\sum_{i=1}^{n} x_i^2}$，求得 *SLP* 值。

表 3- 3 SLP 範例計算之結果與排序

主題 \ 年代	2000	2001	2002	2003	2004	總計	*SLP*	排序
Web(IR,search,structure,www)	-1.2	0.4	0	-0.4	-0.8	-2	-0.632	3
Retrieval Model(S)	7.6	-4.2	0	-0.8	10.4	13	4.1109	1
Classification、Categorisation(Classifiers)	7.2	0.6	0	3.4	-1.2	10	3.1623	2
User Issues	2.4	-1.8	0	-1.2	-2.4	-3	-0.949	4
Evaluation = PERFORMANCE	-4.4	-1.2	0	-1.8	-3.6	-11	-3.478	5

(二) 主題趨勢預測評鑑

「預測集」依趨勢預測指標計算出熱門程度的排序後，需要再透過評鑑指標的比對「驗證集」的熱門主題，才能看出會議文獻和期刊文獻在預測主題趨勢上的相對性效果。

Tseng, Lin, Kuo, & Lee(2008)在網路計量與探勘學術研討會中曾提出的兩種趨勢評鑑的指標，分別是 NAP(Non-interpolated Average Precision rate)和 Pre@R(Precision rate at Recall position)。NAP 的計算公式為 $NAP = \frac{1}{R}\sum_{i=1}^{R}\frac{i}{Rank_i}$，*R* 為實際熱門主題相關項目的總數，$Rank_i$ 是 *i* 相關數量上的排序位置；Pre@R 則是指 *R* 在排序上位置的準確率，計算方式為 *r/R*，*r* 是指在 *R* 排序上相關項目的數量。雖然 NAP 較 Pre@R 為嚴謹，但本研究易產生主題無法對應之情形，因此將可對應之主題排序全數進行評鑑時，較無法適用 NAP 的評鑑方式，故本研究

另以計算全部排序之相對位差的方式(All Ranking，簡稱 AR)作為本研究主要的評鑑排序方式。

AR 之計算方式為 $AR = \frac{\left|Rank_i - Rank_j\right|}{N}$，$Rank_i$ 和 $Rank_j$ 為相同主題在兩種排序下的位置，亦及比對兩種排序之相對位差後，再取其位差絕對值除以主題數量得出平均相對位差值(AR)。以表 3-4 為例，以 S2 相對於 S1 之 AR 值而言，V 主題之相對位差值為｜1－6｜ ＝ 5，W 主題之相對位差值為｜2－7｜ ＝ 5，而 A 至 Z 主題相對位差絕對值總和則為 50，則平均相對位差得分即為 5。

表 3- 4 評鑑排序方式之範例

Rank	S1	S2	S3
1	**A**	V	**A**
2	**B**	W	V
3	**C**	X	**B**
4	**D**	Y	W
5	**E**	Z	**C**
6	V	**A**	X
7	W	**B**	**D**
8	X	**C**	Y
9	Y	**D**	**E**
10	Z	**E**	Z
NAP	1.00	0.35	0.68
Pre@R	1.00	0.00	0.60
All Ranking	0	5	2

資料來源:改編自 *Tseng, Y.-H., Lin, Y.-I., Kuo, C.-H., & Lee, Y.-Y. (2008).*

AR 和 NAP 分別為兩種類似的評鑑排序方法，其差別在於 AR 的計算方式較為嚴謹，其全部的主題排序位置皆視為同等重要，亦即無論排序在前或後的主題，其對應至驗證集的順序都要被考量進去；而 NAP 的計算方式則較為寬鬆，僅考量前面特定主題的是否有相同的主題而已，且在前面幾個主題範圍內的排序順位皆不再比較其先後順序，亦即有被納入之主題即被視為得分。

以表 3-4 來說，假設表格中 A-E 與 V-Z 代表十種不同主題，其中以 A-E 為

最熱門主題，S1 至 S3 則分別代表三種不同的排序方式。S1 代表最優的排序方法，可依照熱門程度將前五種熱門主題排序為最前端；S2 代表最差的排序方式，將熱門主題 A-E 排在最後五項；S3 則是將五種熱門主題 A-E 分散至各處(Tseng, Lin, Kuo, & Lee, 2008)。在客觀的認定下，三種排序方式依照優劣可分為 S1 為最佳排序方式、S3 次之、S2 最差，其評鑑三種排序方法的兩種評鑑指標之計算過程如下表 3-5，然而這兩個評鑑指標最終所計算出來的結果，在解讀上的意義為 NAP 的結果與其排序優劣成正比、AR 的結果則是成反比，亦即 NAP 數值越大越好，AR 的數值則是越小越好。此乃因 NAP 的計算結果為 S2 和 S3 在比對 S1 時，其相符程度越高者數值越高，但是 AR 的計算值則是 S2 和 S3 在比對 S1 時的相對差距，亦即兩者之間的差距越小越好，故在解讀時須特別注意兩者數值的不同。

表 3- 5 兩種評鑑排序指標之計算過程

NAP 評鑑之計算範例		
S1	(1/1＋2/2＋3/3＋4/4＋5/5)÷5＝1.0	
S2	(1/6＋2/7＋3/8＋4/9＋5/10)÷5＝0.3547	劣
S3	(1/1＋2/3＋3/5＋4/7＋5/9)÷5＝0.6787	優
AR 評鑑之計算範例		
S1	(1-1+2-2+3-3+4-4+5-5+6-6+7-7+8-8+9-9+10-10) ÷10＝0	
S2	(|1-6|+|2-7|+|3-8|+|4-9|+|5-10|+|6-1|+|7-2|+|8-3|+|9-4|+|10-5|)÷10＝5	劣
S3	(|1-1|+|2-3|+|3-5|+|4-7|+|5-9|+|6-2|+|7-4|+|8-6|+|9-8|+|10-10|)÷10＝2	優

三、 預測集與驗證集之判定與取樣

在比較會議與期刊文獻對主題趨勢預測的相對性效果時，須事先決定可供比對驗證的文獻集，稱之為「驗證集」。因無法以客觀的角度判定具有絕對性的驗證集，本研究以兩種獨立的驗證集作為比較對象，也就是說，本研究之驗證集分別以會議文獻和期刊文獻作為兩組獨立的驗證集，再各自以不同來源的預測集進行趨勢預測，而驗證集的文獻總數量即為熱門程度判斷之依據。

根據文獻探討，Subramanyam(1981)指出會議與期刊文獻之間約有一年左右的時間落差，Zhao and Strotmann(2007)則認為網路文獻與期刊文獻的時間落差大約為 2~3 年，而 WOS 資料庫在計算影響係數(Impact factor)時，計算方法是以該期刊前兩年中所刊登的論文在今年被引用的次數總和，除以該期刊前兩年中出版論文的總數。值愈高，表示平均每篇文章被引用的次數愈高，代表該期刊的影響力愈大。故本研究在主題趨勢預測之驗證集取樣以兩年作為驗證集的年代範圍 ，在預測集的年代範圍則是固定以 10 年為一區段進行預測。

此外，本研究亦想要探討驗證集的年代範圍大小對預測效果之影響，故本研究以接續預測集後的一年、兩年、三年，各年度區間的文獻總數作為主題趨勢預測的標準答案，以作為供預測主題趨勢時可比對的驗證集。在探討預測集年代範圍大小對主題預測之影響的部份，預測集的年代範圍採用 4、6、8、10 年的年代區間，預測驗證集為 2 年的情況下加以探討，最後也會分別依照兩種不同的預測指標對主題預測之影響進行比較分析。

四、 資料分析方法

在探討資訊檢索領域文獻的主題分佈與研究趨勢變化時，是以各主題的篇數按年代排列與基礎的描述統計進行探討，而在主題萃取系統的歸類方式上是以共現字分析作為分析方法。

在實驗進行的部份則是以兩種不同類型文獻(會議與期刊)分別進行交叉預

測後，取得各組實驗後的相對位差平均數，再以雙因子變異數分析檢定各組實驗之平均數是否有達到信心水準 95%下的顯著性差異，以探討兩自變數間是否存在主要效果與交互效果，

表 3-6 為本研究之研究目的與問題及其對應的研究方法和檢定方式，由表中可知，本研究主要是以共現字分析與實驗法作為主要的研究方法，並以描述統計和雙因子變異數分析檢定其差異性。

表 3- 6 研究目的、研究問題與研究方法對照表

研究目的與問題	研究方法	檢定方式
瞭解資訊檢索領域文獻的主題分佈與研究趨勢變化。 1. 資訊檢索領域的研究主題分佈為何？ 2. 近年來，資訊檢索領域的研究趨勢有何變化？	描述統計 共現字分析	
比較會議與期刊文獻進行預測研究趨勢的差異。 1. 採用會議文獻和期刊文獻作為研究樣本之交叉預測的相對性效果為何？會議文獻是否比期刊文獻較為準確？ 2. 交叉預測之結果是否隨著年代不同而有差異？	實驗法	雙因子 變異數分析
瞭解會議與期刊文獻在預測主題趨勢上的穩定性。 1. 主題趨勢預測的情形是否隨著預測集的改變而有所不同？ 2. 主題趨勢預測的情形是否隨著驗證集的改變而有所不同？	實驗法	雙因子 變異數分析
瞭解趨勢預測指標對預測研究趨勢的影響情形。 1. 不同的趨勢預測指標應用於預測研究趨勢時有何變化？ 2. 預測指標是否會改變主題預測之結果？	實驗法	雙因子 變異數分析

第四節 研究流程與實驗規劃

本研究之研究流程如下圖 3-2，在相關的研究背景與產生研究意識的情況下發展出研究目的與研究問題，收集文獻進行探討後擬定研究架構與方法，在根據研究設計與規劃進行主題趨勢預測和實驗，最後分析統計資料並撰寫論文。

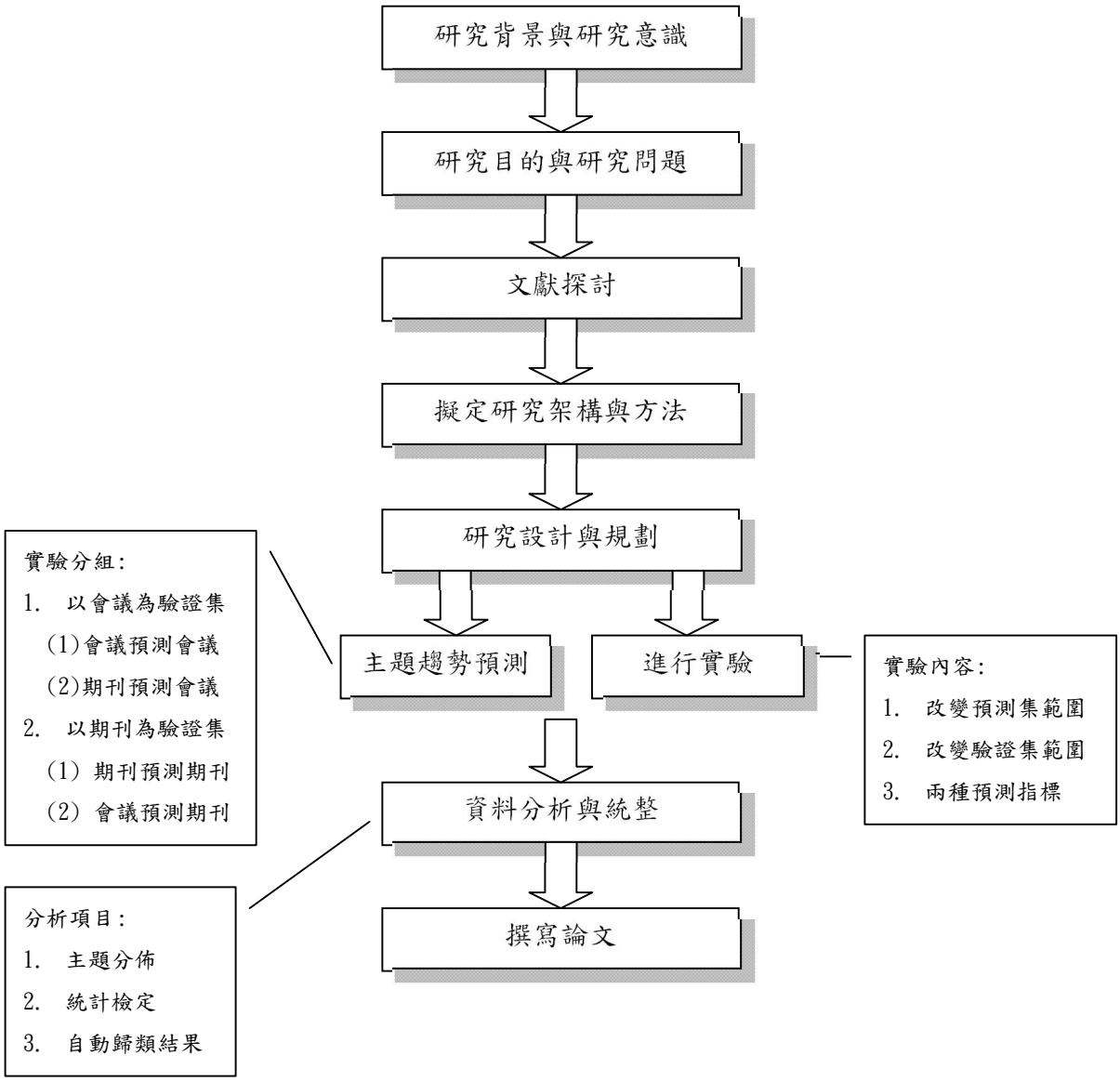

圖 3- 2 研究流程圖

本研究以實驗法針對兩種資料類型對交叉預測主題趨勢之相對性效果、預測集與驗證集年代範圍、預測指標之應用等進行比較與探討，再以統計檢定驗證研究假設，以下分為實驗規劃、分組與流程分別敘述如下：

一、 實驗規劃

本研究將各文獻分為「預測集」與「驗證集」，再以各組「預測集」預測「驗證集」之主題，圖 3-3 表示各組「預測集」與「驗證集」所跨越的年代分佈與預測路徑。

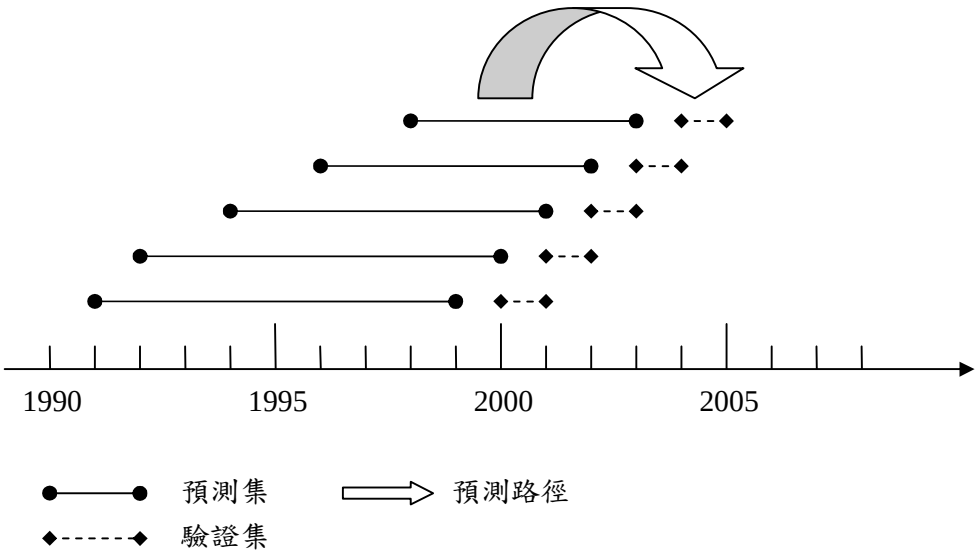

圖 3- 3「預測集」與「驗證集」預測橫跨年代示意圖

二、 實驗之自變項與依變項

(一) 兩種文獻類型之主題交叉趨勢預測

此階段之實驗分別以第一種主題整併、第二種主題整併、自動化歸類等三種不同的歸類情況下進行分析比較，而自變項與依變項之項目皆相同。亦即以三種不同的文獻規模進行主題趨勢預測，以觀察文獻在不同的整併方法下，所呈現的主題趨勢預測情形，以下為自變項與依變項之細部說明：

1. 自變項

(1) 文獻類型：會議預測會議、期刊預測會議、會議預測期刊、期刊預測期刊。

(2) 年代範圍：以 1991、1992、1993、1994、1995、1996 為預測集起始年代，其後 10 年為預測集之範圍，再以預測集後 2 年為驗證集範圍。

2. 依變項：主題趨勢預測排序相對位差得分

(二) 預測集年代範圍大小之預測

預測集的年代範圍大小是否影響預測結果，為本研究所欲探討的項目之一，為避免影響實驗之其他因素，故本研究僅採自動化歸類後之結果進行實驗，以排除在歸類上人工主觀之因素。此實驗之自變項與依變項如下：

1. 自變項：

(1) 預測集年代範圍：預測集為 4 年、6 年、8 年、10 年。

(2) 年代區間：以 1990、1991、1992、1993、1994、1995、1996 為預測集起始年代。

2. 依變項：主題趨勢預測排序相對位差得分。

表 3-7 為預測集年代範圍大小之各組詳細的年代起始範圍，由表中可看出每次的預測下，因起始年代不同的關係都會進行 7 次實驗(G1-G7)，而每次實驗所的之結果皆會取得平均數後分別進行雙因子變異數分析，以檢定是否各次實驗之間有達到顯著性差異。

表 3- 7 預測集年代範圍大小之分組詳細年代表

代碼 ＼ 組別		實驗組 1	實驗組 2	實驗組 3	控制組
組別代碼		4v2	6v2	8v2	10v2
G-1	預測集年代	1990~1993	1990~1995	1990~1997	1990~1999
	驗證集年代	(1994、1995)	(1996、1997)	(1998、1999)	(2000、2001)
G-2	預測集年代	1991~1994	1991~1996	1991~1998	1991~2000
	驗證集年代	(1995、1996)	(1997、1998)	(1999、2000)	(2001、2002)
G-3	預測集年代	1992~1995	1992~1997	1992~1999	1992~2001
	驗證集年代	(1996、1997)	(1998、1999)	(2000、2001)	(2002、2003)
G-4	預測集年代	1993~1996	1993~1998	1993~2000	1993~2002
	驗證集年代	(1997、1998)	(1999、2000)	(2001、2002)	(2003、2004)
G-5	預測集年代	1994~1997	1994~1999	1994~2001	1994~2003
	驗證集年代	(1998、1999)	(2000、2001)	(2002、2003)	(2004、2005)
G-6	預測集年代	1995~1998	1995~2000	1995~2002	1995~2004
	驗證集年代	(1999、2000)	(2001、2002)	(2003、2004)	(2005、2006)
G-7	預測集年代	1996~1999	1996~2001	1996~2003	1996~2005
	驗證集年代	(2000、2001)	(2002、2003)	(2004、2005)	(2006、2007)

(三) 驗證集年代範圍大小之預測

驗證集的年代範圍大小是否影響預測結果，亦為本研究所欲探討的項目之一，為避免影響實驗之其他因素，故本研究亦僅採自動化歸類後之結果進行實驗，以排除在歸類上人工主觀之因素。由表 3-8 可見此實驗之自變項與依變項如下：

1. 自變項：

(1) 驗證集年代範圍：驗證集為 1 年、2 年、3 年。

(2) 年代區間：以 1990、1991、1992、1993、1994、1995 預測集起始年代。

2. 依變項：主題趨勢預測排序相對位差得分。

表 3-8 為驗證集年代範圍大小之各組詳細的年代起始範圍，由表中可看出每次的預測下，因起始年代不同的關係都會進行 6 次實驗(H1-H6)，而每次實驗所的之結果皆會在取得平均數後分別進行 雙因子變異數分析，以檢定是否各次實驗

之間有達到顯著性差異。

表 3- 8 驗證集年代範圍大小之分組詳細年代表

代碼 \ 組別		實驗組 1	控制組	實驗組 2
組別代碼		10 v 1	10 v 2	10 v 3
H-1	預測集年代	1990~1999	1990~1999	1990~1999
	驗證集年代	2000	2000、2001	2000、2001、2002
H-2	預測集年代	1991~2000	1991~2000	1991~2000
	驗證集年代	2001	2001、2002	2001、2002、2003
H-3	預測集年代	1992~2001	1992~2001	1992~2001
	驗證集年代	2002	2002、2003	2002、2003、2004
H-4	預測集年代	1993~2002	1993~2002	1993~2002
	驗證集年代	2003	2003、2004	2003、2004、2005
H-5	預測集年代	1994~2003	1994~2003	1994~2003
	驗證集年代	2004	2004、2005	2004、2005、2006
H-6	預測集年代	1995~2004	1995~2004	1995~2004
	驗證集年代	2005	2005、2006	2005、2006、2007

(四) 趨勢預測指標不同之預測

預測指標之不同是否影響預測結果，亦為本研究所欲探討的項目之一，由於本研究進行主題趨勢預測時，即會採用兩種預測指標進行預測，故此階段僅以前述進行自動化歸類的主題趨勢預測實驗時，所取得之數據進行假設檢定，而自變項與依變項如下：

1. 自變項：

(1) 趨勢預測指標：API、SLP。

(2) 年代區間：以 1990、1991、1992、1993、1994、1995、1996 預測集起始年代。

2. 依變項：主題趨勢預測排序相對位差得分。

表 3-9 為兩種趨勢預測指標之各組詳細的年代起始範圍，由表中可看出每次的預測下，因起始年代不同的關係都會進行 7 次實驗(I1-I7)，而每次實驗所的之結果皆會在取得平均數後分別進行雙因子變異數分析，以檢定是否各次實驗之間有達

到顯著性差異。

表 3- 9 兩種趨勢預測指標之實驗各組詳細起始年代

預測指標 / 組別代碼	預測指標為 API		預測指標為 SLP	
I-1	1990~1999	2000、2001	1990~1999	2000、2001
I-2	1991~2000	2001、2002	1991~2000	2001、2002
I-3	1992~2001	2002、2003	1992~2001	2002、2003
I-4	1993~2002	2003、2004	1993~2002	2003、2004
I-5	1994~2003	2004、2005	1994~2003	2004、2005
I-6	1995~2004	2005、2006	1995~2004	2005、2006
I-7	1996~2005	2006、2007	1996~2005	2006、2007

三、 實驗流程

圖 3-4 為本研究實驗之詳細流程，由圖中可看出各次實驗進行之細節，其中會議文獻與期刊文獻皆以三種歸類方式進行主題預測。也就是說，本研究將文獻按照第一種主題整併、第二種主題整併、自動化歸類三種情形先進行文獻的主題統整與對應，接著再以兩種預測指標 API 和 SLP 進行主題趨勢的預測，最後採用統計學上的雙因子變異數分析進行顯著性差異的檢定。

而在進行預測集和驗證集的範圍大小實驗時，為了避免人工因素的干擾影響實驗之進行，因此僅採用自動化歸類的結果進行實驗，分別改變預測集和驗證集的年代範圍，再以兩種預測指標 API 和 SLP 進行主題趨勢的預測，以觀察不同的預測集和驗證集之範圍是否會影響主題趨勢預測的結果，最後同樣採用統計學上的雙因子變異數分析進行顯著性差異的檢定。而預測指標是否對預測主題趨勢有所影響，則是將上述之實驗過程擷取兩種不同的趨勢指標做分析探討，而顯著性差異的檢定則採用自動化歸類後的趨勢結果進行分析。

圖 3-5 即為預測結果之範例圖，由圖中可知本研究以系統驗算出以 API 為預測指標下 1995 年至 2004 年為預測集、2005 年和 2006 年為驗證集，所得到之會議預測會議與期刊預測會議(CC/JC)預測結果排序，最後再由排序之相對位差計算出最後預測結果得分。

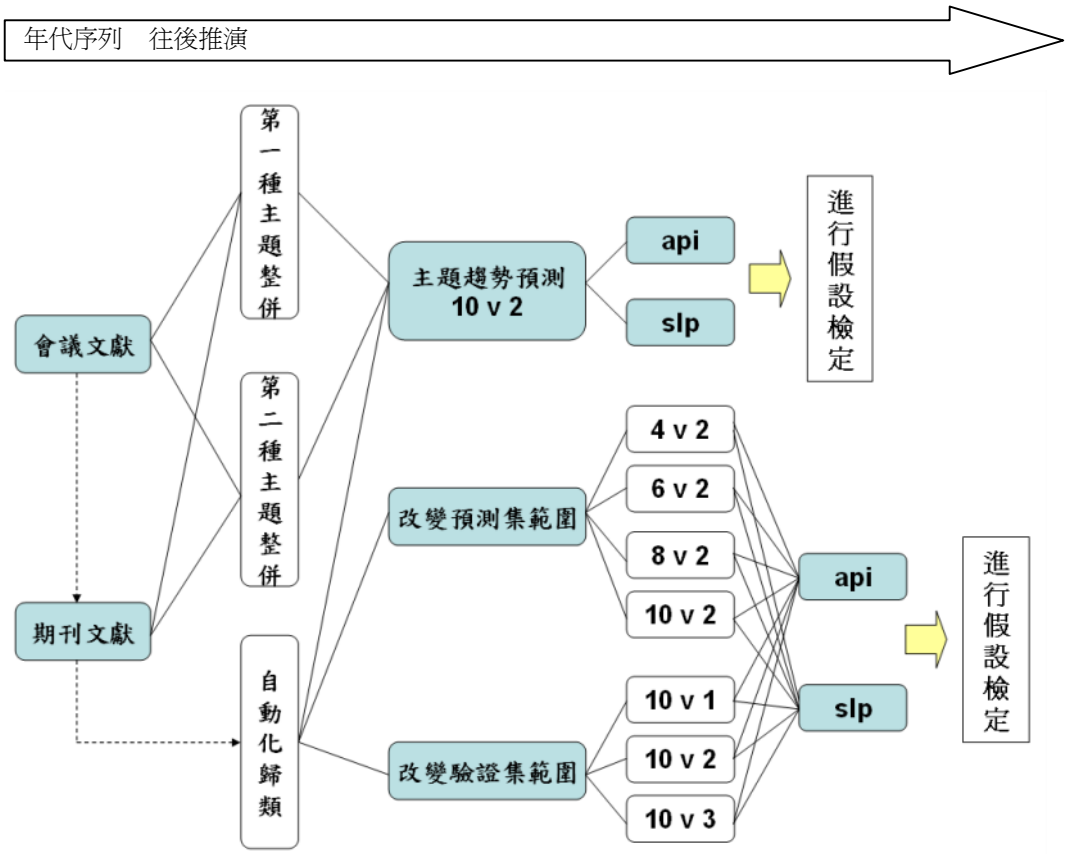

圖 3- 4 實驗流程圖

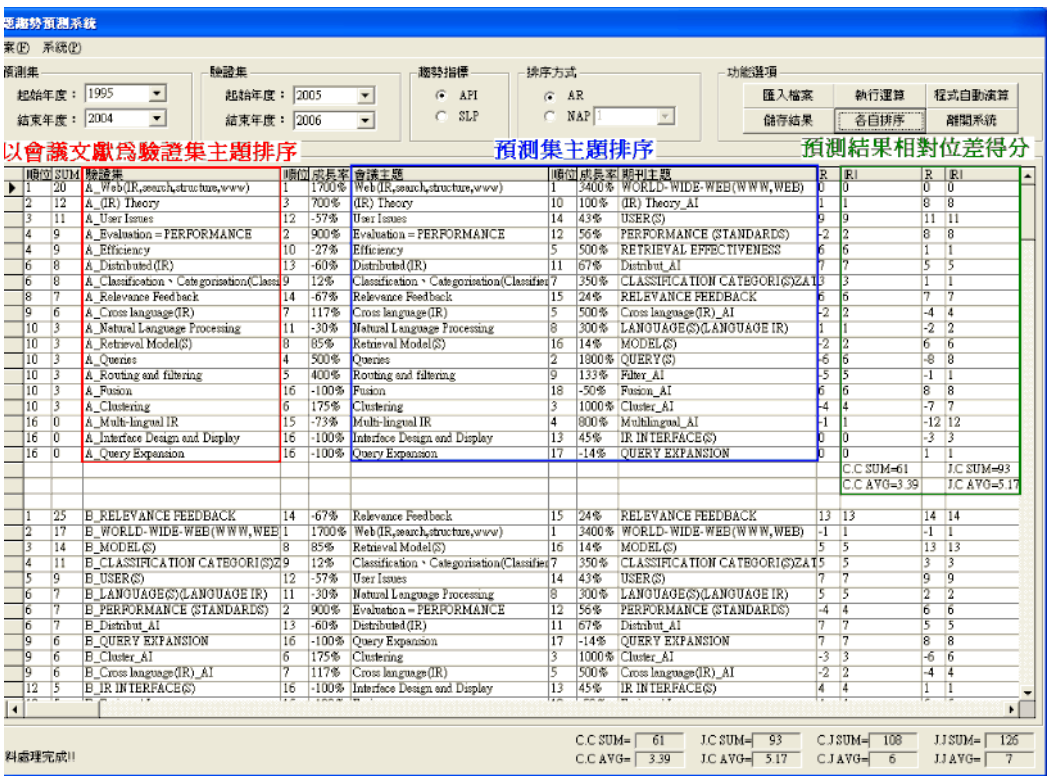

圖 3- 5 第一種主題整併 API 為趨勢指標下 1995~2004 年 CC/JC 主題趨勢預測

四、 資料處理步驟與流程

(一) 會議文獻與期刊文獻整理流程

本研究針對期刊文獻的主題整理流程如圖 3-6，即為先將期刊文獻出現 10 次以上之 ID 詞彙統整出年代序列，再將主題對應至 SIGIR 會議文獻，若無 SIGIR 可對應但卻又 SIGIR 屬於重要探討之主題，本研究再以主題萃取系統將期刊文獻各年代進行歸類後以人工之方式找出文獻之主題年代與篇數 。

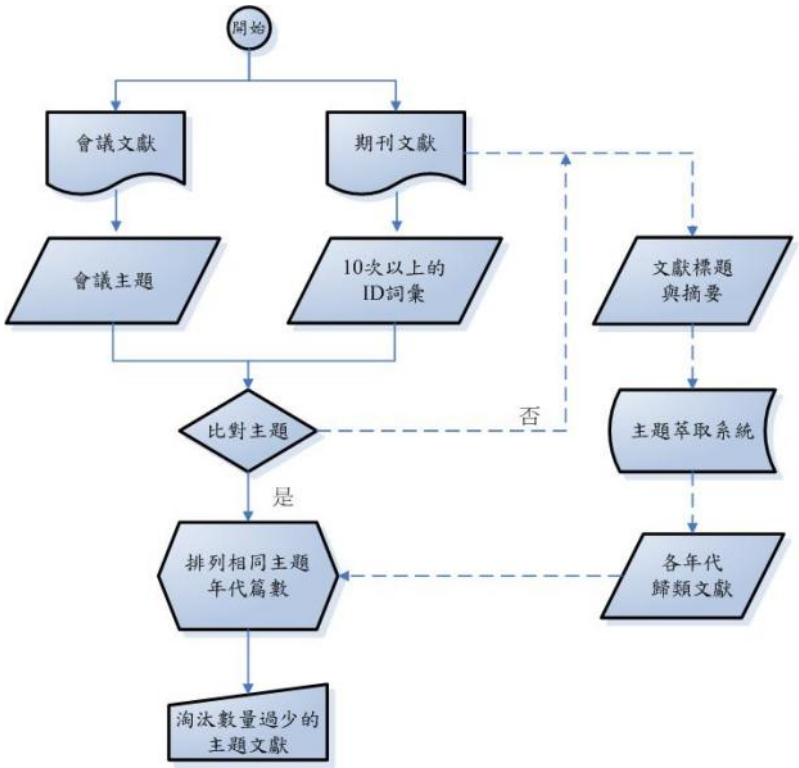

圖 3- 6 期刊主題整理流程圖

會議文獻的主題由每年度的評選委員審定，故並無統一的架構或詞彙，需由人工統整會議主題詞彙，且為求嚴謹本研究亦採用人工逐一比對會議文獻與會議主題是否相符，並由一位具有資工系博士背景之教授協助進行檢視原始分類 。圖 3-7 即本研究針對會議文獻是否需要重新分類之判斷流程圖 ，首先將會議文獻中具有雙重主題之文獻重新選定主題，再將會議文獻中具有僅該年度特有主題之特性的文獻進行重新分類，因該年度特有之主題通常為該年度委員無法歸類或該年度篇數過少等特殊文獻，故需將其重新分類，而無上述兩種情形的一般文獻，則

僅比對文獻是否符合該年度所給予的主題即可，最後再以人工的方式將文獻主題詞進行詞彙統整。

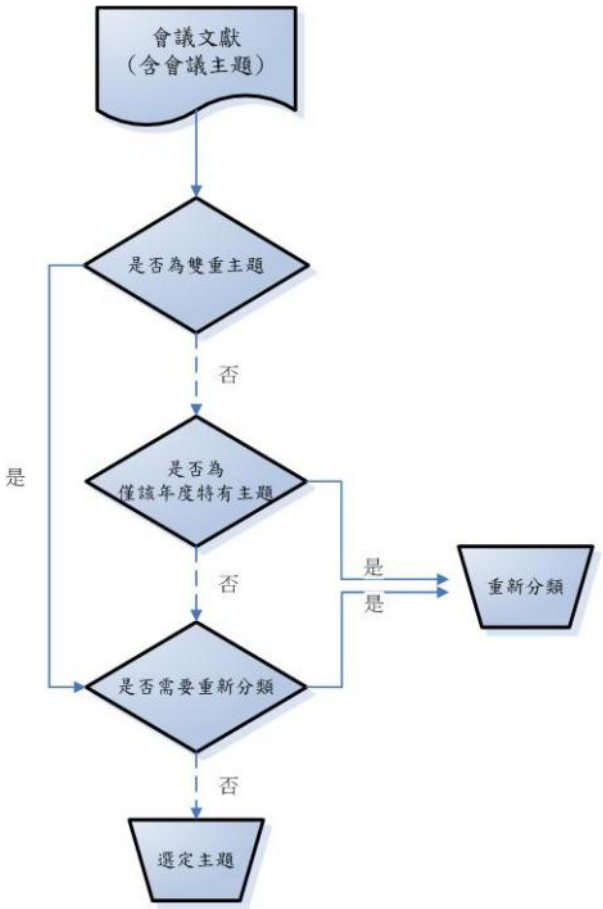

圖 3- 7 判斷會議文獻重新分類流程圖

(二) 會議文獻與期刊文獻主題歸類

本研究在期刊文獻的主題詞部份係採用資料庫內的 ID 詞彙，因每篇期刊文獻的 ID 詞彙數量並不一致，故每篇期刊文獻皆可能出現重複分類的情形，但每個 ID 詞彙在同一篇期刊文獻中只會出現一次，因此本研究將各次的 ID 詞彙均納入分析，並將每次 ID 詞彙視為一篇文獻的主題，亦即在各年度中 ID 詞彙出現的次數即代表該主題出現的篇數；而會議文獻的部份，因各篇會議文獻均僅有一個主題，無重複分類的情形，故在年代序列中各篇會議文獻的主題次數即代表篇數。在取出可對應之主題後，確認各主題文獻數量後即可進行下列各步驟的歸類，而無法於本研究找到相對應之主題則不再進行歸類，僅依前段所敘述各別探

討之，歸類方式共分為以下三個階段進行。

1. 第一種主題整併

第一種主題整併即將文獻在文字上「完全相同」之詞彙主題先進行整合，像是單複數、縮寫、不同詞性等的文字，為求嚴謹先於文字上整合後的結果進行初步比較，再進一步將同義詞(Synonyms)與相似詞(Ambiguity)的字彙進行合併歸類。因此，在第一種主題整併後共找出 18 個可對應的主題，進行後續分析的會議文獻共 479 篇；期刊文獻共 663 篇，其歸類結果如附錄一的表三。

2. 第二種主題整併

第二種主題整併合併則是以第一種主題整併為基礎，再進一步將各主題之廣義詞、狹義詞和同類詞進行合併，第二種主題整併後總共找出 11 個主題，進行後續分析的會議文獻共 609 篇；期刊文獻共 1090 篇，在數量上有較大的差距，其歸類結果如附錄一的表四。

3. 主題萃取系統自動化歸類

系統輔助主題歸類之方式乃以曾元顯教授所開發之主題萃取系統進行歸類判斷，首先將欲分析之會議與期刊樣本文獻的年代、標題與摘要下載至 ACCESS 資料庫中，再以 perl 程式將文獻進行共現字分析再加以歸類。本研究所採用此主題系統進行協助歸類之方式為兩種，一是下載資訊檢索領域的核心期刊文獻後，將其各年度各自進行歸類，作為輔助本章第一節第三項所述之補足無法對應至會議主題的期刊文獻數量；二為將會議文獻與期刊文獻全數進行自動化歸類，其目的在於減少本研究於第二種主題整併後可能產生的人為主觀因素，其詳細結果於本章第二節與第三節詳細敘述之。本研究利用主題萃取系統自動歸類則可找出 38 個主題，進行後續分析的會議文獻共 477 篇；期刊文獻共 502 篇，其歸類結果如附錄二與附錄三。

圖 3-8 為各階段歸類之原則與流程，由圖中可知，第一種和第二種主題整併的差異之處，即為第二種主題整併多了一個合併廣義詞、狹義詞和相同類別詞彙的步驟。

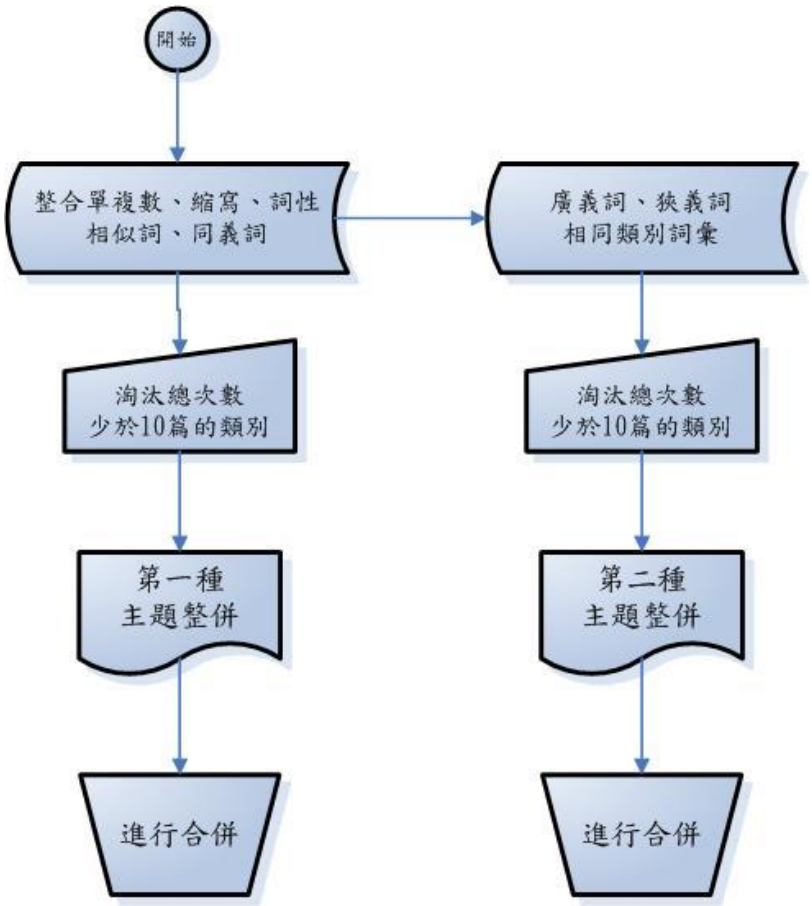

圖 3- 8 各階段歸類之原則與流程

第五節 研究對象與工具

本研究所分析的對象乃是以期刊文獻與會議文獻為主，其研究對象之選取方式及所使用之研究工具，分別敘述如下：

一、 研究對象

(一) 期刊文獻：以核心期刊作為研究樣本

本研究以檢索自 Web of Science 資料庫中，收錄主題為資訊檢索的所有文獻，依照期刊數量排列後取出數量最多的前七名期刊[3]與新加坡學者於 1999 年的研究中找出資訊檢索領域的核心期刊 (Ding et al.，1999b)之交集，最後決定以五本核心期刊中所收錄的資訊檢索文獻作為研究對象，此五本核心期刊分別是：Journal of information science 、Information processing & management 、JASIST&JASIS、Journal of documentation、Information retrieval。最後，本研究進行期刊文獻的樣本檢索條件如下：

(1) 檢索語為：Topic[4]= “Information Retrieval”[5]

(2) 年代限制：1990 至 2007 年

(3) 期刊名稱限制：Journal of information science 、Information processing & management 、 JASIST&JASIS 、 Journal of documentation、Information retrieval。

最後共得文獻總數為 1037 篇，表 3-10 為各期刊所佔篇數與比例分佈，其中以 INFORMATION PROCESSING& MANAGEMENT 佔全部樣本文獻的 38.48%，而 JASIS&JASIST 佔全部樣本文獻的 36.55%，分居第一、二名，圖 3-9 則是樣本文獻中各本期刊所佔之比例圓餅圖。

---

[3] 期刊依文獻數量多寡排名依序爲 INFORMATION PROCESSING & MANAGEMENT、JASIS&JASIST、JOURNAL OF DOCUMENTATION、ONLINE INFORMATION REVIEW、JOURNAL OF DOCUMENTATION、INFORMATION RETRIEVAL、ACM TRANSACTIONS ON INFORMATION SYSTEMS、JOURNAL OF INFORMATION SCIENCE

[4]檢索欄位 Topic 即同時跨欄位檢索 Title + Abstract + Author Keywords + Keywords Plus®。

[5]檢索詞“Information Retrieval”，加註雙引號即代表檢索詞中的所有單字視爲一個詞彙。

表 3- 10 期刊文獻樣本中各期刊所佔之篇數與比例

期刊名稱	文獻篇數	所佔百分比
INFORMATION PROCESSING & MANAGEMENT	399	38.48%
INFORMATION RETRIEVAL	85	8.20%
JOURNAL OF DOCUMENTATION	102	9.84%
JOURNAL OF INFORMATION SCIENCE	72	6.94%
JASIS&JASIST	379	36.55%
總計	1037	100%

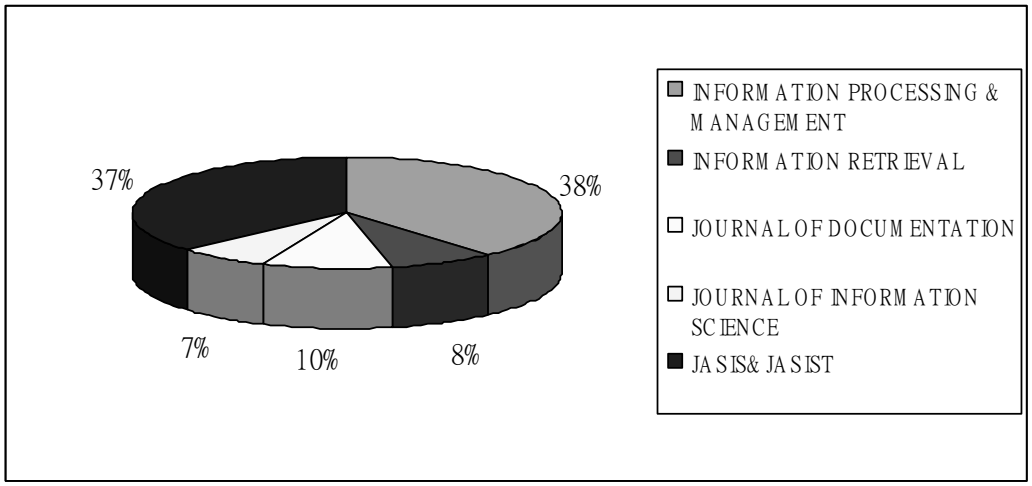

圖 3- 9 樣本文獻中各本期刊所佔之比例圓餅圖

(二) 會議文獻：SIGIR 歷年來所收錄的會議論文為主要研究樣本 。

本研究以資訊檢索領域為研究範圍 ，因此會議文獻的研究樣本經由專家建議取自資訊檢索領域中最具有代表性的 ACM SIGIR 會議為研究樣本。

SIGIR 會議從 1978 年至 2007 年，共舉辦 30 屆，而本研究所指的 SIGIR 會議文獻乃是會議論文集所收錄發表於會議中的文獻，但不包括會議前或會議中的專欄文獻(Keynote)與摘要(Summery)，及會議論文集後半部的海報(Posters)與示範(Demos)文獻，僅以會議論文集中所發表的研究文獻為研究樣本 。

本研究下載年代範圍界定於 1990 至 2007 年的文獻作為分析對象，圖 3-10 為 SIGIR 各年度的文獻分佈長條圖，由圖中可知近年來 SIGIR 會議文獻有逐年增長之趨勢。

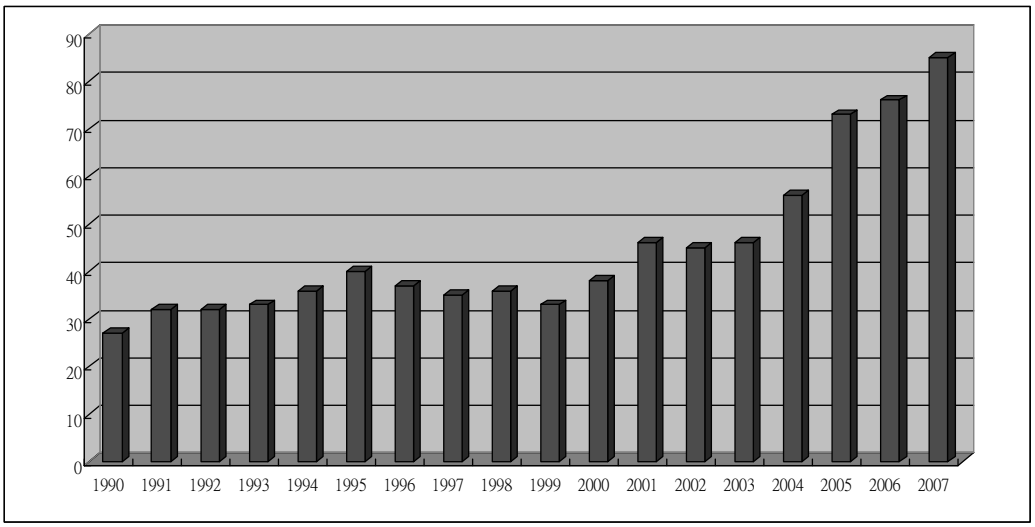

圖 3- 10 SIGIR 會議文獻 1990~2007 年各年度文獻篇數

(三) 研究樣本之調整

由於第一種主題與第二種主題整併之期刊文獻的部份，必須藉由關鍵字詞彙進行主題歸類與分析，而期刊文獻樣本於 1990 年以前的文獻均無關鍵字詞彙，故刪除 1990 年的文獻，從 1991 年至 2007 年進行主題歸類與分析，而會議文獻同樣刪除 1990 年之文獻，將其年代調整至與期刊文獻相同。在系統自動化歸類的部分，因不需採用關鍵字詞彙作為歸類的依據，故仍期刊文獻與會議文獻仍以 1990 年至 2007 年的文獻進行分析。

此外，在期刊文獻樣本的 1037 篇文獻中共取得 3154 次的關鍵字詞彙，雖然每篇文獻有可能有多個關鍵字詞彙，但每篇文獻的關鍵字詞彙均不會重複出現，故本研究將各次的關鍵字詞彙均視為一篇文獻，然而關鍵字詞彙的數量過多，故最後僅擷取出現次數超過 10 次以上之關鍵字詞彙做為分析對象，總計為 1845 次的關鍵字詞彙。

二、 研究工具

(一) 資料庫

1. SCIE、SSCI 資料庫

透過 Web of Science ®可以直接使用 Thomson Scientific 的三大引文資料庫 Science Citation Index Expanded ™, Social Sciences Citation Index ®, Arts & Humanities Citation Index®。資料可以一直回溯到 1900 年。這一豐富的綜合性資訊來自於全球 9,300 多份權威的、高影響力的學術期刊。透過獨特的被引文獻檢索，可以用一篇文章、一個專利號、一篇會議文獻或者一本書作爲檢索詞，檢索這些文獻被引用的情況，也可以輕鬆地回溯某一研究文獻的起源與歷史(Cited References)或者追蹤其最新的進展(Citing Articles)。本研究以 Web of Science 的 SCIE 和 SSCI 資料庫做為檢索對象，以下載期刊文獻進行分析。

2. ACM Digital Library

ACM 全名為 Association for Computing Machinery，是全世界最大的教育和電腦科學社群，傳遞最新先進的電腦相關資訊，ACM 提供電腦領域最首要的數位圖書館，服務它的成員並扮演提供時代尖端的出版商 、研討會、職業資源。ACM 數位圖書館是一個收錄大量 ACM 期刊、時事通訊和會議文獻的引用文獻與全文資料庫。SIGIR 會議文獻即完整收錄於 ACM 數位圖書館中。

(二) 微軟(Microsoft)Office 軟體 - Excel 試算表

微軟公司所發展的 Excel 整合性試算表套裝軟體，主要提供了電子試算表、繪製統計圖表及資料處理等三種應用功能 ，以便使用者進行工作表之建立、編輯、運算等工作，並能依資料提供繪製統計圖表，例如：直條圖、折線圖等；資料處理方面，提供了資料內容之排序、篩選、彙總、樞紐統計等簡易分析作業。

(三) Visual Basic 程式

由於本研究進行主題趨勢預測時，需透過公式計算趨勢指標及相對位差得分，為避免人工計算之錯誤發生，本研究委由程式纂寫人員將各種計算公式以 Visual Basic 語法編寫程式，以利本研究將各組文獻數量載入程式進行計算 。

(四) SPSS 統計分析軟體

本研究使用 SPSS 12.0 統計分析軟體進行統計分析，SPSS 是 Statistical Package for the Social Science 的縮寫，為一用於社會科學研究上的統計分析軟體，目前 SPSS 與 SAS 同為世界上公認最優秀的統計分析軟體(楊世瑩，2005)。其資料分析方法為獨立樣本雙因子變異數分析，以瞭解兩個自變項對依變項的影響，由於本研究為一個依變項，故採用單變量的雙因子變異數分析 。

(五) 主題萃取系統

曾元顯所開發的主題萃取系統可以應用於書目對分析 、控制詞彙之趨勢分析、少量文件之歸類分析、自由詞彙之共現字分析等。該系統首先將文件下載後，進行切割、分欄擷取、正規化、解剖儲存至關聯式資料庫中，以便於管理與運用，並進行初步量化分析與圖表製作，以便掌握其統計特性，並對其有初步之認識。再以文字探勘技術，分別對下載的文件進行主題自動偵測與辨識的工作，必要時須將「詞彙」或「文件」歸類成「概念」、「概念」歸類成「主題」、「主題」歸類成「次領域」、「次領域」再歸類為「大領域」，以便進行概括性的趨向分析 。根據自動辨識的主題，進一步分析並預測其未來趨勢，並自動擷取各類別特徵詞彙或標題詞彙，以協助人工判讀。最後將上述主題偵測與分析結果，以主題樹或主題地圖方式呈現，並結合其他結構化資料，如：出刊年代、國家、期刊等，以提供豐富多樣的判讀素材(曾元顯，2007)。

第四章　研究結果與分析

本研究旨在比較會議文獻與期刊文獻對預測主題趨勢之相對性效果，以實驗之方式針對不同類型的文獻進行預測結果的比較。本章先說明文獻主題之判斷過程，並將兩種文獻各自歸類後的主題進行比較分析，再針對兩種文獻對預測主題之效果進行統計檢定，以驗證本研究之假設，最後統整各文獻類型之主題探討與統計檢定後之結果。

第一節　文獻主題判斷與歸類

本節敘述期刊文獻與會議文獻的主題判斷過程與依據，並綜合兩者之間所取得的結果進行比較，期刊文獻的部份主要取自資料庫所下載的關鍵字做為分析對象，會議文獻的部份則以各年度會議主題為判斷依據。

一、　期刊文獻主題判斷

期刊文獻的主題判斷，原先預計採用系統將文獻初步歸類，再由人工逐一檢視歸類結果，並進行歸類判斷，但由於人工歸類判斷易出現歸類不一致的情形，且無法取得較具公信力之依據，故改由 Web of Science 資料庫本身所下載取得的關鍵字(ID 詞彙)作為本研究的主題依據。

本研究之研究樣本 1037 篇文獻中，共有 243 篇文獻沒有關鍵字，且 1990 年的文獻全部都沒有收錄關鍵字，故第一種主題與第二種主題整併之分析的期刊文獻樣本年代為 1991 至 2007 年，文獻數量共計 794 篇，會議文獻亦於本階段只取 1991 至 2007 年之文獻，以期與期刊文獻年代一致。表 4-1 為各年度缺少關鍵字詞彙與 DE 詞彙的情形，在關鍵字詞彙的部分扣除 1999 年的空值數量後，其他年度的關鍵字缺少平均值為 0.22，標準差為 0.06，亦即 1991 年至 2007 年平均各年度約有 22%的文獻缺少關鍵字，其中又以 1998 至 1999 年最為明顯，而缺少

DE 詞彙的比率普遍偏高，空值比率由 2003 年至 2007 年始降至 50%以下。

表 4- 1 各年代缺少關鍵字詞彙與 DE 詞彙的文獻分佈比率

年代	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
總篇數	14	34	36	39	37	34	42	46	60	50	77	76	69	66	81	73	112	91
缺少 ID 篇數	14	7	6	7	8	8	11	9	17	18	19	15	14	7	23	18	21	21
百分比	100%	21%	17%	18%	22%	24%	26%	20%	28%	36%	25%	20%	20%	11%	28%	25%	19%	23%
缺少 DE 篇數	14	34	33	39	35	34	40	46	59	46	54	38	36	25	29	18	32	27
百分比	100%	100%	92%	100%	95%	100%	95%	100%	98%	92%	70%	50%	52%	38%	36%	25%	29%	30%

每篇期刊文獻皆可能有數個的關鍵字，本研究參考蔡明月與賴淑靜於 2004 年針對資訊計量學的主題分佈研究所採用之方法，將每一篇所包含的主題均各別計算一次，呈現各主題出現於年代分佈的情形。由表 4-2 可知在 794 篇文獻中總共出現 3154 次主題，平均每篇文獻探討 3.97 個主題；此外總計出現 729 種不同的主題，平均每個主題被 4.3 篇文獻所探討。超過 10 篇以上共同探討的主題有 55 個，總計 1845 篇文獻皆有探討，佔全部的 58.5％。

表 4- 2 ID 主題詞彙次數分佈比率

出現次數	主題種類	總出現次數	出現次數百分比
1	404	404	12.81%
2	122	244	7.74%
3	57	171	5.42%
4	36	144	4.57%
5	23	115	3.65%
6	11	66	2.09%
7	9	63	2.00%
8	6	48	1.52%
9	6	54	1.71%
≧10	55	1845	58.50%
總計	729	3154	100.00%

表 4-3 則為關鍵字詞彙各年度的分佈情形，計算各關鍵字詞彙平均出現的主題種類與次數後可發現，除了 1990 年沒有關鍵字詞彙外，各年度中平均每篇文獻所包含的 ID 主題種類約 2 個左右，平均每個詞彙出現的次數也約 1~2 次。

表 4- 3 資訊檢索領域中核心期刊文獻每年主題分佈情況

年代	總篇數	ID 詞彙篇數	ID 詞彙次數	主題種類(個)	主題種類/篇	出現次數/個
1990	14	0	0	0	0.00	0.00
1991	34	27	82	51	1.89	1.61
1992	36	30	118	69	2.30	1.71
1993	39	32	119	70	2.19	1.70
1994	37	29	106	70	2.41	1.51
1995	34	26	88	56	2.15	1.57
1996	42	31	127	79	2.55	1.57
1997	46	37	163	106	2.86	1.54
1998	60	43	178	87	2.02	2.05
1999	50	32	123	75	2.34	1.64
2000	77	58	232	131	2.26	1.77
2001	76	61	213	122	2.00	1.75
2002	69	55	222	122	2.22	1.82
2003	66	59	260	153	2.59	1.70
2004	81	58	227	127	2.19	1.79
2005	73	55	231	138	2.51	1.67
2006	112	91	365	187	2.05	1.95
2007	91	70	300	173	2.47	1.73
總計	1037	794	3154		39.02	29.08

由於 794 個 ID 詞彙數量過多，故本研究僅採用出現次數 10 次以上之 ID 詞彙，總計共有 55 個主題。而其中有部份的詞彙仍屬重覆，像是 SYSTEMS 和 SYSTEM、WEB 和 WORLD-WIDE-WEB 等，經由人工檢視合併後剩下 49 個主題詞彙如表 4-4，由表中可看出，除了過於一般性的詞彙如 INFORMATION-RETRIEVAL(IR)、RETRIEVAL 和 SCIENCE 等，數量最多的前五名主題分別是 MODEL(S)、SYSTEM(S)、RELEVANCE、DESIGN、SEEKING。

表 4- 4 ID 主題詞彙出現 10 次以上之數量分佈

編號	主題詞彙	ID 詞彙出現次數	總次數%
1	INFORMATION-RETRIEVAL(IR)	297	9.42%
2	SYSTEM(S)	146	4.63%
3	MODEL(S)	103	3.27%

4	RELEVANCE	85	2.69%
5	RETRIEVAL	82	2.60%
6	DESIGN	81	2.57%
7	SCIENCE	77	2.44%
8	SEEKING	73	2.31%
9	USER(S)	65	2.06%
10	WEB(WORLD-WIDE-WEB)	62	1.97%
11	BEHAVIOR	48	1.52%
12	DOCUMENT-RETRIEVAL	47	1.49%
13	INFORMATION	41	1.30%
14	TEXT	36	1.14%
15	PERFORMANCE	35	1.11%
16	KNOWLEDGE	34	1.08%
17	DATABASE(S)	34	1.08%
18	QUERIES	26	0.82%
19	ALGORITHM	25	0.79%
20	SEARCH	23	0.73%
21	FEEDBACK	23	0.73%
22	INTERFACE	22	0.70%
23	TEXT RETRIEVAL	22	0.70%
24	ACCESS	21	0.67%
25	CLASSIFICATION	20	0.63%
26	STRATEGIES	19	0.60%
27	QUERY EXPANSION	18	0.57%
28	INTERNET	17	0.54%
29	ONLINE	17	0.54%
30	SELECTION	17	0.54%
31	ASK	16	0.51%
32	NETWORKS	16	0.51%
33	FRAMEWORK	15	0.48%
34	THESAURUS	14	0.44%
35	HYPERTEXT	14	0.44%
36	RELEVANCE FEEDBACK	13	0.41%
37	TACTICS	13	0.41%
38	PATTERNS	12	0.38%
39	NEEDS	12	0.38%

40	SUBJECT ACCESS	11	0.35%
41	DEFINITION	11	0.35%
42	DOCUMENTS	11	0.35%
43	TERMS	11	0.35%
44	WORDS	10	0.32%
45	PROBABILISTIC MODEL	10	0.32%
46	DOCUMENT-RETRIEVAL SYSTEMS	10	0.32%
47	JUDGMENTS	10	0.32%
48	KEYS	10	0.32%
49	ONLINE CATALOGS	10	0.32%
總計		1845	58.50%

二、 會議文獻主題判斷

SIGIR 每年度的會議論文集皆會由該年度的委員們將各篇會議文獻分類至各主題中，此為本研究歸類會議文獻的主要分類依據。首先自網頁上將各篇會議的主題(Session title)、標題(Title)、摘要(Abstract)下載至 EXCEL 試算表後載入 ACCESS 資料庫進行整理與分析，並比對每年度的文獻數量是否正確無誤，將不需要的文獻刪除，如：Keynote Address、Invited Speaker、TREC Overview 等皆不予收錄。最後將 1990 年至 2007 年的 SIGIR 會議文獻與其主題統整如表 4-5，由表中可知 SIGIR 會議文獻一共取出 280 個主題，各年代中平均每個主題約有 2~4 篇文獻。

表 4- 5 會議文獻各年度文獻主題平均數量分佈

年代	文獻篇數	主題數量(個)	平均每個主題文獻數
1990	27	8	3.38
1991	32	11	2.91
1992	32	17	1.88
1993	33	14	2.36
1994	36	15	2.40
1995	40	15	2.67
1996	37	17	2.18
1997	35	13	2.69
1998	36	15	2.40
1999	33	12	2.75

2000	38	16	2.38
2001	46	12	3.83
2002	45	15	3.00
2003	46	17	2.71
2004	56	17	3.29
2005	73	17	4.29
2006	76	24	3.17
2007	85	25	3.40
總計	806	280	51.68

表 4-6 為 SIGIR 會議文獻中出現次數 10 篇以上之主題文獻分佈，由表中可知，數量最多的前五個主題分別是 Evaluation、Efficiency、Classification&Categorisation、Cross language、Clustering。

表 4- 6 SIGIR 會議文獻主題數量分佈

編號	主題	文獻篇數	總篇數%
1	Evaluation	28	2.54%
2	Efficiency	25	3.10%
3	Classification、Categorisation	25	3.10%
4	Cross language	24	2.98%
5	Clustering	22	2.73%
6	Natural Language Processing	21	2.61%
7	Formal Models	16	1.99%
8	Distributed	16	1.99%
9	Summarization	15	1.86%
10	Theory	13	1.61%
11	Multimedia Information Retrieval	13	1.61%
12	User Studies	13	1.61%
13	Image Retrieval	12	1.49%
14	Filtering	11	1.36%
15	Asian	11	1.36%
16	Web search	10	1.24%
17	Question answering	10	1.24%
18	Retrieval Models	10	1.24%
總計		295	35.67%

三、 無法對應主題之補足與探討

本研究將期刊文獻以關鍵字詞彙(ID 詞彙)與會議主題詞彙經過人工初步整理詞彙後，再進一步將兩者進行比較即可發現，兩者之間有部份可以對應出相同主題，也有無法找到相對應的主題詞彙產生，可以對應出的相同主題於本章第三節與年代序列一併探討，以下分別僅就尚未歸類前即無法對應之主題分別探討之。

(一) 期刊文獻主題無法找到相對應之會議主題

本研究取期刊文獻於資料庫中的關鍵字欄位，然因關鍵字詞彙量較多，故先初步以人工整理方式將辭彙進行初步統整，同時將單複數、詞性不同等相同詞彙進行初步整合，且僅比對在本研究範圍中至少出現 10 次以上之辭彙。

經由人工比對後可發現，表 4-7 列出 ID 詞彙出現次數超過 30 次且未無法比對至 SIGIR 會議文獻的文獻數量，由表中可知大致上可分為三部分：

(1) 一般性詞彙：無法解讀該主題所探討之內容，如 INFORMATION-RETRIEVAL、SCIENCE、INFORMATION。

(2) 意義相近之詞彙：如期刊文獻的 SEARCH(ING) BEHAVIOR 與 SIGIR 會議文獻的相似主題 User behavior。

(3) 特有之主題：像是期刊文獻中的 LIBRARY、CATALOG、BIBLIOMETRICS 等議題確實無法於 SIGIR 文獻中找到較相近主題。

表 4-7 期刊文獻無法對應至會議文獻之主題

主題類別	ID 詞彙	出現次數	總次數
INFORMATION-RETRIEVAL	INFORMATION-RETRIEVAL	285	379
	IR	12	
	RETRIEVAL	82	
DOCUMENT-RETRIEVAL	DOCUMENT-RETRIEVAL	47	68
	DOCUMENTS	11	
	DOCUMENT-RETRIEVAL SYSTEMS	10	
TACTICS、STRATEGIES	TACTICS	13	32
	STRATEGIES	19	
INFORMATION-SEEKING	SEEKING	73	32
	INFORMATION-SEEKING	8	
SEARCH	SEARCH(ING) BEHAVIOR	6	44
	SEARCH PROCESS	5	
	SEARCH	23	
	SEARCHERS SELECTION	5	
	SEARCH ENGINES	5	
ALGORITHM(S)	ALGORITHM(S)	27	38
	GENETIC ALGORITHMS	5	
	STEMMING ALGORITHM(S)	6	
LIBRARY、CATALOG	DIGITAL LIBRARIES	6	42
	LIBRARY(S)	11	
	CATALOG(S)	9	
	ONLINE CATALOG(S)(DISPLAYS)	16	
BIBLIOMETRICS	SCIENTOMETRICS	3	40
	BIBLIOMETRICS(CLASSIFICATION DATABASES RETRIEVAL)	7	
	COCITATION(S) (ANALYSIS)	9	
	CITATION(S)(ANALYSIS)(RETRIEVAL)	14	
	CO-WORD ANALYSIS	3	
	CO-OCCURRENCE (DATA)	4	
DESIGN			81
SCIENCE			77
INFORMATION			41
KNOWLEDGE			34

(二)　會議文獻主題無法找到相對應之期刊主題

經由人工比對 SIGIR 會議主題至期刊文獻中後可發現，共有 73 個主題詞彙未曾出現於期刊文獻的 ID 詞彙中，再進一步將 73 個主題詞彙進行歸類整理後發現大致上可分為 11 個主題，如表 4-8 無法對應到的會議主題名稱之統整，分別是 Question and answering、Routing and filtering、Cross language(IR)、Multi-lingual IR、Summarization、Clustering、(IR) Theory、Distributed(IR)、Structured data、Structured documents & XML 和 Machine learning for IR。

雖然這 11 個主題無法對應至期刊文獻中，但並不表示期刊文獻中沒有這些主題的文獻，而是在 ID 詞彙中沒有出現這些主題詞，因此本研究再利用主題萃取系統將期刊文獻各年度各自歸類，並取出無法對應之主題，補足此部份不足之處，其結果如附錄一之表二。

表 4- 8 會議文獻無法對應到期刊文獻之主題表

會議主題名稱	篇數	總計
Question and answering	20	20
Spam	3	52
Recommendation: use and abuse	3	
Topic Detection and Tracking	6	
Novelty and topic change	3	
Routing	2	
Routing and filtering	3	
Collaborative Filtering	3	
Filtering	14	
filtering & collaborative	4	
Fusion	11	
Cross language(IR)	25	25
Multi-lingual IR	4	19
Asian	11	
Processing for Japanese Text	2	
MIR	2	
Text Summarization	2	26
Summarization & Sentence Selection	24	

Clustering	22	22
(IR) Theory	22	22
Distributed Systems	4	29
Distributed(IR)	25	
Integrating Structured and Unstructured Information	3	9
Structured data	3	
Using Structure	3	
Text Structures	3	12
XML retrieval	3	
Structured documents	2	
Structured Text	2	
Document Structure	2	
Machine learning for IR	6	10
Learning	4	

(三) 綜合討論

由上述可知，會議文獻與期刊文獻的主題比對，在無法對應的主題類目大致上可分為三部份：一般詞彙、相近詞彙、特殊主題詞彙。一般詞彙像是 IR、Science、Information 等，無法從字面上判斷該文獻主題所探討的議題；相近詞彙如 SEARCH(ING) BEHAVIOR 與 User behavior；特殊主題詞彙則是指 Digital Library、Scientometrics、Search Behavior 等具有明顯可看出文獻主題之詞彙，以下以這三部份的詞彙，其無法對應之原因進行探討。

一般詞彙在 SIGIR 會議文獻中並未被看見，可能是因為 SIGIR 會議文獻的主題詞通常具有代表性，較少出現一般詞彙，而期刊文獻在給予主題詞時可能以最廣的範圍給定主題，所以會出現 IR、Science 等較廣的辭彙。用語上的差異也是造成兩邊主題無法對應的原因之一，其詞彙之統整如表 4-9，由表中可知在 ID 詞彙中的 USER 無法由 SIGIR 會議文獻主題中找到，但卻與 SIGIR 會議文獻主題的類別 Personalization 很相似。而特殊主題詞彙則可能代表某些主題原先於會議文獻與期刊文獻中即有各自偏重之主題，亦可能是這些類別的研究人員大多喜好投稿其他類型的期刊或會議而非本研究中所探討的核心期刊與 SIGIR 會議。

表 4- 9 會議文獻與期刊文獻主題相似但詞彙不同之主題

SIGIR 會議文獻主題	期刊文獻主題
Link analysis	WORLD-WIDE-WEB(WWW,WEB)
Passage Retrieval	MODEL(S)
Mathematical Models	
Data Structures	RETRIEVAL EFFECTIVENESS
Dimensionality reduction	
Signatures	INDEX TERMS
Linguistic Analysis	LANGUAGE(S)(LANGUAGE IR)
Personalization	USER(S)
Query and Profile Modification	QUERY(S)
Query Refinement	
INFORMATION-RETRIEVAL SYSTEMS	SYSTEM(S)
VOCABULARY PROBLEM	TERM(S)

在 SIGIR 會議文獻無法找到相對應的期刊 ID 詞彙亦不少，然而對照 DE 詞彙和主題萃取系統輔助歸類後卻可發現，期刊文獻中仍有這些主題文獻存在，此即表示 ID 詞彙仍有部份無法表達出文獻的主題，但 DE 於本研究的研究樣本中，其遺漏值仍佔大部分，此亦顯示出資料庫本身仍有很大的空間可以改善 ID 詞彙與 DE 詞彙。

(四)　無法對應主題之補足

本研究在比對兩者文獻後發現，部份主題是無法相對應上的，為求嚴謹本研究以主題萃取系統將各年代期刊文獻進行歸類後 ，再逐一由人工檢視挑出 ID 詞彙無法對應的會議主題，補足 ID 詞彙缺漏之遺憾。

此部份即先將下載的 794 篇文獻將每個年代依序分開存檔，再透過曾元顯教授所開發的主題萃取系統，將每年度的期刊文獻集進行共現字歸類，即可得到 1991 年至 2007 年共 17 個檔案各自歸類之情形，再利用無法對應到期刊文獻的 SIGIR 會議文獻主題詞，以人工方式比對主題萃取系統所歸類的期刊文獻，即可

找出會議主題無法比對到的期刊文獻，利用自動化歸類之判讀方式如圖 4-3，由歸類後的文獻中找出需要的主題詞彙與文獻數量。

圖 4- 1 主題系統萃取歸類期刊文獻之範例

第二節　文獻主題與年代分佈

本研究將會議文獻與期刊文獻相同的主題文獻進行統計後，依年代序列分別依照主題相關度將其並列，以便看出相近主題下的文獻數量增長情形，然因採用人工之方式進行歸類時，皆需介入人工之判斷，因此無法完全達到公正客觀的角度，故最後再以系統進行會議文獻與期刊文獻一併歸類，一方面可比較與人工判斷之間的差異性，另一方面亦可補強前述之公正客觀的特性。

一、　主題整併之結果

在進行歸類前即可看出共有 33 個主題為字面上相同之主題詞彙，但有部分主題因數量過少且文獻集中於特定年代中，無法看出年代變化，因此於第一種主題整併後先將此類文獻群淘汰，僅以數量較多或年代分散之主題類別進行比較分析。

(一) 第一種主題整併結果

由附錄一的表三可知其各年代文獻數量的分佈，表中期刊文獻之主題是以字母皆為大寫所表示；會議文獻的主題則是以僅有開頭字母為大寫表示之；期刊文獻之主題前後加上『_』則表示以主題萃取系統進行輔助歸類之主題。

1.　整體趨勢之探討

將此區的主題整併後會議文獻共 479 篇，期刊文獻共 663 篇，總計 18 個主題，以 3 個年代區間來觀看 18 個主題的文獻總數量趨勢，則可看出在前後兩階段各主題的文獻消長情形，圖 4-2 和圖 4-3 為會議文獻的 18 個主題趨勢；圖 4-4 和圖 4-5 為期刊文獻的 18 個主題趨勢。在會議文獻的部份可發現，主題文獻數量的成長曲線大致上可分為四種：

(1) 逐年增長之主題：Web、Retrieval Model(S)、Evaluation、Routing and filtering、Distributed(IR) 、Fusion、(IR) Theory

(2) 第一區間下降而後成長：Classification&Categorisation(Classifiers) 、

Queries、Relevance Feedback

(3) 第一區間成長而後趨於平穩或下降：User Issues、Efficiency、Cross language(IR) 、Multi-lingual IR、Clustering

(4) 逐年下降之主題：Query Expansion、Interface Design and Display、Natural Language Processing

其中 Web 是成長率最為驚人的主題，Retrieval Model(S)和 Evaluation 次之，(IR) Theory 也是在近年來成長較快的主題。

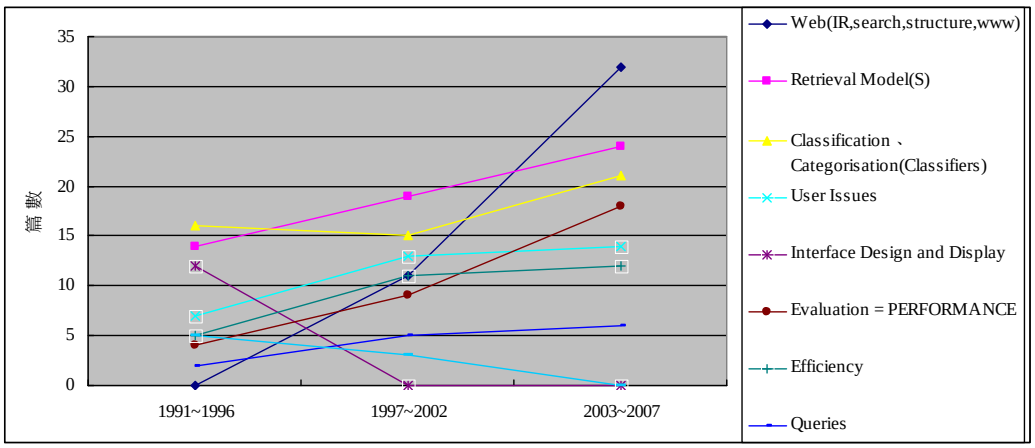

圖 4- 2 會議文獻第一種主題整併後 18 個主題整體趨勢圖 I

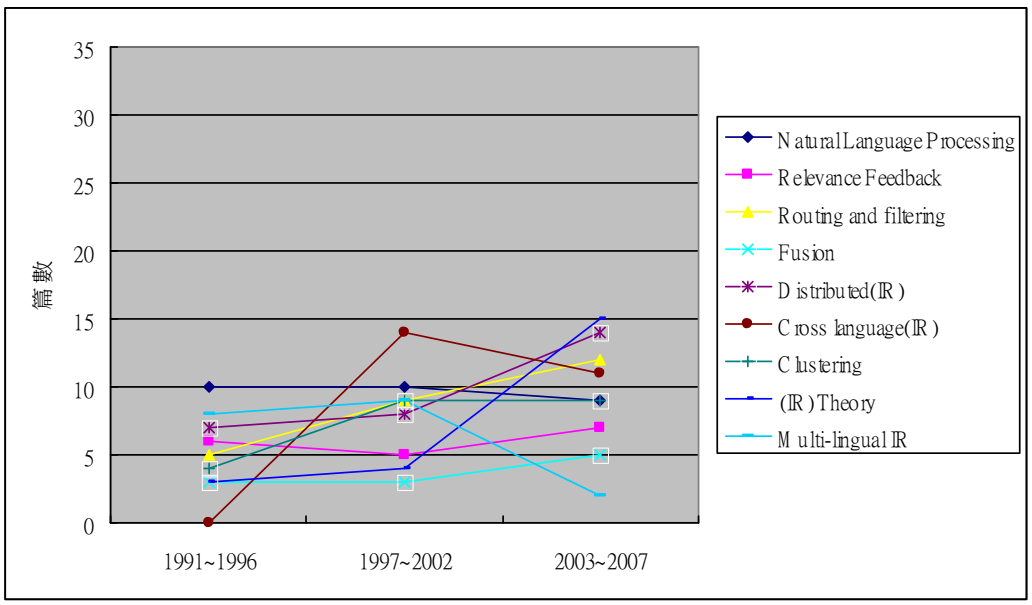

圖 4- 3 會議文獻第一種主題整併後 18 個主題整體趨勢圖 II

在期刊文獻的部份則可發現，大部分的主題文獻數量都是呈現正成長，除了MODEL(S)和 RELEVANCE FEEDBACK 在近年度有下降的趨勢。

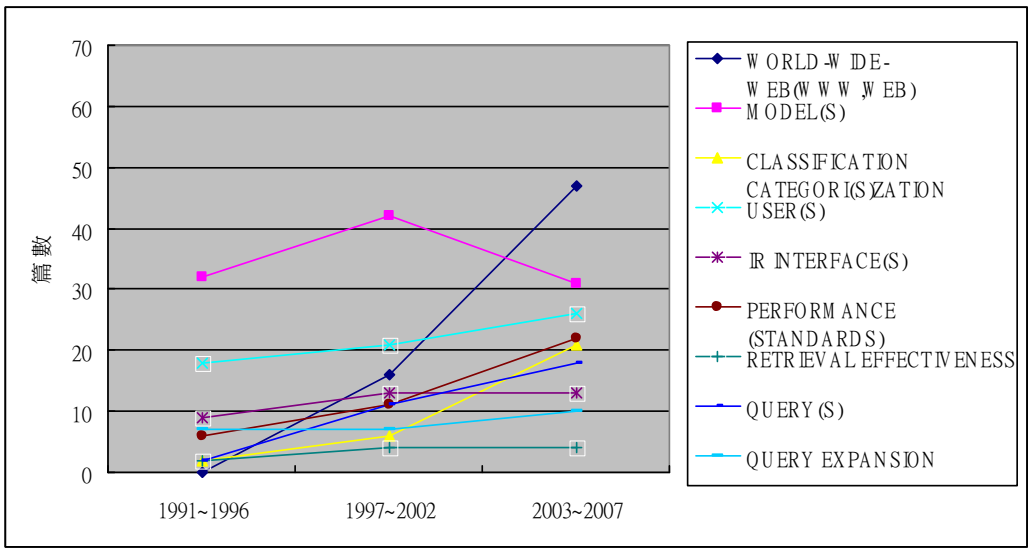

圖 4-4 期刊文獻第一種主題整併後 18 個主題整體趨勢圖 I

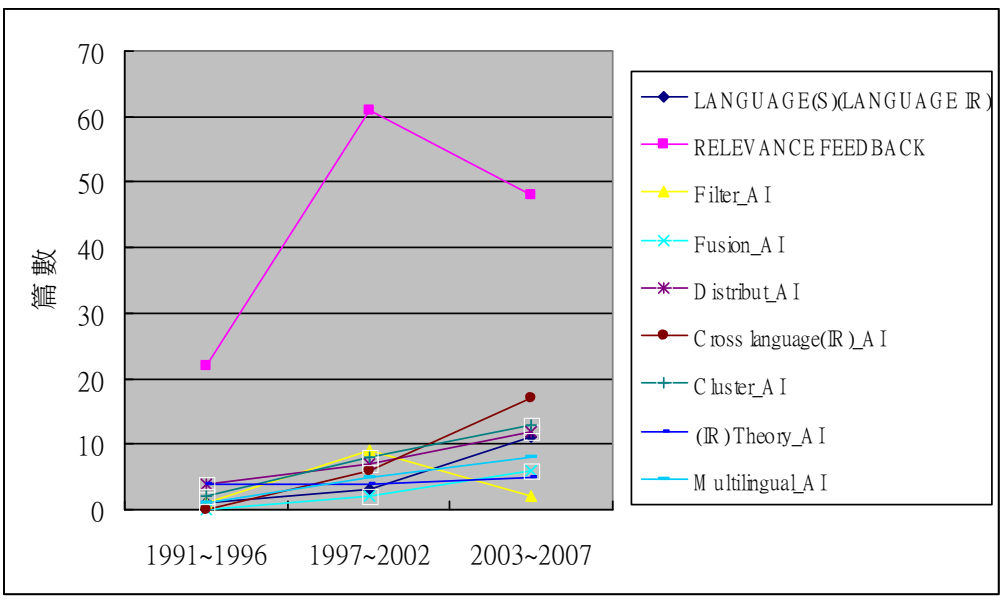

圖 4-5 期刊文獻第一種主題整併後 18 個主題整體趨勢圖 II

以上是以整體的觀點來看會議和期刊文獻各主題的文獻數量分佈，藉此可知文獻數量在三個年代區段的增長情形，但尚未比較相同主題下會議文獻和期刊文獻在各年之間的數量差異與分佈。

2. 各主題文獻數量分佈與趨勢探討

由相同主題下比較會議和期刊文獻的數量差異與分佈，則可以看出文獻出現的先後順序以及成長曲線，以下分別是 18 個主題在會議文獻與期刊文獻主題在相同主題下，各年度所呈現的文獻數量分佈：

(1) Web 和 WORLD-WIDE-WEB

由圖 4-6 可看出有關 WWW 主題之期刊和會議文獻的出現年代同樣都始於 2000 年迄今，且在數量上的分佈比率也相近，在 2000 年至 2004 年的期刊文獻的成長率高於會議文獻，2005 年以後則是會議文獻追上期刊文獻的數量，但於 2007 年後會議文獻又開始下降，反而是期刊文獻有上升的趨勢。在文獻數量的比例上，會議文獻共出現 43 篇，出現比率佔此區會議文獻的 8.51%；期刊文獻則有 63 次的 ID 詞彙出現，佔此區期刊文獻篇數的 7.61%。

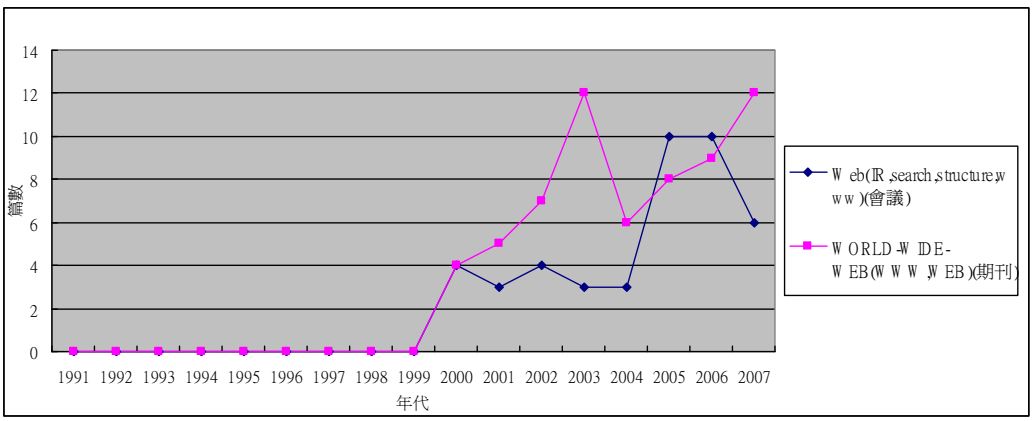

圖 4- 6 以 Web 為主題的會議和期刊文獻各年度數量折線圖

(2) Retrieval Model(S) 和 MODEL(S)

期刊文獻 MODEL(S)的類別共有 105 篇文獻，佔此區期刊文獻的 15.84%，而會議文獻則是分佈於 4 種次主題之下，若以完全相近的主題詞進行比較則有 Information Retrieval Models、Retrieval Model(S)、Formal Models 三種主題共 48 篇文獻，佔此區會議文獻的 11.90%，在此主題下兩者的文獻數量都算偏多，由圖 4-7 可看出兩者之間在數量曲線上都有兩次的高峰期出現，分別是期刊文獻的 1997 年至 1999 年與 1999 年至 2003 年，而會議文獻同樣也於 2000 至 2002 年與

2002 年至 2005 年有相同的高峰期出現。

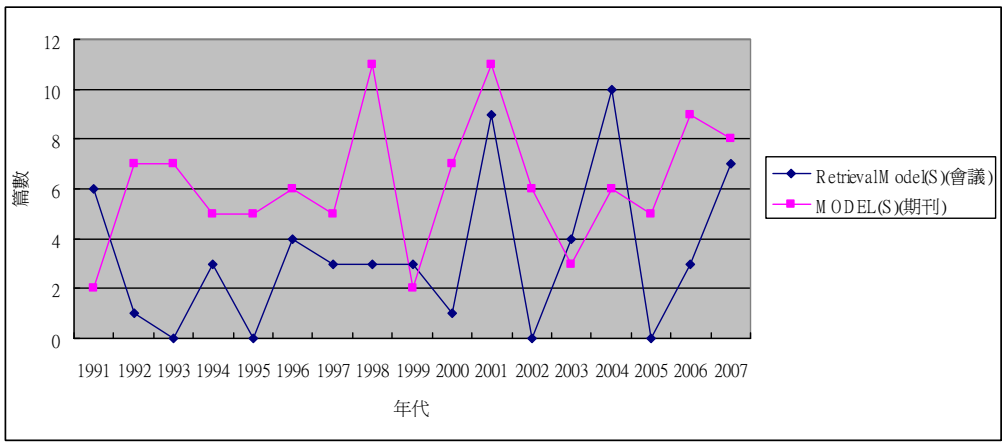

圖 4- 7 以 Model(S)為主題的會議和期刊文獻各年度數量折線圖

(3) Text Categoris(z)ation 和 TEXT CATEGORI(S)ZATION

此主題將 Categorisation 和 Classification 視為同一主題，會議文獻共出現 49 篇，期刊文獻則出現 29 篇，相較之下期刊文獻明顯較少。由圖 4-8 可看出，在年代的部份，會議與期刊文獻出現的時間幾乎相同，都可於 1992 年發現此主題的文獻。且 1999 年以前會議文獻的數量都比期刊文獻多，2000 年對會議文獻來說是此主題的第二波文獻持續成長的階段，而期刊文獻則是在 2001 之後才開始陸續增多，晚了會議文獻約一年的時間，且兩者都於 2004 年走向谷底，而 2006 年之後期刊文獻又追上會議文獻的數量，反而是會議文獻的成長開始趨於平緩。

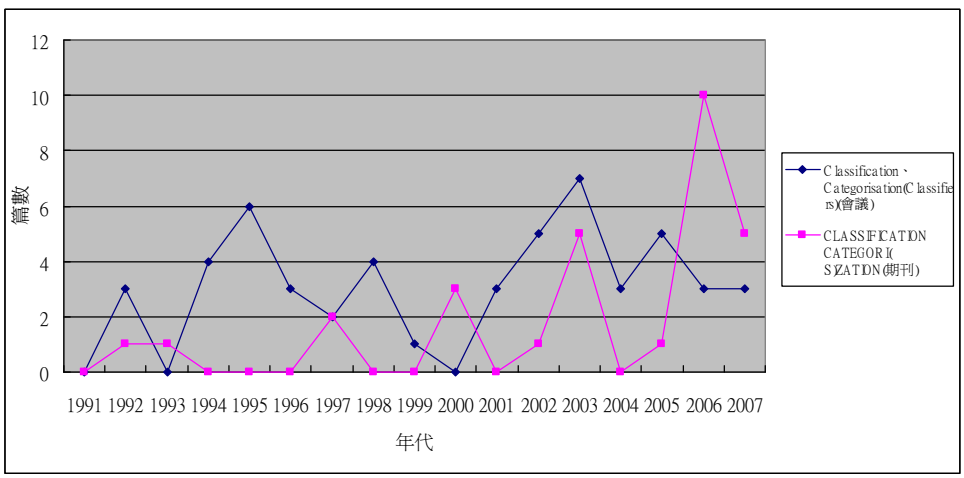

圖 4- 8 以 Text Categoris(z)ation 為主題的會議和期刊文獻各年度數量折線圖

(4) User Issues 和 USER(S)

由此主題的文獻數量上可看出，期刊文獻比會議文獻多一些，期刊文獻總共有 65 篇，佔此區期刊文獻的 9.80%；會議文獻共有 34 篇，佔此區會議文獻的 7.10%。由圖 4-9 可看出，期刊文獻比會議文獻較早約 3 年出現，期刊文獻於 1993 年後數量便開始下降，但此時卻正好是會議文獻開始出現並成長的階段；又期刊文獻於 1998 年出現第一次高峰後即趨於平穩，於 2004 年出現第二次的高峰期，然會議文獻則是於 1998 年後各年間的文獻數量起伏不定，直至 2005 年才又有高峰出現，但兩者都同樣於 2006 年後持平穩的成長水準。

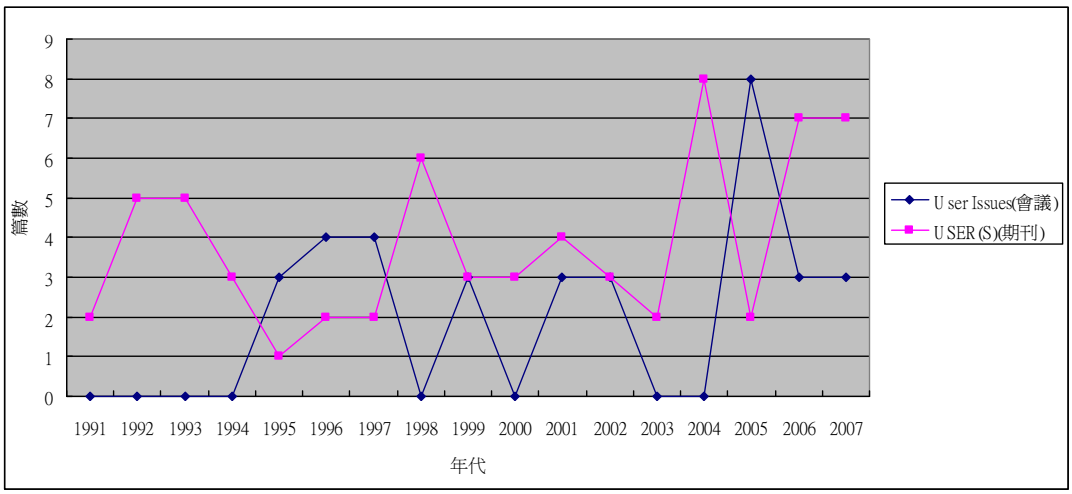

圖 4- 9 以 User 為主題的會議和期刊文獻各年度數量折線圖

(5) Interfaces 和 IR INTERFACE(S)

此主題下可看出有 3 個次主題都很相似，若將 3 個次主題合併來看可知，在數量上會議文獻共有 12 篇，佔此區會議文獻的 2.51%；期刊文獻則有 35 篇，佔此區期刊文獻的 5.28%，相差有 1 倍之多。但在年代分佈上則呈現有趣的現象，由圖 4-10 可看出會議文獻的年代分佈為 1991 年至 1995 年而後則未再出現，但與前一相近主題「User Issues」恰巧可銜接上，由此可知，會議文獻的詞彙已從早年的 Interfaces 在 1995 年以後轉變為 User Issues；期刊文獻則分佈於 1994 之後年，雖較晚於會議文獻，但卻持續有相關文獻產出，與會議文獻有很大的不同。

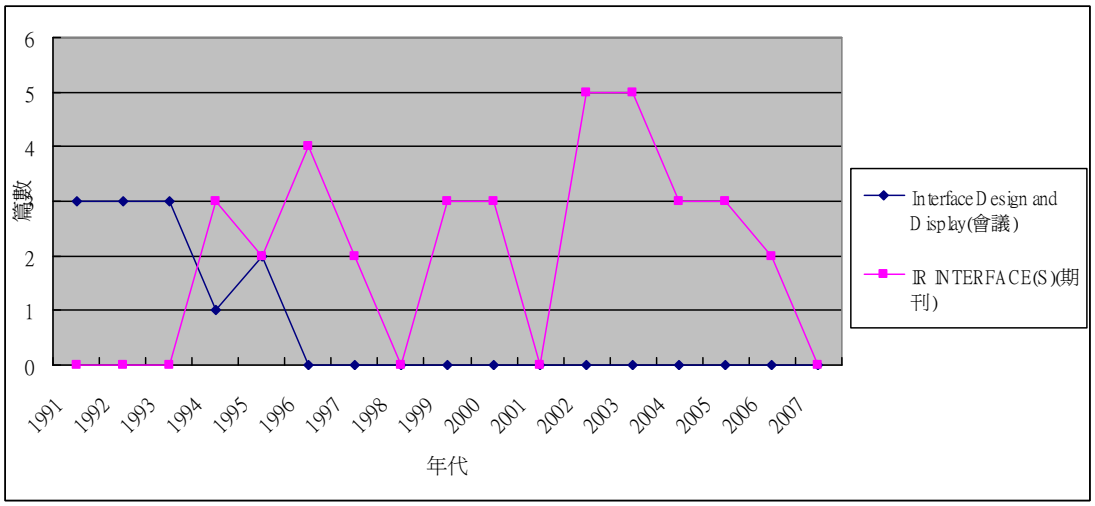

圖 4- 10 以 Interfaces 為主題的會議和期刊文獻各年度數量折線圖

(6) Evaluation 和 PERFORMANCE

會議文獻的主題詞「Evaluation」在期刊文獻中並無找到完全相對應的 ID 詞彙，僅有相近的主題詞「PERFORMANCE」與「SYSTEM PERFORMANCE」可比較之，由此可知會議文獻與期刊文獻在主題詞的用法上會有些許不同。此外兩者在文獻數量上相差不遠，會議文獻若將「Evaluation」和「Performance Evaluation」合併觀之，則共有 31 篇文獻，佔此區會議文獻的 6.47%；期刊文獻則有 39 篇，佔此區期刊文獻的 5.88%。而其年代分佈可由圖 4-11 看出，期刊文獻比會議文獻早約 1 年出現，會議文獻的部份從 1995 年至 1999 年間都無文獻產出，2000、2001、2002 年有少量文獻的產出，直到 2004 年以後才又有持續性的成長；期刊文獻的部份則是從 1994 年後均陸續有文獻產出，於 2003 年達到高峰後又往下降，至 2005 年之後才又開始持續攀升。

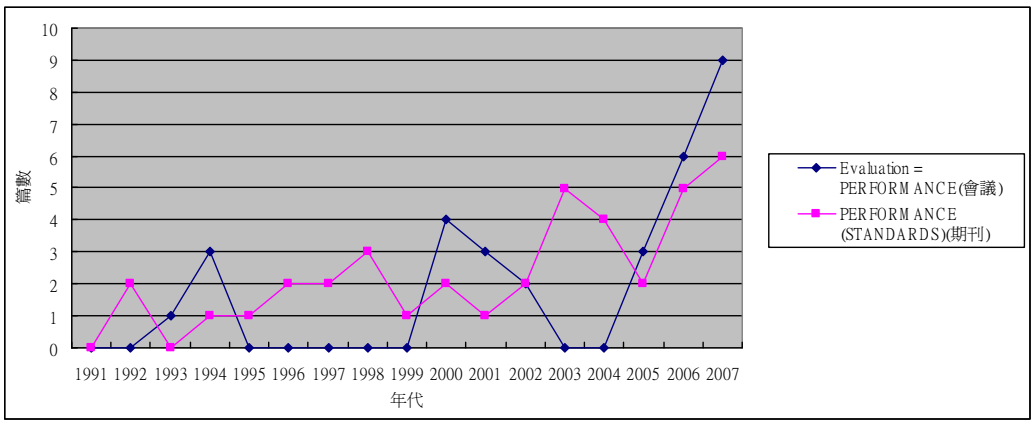

圖 4- 11 以 Evaluation 為主題的會議和期刊文獻各年度數量折線圖

(7) Efficiency 和 RETRIEVAL EFFECTIVENES

該主題下期刊文獻出現的時間較早於會議文獻，期刊文獻於 1992 年即出現，但會議至 1995 年才開始有文獻產生，而數量上會議文獻則是多於期刊文獻，會議文獻共有 28 篇，佔此區會議文獻的 5.85%；期刊文獻則出現 10 篇，佔此區期刊文獻的 1.51%，兩者文獻在數量上都偏少。圖 4-12 則是此主題的文獻年代分佈，雖然期刊比會議文獻早出現 3 年，但會議文獻從 1995 年至 2007 年的篇數大多高於期刊文獻，且兩者數量上都有呈現不連續的狀況。

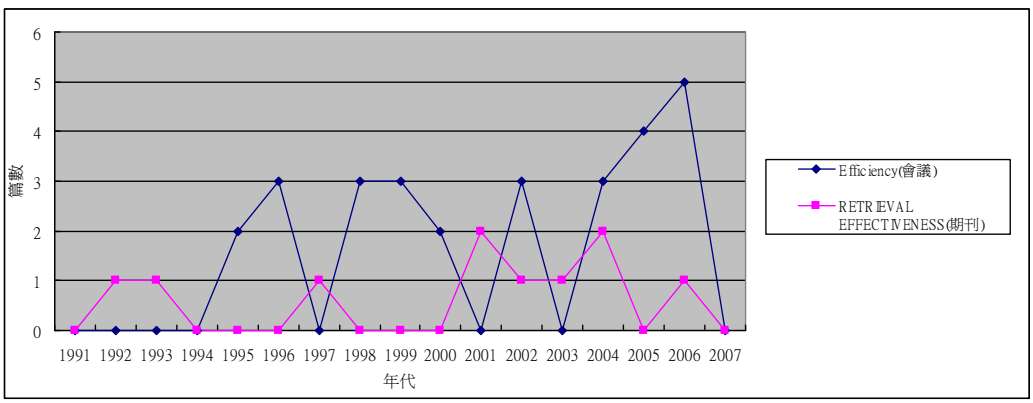

圖 4- 12 以 Efficiency 為主題的會議和期刊文獻各年度數量折線圖

(8) Queries 和 QUERY(S)

會議文獻共出現 13 篇，佔此區會議文獻的 2.71%；期刊文獻則出現 31 篇，佔此區期刊文獻的 4.68%，兩者數量都偏少，由圖 4-13 可看出期刊文獻於 2001 年之後才有明顯的文獻成長，而會議文獻的部份更是僅有 1993、2002、2006、2007 這四個年度有文獻產生。

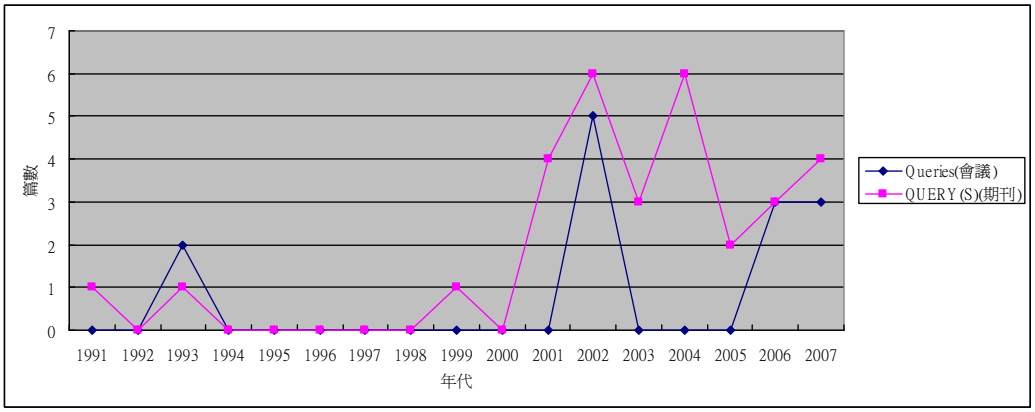

圖 4- 13 以 Queries 為主題的會議和期刊文獻各年度數量折線圖

(9) Query Expansion 和 QUERY EXPANSION

在此主題下，會議主題共有 8 篇，佔此區會議文獻的 1.67%；期刊文獻則出現 24 篇，佔全部期刊文獻的 3.62%，在數量上都偏少。由圖 4-14 可看出在年代分佈上，會議文獻晚期刊文獻約一年的時間出現，且於 1998 年後即未再出現；而期刊文獻則是少量的分佈在各年度中，直到近兩年才有較多的文獻產生。

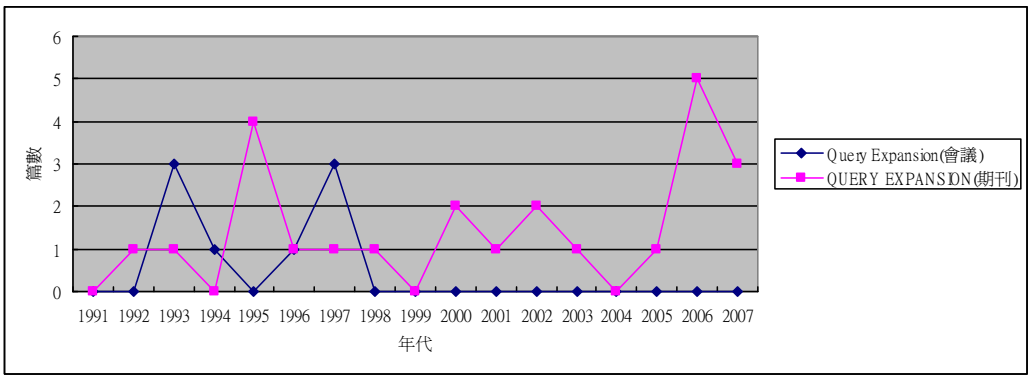

圖 4- 14 以 Query Expansion 為主題的會議和期刊文獻各年度數量折線圖

(10) Natural Language Processing 和 LANGUAGE(S)(LANGUAGE IR)

「Language」在會議文獻的主題中有較多相似主題的出現，其中又以「Natural Language Processing」有 21 篇為數量最多者，而此主題下的會議文獻共 29 篇，佔此區會議文獻的 6.05%；期刊文獻則有 15 篇，佔此區期刊文獻的 2.26%。兩者在數量上都沒有很多，由圖 4-15 可看出在會議文獻比期刊文獻早約二年，且早期的期刊文獻數量皆很少，直到 2004 年之後才有較多的文獻出現，而會議文獻則是陸續都有出現，但與期刊文獻同樣於 2001 年至 2003 年都無文獻產生。

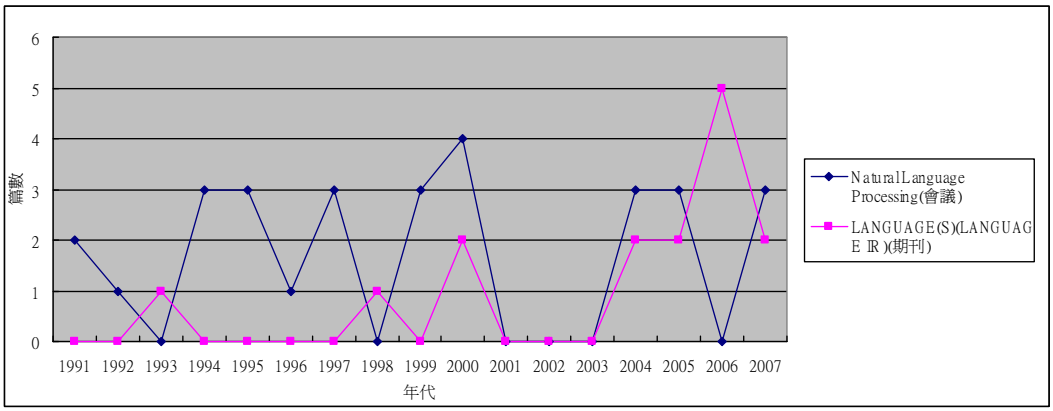

圖 4- 15 以 Language 為主題的會議和期刊文獻各年度數量折線圖

(11) Relevance Feedback 和 RELEVANCE FEEDBACK

在此主題下可看出會議文獻主題僅集中於三個主題下，但期刊文獻的主題分佈則很分散，有許多相近的次主題出現，若將次主題皆合併觀之，則兩者數量差距甚大，期刊文獻高達 131 篇，佔此區期刊文獻的 19.76%，其比例為此區期刊文獻之冠；但會議文獻卻僅有 18 篇，佔此區會議文獻的 3.76%。由圖 4-16 可看出期刊文獻於各年代的篇數都高於會議文獻，由此可知期刊文獻要比會議文獻較重視此主題。

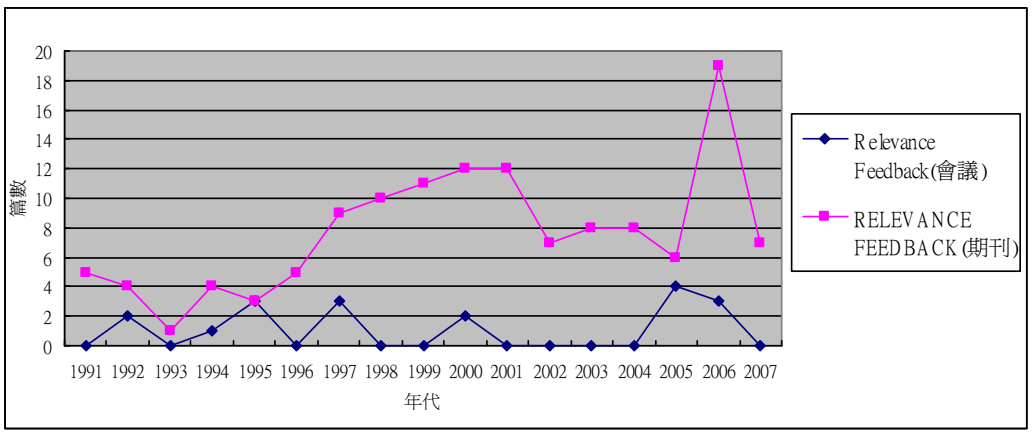

圖 4- 16 以 Relevance Feedback 為主題的會議和期刊文獻各年度數量折線圖

(12) Routing and filtering 和 Filter

此主題的期刊文獻因沒有相對應的 ID 詞彙，故以主題萃取系統進行輔助歸類後在由人工檢視取出此主題下的文獻篇數。在此主題下，會議文獻總共有 26 篇，佔此區會議文獻的 5.43%；期刊文獻則有 12 篇，佔此區期刊文獻的 1.81%。

由圖 4-17 可看出會議比期刊早約 1~2 年的時間，會議文獻於 2002 年達到高峰，但後三年的文獻都持續下降；期刊文獻的部份則是斷斷續續的出現，且每年的數量都不多。

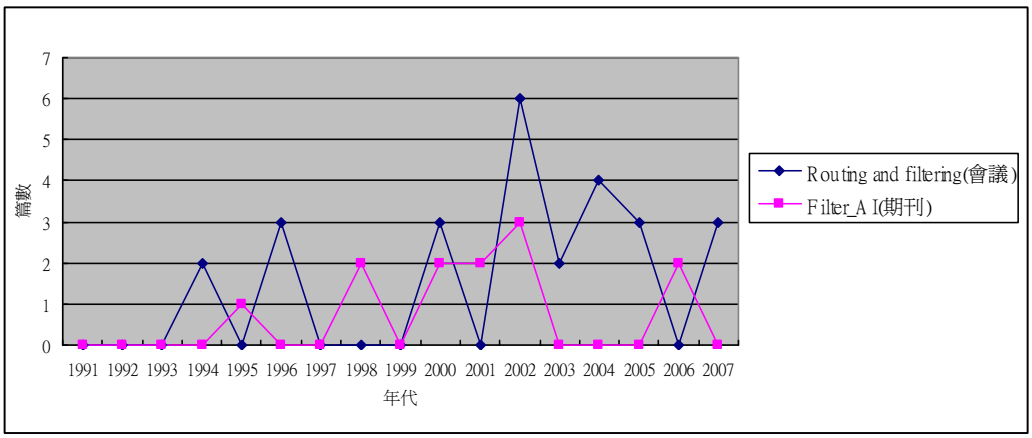

圖 4- 17 以 Routing and filtering 為主題的會議和期刊文獻各年度數量折線圖

(13) Fusion 和 Fusion

此主題的期刊文獻數量也是藉由主題萃取系統的輔助，將期刊文獻中關於此主題的文獻找出。由圖 4-18 可以明顯看出兩者的文獻數量都很少，在年代的部份，會議文獻比期刊文獻早約 4 年左右，但兩者都於 2000 年後沒有文獻產生，直到 2004 年起才有期刊文獻產生，而會議文獻更是到 2006 年後才有文獻出現。

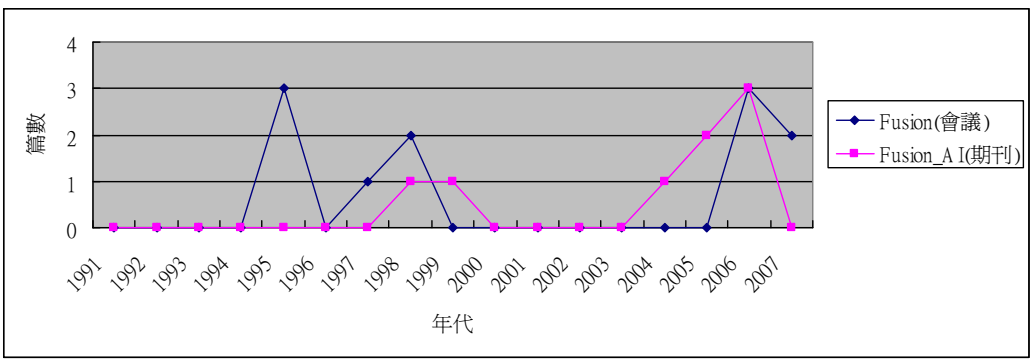

圖 4- 18 以 Fusion 為主題的會議和期刊文獻各年度數量折線圖

(14) Distributed(IR) 和 Distribute

此主題的期刊文獻亦是透過主題萃取系統的輔助，在數量分佈上，會議文獻共有 29 篇，佔此區會議文獻的 6.05%；期刊文獻則有 23 篇，佔此區期刊文獻的

3.47%。由圖 4-19 可看出在年代分佈上以會議文獻於 1991 年首次出現 Distributed Systems 主題為最早，比期刊文獻早約 2 年的時間，但會議文獻亦是斷斷續續的出現，期刊文獻則是於 2005 年後有成長的趨勢。

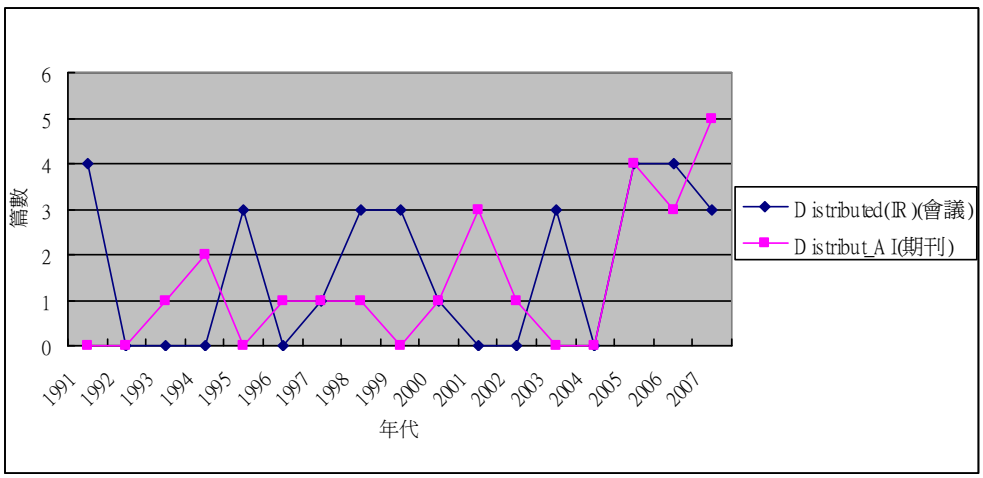

圖 4- 19 以 Distributed 為主題的會議和期刊文獻各年度數量折線圖

(15) Cross language(IR) 和 Cross language

此主題的期刊文獻亦是透過主題萃取系統的輔助，會議與期刊文獻數量相差不大，會議文獻共有 25 篇，佔此區會議文獻的 5.22%；期刊文獻則有 23 篇，佔此區期刊文獻的 3.47%。在年代上可由圖 4-20 可看出，會議文獻約於 1997 年起有文獻產生，比期刊文獻早約兩年，但期刊文獻於 2004 年追上會議文獻的數量，其後則是持平的穩定出現，而會議文獻則是於 2002 年達到高峰後即持平或下降的趨勢。

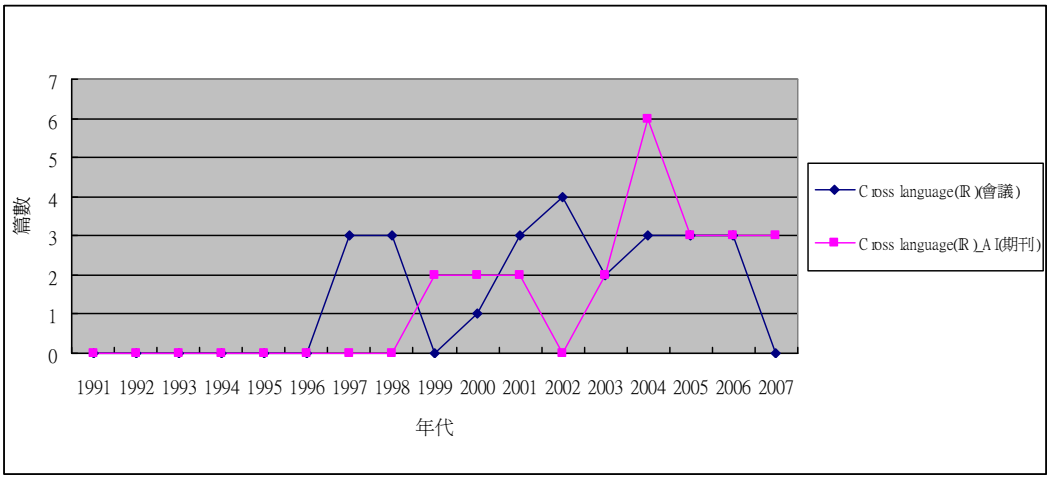

圖 4- 20 以 Cross language 為主題的會議和期刊文獻各年度數量折線圖

(16) Clustering 和 Cluster

此主題的期刊文獻亦是透過主題萃取系統的輔助，數量上兩者相差不大，會議文獻共有 22 篇，佔此區會議文獻的 4.59%；期刊文獻則有 23 篇，佔此區期刊文獻的 3.47%。在年代的分佈上，可由圖 4-21 看出會議文獻約於 1992 年後才開始有此主題之文獻出現，且比期刊文獻早約兩年。會議文獻於各年度中呈現斷斷續續的狀態，並於 2004 年達到高峰；期刊文獻則是於 2000 年後才有穩定的文獻產生，雖然 2004 年沒有期刊文獻，但 2005 年之後又有成長的趨勢。

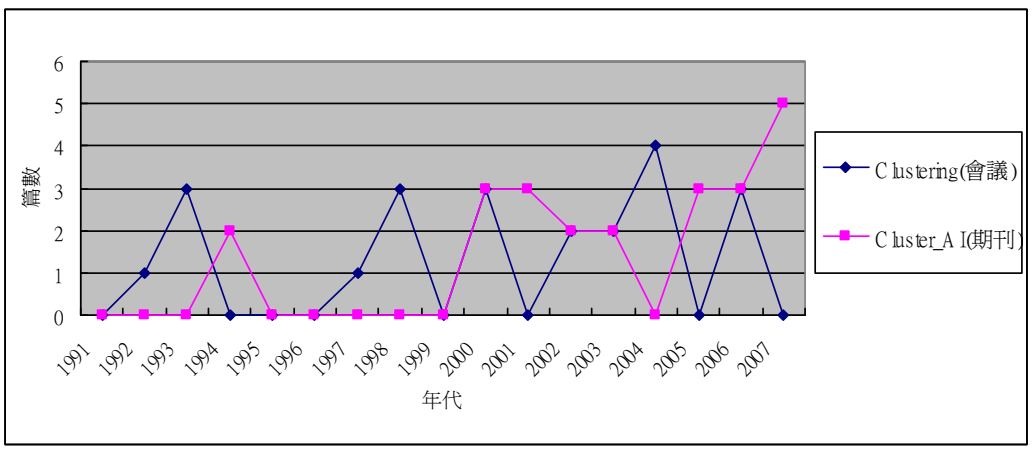

圖 4- 21 以 Cluster 為主題的會議和期刊文獻各年度數量折線圖

(17) (IR) Theory 和 Theory

此主題的期刊文獻亦是透過主題萃取系統的輔助，會議文獻共有 22 篇，佔

此區會議文獻的 4.59%；期刊文獻則有 13 篇，佔此區期刊文獻的 1.96%。期刊文獻的數量偏少，但兩者出現的年代相近，會議文獻僅比期刊文獻早一年，兩者同樣於 1997 至 1999 年間並無任何文獻產生，也都於 2005 年達到高峰，由圖 4-22 為此主題下各年度文獻數量的成長情形。

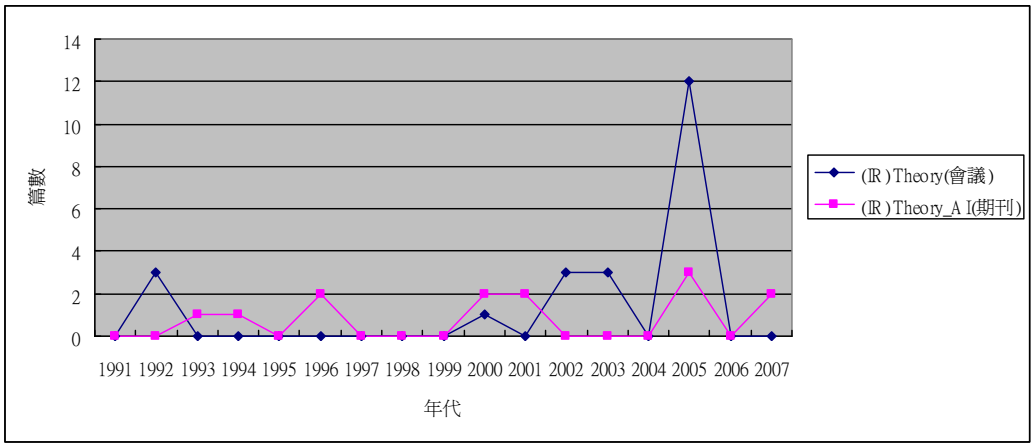

圖 4- 22 以 Theory 為主題的會議和期刊文獻各年度數量折線圖

(18) Multi-lingual IR

此主題的期刊文獻亦是透過主題萃取系統的輔助，此主題的會議文獻共有 19 篇，佔此區會議文獻的 3.97%；期刊文獻則有 14 篇，佔此區期刊文獻的 2.11%，兩者文獻皆不多，但在文獻年代上的差距較大，會議文獻於 1992 年即出現，期刊文獻則要到 1996 年才出現，兩者相差約四年之久，由圖 4-23 為此主題下各年度文獻數量的成長情形。

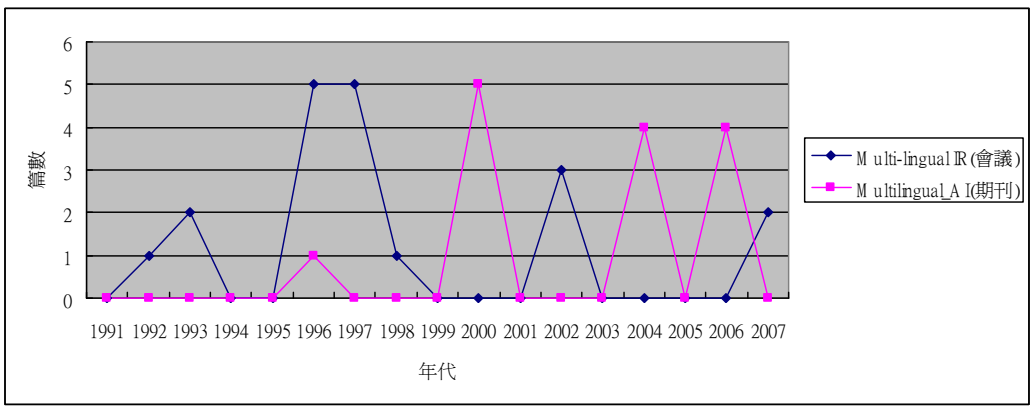

圖 4- 23 以 Multi-lingual IR 為主題的會議和期刊文獻各年度數量折線圖

以上所敘述的 18 個類別，即為第一種主題整併後可比較之主題類別，其統整如表 4-10，由表中的文獻數量與百分比可知，期刊文獻數量最多的主題分別是 Web、Language(Retrieval)(Formal) Models、User Interaction、User Studies、Evaluation&Performance Evaluation、Relevance Feedback；會議文獻數量最多的主題分別是 Web、Language(Retrieval)(Formal) Models、Text Classification& Categorisation(Classifiers)、User Interaction&User Studies、Evaluation&Performance Evaluation。由兩者數量百分比的差距則可看出，差距最大的前五名依序是：

(1) Relevance Feedback

(2) Text Classification&Categorisation(Classifiers)

(3) Efficiency&Scalability

(4) Language(Retrieval)(Formal) Models

(5) Language& NLP(Natural Language Processing)

其中期刊文獻比會議文獻多的主題是 Relevance Feedback 和 Language(Retrieval)(Formal) Models，而會議文獻比期刊文獻多的主題則是 Text Classification&Categorisation(Classifiers)、Efficiency&Scalability 和 Language& NLP(Natural Language Processing)。

兩者數量差距過大的主題亦是很值得探討的議題，在第一階段尚未歸類前即可看出期刊比會議文獻數量上明顯較多的主題有七個，分別是：MODEL(S)、QUERY(S)&QUERY EXPANSION、USER(S)、RELEVANCE、SYSTEM(S)、DATABASE(S)、KNOWLEDGE。此亦代表兩種情形之可能，一為期刊主題的 ID 詞彙含義過廣，導致多篇文獻皆可取用該 ID 詞彙，另一可能性為期刊文獻比會議文獻要重視此主題，如 KNOWLEDGE 即為會議文獻極少出現之主題。

而在初始年代先後的差異上，會議文獻比期刊文獻早的一共有 9 個主題，期刊文獻比會議文獻早的則有 6 個，兩者同時出現的有 3 個，由此可知大部分的主題仍然是以會議文獻較早出現。

表 4- 10 第一種主題整併後會議文獻與期刊文獻統整比較表

編號	主　題	會議篇數/百分比	期刊篇數/百分比	會議與期刊文獻初始年代 (會議早於期刊文獻●) (期刊早於會議文獻△) (會議文獻與期刊文獻年代相同=)
1	Web(IR,search,structure,www)	43	63	=
		8.98%	9.50%	
2	Language(Retrieval)(Formal) Models	57	105	=
		11.90%	15.84%	
3	Text Classification&Categorisation(Classifiers)	52	29	=
		10.86%	4.37%	
4	User Interaction&User Studies	34	65	△
		7.10%	9.80%	
5	Search Interfaces(Issues)	12	35	●
		2.51%	5.28%	
6	Evaluation&Performance Evaluation	31	39	△
		6.47%	5.88%	
7	Efficiency&Scalability	28	10	△
		5.85%	1.51%	
8	Queries&Query Processing	13	31	△
		2.71%	4.68%	
9	Query Expan(d)sion	8	24	△
		1.67%	3.62%	
10	Language& NLP(Natural Language Processing)	29	15	●
		6.05%	2.26%	
11	Relevance Feedback	18	131	△
		3.76%	19.76%	
12	Routing and filtering	26	12	●
		5.43%	1.81%	
13	Fusion	11	8	●
		2.30%	1.21%	
14	Distributed Systems(IR)	29	23	●
		6.05%	3.47%	
15	Cross language(IR)	25	23	●
		5.22%	3.47%	
16	Clustering	22	23	●
		4.59%	3.47%	
17	(IR) Theory	22	13	●
		4.59%	1.96%	
18	Multi-lingual IR	19	14	●
		3.97%	2.11%	
	總計篇數	479	663	

(二) 第二種主題整併結果

第二種主題整併後已將較細的分類進行合併，因此可明顯看出數量偏多或偏少之主題，亦可較容易看出數量上差距過大或年代過於分散的主題，由於合併歸類至第二種主題整併後，尚有三個類別的文獻數量仍少於十篇，故亦不再進行趨勢分析，分別是 Compression、Multimedia(IR)、Summarization&Sentence Selection(multidocuments and new applications)。

1. 整體趨勢之探討

第二種主題整併後排除少量文獻的類別，即出現 11 個大類可以進行趨勢預測分析，將全部年代分成 3 個區段則可看出文獻主題的數量趨勢，圖 4-24 為會議文獻的 11 個主題文獻數量趨勢，其成長曲線大致上可分為四種：

(1) 逐年增長之主題：

- Classification&Categorisation(Classifiers)；Clustering；Filtering；Fusion
- Web；Hypertext；Link analysis
- Performance Evaluation；Test collections
- Indexing(structures)；Term Relationships；Statistics；Question and answering
- Distributed(IR)(Systems)
- Efficiency；Scalability

(2) 第一區間下降而後成長：

- Semantics；LSI；Probabilistic Models；(IR) Theory；Logic；Inference Networks
- User behavior；Human；Interface Design and Display；Visualization

(3) 第一區間成長而後趨於平穩或下降：Natural Language Processing；Cross language；MIR

(4) 逐年下降之主題：

- Relevance Feedback；Query；Query Expansion
- DB&IR(Integration)(Structures)；Systems；Knowledge Based Information Retrieval

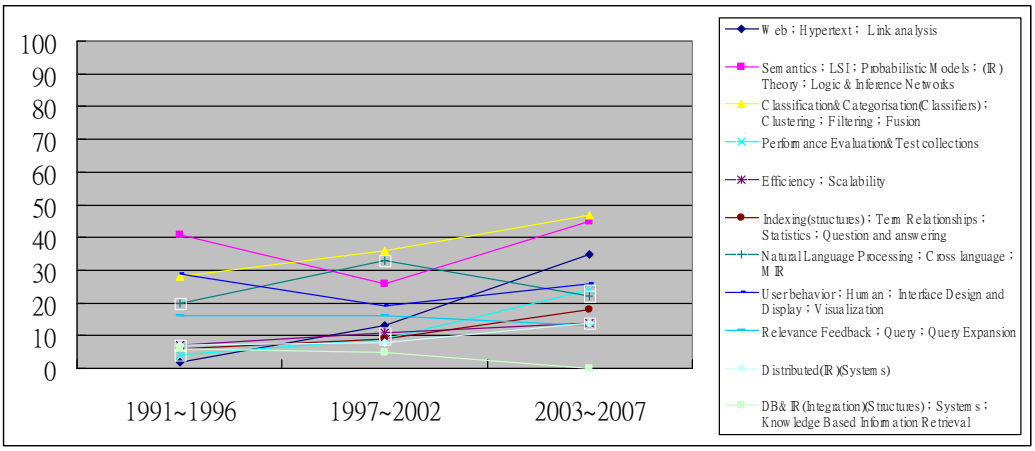

圖 4- 24 會議文獻第二種主題整併後 11 個主題整體趨勢圖

圖 4-25 則為期刊文獻的 11 個主題文獻數量趨勢，其成長曲線大致上可分為二種：

(1) 逐年增長之主題：

- USER(S)；BEHAVIOR(S)；HUMAN；SEARCH INTERFACE ；VISUALIZATION
- WEB；HYPERMEDIA、HYPERTEXT
- CATEGORI(S)ZATION；Cluster；Filter；Fusion
- LANGUAGE；Cross language；Multilingual
- PERFORMANCE；TEST COLLECTIONS
- _Distributed_
- EFFICIENT(EFFICIENCY)

(2) 第一區間成長而後趨於平穩或下降：

- DATABASE(S)；SYSTEM(S)；KNOWLEDGE
- RELEVANCE FEEDBACK；QUERY EXPANSION
- LATENT SEMANTIC (SEMANTICS)；PROBABILISTIC MODEL；(IR) Theory；INFERENCE(S)
- INDEX TERMS ；TERM(S) DEPENDENCE ；VOCABULARY ；_Question and answering_

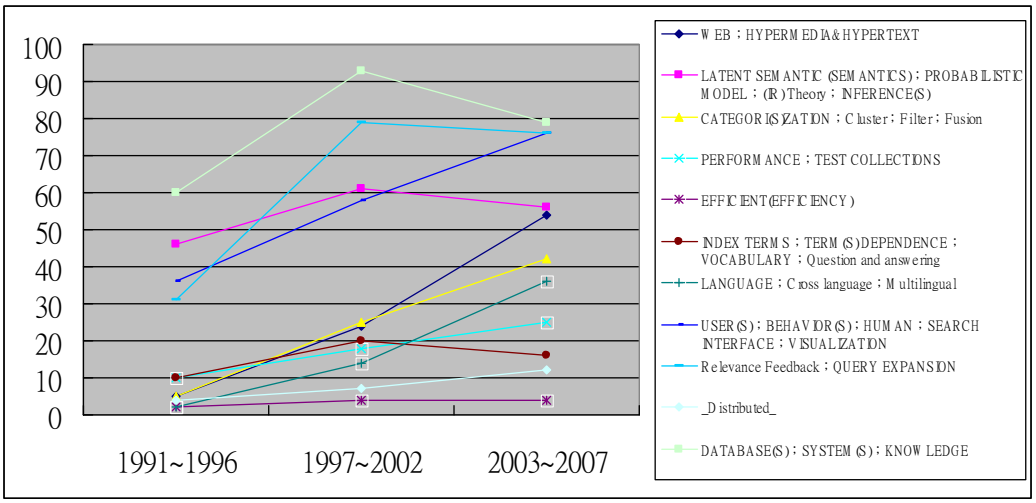

圖 4- 25 期刊文獻第二種主題整併後 11 個主題整體趨勢圖

2. 各主題趨勢之探討

(1) Web；Hypertext；Link analysis

此主題於第一階歸類時即可看出在 Web 的主題下，會議和期刊文獻的出現年代較為一致，但在相關類別 Hypertext 主題下，期刊比會議文獻要早很多，然而在 Hypertext 主題下會議文獻的年代分部較為分散，僅於 1991 年和 2000 年各有兩篇文獻出現，但期刊文獻則是於各年代中皆陸續出現，且可由圖 4-26 看出此主題的文獻於 1999 年之後有大幅的成長。

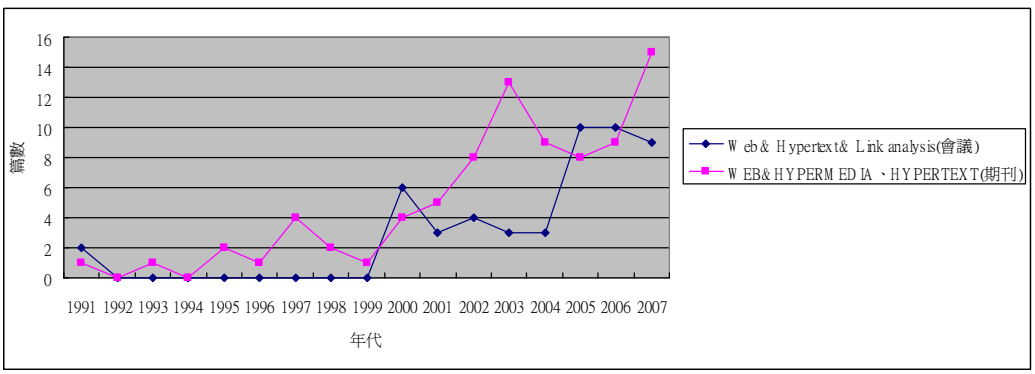

圖 4- 26 Web；Hypertext；Link analysis 的會議和期刊文獻各年度數量折線圖

(2) Semantics；LSI；Probabilistic Models；(IR) Theory；Logic；Inference Networks

此類別的文獻於 1991 年至 2007 年都有相關文獻產生，且會議文獻與期刊文獻都沒有中斷的情形出現，由此可知在此主題下兩者的文獻均不少，由圖 4-27

可看出此大類下的文獻數量成長趨勢。

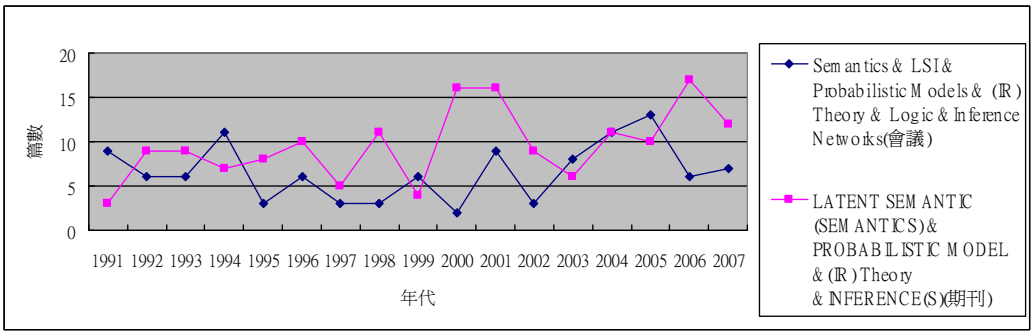

圖 4- 27 Semantics；LSI；Probabilistic Models；(IR) Theory；Logic；Inference Networks 的會議和期刊文獻各年度數量折線圖

(3) Classification&Categorisation(Classifiers)；Clustering；Filtering；Fusion

由圖 4-28 看出，此大類下的會議文獻和期刊文獻在 2002 年以前的文獻成長幅度都很相似，直到 2002 年以後的文獻成長開始出現比較不一樣的趨勢，尤其可看出期刊文獻從 2004 年後的成長幅度大幅提升，並於 2006 年超越會議文獻的數量。

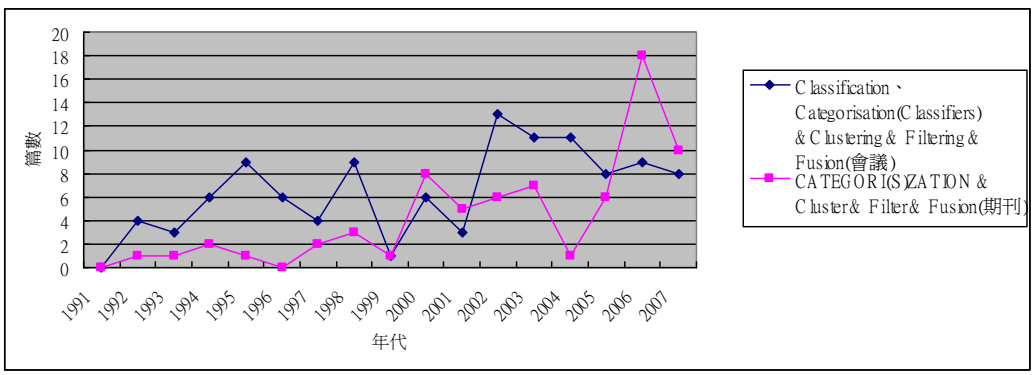

圖 4- 28 Classification&Categorisation(Classifiers)；Clustering；Filtering；Fusion 的會議和期刊文獻各年度數量折線圖

(4) Performance Evaluation；Test collections

此主題的會議文獻於 1995 年至 1999 年並未有文獻產生，但期刊文獻則是幾乎每年都有，在數量上則是以期刊文獻比會議文獻多，但其差距並不大，由圖 4-29 此主題下各年度文獻成長的趨勢圖，由圖中可看出兩者文獻的數量同時於 2000 年後開始下降，也同時於 2006 年大幅提升。

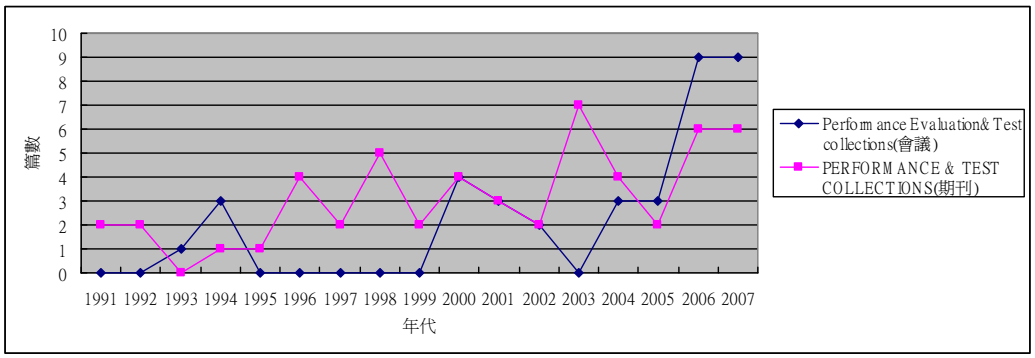

圖 4- 29 Performance Evaluation；Test collections 的會議和期刊文獻各年度數量折線圖

(5) Efficiency；Scalability

由圖 4-30 可看出此主題下的會議在每年度的文獻數量經常出現斷斷續續的情況，而期刊文獻亦是如此，且兩者數量都偏少。

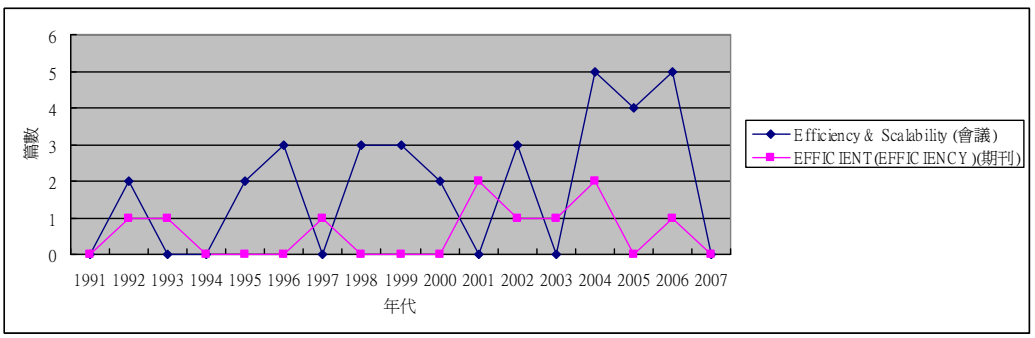

圖 4- 30 Efficiency；Scalability 的會議和期刊文獻各年度數量折線圖

(6) Indexing(structures)；Term Relationships；Statistics；Question and answering

由圖 4-31 可看出會議和期刊文獻在此主題的文獻成長曲線，於 2003 年以前的文獻成長情形可發現，期刊文獻的成長率起伏高低大致上均比會議文獻早約 1~3 年，如 1991 年至 1995 年的期刊文獻與 1992 年至 1996 年的會議文獻、1995 年至 1999 年的期刊文獻與 1998 年至 2002 年的會議文獻、1999 年至 2003 年的期刊文獻與 2002 年至 2004 年的會議文獻。

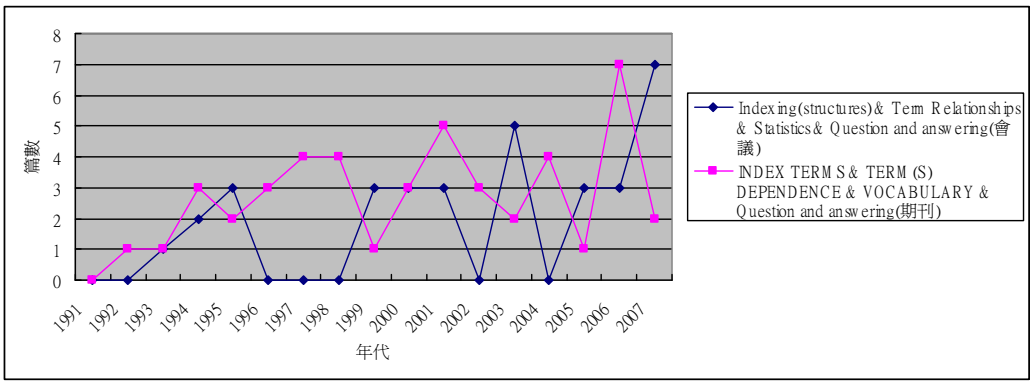

圖 4- 31 Indexing(structures)；Term Relationships；Statistics；Question and answering 的會議和期刊文獻各年度數量折線圖

(7) Natural Language Processing；Cross language；MIR

此主題的會議與期刊文獻數量分部呈現很有趣的情形，由圖 4-32 可看出大致上以 1999 年為界線，1999 年以前會議文獻的數量平均比期刊文獻要多，除了 2002 年沒有刊文獻以外，1999 年之後期刊文獻都超越會議文獻的數量。

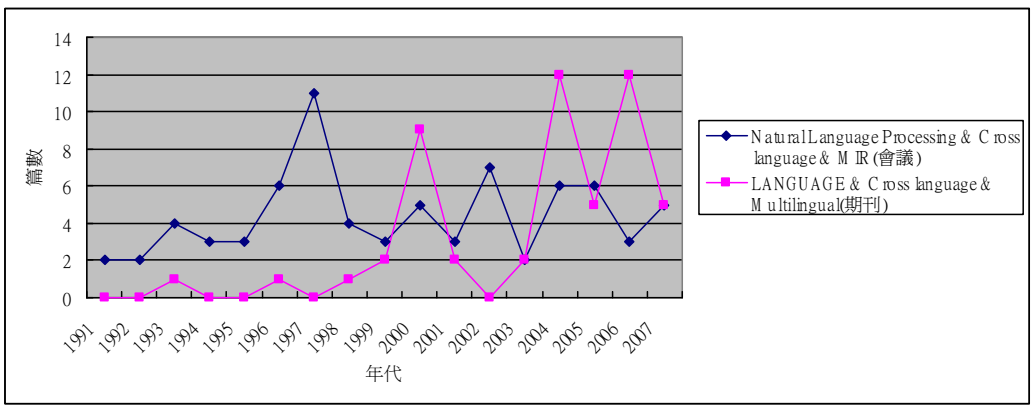

圖 4- 32 Natural Language Processing；Cross language；MIR 的會議和期刊文獻各年度數量折線圖

(8) User behavior；Human；Interface Design and Display；Visualization

由圖 4-33 可看出此主題下的會議與期刊文獻成長趨勢，其中可發現會議文獻與期刊文獻成長率的相似曲線中可看出大約早 1 年，如 1997 年至 2000 年的會議文獻與 1998 年至 2001 年的期刊文獻、2002 年至 2006 年的會議文獻與 2003 年至 2007 年的期刊文獻。

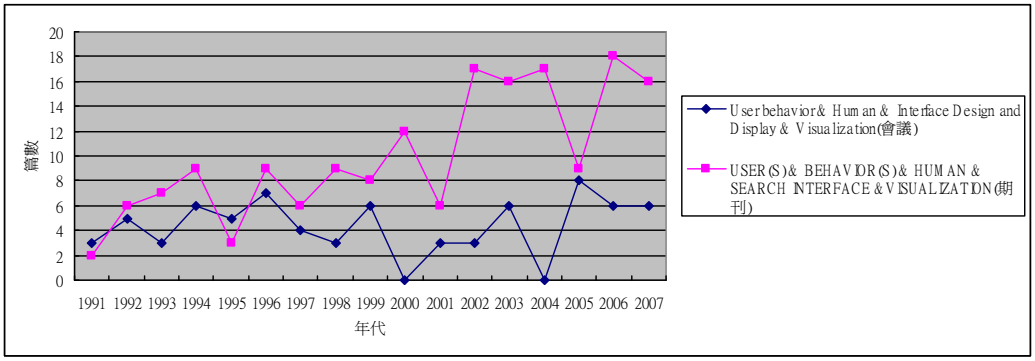

圖 4- 33 User behavior；Human ；Interface Design and Display；Visualization 的會議和期刊文獻各年度數量折線圖

(9) Relevance Feedback；Query；Query Expansion

由圖 4-34 可看出此主題下會議文獻與期刊文獻在數量上以期刊文獻較多，其成長曲線也有類似的情形可以看出，如會議文獻從 1991 年至 2000 年與期刊文獻從 1993 年至 2004 年，兩者之間的差距早期約相差 2 年，但會議文獻從 1997 年至 1999 年的文獻成長開始走下坡，然期刊文獻卻是從 2001 年至 2003 年才走下坡，兩者之間的差距達到 4 年之久。

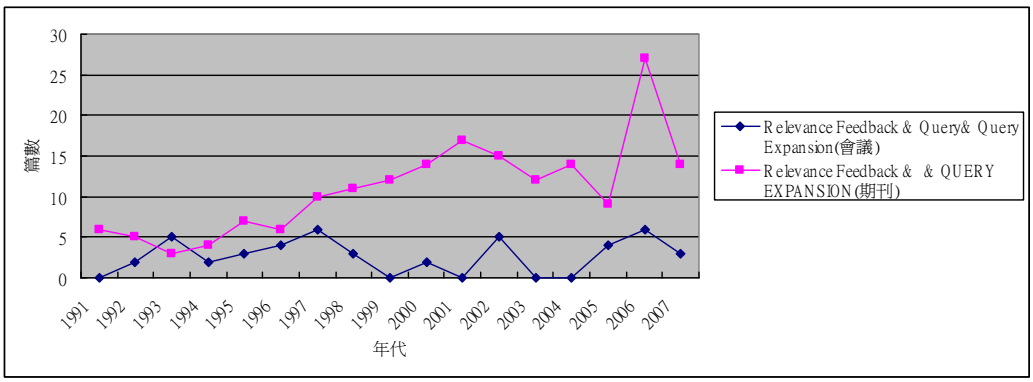

圖 4- 34 Relevance Feedback；Query；Query Expansion 的會議和期刊文獻各年度數量折線圖

(10) Distributed(IR)(Systems)

由圖 4-35 可看出此主題下會議文獻與期刊文獻在數量上都偏少，造成成長曲線起伏很大，雖然兩者之間數量上差距不大，但仍不易看出兩者之間的關連性。

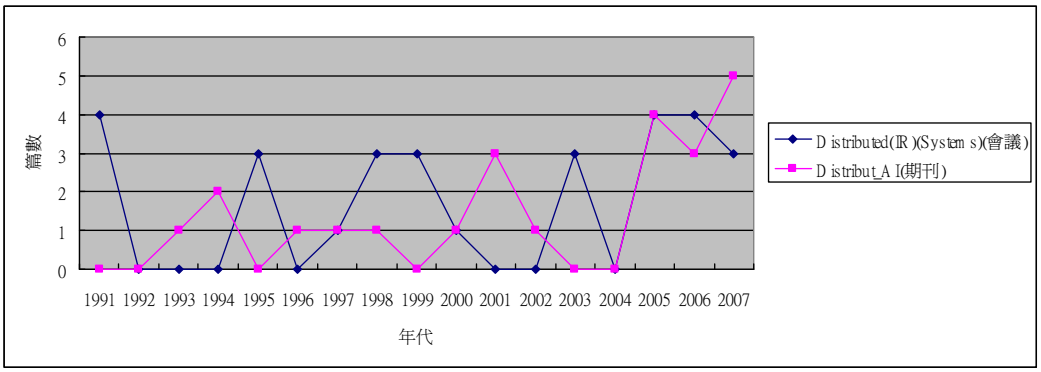

圖 4- 35 Distributed(IR)(Systems) 的會議和期刊文獻各年度數量折線圖

(11) DB&IR(Integration)(Structures)；Systems；Knowledge Based Information Retrieval

由圖 4-36 可看出，期刊文獻比會議文獻要多出很多，會議文獻出現的時間僅 1992 年至 1993 年、2000 年與 2001 年，再查附錄五即可發現兩者數量上的差距更多達 200 多篇。

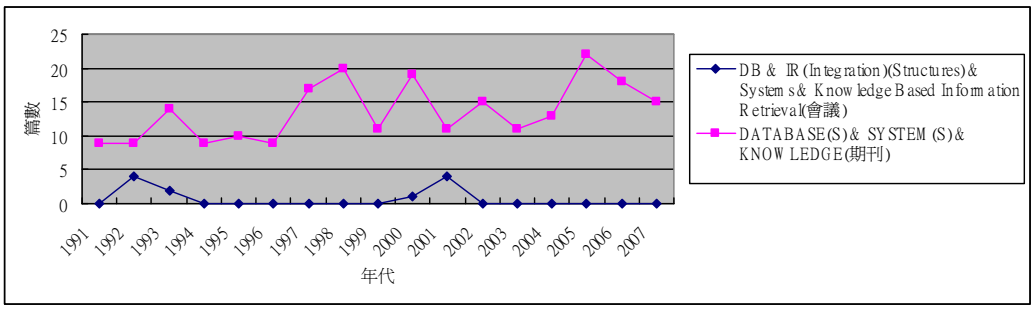

圖 4- 36 DB&IR(Integration)(Structures)；Systems；Knowledge Based Information Retrieval 的會議和期刊文獻各年度數量折線圖

以上 11 個類別統整如表 4-11，由表中可知，會議文獻與期刊文獻所佔比例較多的前五名主題中，有 3 個主題是相同的，分別是 Web；Hypertext；Semantics；LSI；Probabilistic Models；(IR) Theory；Logic；Inference Networks、User behavior；Human；Interface Design and Display；Visualization。而兩者之間數量百分比差距最大的前三個主題分別是:

(1) DB&IR(Integration)(Structures)；Systems；Knowledge Based Information Retrieval

(2) Classification；Categorisation(Classifiers)；Clustering；Filtering；Fusion；Efficiency；Scalability

(3) Relevance Feedback；Query；Query Expansion

由初始年代上的差異則可看出，兩者幾乎同時出現的主題共有 5 個；會議比期刊文獻早的主題有 2 個；期刊比會議文獻早的主題有 4 個，因此在第二種主題整併的方式下來看，兩者同時出現的比例較高。

表 4- 11 第二種主題整併後會議文獻與期刊文獻統整比較表

編號	主題詞彙	會議篇數	期刊篇數	會議與期刊文獻初始年代 (會議早於期刊文獻●) (期刊早於會議文獻△) (會議文獻與期刊文獻年代相同=)
1	Web；Hypertext	53	83	=
		8.51%	7.61%	
2	Semantics；LSI；Probabilistic Models；(IR) Theory；Logic；Inference Networks	118	163	=
		18.94%	14.94%	
3	Classification；Categorisation(Classifiers)；Clustering；Filtering；Fusion	111	72	=
		17.82%	6.60%	
4	Performance Evaluation；Test collections	37	52	△
		5.94%	4.77%	
5	Efficiency；Scalability	32	10	=
		5.14%	0.92%	
6	Indexing(structures)；Term Relationships；Statistics；Question and answering	33	46	△
		5.30%	4.22%	
7	Natural Language Processing；Cross language；MIR	75	52	●
		12.04%	4.77%	
8	User behavior；Human；Interface Design and Display；Visualization	73	170	=
		11.72%	15.58%	
9	Relevance Feedback； Query；Query Expansion	48	186	△
		7.70%	17.05%	
10	Distributed(IR)(Systems)	29	25	●
		4.65%	2.29%	
11	DB &IR(Integration)(Structures)；Systems；Knowledge Based Information Retrieval	14	232	△
		2.25%	21.26%	
	總計	623	1091	

(三) 兩種主題整併結果比較

第一種主題整併乃是以較細的主題進行觀察，而第二種主題整併則是以較宏觀的角度觀察，兩者之間的整體主題趨勢之比較分別以會議文獻與期刊文獻兩種角度分別探討如下。

1. 會議文獻兩種主題整併結果比較

表 4-12 為會議文獻兩種主題整併之比較，由表中可知，第二種主題整併下有部分主題是第一種主題整併所未曾出現的，是透過許多相似的主題所合併而成，例如 User behavior；Human；Interface Design and Display；Visualization。

此外，有些小主題透過整併後，影響大主題的趨勢走向。例如第二種主題整併下的 Filtering 將原先位於第Ⅱ區的 Classification&Categorisation(Classifiers)提升至第Ⅰ區；而原先位於第Ⅱ區的 Queries 則是於第二種主題整併下因 Relevance Feedback 下降至第Ⅳ區；Efficiency 也因 Scalability 而被提升至第Ⅰ區。

表 4- 12 會議文獻第一種與第二種主題整併結果比較表

主題整併方式 / 趨勢走向	第一種主題整併	第二種主題整併
第Ⅰ區 (逐年增長之主題)	Web Retrieval Model(S) Evaluation Routing and filtering Distributed(IR) Fusion (IR) Theory	Classification&Categorisation(Classifiers)；Clustering；Filtering；Fusion
		Web；Hypertext；Link analysis
		Performance Evaluation；Test collections
		Indexing(structures)；Term Relationships；Statistics；Question and answering
		Distributed(IR)(Systems)
		Efficiency；Scalability
第Ⅱ區 (第一區間下降而後成長)	Classification&Categorisation(Classifiers) Queries Relevance Feedback	Semantics；LSI；Probabilistic Models；(IR) Theory；Logic；Inference Networks
		User behavior；Human；Interface Design and Display；Visualization
第Ⅲ區 (第一區間成長而後趨於平穩或下降)	User Issues Efficiency Cross language(IR) Multi-lingual IR Clustering	Natural Language Processing；Cross language；MIR
第Ⅳ區 (逐年下降之主題)	Query Expansion Interface Design and Display Natural Language Processing	Relevance Feedback；Query；Query Expansion
		DB&IR(Integration)(Structures)；Systems；Knowledge Based Information Retrieval

2. 期刊文獻兩種主題整併結果比較

由於期刊文獻於第一種主題整併下之趨勢皆大多為逐年成長，故僅須找出第二種主題整併下所造成之差異方可瞭解主題整併下細部主題所造成之差異 。表 4-13 為期刊文獻於兩種主題整併之結果比較 ，由表中可知 QUERY EXPANSION 原先為第 I 區逐年成長之主題，然而受到 RELEVANCE FEEDBACK 主題逐年下降之影響，造成主題趨勢走向有所轉變 ；而_(IR)Theory_和 INFERENCE(S)原先也是第 I 區之主題，然而於第二種主題整併時受到 LATENT SEMANTIC (SEMANTICS)和 PROBABILISTIC MODEL 的影響而趨勢轉變為第Ⅱ區。

表 4- 13 期刊文獻第一種與第二種主題整併結果比較表

主題整併方式 / 趨勢走向	第一種主題整併	第二種主題整併
第 I 區 (逐年增長之主題)	全部主題幾乎都是	USER(S) ； BEHAVIOR(S) ； HUMAN ； SEARCH INTERFACE ；VISUALIZATION
		WEB；HYPERMEDIA、HYPERTEXT
		CATEGORI(S)ZATION ； _Cluster_ ； _Filter_ ； _Fusion_
		LANGUAGE；_Cross language_；_Multilingual_
		PERFORMANCE；TEST COLLECTIONS
		Distributed
		EFFICIENT(EFFICIENCY)
第Ⅱ區 (第一區間成長 而後趨於平穩或下降)	MODEL(S) RELEVANCE FEEDBACK	DATABASE(S)；SYSTEM(S)；KNOWLEDGE
		RELEVANCE FEEDBACK；QUERY EXPANSION
		LATENT SEMANTIC (SEMANTICS) ； PROBABILISTIC MODEL ； _(IR)Theory_ ； INFERENCE(S)
		INDEX TERMS ； TERM(S) DEPENDENCE ； VOCABULARY；_Question and answering_

二、 自動化歸類主題與年代分佈

本研究利用系統輔助歸類之用意在於排除歸類時因主觀因素之影響，亦可加強本研究之信度。會議文獻與期刊文獻合計總共有 1843 篇，利用多階層歸類的主題萃取系統進行歸類後，共取出三個階層的歸類結果，第一階層總計 192 個類別，第二階層總計 18 個類別，第三階層則為 1 個類別。

為了方便人工判讀，本研究以第二階層歸類為主，輔以第一階層的歸類結果進行解讀。除了擷取部份較明顯主題類別外，亦參考 Smeaton et al.(2002)於文獻中所彙整歸類的 29 個類別，最後共取出 38 個主題，總計會議文獻共 477 篇；期刊文獻共 502 篇，如附錄三。圖 4-37 為系統萃取後判讀結果範例說明，本研究所取出的 38 個主題，其判斷依據有三項主要條件：

1. 主題萃取詞彙有明顯的主題類別可判斷
2. 主題群組內之文獻相似度較高
3. 最細部之主題詞與上層類別主題詞一致

而在本研究編號 34 的主題類別 interactive&information-seeking，因 information-seeking 為複合詞彙，故於歸類結果中另外出現 information seeking 的類別，本研究將兩者予以合併，所以此類別的最後文獻數量總計為 48 篇。最後再以人工比對主題類別所屬之文獻標題，以確認該文獻是屬於會議或期刊文獻，最後依年代序列將各主題之文獻數量加以統整，如附錄七。

圖 4- 37 主題萃取系統歸類後結果判讀示意圖

(一) 自動化歸類結果

由表 4-14 中所列 38 個主題的文獻數量百分比可知，期刊文獻中數量佔最多的前五名主題分別是 interactive&information-seeking、query expansion、highly relevant、cluster、citation；會議文獻中數量佔最多的前五名主題分別是 highly relevant、query expansion、cluster、language model、track。依兩者文獻數量百分比的差距可知，會議與期刊文獻數量上差距最大的前五名，分別是：

(1) interactive&information-seeking

(2) citation

(3) track

(4) question answer

(5) recall and precision

其中期刊比會議文獻多的主題是 interactive&information-seeking、citation、recall and precision；會議比期刊文獻多的主題是 track、question answer。比對先前所整理的無法對應之主題即可發現，期刊文獻的確在 interactive& information-seeking 和 citation 是較多的主題，而會議文獻也可看出 question answer 的確是期刊文獻少有之主題。

在兩者文獻初始年代上的差異，會議比期刊文獻早出現的主題共有 24 個，期刊比會議文獻早出現的共有 9 個，兩者於同一年代出現的則有 5 個。由此可知會議比期刊文獻較早出現的主題仍佔多數，但並非全部的主題文獻皆是如此。由表中更可以發現一個有趣的現象，即是在自動化歸類下，兩者文獻數量百分比的差距比前兩種主題整併的情況下要小很多。

最後，從所擷取之主題詞上也可發現，自動化歸類中有 20 個詞彙是無法在第一種主題整併中被看見，也就是說，自動化歸類的方式得以找出第一種主題整併下所無發現的 20 個可對應之主題詞彙。

表 4- 14 自動化歸類下系統萃取 38 個類別主題文獻統整表

編號	主題類別	會議篇數/百分比	期刊篇數/百分比	會議與期刊文獻初始年代 (會議早於期刊文獻●) (期刊早於會議文獻△) (會議文獻與期刊文獻年代相同=)
1	citation	3	21	●
		0.63%	4.18%	
2	visualization	10	20	●

		2.10%	3.98%	
3	link analysis	15	6	△
		3.14%	1.20%	
4	genetic algorithm	7	12	●
		1.47%	2.39%	
5	markov	8	2	●
		1.68%	0.40%	
6	highly relevant	22	25	●
		4.61%	4.98%	
7	morphological	14	14	=
		2.94%	2.79%	
8	logic	9	12	△
		1.89%	2.39%	
9	space	8	16	●
		1.68%	3.19%	
10	retrieval model/probabilistic retrieval	22	21	△
		4.61%	4.18%	
11	query expansion	29	36	=
		6.08%	7.17%	
12	cluster	31	25	●
		6.50%	4.98%	
13	language model	22	15	●
		4.61%	2.99%	
14	question answer	18	3	●
		3.77%	0.60%	
15	latent semantic index	13	14	=
		2.73%	2.79%	
16	collaborative filter	12	2	△
		2.52%	0.40%	
17	music	1	9	●
		0.21%	1.79%	
18	text categorization	17	8	●
		3.56%	1.59%	
19	cross-language	21	21	●
		4.40%	4.18%	
20	image retrieval	20	11	●
		4.19%	2.19%	
21	text classification	13	5	●

		2.73%	1.00%	
22	boolean	5	5	●
		1.05%	1.00%	
23	recall and precision	4	18	△
		0.84%	3.59%	
24	data fusion	21	18	●
		4.40%	3.59%	
25	filter	14	10	△
		2.94%	1.99%	
26	invert file	8	6	●
		1.68%	1.20%	
27	space efficiency	9	7	●
		1.89%	1.39%	
28	signature file	8	4	●
		1.68%	0.80%	
29	opac	3	14	△
		0.63%	2.79%	
30	search tactic	6	11	●
		1.26%	2.19%	
31	intelligent	4	11	=
		0.84%	2.19%	
32	text summarization	12	3	●
		2.52%	0.60%	
33	xml retrieval / database	11	17	=
		2.31%	3.39%	
34	interactive&information-seeking	7	41	△
		1.47%	8.17%	
35	hypertext	8	19	●
		1.68%	3.78%	
36	track	22	1	●
		4.61%	0.20%	
37	word segmentation	9	16	△
		1.89%	3.19%	
38	structur query	11	3	●
		2.31%	0.60%	
	總計	477	502	

(二)與先前研究之比較

Smeaton et al.(2002)和 Ding et al.(2001)曾分別針對資訊檢索領域文獻進行研究，與本研究之比較如表 4-15。由表中可知，在年代的部份，Smeaton et al.橫跨近 25 年的會議文獻；Ding et al.則是研究 10 年的期刊文獻；而本研究則是接續先前之研究，探討近 17 年的會議與期刊文獻。此外，由於 Smeaton et al.和 Ding et al.的研究皆是以取自標題、摘要等欄位的自由詞彙所進行的歸類研究，故僅以同樣取標題與摘要等自由詞彙的自動化歸類結果進行比較，不再以控制詞彙的第一階段與第二階段整併結果進行比較。

表 4- 15 本研究與先前研究之比較

同異處 / 研究者	分析文獻類型	欄位擷取	研究年代範圍	系統歸類方式
Smeaton et al.(2002)	會議文獻	標題與摘要	1978~2002	Clustan Graphics
Ding et al. (2001)	期刊文獻	標題與摘要	1987~1997	SPSS
本研究	會議與期刊文獻	標題與摘要 (自動化歸類)	1990~2007	主題萃取系統

由於各研究的年代範圍有所差異，故以下分為相同年代之比較及接續先前研究年代範圍之結果兩部份進行探討：

1. 相同年代之比較

在 Ding et al.(2001)的研究中，將 10 年的期刊文獻分為兩群，分別是 1987 年至 1991 年、1992 年至 1997 年，故本研究亦按照 Ding et al.的年代進行分區，統計各區文獻數量，其結果如表 4-16。由表中可知，在第一階段 1992 年至 1997 年的主題以 interactive&information-seeking、hypertext 數量最多；第二階段 1998 年至 2002 年則是 interactive&information-seeking 持續有較多的文獻產出，但 highly relevant、query expansion、cluster 逐漸成為主流；第三階段則可看出近幾年來 query expansion 和 cluster 仍然是主流，但 cross-language 和 data fusion 有後續追上的趨勢。比對 Ding et al.的研究結果可發現，hypertext 主題詞出現在 1992 年至 1997 年的頻率為 62 次，而 information-seeking 則是被歸類在新興詞彙的第三名，由此可知這兩個熱門主題都可由先前的研究中可得知。

表 4- 16 自動化歸類與先前研究相同年代比較表

主題 \ 年代	與 Ding et al. 相同主題	期刊文獻		
		1992~1997	1998~2002	2003~2007
citation		3	11	7
visualization	△	4	9	6
link analysis		1	2	2
genetic algorithm	△	2	6	4
markov		0	1	1
highly relevant		2	13	9
morphological		5	1	8
logic		4	3	0
space		3	5	8
retrieval model/probabilistic retrieval		1	8	11
query expansion		5	12	18
cluster		0	12	13
language model		3	2	10
question answer		1	0	2
latent semantic index		3	2	9
collaborative filte		0	1	1
music		0	2	7
text categorization		1	4	3
cross-language		0	5	16
image retrieval	△	3	5	3
text classification		0	0	5
boolean		2	2	1
recall and precision		8	6	3
data fusion		2	4	12
filter		0	8	1
invert file		1	1	4
space efficiency		0	4	3
signature file		1	1	1
opac		8	0	6
search tactic		3	2	6
intelligent		2	3	3
text summarization		1	0	2
xml retrieval / database		6	4	5
interactive&information-seeking	△	12	17	10
hypertext	△	11	7	1
track	△	0	1	0
word segmentation		2	10	4
structur query		0	1	2

2. 接續先前研究之結果

首先將自動化歸類結果取出先前研究之後五年文獻主題總數，以驗證先前研究之結果，亦即將 2003 年至 2007 年的會議文獻和 1998 年至 2002 年的期刊文獻，分別比較 Smeaton et al.(2002)和 Ding et al.(2001)針對 IR 領域進行研究的文獻。由表 4-17 可知，在會議文獻的部份，其中可對應至 Smeaton et al.的研究主題共有 12 個；而與 Ding 的可對應之主題則有 6 個，造成主題之間可對應數量上差異的主要原因有兩個，一為 Ding 的研究將關鍵字取出後再對應至 LCSH 的標題表中，但本研究僅針對自由詞彙直接歸類，二則是本研究於自動歸類之結果取出重要主題群組時，參考 Smeaton et al.之歸類結果，故可對應之主題會較多。

在研究結果的部份，Smeaton et al.(2002)所提出之後續熱門研究主題中，有兩個與 2003 年至 2007 年文獻總數最多的前五名主題相符合，分別是「Language of Model」和「Question-Answering」。Ding 的研究僅提出詞彙為正負成長率及新興之詞彙列表，並未預測後續研究的可能發展，故只比對 Ding et al.(2001)的研究中提出有正成長率及新興詞彙，其中「interactive&information-seeking」與 1998 年至 2002 年文獻總數量排名相同。

表 4- 17 自動化歸類接續先前研究之主題對應表

自動化歸類之主題	2003~2007 年會議文獻總數	Smeaton et al. 預測熱門主題	與 Smeaton et al. 相同主題	1998~2002 年期刊文獻總數	與 Ding et al. 相同主題
citation	0			11	
visualization	2		△	9	△
link analysis	8			2	
genetic algorithm	3			6	△
markov	6			1	
highly relevant	9			13	
morphological	5			1	
logic	1			3	
space	2			5	
retrieval model/probabilistic retrieval	8	●	△	8	
query expansion	7			12	
cluster	13		△	12	
language model	10	●	△	2	
question answer	10	●	△	0	
latent semantic index	5			2	
collaborative filte	9			1	

music	0			2	
text categorization	5			4	
cross-language	9	●	△	5	
image retrieval	10		△	5	△
text classification	7			0	
boolean	0		△	2	
recall and precision	2			6	
data fusion	10			4	
filter	5		△	8	
invert file	1			1	
space efficiency	3			4	
signature file	0		△	1	
opac	0			0	
search tactic	1			2	
intelligent	1			3	
text summarization	4		△	0	
xml retrieval / database	8			4	
interactive&information-seeking	4			17	△
hypertext	2		△	7	△
track	9			1	△
word segmentation	1			10	
structur query	2			1	

第三節　會議與期刊文獻主題趨勢預測差異

本研究根據第三章所述之預測方法，將會議與期刊文獻各自取出預測集與驗證集以進行主題預測之排序並計算相對位差，再依相對位差的平均數進行差異分析之檢定。

本研究假設如下所列：

假設一：兩種文獻類型交叉預測於不同年代序列，在主題趨勢預測的結果上有顯著差異。

假設二：不同預測集年代範圍於不同年代序列，在主題趨勢預測的結果上有顯著差異。

假設三：不同驗證集年代範圍於不同年代序列，在主題趨勢預測的結果上有顯著差異。

假設四：不同預測指標兩種於不同年代序列，在主題趨勢預測的結果上有顯著差異。

本節依上述之研究假設，分別進行以下檢定分析：

(一) 不同文獻類型的主題預測在各年代序列上之差異分析

(二) 不同預測集年代範圍在主題預測上之差異分析

(三) 不同驗證集年代範圍在主題預測上之差異分析

(四) 不同預測指標在主題預測上之差異分析

一、　不同文獻類型的主題預測在各年代序列上之差異分析

為了檢驗研究假設(一)，本小節以「年代區間」與「不同文獻類型」為自變項，以「趨勢預測排序相對位差得分」為依變項，將預測集年代固定為 10 年、預測集後 2 年固定為驗證集年代，再分別以兩種預測指標進行主題趨勢預測後所得之結果的平均數差異考驗，以瞭解不同文獻類型的主題預測在各年代序列上的表現於 95%信心水準之下的差異考驗。

自變項：(1)文獻類型：會議預測會議、期刊預測會議、會議預測期刊、期刊預測期刊

(2)年代範圍：
1991~2000 年文獻預測 2001、2002 主題
1992~2001 年文獻預測 2002、2003 主題
1993~2002 年文獻預測 2003、2004 主題
1994~2003 年文獻預測 2004、2005 主題
1995~2004 年文獻預測 2005、2006 主題
1996~2005 年文獻預測 2006、2007 主題

依變項：(1)主題趨勢預測排序相對位差得分

以下分別以三種不同的歸類基礎，再分別以兩種預測指標 API 和 SLP 所得之主題預測排序之相對位差得分，進行差異考驗：

(一) 第一種主題整併

1. 以 **API** 為預測指標的差異考驗

表 4- 18 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

文獻類型 / 預測集年代區間	Type=1 會議預測會議 (108)	Type=2 期刊預測會議 (108)	Type=3 會議預測期刊 (108)	Type=4 期刊預測期刊 (108)	邊緣平均數
1991~2000(72)	4.333	4.944	5.278	7.222	5.444
1992~2001(72)	5.056	4.500	5.444	5.667	5.167
1993~2002(72)	4.667	4.778	5.222	5.111	4.944
1994~2003(72)	3.667	4.778	5.389	5.611	4.861
1995~2004(72)	3.389	5.167	6.000	7.000	5.389
1996~2005(72)	4.556	4.278	5.778	6.056	5.167
邊緣平均數	4.278	4.741	5.519	6.111	
註：括號內為個數					

在第一種主題整併下以 API 為預測指標所得之不同文獻類型預測得分平均數可知，會議預測會議平均得分為 4.278、期刊預測會議平均得分為 4.741、會議預測期刊平均得分為 5.519、期刊預測期刊平均得分為 6.111。

表 4- 19 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.251	23	408	.197

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+type+year+type * year

利用 Levene's Test 判定變異數同質性之檢定，其結果為 F=1.251，P 值=0.197 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 20 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	事後比較
文獻類型(type)	214.620	3	71.540	4.643*	.003	Type1 < Type4
年代區間(year)	19.380	5	3.876	.252 n.s	.939	
type * year(交互作用)	96.546	15	6.436	.418 n.s	.974	
Error(誤差)	6286.111	408	15.407			
全體	18128.000	432				
*p<.05 n.s=不顯著						

經雙因子變異數分析檢定，對於不同文獻類型預測主題趨勢與年代序列之交互作用顯著性為 0.974 > 0.05，未達顯著性，但四種文獻類型預測上之顯著性為 0.003 < 0.05，已達顯著性，進行事後比較後發現為第一組(會議預測會議)之平均數 < 第四組(期刊預測期刊)平均數，亦即會議預測會議之效果優於期刊預測期刊，達百分之九十五信心水準的顯著性差異，且不因年代的差異而有所影響。

2. 以 **SLP** 為預測指標的差異考驗

表 4- 21 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之各項細格平均數與邊緣平均數

文獻類型 / 預測集年代區間	Type=1 會議預測會議 (108)	Type=2 期刊預測會議 (108)	Type=3 會議預測期刊 (108)	Type=4 期刊預測期刊 (108)	邊緣平均數
1991~2000(72)	4.667	4.944	5.833	5.333	5.194
1992~2001(72)	5.000	4.000	5.389	4.833	4.806
1993~2002(72)	4.056	4.722	4.944	3.167	4.222
1994~2003(72)	3.889	5.111	5.611	4.611	4.806
1995~2004(72)	5.167	4.944	5.889	4.889	5.222
1996~2005(72)	4.278	4.556	4.722	5.333	4.722
邊緣平均數	4.509	4.713	5.398	4.694	
註：括號內為個數					

在第一種主題整併下以 SLP 為預測指標所得之不同文獻類型預測得分平均數可知，會議預測會議平均得分為 4.509、期刊預測會議平均得分為 4.713、會議預測期刊平均得分為 5.398、期刊預測期刊平均得分為 4.694。

表 4- 22 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable:主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.109	23	408	.332

Tests the null hypothesis that the error variance of the dependent variable is equal across gro ups.
a Design: Intercept+type+year+type * year

利用 Levene's Test 判定變異數同質性之檢定，其結果為 F=1.109，P 值=0.332 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 23 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
文獻類型(type)	49.435	3	16.478	1.089 n.s	.353
年代區間(year)	48.157	5	9.631	.637 n.s	.672
type * year(交互作用)	68.731	15	4.582	.303 n.s	.995
誤差	6171.000	408	15.125		
全體	16410.000	432			
n.s=不顯著					

經雙因子變異數分析檢定，對於不同文獻類型預測主題趨勢與年代序列之交互作用顯著性為 0.995 > 0.05，未達顯著性，且文獻類型與年代區間皆未達顯著性。

(二) 第二種主題整併

1. 以 **API** 為預測指標的差異考驗

表 4- 24 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之各項細格平均數與邊緣平均數

文獻類型 / 預測集年代區間	Type=1 會議預測會議 (66)	Type=2 期刊預測會議 (66)	Type=3 會議預測期刊 (66)	Type=4 期刊預測期刊 (66)	邊緣平均數
1991~2000(44)	3.545	2.545	4.455	4.000	3.636
1992~2001(44)	4.818	3.000	5.000	3.727	4.136
1993~2002(44)	4.182	2.727	4.364	3.091	3.591
1994~2003(44)	4.000	2.091	4.182	3.545	3.455
1995~2004(44)	3.182	2.636	3.727	4.273	3.455
1996~2005(44)	2.545	3.273	3.636	4.364	3.455
邊緣平均數	3.712	2.712	4.227	3.833	
註：括號內為個數					

在第二種主題整併下以 API 為預測指標所得之不同文獻類型預測得分平均數可知，會議預測會議平均得分為 3.712、期刊預測會議平均得分為 2.712、會議預測期刊平均得分為 4.227、期刊預測期刊平均得分為 3.833。

表 4- 25 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
.582	23	240	.938

Tests the null hypothesis that the error variance of the dependent variabl e is equal across groups.
a Design: Intercept+type+year+type * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=0.582，P 值=0.938 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 26 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	事後比較
文獻類型(type)	82.303	3	27.434	4.072*	.008	Type2 < Type3
年代區間(year)	15.394	5	3.079	.457 n.s	.808	
type * year(交互作用)	55.333	15	3.689	.547 n.s	.912	
誤差	1617.091	240	6.738			
全體	5232.000	264				
*p<.05 n.s=不顯著						

a Computed using alpha = .05
b R Squared = .086 (Adjusted R Squared = -.001)

經雙因子變異數分析檢定，對於不同文獻類型預測主題趨勢與年代序列之交互作用顯著性為0.912 > 0.05，未達顯著性，但文獻類型之顯著性為0.008 < 0.05有顯著，故進行事後比較後發現第二組(期刊預測會議)平均數 < 第三組(會議預測期刊)平均數，亦即期刊預測會議之效果優於會議預測期刊 ，達95%的信心水準的顯著性差異，且不因年代的差異而有所影響。

2. 以 **SLP** 為預測指標的差異考驗

表 4- 27 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之各項細格平均數與邊緣平均數

文獻類型 / 預測集年代區間	Type=1 會議預測會議 (66)	Type=2 期刊預測會議 (66)	Type=3 會議預測期刊 (66)	Type=4 期刊預測期刊 (66)	邊緣平均數
1991~2000(44)	3.091	2.727	5.091	1.455	3.091
1992~2001(44)	4.636	3.273	5.000	2.182	3.773
1993~2002(44)	3.091	3.273	4.545	1.091	3.000
1994~2003(44)	3.727	3.364	4.818	2.273	3.545
1995~2004(44)	2.545	2.091	4.000	2.455	2.773
1996~2005(44)	2.364	3.091	4.545	2.545	3.136
邊緣平均數	3.242	2.970	4.667	2.000	
註：括號內為個數					

在第二種主題整併下以 SLP 為預測指標所得之不同文獻類型預測得分平均數可知，會議預測會議平均得分為 3.242、期刊預測會議平均得分為 2.970、會議預測期刊平均得分為 4.667、期刊預測期刊平均得分為 2.000。

表 4- 28 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.511	23	240	.067

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+type+year+type * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.511，P 值 =0.067 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 29 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	事後比較
文獻類型(type)	240.530	3	80.177	16.826*	.000	Type1 < Type3 Type2 < Type3 Type4 < Type3 Type4 < Type1
年代區間(year)	30.076	5	6.015	1.262 n.s	.281	
type * year(交互作用)	49.015	15	3.268	.686 n.s	.798	
誤差	1143.636	240	4.765			
全體	4200.000	264				
*p<.05 n.s=不顯著	1463.258	263				

a Computed using alpha = .05
b R Squared = .218 (Adjusted R Squared = .144)

經雙因子變異數分析檢定，對於不同文獻類型預測主題趨勢與年代序列之交互作用顯著性為 0.798 > 0.05，未達顯著性，但文獻類型之顯著性為 0.000 < 0.05 有顯著，故進行事後比較後發現第一組(會議預測會議)平均數 < 第三組(會議預

測期刊)平均數、第二組(期刊預測會議)平均數 < 第三組(會議預測期刊)平均數、第四組(期刊預測期刊)平均數 < 第三組(會議預測期刊)平均數、第四組(期刊預測期刊)平均數 < 第一組(會議預測會議)平均數。

亦即會議預測會議和期刊預測會議皆優於會議預測期刊；期刊預測期刊優於會議預測期刊和會議預測會議，也就是說期刊預測期刊所得之效果最佳、期刊預測會議次之。

(三) 自動化歸類

1. 以 **API** 為預測指標的差異考驗

表 4- 30 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之各項細格平均數與邊緣平均數

文獻類型 / 預測集年代區間	Type=1 會議預測會議 (266)	Type=2 期刊預測會議 (266)	Type=3 會議預測期刊 (266)	Type=4 期刊預測期刊 (266)	邊緣平均數
1990~1999(152)	9.132	10.947	11.026	10.842	10.487
1991~2000(152)	10.368	10.658	11.316	10.816	10.789
1992~2001(152)	11.605	12.026	12.053	9.211	11.224
1993~2002(152)	9.737	9.763	12.421	9.658	10.395
1994~2003(152)	10.132	10.974	12.395	9.921	10.855
1995~2004(152)	9.789	11.658	11.605	10.000	10.763
1996~2005(152)	11.895	12.368	11.368	12.053	11.921
邊緣平均數	10.380	11.199	11.741	10.357	
註：括號內為個數					

在自動化歸類下以 API 為預測指標所得之不同文獻類型預測得分平均數可知，會議預測會議平均得分為 10.380、期刊預測會議平均得分為 11.199、會議預測期刊平均得分為 11.741、期刊預測期刊平均得分為 10.357。

表 4- 31 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.450	27	1036	.065

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+type+year+type * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.450，P 值=0.065 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 32 四種文獻類型以 API 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
文獻類型(type)	361.786	3	120.595	2.156 n.s	.092
年代區間(year)	243.759	6	40.627	.726 n.s	.628
type * year(交互作用)	447.346	18	24.853	.444 n.s	.978
誤差	57952.158	1036	55.938		
全體	185864.000	1064			
n.s = 無顯著					

a Computed using alpha = .05
b R Squared = .018 (Adjusted R Squared = -.008)

經雙因子變異數分析檢定，對於不同文獻類型預測主題趨勢與年代序列之交互作用顯著性為 0.978 > 0.05，未達顯著性，表示其交互作用不顯著。

2. 以 **SLP** 為預測指標的差異考驗

表 4- 33 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之各項細格平均數與邊緣平均數

文獻類型 / 預測集年代區間	Type=1 會議預測會議 (266)	Type=2 期刊預測會議 (266)	Type=3 會議預測期刊 (266)	Type=4 期刊預測期刊 (266)	邊緣平均數
1990~1999(152)	8.737	12.184	10.947	9.132	10.250
1991~2000(152)	9.605	12.184	12.763	8.921	10.868
1992~2001(152)	10.132	11.868	12.263	7.211	10.368
1993~2002(152)	9.711	10.316	12.605	9.000	10.408
1994~2003(152)	9.000	11.868	11.947	8.079	10.224
1995~2004(152)	8.895	11.763	9.763	8.947	9.842
1996~2005(152)	9.132	12.079	10.763	10.079	10.513
邊緣平均數	9.316	11.752	11.579	8.767	
註：括號內為個數					

在自動化歸類下以 SLP 為預測指標所得之不同文獻類型預測得分平均數可知，會議預測會議平均得分為 9.316、期刊預測會議平均得分為 11.752、會議預測期刊平均得分為 11.579、期刊預測期刊平均得分為 8.767。

表 4- 34 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.382	27	1036	.093

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept+type+year+type * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.382，P 值=0.093 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 35 四種文獻類型以 SLP 為預測指標在不同年代區間呈現對趨勢預測排序 相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	事後比較
文獻類型(type)	1875.639	3	625.213	10.506*	.000	Type1 < Type2 Type1 < Type3 Type4 < Type2 Type4 < Type3
年代區間(year)	88.602	6	14.767	.248 n.s	.960	
type * year(交互作用)	534.729	18	29.707	.499 n.s	.959	
誤差	61652.158	1036	59.510			
全體	178204.000	1064				
*P<.05 n.s=不顯著						

a Computed using alpha = .05
b R Squared = .039 (Adjusted R Squared = .014)

經雙因子變異數分析檢定，對於不同文獻類型預測主題趨勢與年代序列之交互作用顯著性為 0.959 > 0.05，未達顯著性，表示其交互作用不顯著，但在文獻類型(type)的顯著性 P 值=0.000 < 0.05，有顯著，進行事後比較可發現第一組和第四組文獻類型預測(會議預測會議、期刊預測期刊)平均數均 < 第二組和第三組文獻類型預測(期刊預測會議、會議預測期刊)平均數。

亦即在主題趨勢預測的表現上以會議預測會議和期刊預測期刊優於期刊預測會議和會議預測期刊，也就是說在此情況下相同類型的文獻預測皆優於不同類型的文獻預測，達 95% 信心水準有差異顯著，且不因年代而有所影響。

二、 不同預測集年代範圍在主題預測上之差異分析

為了檢驗研究假設(二)，本小節以「不同預測集年代範圍」與「各年度區間」為自變項，以「趨勢預測排序相對位差得分」為依變項，將預測集年代範圍設定為 4 年、6 年、8 年、10 年；預測集後 2 年固定為驗證集年代，再分別以兩種預測指標進行四種不同文獻類型的交叉預測，計算主題文獻趨勢預測後所得之結果的平均數差異考驗，以瞭解不同預測集的年代範圍對主題文獻預測於 95%信心水準之下的差異考驗。

自變項：(1)預測集年代範圍：預測集為 4 年、6 年、8 年、10 年。

(2)年代區間：以 1990、1991、1992、1993、1994、1995、1996 預測集起始年代。

依變項：(1)主題趨勢預測排序相對位差得分。

(一) 以會議文獻預測會議主題之不同預測集年代範圍差異考驗

1. 以 **API** 為預測指標

表 4- 36「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測集範圍 / 預測集年代區間	預測集為 4 年 (266)	預測集為 6 年 (266)	預測集為 8 年 (266)	預測集為 10 年 (266)	邊緣平均數
1990~1999(152)	10.711	11.000	8.079	9.132	9.730
1991~2000(152)	10.658	11.053	6.947	10.368	9.757
1992~2001(152)	10.737	8.895	9.184	11.605	10.105
1993~2002(152)	9.342	6.553	10.711	9.737	9.086
1994~2003(152)	8.816	10.447	9.395	10.132	9.697
1995~2004(152)	11.263	8.842	10.737	9.789	10.158
1996~2005(152)	10.500	9.763	10.342	11.895	10.625
邊緣平均數	10.289	9.508	9.342	10.380	
註：括號內為個數					

在以 API 為趨勢預測指標下，以會議預測會議主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，預測集為 4 年平均得分為 10.289；預測集為 6 年平均得分為 9.508；預測集為 8 年平均得分為 9.342；預測集為 10 年平均得分為 10.380。

表 4- 37「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.045	27	1036	.403

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept+add_so+year+add_so * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.045，P 值=0. 403 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 38「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
預測集範圍(add_so)	224.887	3	74.962	1.430 n.s	.232
年代(year)	210.549	6	35.091	.670 n.s	.674
add_so * year(交互作用)	1235.060	18	68.614	1.309 n.s	.173
誤差	54290.105	1036	52.404		
全體	159816.000	1064			
n.s=不顯著					

a Computed using alpha = .05
b R Squared = .030 (Adjusted R Squared = .005)

經雙因子變異數分析檢定，對於「不同預測集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.173 > 0.05，未達顯著性，表示其交互作用不顯著。

2. 以 **SLP** 為預測指標

表 4- 39「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測集範圍 / 預測集年代區間	預測集為 4 年 (266)	預測集為 6 年 (266)	預測集為 8 年 (266)	預測集為 10 年 (266)	邊緣平均數
1990~1999(152)	10.395	10.184	8.500	8.737	9.454
1991~2000(152)	10.868	11.842	7.816	9.605	10.033
1992~2001(152)	10.421	9.211	9.289	10.132	9.763
1993~2002(152)	10.237	7.974	10.079	9.711	9.500
1994~2003(152)	10.605	10.526	9.342	9.000	9.868
1995~2004(152)	10.658	9.842	10.237	8.895	9.908
1996~2005(152)	11.316	9.553	9.263	9.132	9.816
邊緣平均數	10.643	9.876	9.218	9.316	
註：括號內為個數					

在以 SLP 為趨勢預測指標下，以會議預測會議主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，預測集為 4 年平均得分為 10.643；預測集為 6 年平均得分為 9.876；預測集為 8 年平均得分為 9.218；預測集為 10 年平均得分為 9.316。

表 4- 40「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.064	27	1036	.377

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept+add_so+year+add_so * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.064，P 值=0. 377 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 41「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
預測集範圍(add_so)	341.511	3	113.837	1.862 n.s	.134
年代(year)	41.408	6	6.901	.113 n.s	.995
add_so * year(交互作用)	534.923	18	29.718	.486 n.s	.965
誤差	63344.474	1036	61.143		
全體	165682.000	1064			
n.s=不顯著					

a Computed using alpha = .05

b R Squared = .014 (Adjusted R Squared = -.011)

經雙因子變異數分析檢定，對於「不同預測集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.965 > 0.05，未達顯著性，表示其交互作用不顯著。

(二) 以期刊文獻預測會議主題之不同預測集年代範圍差異考驗

1. 以 **API** 為預測指標

表 4- 42「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測集範圍 / 預測集年代區間	預測集為 4 年 (266)	預測集為 6 年 (266)	預測集為 8 年 (266)	預測集為 10 年 (266)	邊緣平均數
1990~1999(152)	12.605	9.316	12.026	10.947	11.224
1991~2000(152)	9.974	10.368	11.553	10.658	10.638
1992~2001(152)	9.421	10.526	11.605	12.026	10.895
1993~2002(152)	11.316	11.500	12.421	9.763	11.250
1994~2003(152)	11.500	11.105	12.421	10.974	11.500
1995~2004(152)	10.026	11.553	9.947	11.658	10.796
1996~2005(152)	10.632	11.526	10.789	12.368	11.329
邊緣平均數	10.782	10.842	11.538	11.199	

註：括號內為個數

在以 API 為趨勢預測指標下，以期刊預測會議主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，預測集為 4 年平均得分為 10.782；預測集為 6 年平均得分為 10.842；預測集為 8 年平均得分為 11.538；預測集為 10 年平均得分為 11.199。

表 4- 43「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable:主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.295	27	1036	.144

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept+add_so+year+add_so * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.295，P 值=0.144 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 44「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
預測集範圍(add_so)	98.053	3	32.684	.505 n.s	.679
年代(year)	90.799	6	15.133	.234 n.s	.966
add_so * year(交互作用)	706.329	18	39.240	.606 n.s	.897
誤差	67048.158	1036	64.718		
全體	198808.000	1064			
n.s=不顯著					

a Computed using alpha = .05
b R Squared = .013 (Adjusted R Squared = -.013)

經雙因子變異數分析檢定，對於「不同預測集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.897 > 0.05，未達顯著性，表示其交互作用不顯著。

2. 以 **SLP** 為預測指標

表 4- 45「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測集範圍 / 預測集年代區間	預測集為 4 年 (266)	預測集為 6 年 (266)	預測集為 8 年 (266)	預測集為 10 年 (266)	邊緣平均數
1990~1999(152)	12.711	9.447	11.658	12.184	11.500
1991~2000(152)	10.079	10.895	12.263	12.184	11.355
1992~2001(152)	9.395	11.500	12.553	11.868	11.329
1993~2002(152)	11.500	11.763	12.184	10.316	11.441
1994~2003(152)	11.737	11.474	12.132	11.868	11.803
1995~2004(152)	10.105	12.289	10.684	11.763	11.211
1996~2005(152)	11.737	11.289	10.789	12.079	11.474
邊緣平均數	11.038	11.237	11.752	11.752	
註：括號內為個數					

在以 SLP 為趨勢預測指標下，以期刊預測會議主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，預測集為 4 年平均得分為 11.038；預測集為 6 年平均得分為 11.237；預測集為 8 年平均得分為 11.752；預測集為 10 年平均得分為 11.752。

表 4- 46「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨

勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
.943	27	1036	.548

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+add_so+year+add_so * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.943，P 值=0.548 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 47「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
預測集範圍(add_so)	105.777	3	35.259	.550 n.s	.648
年代(year)	31.656	6	5.276	.082 n.s	.998
add_so * year(交互作用)	697.532	18	38.752	.605 n.s	.898
誤差	66363.763	1036	64.058		
全體	206559.000	1064			

n.s=不顯著

a Computed using alpha = .05
b R Squared = .012 (Adjusted R Squared = -.013)

經雙因子變異數分析檢定，對於「不同預測集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.898 > 0.05，未達顯著性，表示其交互作用不顯著。

(三) 以會議文獻預測期刊主題之不同預測集年代範圍差異考驗

1. 以 **API** 為預測指標

表 4- 48「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測集範圍 / 預測集年代區間	預測集為 4 年 (266)	預測集為 6 年 (266)	預測集為 8 年 (266)	預測集為 10 年 (266)	邊緣平均數
1990~1999(152)	10.789	9.553	10.395	11.026	10.441
1991~2000(152)	9.921	10.026	11.605	11.316	10.717
1992~2001(152)	9.921	10.684	9.342	12.053	10.500
1993~2002(152)	10.158	9.158	10.711	12.421	10.612
1994~2003(152)	9.342	10.921	12.158	12.395	11.204
1995~2004(152)	10.289	13.158	14.000	11.605	12.263
1996~2005(152)	14.553	12.158	12.553	11.368	12.658
邊緣平均數	10.711	10.808	11.538	11.741	
註：括號內為個數					

在以 API 為趨勢預測指標下，以會議預測期刊主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，預測集為 4 年平均得分為 10.711；預測集為 6 年平均得分為 10.808；預測集為 8 年平均得分為 11.538；預測集為 10 年平均得分為 11.741。

表 4- 49「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.791	27	1036	.008

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+add_so+year+add_so * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.791，P 值=0.008 < 0.05，有達到顯著性，代表變異數不具同質性，無法繼續檢定。

2. 以 **SLP** 為預測指標

表 4- 50「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測集範圍 / 預測集年代區間	預測集為 4 年 (266)	預測集為 6 年 (266)	預測集為 8 年 (266)	預測集為 10 年 (266)	邊緣平均數
1990~1999(152)	10.684	10.158	10.605	10.947	10.599
1991~2000(152)	10.868	11.026	11.211	12.763	11.467
1992~2001(152)	9.553	10.684	10.184	12.263	10.671
1993~2002(152)	10.895	10.316	10.974	12.605	11.197
1994~2003(152)	10.658	11.421	12.000	11.947	11.507
1995~2004(152)	10.526	13.263	13.711	9.763	11.816
1996~2005(152)	14.053	12.158	12.263	10.763	12.309
邊緣平均數	11.034	11.289	11.564	11.579	
註：括號內為個數					

在以 SLP 為趨勢預測指標下，以會議預測期刊主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，預測集為 4 年平均得分為 11.034；預測集為 6 年平均得分為 11.289；預測集為 8 年平均得分為 11.564；預測集為 10 年平均得分為 11.579。

表 4- 51「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
.670	27	1036	.899

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+add_so+year+add_so * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=0.670，P 值=0.899 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 52「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
預測集範圍(add_so)	53.387	3	17.796	.246 n.s	.864
年代(year)	337.759	6	56.293	.779 n.s	.587
add_so * year(交互作用)	1000.218	18	55.568	.769 n.s	.739
誤差	74895.684	1036	72.293		
全體	213754.000	1064			
n.s=不顯著	76287.049	1063			

a Computed using alpha = .05
b R Squared = .018 (Adjusted R Squared = -.007)

經雙因子變異數分析檢定，對於「不同預測集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.739 > 0.05，未達顯著性，表示其交互作用不顯著。

(四) 以期刊文獻預測期刊主題之不同預測集年代範圍差異考驗

1. 以 **API** 為預測指標

表 4- 53「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測集範圍 / 預測集年代區間	預測集為 4 年 (266)	預測集為 6 年 (266)	預測集為 8 年 (266)	預測集為 10 年 (266)	邊緣平均數
1990~1999(152)	10.158	8.500	10.026	10.842	9.882
1991~2000(152)	8.868	10.132	11.474	10.816	10.322
1992~2001(152)	8.921	10.263	10.553	9.211	9.737
1993~2002(152)	9.132	11.368	10.211	9.658	10.092
1994~2003(152)	10.816	10.842	9.237	9.921	10.204
1995~2004(152)	10.947	9.237	10.789	10.000	10.243
1996~2005(152)	10.158	9.184	9.789	12.053	10.296
邊緣平均數	9.857	9.932	10.297	10.357	
註：括號內為個數					

在以 API 為趨勢預測指標下，以期刊預測期刊主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，預測集為 4 年平均得分為 9.857；預測集為 6 年平均得分為 9.932；預測集為 8 年平均得分為 10.297；預測集為 10 年平均得分為 10.357。

表 4- 54 不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.171	27	1036	.251

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+add_so+year+add_so * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.171，P 值=0.251 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 55「不同預測集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
預測集範圍(add_so)	50.951	3	16.984	.330 n.s	.804
年代(year)	45.308	6	7.551	.147 n.s	.990
add_so * year(交互作用)	691.865	18	38.437	.747 n.s	.764
誤差	53332.789	1036	51.480		
全體	162894.000	1064			

n.s=不顯著

a Computed using alpha = .05
b R Squared = .015 (Adjusted R Squared = -.011)

經雙因子變異數分析檢定，對於「不同預測集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.764 > 0.05，未達顯著性，表示其交互作用不顯著。

2. 以 **SLP** 為預測指標

表 4- 56「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測集範圍 / 預測集年代區間	預測集為 4 年 (266)	預測集為 6 年 (266)	預測集為 8 年 (266)	預測集為 10 年 (266)	邊緣平均數
1990~1999(152)	10.263	6.684	9.921	9.132	9.000
1991~2000(152)	8.658	10.500	11.026	8.921	9.776
1992~2001(152)	8.316	10.237	8.816	7.211	8.645
1993~2002(152)	10.737	10.684	8.816	9.000	9.809
1994~2003(152)	11.632	9.474	7.737	8.079	9.230
1995~2004(152)	12.079	9.184	10.263	8.947	10.118
1996~2005(152)	9.474	8.421	8.842	10.079	9.204
邊緣平均數	10.165	9.312	9.346	8.767	
註：括號內為個數					

在以 SLP 為趨勢預測指標下，以期刊預測期刊主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，預測集為 4 年平均得分為 10.165；預測集為 6 年平均得分為 9.312；預測集為 8 年平均得分為 9.346；預測集為 10 年平均得分為 8.767。

表 4- 57「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.211	27	1036	.212

Tests the null hypothesis that the error variance of the depe ndent variable is equal across groups.
a Design: Intercept+add_so+year+add_so * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.211，P 值 =0.212 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 58「不同預測集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
預測集範圍(add_so)	265.281	3	88.427	1.582 n.s	.192
年代(year)	246.669	6	41.112	.735 n.s	.621
add_so * year(交互作用)	1139.278	18	63.293	1.132 n.s	.314
誤差	57909.605	1036	55.897		
全體	153527.000	1064			

n.s=不顯著

a Computed using alpha = .05

b R Squared = .028 (Adjusted R Squared = .002)

經雙因子變異數分析檢定，對於「不同預測集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.314 > 0.05，未達顯著性，表示其交互作用不顯著。

三、 不同驗證集年代範圍在主題預測上之差異分析

為了檢驗研究假設(三)，本小節以「不同驗證集年代範圍」與「各年度區間」為自變項，以「趨勢預測排序相對位差得分」為依變項，將預測集年代範圍設定為 10 年；預測集後 1 年、2 年、3 年設定為驗證集年代，再分別以兩種預測指標進行四種不同文獻類型的交叉預測，計算主題文獻趨勢預測後所得之結果的平均數差異考驗，以瞭解不同驗證集的年代範圍對主題文獻預測是否會顯著性的差異存在。

自變項：(1)驗證集年代範圍：1 年、2 年、3 年。

(2)年代區間：以 1990、1991、1992、1993、1994、1995 預測集起始年代。

依變項：(1)主題趨勢預測排序相對位差得分。

(一) 以會議文獻預測會議主題之不同驗證集年代範圍差異考驗

1. 以 **API** 為預測指標

表 4- 59「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

驗證集範圍 / 預測集年代區間	驗證集為 1 年 (228)	驗證集為 2 年 (228)	驗證集為 3 年 (228)	邊緣平均數
1990~1999(114)	9.947	9.132	9.921	9.667
1991~2000(114)	10.474	10.368	10.789	10.544
1992~2001(114)	10.737	11.605	11.789	11.377
1993~2002(114)	9.605	9.737	10.079	9.807
1994~2003(114)	9.921	10.132	9.579	9.877
1995~2004(114)	10.263	9.789	10.605	10.219
邊緣平均數	10.158	10.127	10.461	
註：括號內為個數				

在以 API 為趨勢預測指標下，以會議預測會議主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，驗證集為 1 年平均得分為 10.158；驗證集為 2 年平均得分為 10.127；驗證集為 3 年平均得分為 10.461。

表 4- 60「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.042	17	666	.409

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+add_ans+year+add_ans * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.042，P 值=0.409 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 61「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
驗證集範圍(add_ans)	15.477	2	7.738	.153 n.s	.858
年代(year)	231.801	5	46.360	.915 n.s	.471
add_ans * year(交互作用)	51.734	10	5.173	.102 n.s	1.000
誤差	33760.737	666	50.692		
全體	105902.000	684			

n.s=不顯著

a Computed using alpha = .05

b R Squared = .009 (Adjusted R Squared = -.017)

經雙因子變異數分析檢定，對於「不同驗證集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 1.000 > 0.05，未達顯著性，表示其交互作用不顯著。

2. 以 **SLP** 為預測指標

表 4- 62「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

驗證集範圍 / 預測集年代區間	驗證集為 1 年 (228)	驗證集為 2 年 (228)	驗證集為 3 年 (228)	邊緣平均數
1990~1999(114)	9.605	8.737	8.579	8.974
1991~2000(114)	10.184	9.605	10.289	10.026
1992~2001(114)	9.684	10.132	10.263	10.026
1993~2002(114)	10.947	9.711	9.789	10.149
1994~2003(114)	8.842	9.000	8.921	8.921
1995~2004(114)	9.684	8.895	9.447	9.342
邊緣平均數	9.825	9.346	9.548	

註：括號內為個數

在以 SLP 為趨勢預測指標下，以會議預測會議主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，驗證集為 1 年平均得分為 9.825；驗證集為 2 年平均得分為 9.346；驗證集為 3 年平均得分為 9.548。

表 4- 63「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.008	17	666	.447

Tests the null hypothesis that the error variance of the dependent variab le is equal across groups.
a Design: Intercept+add_ans+year+add_ans * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.008，P 值=0. 447 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 64「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
驗證集範圍(add_ans)	26.266	2	13.133	.236 n.s	.789
年代(year)	180.170	5	36.034	.649 n.s	.663
add_ans * year(交互作用)	63.646	10	6.365	.115 n.s	1.000
誤差	36989.263	666	55.539		
全體	99944.000	684			

n.s=不顯著

a Computed using alpha = .05
b R Squared = .007 (Adjusted R Squared = -.018)

經雙因子變異數分析檢定，對於「不同驗證集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 1.000 > 0.05，未達顯著性，表示其交互作用不顯著。

(二) 以期刊文獻預測會議主題之不同驗證集年代範圍差異考驗

1. 以 **API** 為預測指標

表 4- 65「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

驗證集範圍 / 預測集年代區間	驗證集為 1 年 (228)	驗證集為 2 年 (228)	驗證集為 3 年 (228)	邊緣平均數
1990~1999(114)	11.079	10.947	10.684	10.904
1991~2000(114)	10.921	10.658	10.921	10.833
1992~2001(114)	12.789	12.026	11.579	12.132
1993~2002(114)	11.579	9.763	10.474	10.605
1994~2003(114)	10.184	10.974	11.842	11.000
1995~2004(114)	11.395	11.658	11.316	11.456
邊緣平均數	11.325	11.004	11.136	
註：括號內為個數				

在以 API 為趨勢預測指標下，以期刊預測會議主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，驗證集為 1 年平均得分為 11.325；驗證集為 2 年平均得分為 11.004；驗證集為 3 年平均得分為 11.136。

表 4- 66「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
.847	17	666	.638

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+add_ans+year+add_ans * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=0.847，P 值=0. 638 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 67「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
驗證集範圍(add_ans)	11.810	2	5.905	.092 n.s	.912
年代(year)	175.257	5	35.051	.546 n.s	.741
add_ans * year(交互作用)	139.822	10	13.982	.218 n.s	.995
誤差	42758.684	666	64.202		
全體	128198.000	684			

n.s=不顯著

a Computed using alpha = .05

b R Squared = .008 (Adjusted R Squared = -.018)

經雙因子變異數分析檢定，對於「不同驗證集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.995 > 0.05，未達顯著性，表示其交互作用不顯著。

2. 以 **SLP** 為預測指標

表 4- 68「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

驗證集範圍 / 預測集年代區間	驗證集為 1 年 (228)	驗證集為 2 年 (228)	驗證集為 3 年 (228)	邊緣平均數
1990~1999(114)	12.105	12.184	12.395	12.228
1991~2000(114)	11.605	12.184	11.711	11.833
1992~2001(114)	12.632	11.868	11.737	12.079
1993~2002(114)	10.921	10.316	11.553	10.930
1994~2003(114)	10.868	11.868	11.684	11.474
1995~2004(114)	12.342	11.763	11.368	11.825
邊緣平均數	11.746	11.697	11.741	

註：括號內為個數

在以 SLP 為趨勢預測指標下，以期刊預測會議主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，驗證集為 1 年平均得分為 11.746；驗證集為 2 年平均得分為 11.697；驗證集為 3 年平均得分為 11.741。

表 4- 69「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
.895	17	666	.581

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+add_ans+year+add_ans * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=0.895，P 值=0. 581 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 70「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
驗證集範圍(add_ans)	.325	2	.162	.003 n.s	.997
年代(year)	124.877	5	24.975	.420 n.s	.835
add_ans * year(交互作用)	95.167	10	9.517	.160 n.s	.999
誤差	39577.053	666	59.425		
全體	133880.000	684			

n.s=不顯著

a Computed using alpha = .05
b R Squared = .006 (Adjusted R Squared = -.020)

經雙因子變異數分析檢定，對於「不同驗證集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.999 > 0.05，未達顯著性，表示其交互作用不顯著。

(三) 以會議文獻預測期刊主題之不同驗證集年代範圍差異考驗

1. 以 **API** 為預測指標

表 4- 71「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

驗證集範圍 / 預測集年代區間	驗證集為 1 年 (228)	驗證集為 2 年 (228)	驗證集為 3 年 (228)	邊緣平均數
1990~1999(114)	11.737	11.026	11.342	11.368
1991~2000(114)	11.263	11.316	12.553	11.711
1992~2001(114)	11.474	12.053	12.553	12.026
1993~2002(114)	10.868	12.421	11.158	11.482
1994~2003(114)	14.684	12.395	12.316	13.132
1995~2004(114)	10.105	11.605	11.342	11.018
邊緣平均數	11.689	11.803	11.877	
註：括號內為個數				

在以 API 為趨勢預測指標下，以會議預測期刊主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，驗證集為 1 年平均得分為 11.689；驗證集為 2 年平均得分為 11.803；驗證集為 3 年平均得分為 11.877。

表 4- 72「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.088	17	666	.361

Tests the null hypothesis that the error variance of the dependent variable is equal acr oss groups.
a Design: Intercept+add_ans+year+add_ans * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.088，P 值=0. 361 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 73「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
驗證集範圍(add_ans)	4.114	2	2.057	.032 n.s	.969
年代(year)	311.333	5	62.267	.957 n.s	.444
add_ans * year(交互作用)	306.237	10	30.624	.471 n.s	.909
誤差	43334.000	666	65.066		
全體	139026.000	684			
n.s=不顯著	43955.684	683			

a Computed using alpha = .05
b R Squared = .014 (Adjusted R Squared = -.011)

經雙因子變異數分析檢定，對於「不同驗證集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.909 > 0.05，未達顯著性，表示其交互作用不顯著。

2. 以 SLP 為預測指標

表 4- 74「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

驗證集範圍 / 預測集年代區間	驗證集為 1 年 (228)	驗證集為 2 年 (228)	驗證集為 3 年 (228)	邊緣平均數
1990~1999(114)	11.079	10.947	11.579	11.202
1991~2000(114)	11.974	12.763	13.000	12.579
1992~2001(114)	12.053	12.263	12.868	12.395
1993~2002(114)	11.789	12.605	11.921	12.105
1994~2003(114)	13.447	11.947	11.921	12.439
1995~2004(114)	9.105	9.763	10.184	9.684
邊緣平均數	11.575	11.715	11.912	
註：括號內為個數				

在以 SLP 為趨勢預測指標下，以會議預測期刊主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，驗證集為 1 年平均得分為 11.575；驗證集為 2 年平均得分為 11.715；驗證集為 3 年平均得分為 11.912。

表 4- 75「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨

勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
.664	17	666	.839

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+add_ans+year+add_ans * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=0.664，P 值=0. 839 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 76「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
驗證集範圍(add_ans)	13.126	2	6.563	.086 n.s	.918
年代(year)	714.749	5	142.950	1.872 n.s	.097
add_ans * year(交互作用)	125.962	10	12.596	.165 n.s	.998
誤差	50853.737	666	76.357		
全體	145884.000	684			

n.s=不顯著

a Computed using alpha = .05
b R Squared = .017 (Adjusted R Squared = -.009)

經雙因子變異數分析檢定，對於「不同驗證集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.998 > 0.05，未達顯著性，表示其交互作用不顯著。

(四) 以期刊文獻預測期刊主題之不同驗證集年代範圍差異考驗

1. 以 **API** 為預測指標

表 4- 77「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

驗證集範圍 / 預測集年代區間	驗證集為 1 年 (228)	驗證集為 2 年 (228)	驗證集為 3 年 (228)	邊緣平均數
1990~1999(114)	10.132	10.842	10.211	10.395
1991~2000(114)	10.921	10.816	10.263	10.667
1992~2001(114)	9.000	9.211	9.868	9.360
1993~2002(114)	9.842	9.658	10.447	9.982
1994~2003(114)	10.053	9.921	9.579	9.851
1995~2004(114)	10.395	10.000	10.632	10.342
邊緣平均數	10.057	10.075	10.167	
註：括號內為個數				

在以 API 為趨勢預測指標下，以期刊預測期刊主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，驗證集為 1 年平均得分為 10.057；驗證集為 2 年平均得分為 10.075；驗證集為 3 年平均得分為 10.167。

表 4- 78「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.215	17	666	.246

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+add_ans+year+add_ans * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.215，P 值=0. 246 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 79「不同驗證集年代範圍」與「各年度區間」以 API 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
驗證集範圍(add_ans)	1.582	2	.791	.016 n.s	.984
年代(year)	124.327	5	24.865	.505 n.s	.772
add_ans * year(交互作用)	60.278	10	6.028	.122 n.s	1.000
誤差	32775.053	666	49.212		
全體	102728.000	684			
n.s=不顯著					

a Computed using alpha = .05
b R Squared = .006 (Adjusted R Squared = -.020)

經雙因子變異數分析檢定，對於「不同驗證集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 1.000 > 0.05，未達顯著性，表示其交互作用不顯著。

2. 以 **SLP** 為預測指標

表 4- 80「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

驗證集範圍 / 預測集年代區間	驗證集為 1 年 (228)	驗證集為 2 年 (228)	驗證集為 3 年 (228)	邊緣平均數
1990~1999(114)	8.526	9.132	8.447	8.702
1991~2000(114)	9.763	8.921	8.000	8.895
1992~2001(114)	7.947	7.211	8.184	7.781
1993~2002(114)	9.395	9.000	8.789	9.061
1994~2003(114)	8.789	8.079	7.368	8.079
1995~2004(114)	9.974	8.947	9.158	9.360
邊緣平均數	9.066	8.548	8.325	
註：括號內為個數				

在以 SLP 為趨勢預測指標下，以期刊預測期刊主題的情況而言，依其預測集範圍大小所得之主題預測平均數得分可知，驗證集為 1 年平均得分為 9.066；驗證集為 2 年平均得分為 8.548；驗證集為 3 年平均得分為 8.325。

表 4- 81「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨

勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.058	17	666	.392

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+add_ans+year+add_ans * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.058，P 值=0. 392 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 82「不同驗證集年代範圍」與「各年度區間」以 SLP 為預測指標在主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
驗證集範圍(add_ans)	65.915	2	32.958	.631 n.s	.532
年代(year)	207.152	5	41.430	.794 n.s	.554
add_ans * year(交互作用)	91.313	10	9.131	.175 n.s	.998
誤差	34772.000	666	52.210		
全體	86270.000	684			

n.s=不顯著

a Computed using alpha = .05
b R Squared = .010 (Adjusted R Squared = -.015)

經雙因子變異數分析檢定，對於「不同驗證集年代範圍」與「各年度區間」進行預測之交互作用顯著性為 0.998 > 0.05，未達顯著性，表示其交互作用不顯著。

四、 不同預測指標在主題預測上之差異分析

為了檢驗研究假設(四)，本小節以「不同趨勢預測指標」與「各年度區間」為自變項，以「趨勢預測排序相對位差得分」為依變項，首先將預測集年代範圍設定為 10 年、預測集後 2 年設定為驗證集年代，再分別以兩種預測指標進行四種不同文獻類型的交叉預測，計算主題文獻趨勢預測後所得之結果的平均數差異考驗，以瞭解不同趨勢預測指標對主題文獻預測是否會顯著性的差異存在 。

自變項：(1)趨勢預測指標：API、SLP。

(2)年代區間：以 1990、1991、1992、1993、1994、1995、1996 預測集起始年代。

依變項：(1)主題趨勢預測排序相對位差得分。

(一) 以會議文獻預測會議主題之不同預測指標差異考驗

表 4- 83「不同預測指標」與「各年度區間」在會議預測會議主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測指標 / 預測集年代區間	預測指標為 API (266)	預測指標為 SLP(266)	邊緣平均數
1990~1999(76)	9.132	8.737	8.934
1991~2000(76)	10.368	9.605	9.987
1992~2001(76)	11.605	10.132	10.868
1993~2002(76)	9.737	9.711	9.724
1994~2003(76)	10.132	9.000	9.566
1995~2004(76)	9.789	8.895	9.342
1996~2005(76)	11.895	9.132	10.513
邊緣平均數	10.380	9.316	
註：括號內為個數			

以會議預測會議主題的情況而言，其依據兩種預測指標所得之主題預測平均數得分可知，預測指標為 API 平均得分為 10.380；預測指標為 SLP 平均得分為 9.316。

表 4- 84「不同預測指標」與「各年度區間」在會議預測會議主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
.673	13	518	.790

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept+trend+year+trend * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=0.673，P 值=0. 790 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 85「不同預測指標」與「各年度區間」在會議預測會議主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
趨勢指標(trend)	150.543	1	150.543	2.912 n.s	.089
年代(year)	204.365	6	34.061	.659 n.s	.683
trend * year(交互作用)	89.365	6	14.894	.288 n.s	.943
誤差	26782.395	518	51.703		
全體	78819.000	532			
n.s=不顯著					

a Computed using alpha = .05
b R Squared = .016 (Adjusted R Squared = -.008)

經雙因子變異數分析檢定，對於「不同預測指標」與「各年度區間」進行預測之交互作用顯著性為 0.943 > 0.05，未達顯著性，表示其交互作用不顯著。

(二) 以期刊文獻預測會議主題之不同驗證集年代範圍差異考驗

表 4- 86「不同預測指標」與「各年度區間」在期刊預測會議主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測指標 / 預測集年代區間	預測指標為 API (266)	預測指標為 SLP(266)	邊緣平均數
1990~1999(76)	10.947	12.184	11.566
1991~2000(76)	10.658	12.184	11.421
1992~2001(76)	12.026	11.868	11.947
1993~2002(76)	9.763	10.316	10.039
1994~2003(76)	10.974	11.868	11.421
1995~2004(76)	11.658	11.763	11.711
1996~2005(76)	12.368	12.079	12.224
邊緣平均數	11.199	11.752	
註：括號內為個數			

以期刊預測會議主題的情況而言，其依據兩種預測指標所得之主題預測平均數得分可知，預測指標為 API 平均得分為 11.199；預測指標為 SLP 平均得分為 11.752。

表 4- 87「不同預測指標」與「各年度區間」在期刊預測會議主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
.915	13	518	.537

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+trend+year+trend * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=0.915，P 值=0. 537 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 88「不同預測指標」與「各年度區間」在期刊預測會議主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
趨勢指標(trend)	40.618	1	40.618	.688 n.s	.407
年代(year)	221.459	6	36.910	.625 n.s	.710
trend * year(交互作用)	56.000	6	9.333	.158 n.s	.987
誤差	30588.605	518	59.051		
全體	100965.000	532			
n.s=不顯著					

a Computed using alpha = .05
b R Squared = .010 (Adjusted R Squared = -.015)

經雙因子變異數分析檢定，對於「不同預測指標」與「各年度區間」進行預測之交互作用顯著性為 0.987 > 0.05，未達顯著性，表示其交互作用不顯著。

(三) 以會議文獻預測期刊主題之不同驗證集年代範圍差異考驗

表 4- 89「不同預測指標」與「各年度區間」在會議預測期刊主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測指標 / 預測集年代區間	預測指標為 API (266)	預測指標為 SLP(266)	邊緣平均數
1990~1999(76)	11.026	10.947	10.987
1991~2000(76)	11.316	12.763	12.039
1992~2001(76)	12.053	12.263	12.158
1993~2002(76)	12.421	12.605	12.513
1994~2003(76)	12.395	11.947	12.171
1995~2004(76)	11.605	9.763	10.684
1996~2005(76)	11.368	10.763	11.066
邊緣平均數	11.741	11.579	
註：括號內為個數			

以會議預測期刊主題的情況而言，其依據兩種預測指標所得之主題預測平均數得分可知，預測指標為 API 平均得分為 11.741；預測指標為 SLP 平均得分為 11.579。

表 4- 90「不同預測指標」與「各年度區間」在會議預測期刊主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
.609	13	518	.848

Tests the null hypothesis that the error variance of the depend ent variable is equal across groups.
a Design: Intercept+trend+year+trend * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=0.609，P 值=0. 848 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 91「不同預測指標」與「各年度區間」在會議預測期刊主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
趨勢指標(trend)	3.476	1	3.476	.049 n.s	.825
年代(year)	238.590	6	39.765	.559 n.s	.763
trend * year(交互作用)	113.169	6	18.862	.265 n.s	.953
誤差	36824.184	518	71.089		
全體	109505.000	532			
n.s=不顯著	37179.419	531			

a Computed using alpha = .05
b R Squared = .010 (Adjusted R Squared = -.015)

經雙因子變異數分析檢定，對於「不同預測指標」與「各年度區間」進行預測之交互作用顯著性為 0.953 > 0.05，未達顯著性，表示其交互作用不顯著。

(四) 以期刊文獻預測期刊主題之不同驗證集年代範圍差異 考驗

表 4- 92「不同預測指標」與「各年度區間」在期刊預測期刊主題趨勢預測排序相對位差得分之各項細格平均數與邊緣平均數

預測指標 / 預測集年代區間	預測指標為 API (266)	預測指標為 SLP(266)	邊緣平均數
1990~1999(76)	10.842	9.132	9.987
1991~2000(76)	10.816	8.921	9.868
1992~2001(76)	9.211	7.211	8.211
1993~2002(76)	9.658	9.000	9.329
1994~2003(76)	9.921	8.079	9.000
1995~2004(76)	10.000	8.947	9.474
1996~2005(76)	12.053	10.079	11.066
邊緣平均數	10.357	8.767	
註：括號內為個數			

以期刊預測期刊主題的情況而言，其依據兩種預測指標所得之主題預測平均數得分可知，預測指標為 API 平均得分為 10.357；預測指標為 SLP 平均得分為 8.767。

表 4- 93「不同預測指標」與「各年度區間」在期刊預測期刊主題趨勢預測排序相對位差得分之 Levene's Test

Levene's Test of Equality of Error Variances(a)

Dependent Variable: 主題趨勢預測排序之相對位差得分

F	df1	df2	Sig.
1.468	13	518	.125

Tests the null hypothesis that the error variance of the dependent vari able is equal across groups.
a Design: Intercept+trend+year+trend * year

利用 Levene's Test 判定變異數同質性之檢定，查看報表結果，F=1.468，P 值=0. 125 > 0.05，未達顯著水準，代表變異數是同質性，可以繼續檢定。

表 4- 94「不同預測指標」與「各年度區間」在期刊預測期刊主題趨勢預測排序相對位差得分之二因子變異數分析摘要表

Tests of Between-Subjects Effects

Dependent Variable: 主題趨勢預測排序之相對位差得分

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	事後比較
趨勢指標(trend)	336.333	1	336.333	6.857*	.009	SLP < API
年代(year)	360.256	6	60.043	1.224 n.s	.292	
trend * year(交互作用)	31.233	6	5.206	.106 n.s	.996	
誤差	25409.132	518	49.052			
全體	74779.000	532				
*p < .05 n.s=不顯著	26136.953	531				

a Computed using alpha = .05
b R Squared = .028 (Adjusted R Squared = .003)

經雙因子變異數分析檢定，對於「不同預測指標」與「各年度區間」進行預測之交互作用顯著性為 0.996 > 0.05，未達顯著性，表示其交互作用不顯著。但在趨勢指標的部分顯著性為.009 < 0.05，在 95%的信心水準下達到顯著性差異，亦即在期刊預測期刊的情況下以 API 和 SLP 進行預測後的結果會有顯著的差異性存在。

第四節　小結

本節將實驗所得到的資料與數據，經統計分析後將其結果綜合整理，並依序以三個面向說明：

一、　主題年代序列之結果

將期刊文獻的 ID 詞彙與會議文獻主題詞相對應，找出無法對應之主題再以自動化歸類結果輔助檢驗後可發現，會議文獻與期刊文獻各自有偏重的主題。會議文獻較注重的主題是 Question and answering、Multi-lingual IR、Summarization、Distributed(IR)；期刊文獻較為注重的主題是 INFORMATION-SEEKING、ALGORITHM(S)、LIBRARY；CATALOG、BIBLIOMETRICS。

在可對應之主題比較上來說，以整體趨勢而言，第二種主題整併下會議主題與期刊主題趨勢較為一致。第一種主題整併的 18 個主題中，會議主題與期刊主題成長趨勢一致的有 8 個，約有 44%的主題趨勢走向相同；第二種主題整併的 11 個主題中則共有 7 個成長趨勢一致，約有 63%的主題趨勢走向相同。由此可知，當主題涵蓋範圍越大時，會議文獻與期刊文獻的主題成長趨勢走向越為一致。

在細部主題文獻數量的比較上來說，第一種主題整併的 18 個主題中，期刊文獻比會議文獻多的主題是 Relevance Feedback 和 Language(Retrieval)(Formal) Models，而會議文獻比期刊文獻多的主題則是 Text Classification、Categorisation(Classifiers)、Efficiency&Scalability、Languag、NLP(Natural Language Processing)。第二種主題整併後的 11 個主題中，期刊文獻中比例最多且差距與會議文獻最大的主題是 DB&IR(Integration)(Structures)、Systems、Knowledge Based Information Retrieval；會議文獻比期刊文獻多的主題是 Natural Language Processing、Cross language、MIR，兩者文獻數量都偏少的主題是 Distributed(IR)(Systems)，各大類主題下有各自的發展特色。在自動化歸類的 38 個主題中，期刊比會議文獻多的主題是 interactive&information-seeking、citation、recall and precision；會議比期刊文獻多的主題是 track、question answer。

在初始年代的差距比較上來看，第一種主題整併的 18 個主題中，會議文獻較多主題早於期刊文獻，總計有 9 個主題；第二種主題整併後的 11 個主題中，兩者同時出現佔多數；在自動化歸類的 38 個主題中則是會議文獻比期刊文獻早

出現的主題數量佔多數。由此可知，雖然會議文獻與期刊文獻在細部主題上的出現的先後順序上並非完全以會議文獻為優先，但會議文獻比期刊文獻出現較早的主題仍佔大多數，且主題範圍越廣，兩者出現的先後順序差異越小。

二、 會議與期刊文獻主題趨勢預測差異

本研究根據本章第二節所述及第三節依研究假設所進行之假設檢定所述結果，綜合說明如下：

(一) 以 API 為預測指標所進行之假設檢定

當預測指標以 API 為計算方式時，其結果如表 4-95 所示，在主題趨勢預測的部份，在第一種主題整併以會議預測會議的效果最佳、期刊預測期刊的效果最差，且兩者之間有達到信心水準為 95%的顯著性差異；第二種主題整併則是以期刊預測會議為最佳、會議預測期刊效果最差，且兩者之間有達到信心水準為 95%的顯著性差異；在自動化歸類的結果則顯示相同文獻類型預測之效果較佳，相異類型的文獻預測效果則較差，但未達到顯著性差異。

表 4- 95 以 API 為預測指標之各項假設檢定之綜合整理

文獻類型 / 組別控制項目	Type=1 會議預測會議	Type=2 期刊預測會議	Type=3 會議預測期刊	Type=4 期刊預測期刊	顯著性差異之事後比較
第一種主題整併 (18)	**4.278**	4.741	5.519	6.111	Type1 < Type4
第二種主題整併 (11)	3.712	**2.712**	4.227	3.833	Type2 < Type3
自動化歸類 (38)	10.380	11.199	11.741	10.357	
	以自動化歸類為基礎之實驗				
預測集 = 4年	10.289	**10.782**	**10.711**	**9.857**	無顯著差異
預測集 = 6年	9.508	10.842	10.808	9.932	
預測集 = 8 年	**9.342**	11.538	11.538	10.297	
預測集 = 10 年	10.380	11.199	11.741	10.357	
	以自動化歸類為基礎之實驗				
驗證集 = 1 年	10.158	11.325	**11.689**	**10.057**	無顯著差異
驗證集 = 2 年	**10.127**	**11.004**	11.803	10.075	
驗證集 = 3 年	10.461	11.136	11.877	10.167	
註：數值越小，代表預測效果越佳					

由第一種主題整併之預測結果可知，會議文獻預測會議文獻優於期刊文獻預測期刊文獻，Cai and Card(2007)曾指出會議文獻之主題比期刊文獻集中，因此有可能是此原因導致會議文獻預測會議文獻之效果較佳。而第二種主題整併後的結果與第一種主題整併之結果有所不同，反而是在期刊預測會議文獻時預測效果較佳，由此可以推測在主題較廣的時候，期刊文獻預測較佔優勢，在主題較細時，會議文獻預測佔優勢。

在預測集年代範圍的假設檢定結果中可看出，當期刊文獻預測會議主題、會議文獻預測期刊主題、期刊文獻預測期刊主題時，預測集年代範圍越大則效果越不好，亦即預測集為 4 年時效果最佳，但是在會議預測會議主題時，反而以預測集為 8 年時效果最好，但其結果均未達到統計上的顯著差異，所以無法證明預測集範圍大小足以明顯影響到預測之效果。

在驗證集範圍的假設檢定中，當預測對象為會議主題時，驗證集為 2 年的效果最佳；當預測對象為期刊主題時，驗證集為 1 年的效果最佳，但其結果均未達到統計上的顯著差異，亦即無法證明預測集範圍大小足以明顯影響到預測之效果。

(二) 以 SLP 為預測指標所進行之假設檢定

當預測指標以 SLP 為計算方式時，其結果如表 4-96 所示，在主題趨勢預測的部分，第一種主題整併下的預測皆無顯著性差異，第二種主題整併則是以期刊預測期刊為最佳、會議預測期刊效果最差，且兩者之間有達到信心水準為 95% 的顯著性差異；期刊預測期刊之效果優於會議預測會議和會議預測期刊，亦有達到信心水準為 95%的顯著性差異，比預測指標為 API 時更加深在第二種主題整併後，期刊在主題趨勢預測上所佔的優勢；在自動化歸類的預測結果則顯示相同文獻類型預測之效果較佳，相異類型的文獻預測效果則較差，此結果與預測指標為 API 時相同，且同樣也有達到信心水準為 95%的顯著性差異。

在預測集年代範圍的假設檢定結果中可看出，當文獻進行交叉預測時，預測集年代範圍越大則效果越不好，亦即當期刊文獻預測會議主題、會議文獻預測期刊主題時，都顯示出預測集為 4 年時效果最佳，而在會議預測會議主題時，反而以預測集為 8 年時效果最好，此與預測指標為 API 時相同，只有在期刊預測期刊時顯示出當預測集範圍越大則效果越佳，此與預測指標為 API 時相異，然其結果

亦均未達到統計上的顯著差異，即無法證明選擇的預測集範圍大小明顯影響到預測之效果。

在驗證集的範圍的假設檢定中，當預測對象為會議主題時，驗證集為 2 年的效果最佳，此與預測指標為 API 時之結果相同；當會議文獻預測期刊主題時，驗證集為 1 年的效果最佳，此與預測指標為 API 時之結果相同；當期刊文獻預測期刊主題時，驗證集範圍則是以 3 年最佳，此與預測指標為 API 時相異，但其結果亦均未達到統計上的顯著差異，即無法證明選擇的預測集範圍大小明顯影響到預測之效果。

表 4- 96 以 SLP 為預測指標之各項假設檢定之綜合整理

文獻類型 / 組別控制項目	Type=1 會議預測會議	Type=2 期刊預測會議	Type=3 會議預測期刊	Type=4 期刊預測期刊	顯著性差異 之事後比較
第一種主題整併 (18)	4.509	4.713	5.398	4.694	無顯著差異
第二種主題整併 (11)	3.242	2.970	4.667	**2.000**	Type4 < Type1 < Type3 Type2 < Type3
自動化歸類 (38)	**9.316**	11.752	11.579	**8.767**	Type4 < Type1 < Type2 Type4 < Type1 < Type3
	以自動化歸類為基礎之實驗				
預測集 = 4年	10.643	**11.038**	**11.034**	10.165	無顯著差異
預測集 = 6年	9.876	11.237	11.289	9.312	
預測集 = 8 年	**9.218**	11.752	11.564	9.346	
預測集 = 10 年	9.316	11.752	11.579	**8.767**	
	以自動化歸類為基礎之實驗				
驗證集 = 1 年	9.825	11.746	**11.575**	9.066	無顯著差異
驗證集 = 2 年	**9.346**	**11.697**	11.715	8.548	
驗證集 = 3 年	9.548	11.741	11.912	**8.325**	
註：數值越小，代表預測效果越佳					

(三) 針對兩種預測指標所進行之假設檢定

在採用不同的預測指標所進行之假設檢定結果歸納如表 4-97，其預測指標在三種情況下的結果均以 SLP 之效果較佳，分別是會議預測會議、會議預測期刊、期刊預測期刊，惟期刊預測會議的情況是以 API 為佳，尤其在期刊預測期刊的情況時，兩種不同的預測指標之差異有達到 95%信心水準下的顯著差異，也就是說在期刊文獻預測期刊主題時，選用之指標對預測主題之結果有顯著性差異存在 。

表 4- 97 兩種不同預測指標之假設檢定統整

	預測指標為 API	預測指標為 SLP	顯著性差異與事後比較
Type=1 會議預測會議	10.380	**9.316**	無顯著差異
Type=2 期刊預測會議	**11.199**	11.752	無顯著差異
Type=3 會議預測期刊	11.741	**11.579**	無顯著差異
Type=4 期刊預測期刊	10.357	**8.767**	SLP < API

註：數值越小，代表預測效果越佳

在預測指標的假設檢定上，僅在期刊文獻預測期刊主題時有達到信心水準為95%的顯著性差異，亦即在此情況下採用 SLP 為預測指標下會比採用 API 作為預測指標的結果較佳。若將本研究中採用兩種指標的所有預測皆拿來比較，則可由表 4-98 看出兩者之差異，在主題預測的方面，會議預測會議和期刊預測期刊都可以發現，預測指標採用 SLP 會比 API 的預測效果佳，但在交叉預測的部分卻無發現此現象，亦即相同類型文獻預測下以 SLP 為預測指標其效果較佳 。

表 4- 98 採用兩種預測指標下的各次預測結果統計表

文獻類型 組別控制項目	Type=1 會議預測會議		Type=2 期刊預測會議		Type=3 會議預測期刊		Type=4 期刊預測期刊	
	API	SLP	API	SLP	API	SLP	API	SLP
第一種主題整併	4.509	**4.278**	4.713	4.741	5.519	5.398	6.111	**4.694**
第二種主題整併	3.712	**3.242**	2.712	2.970	4.227	4.667	3.833	**2.000**
自動化歸類	10.380	**9.316**	11.199	11.752	11.741	11.579	10.357	**8.767**
預測集 =4年	10.289	10.643	**10.782**	**11.038**	**10.711**	**11.034**	**9.857**	10.165
預測集 =6年	9.508	9.876	10.842	11.237	10.808	11.289	9.932	9.312
預測集 =8 年	**9.342**	**9.218**	11.538	11.752	11.538	11.564	10.297	9.346
預測集 =10 年	10.380	9.316	11.199	11.752	11.741	11.579	10.357	**8.767**
驗證集 =1 年	10.158	**9.825**	**11.325**	11.746	11.689	11.575	10.057	**9.066**
驗證集 =2 年	10.127	**9.346**	**11.004**	11.697	11.803	11.715	10.075	**8.548**
驗證集 =3 年	10.461	**9.548**	**11.136**	11.741	11.877	11.912	10.167	**8.325**
註：數值越小，代表預測效果越佳								

在預測集範圍大小的預測上，在兩種預測指標下的效果可於表 4-93 中看出，無論指標採用 API 或 SLP 皆可發現會議預測會議時預測集為 8 年最佳；而交叉類型預測(期刊預測會議、會議預測期刊)的部份則是以預測集為 4 年較佳；惟獨在期刊預測期刊的情況時則是採用 API 和 SLP 之結果會完全相反，亦即在期刊預測期刊時，當預測指標為 API 則預測集範圍越小越佳，但當預測指標為 SLP 則預測集範圍越大越佳。

在驗證集範圍大小的進行預測的結果上則會發現，相同類型文獻預測(期刊預測會議、會議預測期刊)時，採用 SLP 作為預測指標皆比採用 API 做預測指標其預測結果都較佳；在預測不同類型文獻(期刊預測會議、會議預測期刊)時則不一定，期刊預測會議的情況下普遍以 API 為預測指標較佳。

三、 綜合討論

本研究以三種不同的主題整併方式為基礎進行主題預測各有其優缺點，第一種主題整併與第二種主題整併之不同在於第二種主題整併有較宏觀的看法，第一種主題整併則是以較細部的觀察各主題間的變化，但無論是第一種主題整併或第二種主題整併都需透過人為的主觀介入，因此自動化歸類則可以排除這樣的疑慮。

由於本研究所採用之文獻類型兩者之間本身即有差異存在，故仍舊無法完全統整兩者之間的差異，僅能就兩者之間在字面上或主題上可看出相同之主題進行比較分析。而表 4-99 則是以文獻交叉預測後四種情況按主題預測後所得之分數排序，其分數越小者代表準確度越高，由此可看出第三種類型的預測大多排序在後，而第一類型和第四類型的預測排序較前。

表 4- 99 交叉預測之情形依準確度大小排序

預測指標 / 整併形式	以 API 為預測指標	以 SLP 為預測指標
第一種主題整併	1 < 2 < 3 < 4	1 < 4 < 2 < 3
第二種主題整併	2 < 1 < 4 < 3	4 << 2 < 1 <<3
自動化歸類	4 < 1 < 2 < 3	4 <<1 << 3 < 2

註：<<有達到信心水準 95%顯著性差異

表 4-100 則是將四種類型的預測按照驗證集區分，亦即會議文獻與期刊文獻同時預測會議文獻；會議預測和期刊文獻同時預測期刊文獻，由表中可看出當驗證集為會議主題時，兩種預測的準確度各有優劣，且不因預測指標改變而有所不同，而在驗證集為期刊主題時，期刊預測期刊(第四類型)較準確佔大多數。

表 4- 100 依驗證集為區分對象之交叉預測情形

指標類型 / 整併形式	驗證集為會議文獻		驗證集為期刊文獻	
	以 API 為預測指標	以 SLP 為預測指標	以 API 為預測指標	以 SLP 為預測指標
第一種主題整併	1 < 2	1 < 2	3 < 4	4 < 3
第二種主題整併	2 < 1	2 < 1	4 < 3	4 << 3
自動化歸類	1 < 2	1 << 2	4 < 3	4 << 3

註：<<有達到信心水準 95%顯著性差異

第五章 結論與建議

本研究旨在探討會議文獻與期刊文獻進行主題趨勢預測時的相對性效果比較。首先將會議文獻與期刊文獻各自取出主題詞彙進行年代 分佈的趨勢比較，再依照各年度範圍分為預測集與驗證集，將預測集的主題依據趨勢預測指標計算出各主題的成長率，再依其成長率之大小排序，最後比對驗證集與預測集之間的相對位差，做為主題趨勢預測之得分，以展現出會議文獻與期刊文獻在主題趨勢預測上的相對性效果。而本研究採用雙因子變異數分析檢定會議文獻與期刊文獻預測之平均得分，以比較兩者之間的預測結果、預測集與驗證集的年代範圍、預測指標之間的平均得分是否有達到顯著性差異 。本章將總結前述各章節之研究結果，並對未來的相關研究提出進一步的建議。

第一節 結論

本節將針對三部分的研究結果進行總結，分別是資訊檢索領域中會議文獻與期刊文獻在主題分佈上的比較、期刊文獻與會議文獻進行主題趨勢預測之相對性效果、預測集與驗證集年代範圍大小和預測指標對預測效力之影響，以下茲分別說明之。

一、 資訊檢索領域會議文獻與期刊文獻之主題分佈

資訊檢索領域中，1991 年至 2007 年出版的 SIGIR 會議文獻與五本核心期刊文獻的主題比較上可發現，兩者之間的主題有無法比對之情形存在，且兩者各自有偏重發展之主題，以下分別就各項特點進行說明。

(一) 會議文獻與期刊文獻主題詞彙上有所差異

會議文獻與期刊文獻對主題詞的用法有所差異，如期刊文獻所探討關於「使用者」(USER)的議題，會議文獻則會使用「個人化」(Personalization)來描述相關議題，如此一來對於後續進行主題研究趨勢之預測，即會造成相似主題間的比對問題。

本研究亦發現，期刊文獻的主題詞較易出現廣義的詞彙，而會議文獻則較易出現狹義的詞彙。而期刊文獻在資料庫中的 ID 主題詞仍然有權威控制上的問題，亦即本研究從 ID 詞彙中找出許多同義詞彙，如英文的單複數詞彙、英文詞性、縮寫詞彙等。

(二) 會議文獻與期刊文獻各自有較關注探討的主題

本研究發現以無法相對應之主題來看，會議文獻較注重的主題是 Question and answering、Multi-lingual IR、Summarization、Distributed(IR)；期刊文獻較為注重的主題是 INFORMATION-SEEKING、ALGORITHM(S) 、LIBRARY；CATALOG、BIBLIOMETRICS。以可對應之主題來說，在主題萃取系統自動歸類之結果可發現，對會議文獻來說，期刊文獻較少出的主題分別是：track、question answer；對期刊文獻來說，會議文獻較少出的主題則是：interactive ；information-seeking、citation。

(三) 主題涵蓋範圍越大，整體趨勢走向越為一致

以整體趨勢而言，第二種主題整併下比第一種主題整併下的會議主題與期刊主題趨勢走向較為一致，亦即當主題涵蓋範圍越大時，會議文獻與期刊文獻的主題成長趨勢走向越為一致。

(四) 大部分的主題在會議文獻比在期刊文獻較早出現

會議文獻與期刊文獻在年代分佈上的情形則可看出，會議文獻在第一種主題整併與自動化歸類的結果下，都早於期刊文獻的主題數量佔多數，在第二種主題整併下兩者之間的初始年代則大多相同。由此可知，雖然本研究中仍有少部分期刊主題的首次出現年代早於會議文獻，但大多數的情況下會議文獻仍早於期刊文獻，此亦符合資訊循環圖中會議文獻較早於期刊文獻的概念。

(五) 自動化歸類下的樣本數量較為平均

本研究中自動化歸類下的主題分佈比其他主題整併的主題分佈較為平均，樣本數量上的差距也較小，又可排除人為判斷的主觀性，因此主題萃取系統是很適合拿來進行其他實驗的歸類方式。

二、 資訊檢索領域會議文獻與期刊文獻預測主題趨勢之相對效果

本研究以會議文獻和期刊文獻分別進行主題趨勢預測，並將兩種文獻各自為驗證集的情況下進行交叉預測，以觀察會議文獻和期刊文獻在主題趨勢預測上的相對性效果，並採雙因子變異數分析進行檢定，以下分別就檢定後之結果做出結論。

(一) 主題範圍越寬廣，期刊文獻預測越佔優勢

在第一種主題整併的情況，在 95%信心水準下的統計檢定可發現，會議文獻預測會議主題之效果優於期刊文獻預測期刊主題之效果。第二種主題整併後的情況則發現，以期刊文獻預測較佔優勢，期刊文獻預測期刊主題的效果優於其他各項預測，但由於在此情況下的會議文獻與期刊文獻數量差距多達 481 篇，因此是否為數量的差距而造成此結果仍無法確定。主題萃取系統自動歸類後的結果則發現，相同文獻類型的預測效果均優於相異類型的文獻預測。

(二) 相同類型文獻預測之結果優於交叉預測

若依驗證集為區分對象所進行的交叉預測結果比對則可發現，當驗證集以會議主題為準時，會議文獻預測會議主題的效果優於期刊文獻預測會議主題的比例較高，也就是說，進行會議文獻的主題趨勢預測要以會議文獻為佳；當驗證集以期刊主題為準時，期刊文獻預測期刊主題的效果優於會議文獻預測期刊主題，也就是說，進行期刊文獻的主題趨勢預測要以期刊文獻為佳，此亦呼應了在自動化歸類結果下，當預測集與驗證集的文獻類型相同時，其預測效果都會比較好。

(三) 會議文獻比期刊文獻出現時間較早，但預測效果並未較佳。

雖然會議文獻的初始年代較早於期刊文獻的比例較高，容易讓人認為以會議文獻預測期刊文獻之主題為可行之方式，但根據本研究的主題預測結果顯示，交叉預測的效力仍未優於相同文獻類型預測之效果，故在進行主題趨勢預測時，仍以相同類型之文獻預測較佳。另外，會議文獻預測效果並未較佳之原因，亦可能為會議文獻各年度出現的主題容易有斷層，而期刊文獻的樣本較多均呈現各年度都有的穩定趨勢，此亦有可能造成會議文獻在預測上的效果並未較佳 。

三、 預測集、驗證集年代範圍大小和預測指標對預測效力之影響

進行主題趨勢預測時，本研究皆採用預測集為 10 年、驗證集為 2 年，並以自動化歸類之結果進行各項預測，而預測集、驗證集年代範圍大小與預測指標的不同是否影響預測結果，亦為本研究所欲瞭解之項目 。故本研究針對改變預測集、驗證集年代範圍大小與預測指標之預測結果進行檢定，以下分別就各項檢定結果敘述之。

(一) 改變預測集年代範圍並未造成顯著性差異

本研究根據系統自動歸類後的結果進行實驗發現 ，在期刊文獻預測會議主題、會議文獻預測期刊主題時，預測集年代範圍越大則效果越不好，亦即當預測集與驗證集的文獻類型不同時，其預測結果顯示出預測集為 4 年時效果最佳，但因未達到信心水準 95%顯著性差異，故無法推斷是否預測集的範圍大小會對預測效力有顯著性影響。

(二) 改變驗證集年代範圍並未造成顯著性差異

驗證集的部份則是在會議預測會議與期刊預測會議時都顯示期驗證集為 2 年時效果最佳，而會議文獻預測期刊主題時，則是以 1 年時的效果最佳，但因未達到信心水準 95%顯著性差異，故亦無法推斷是否驗證集的範圍大小會對預測效力有顯著性影響。

(三) 在期刊預測期刊的情況下，不同的預測指標達到顯著性差異

在預測指標的選擇上，本研究針對自動化歸類的主題趨勢預測結果下進行假設檢定後發現，在期刊預測期刊的情況時，兩種不同的預測指標之差異有達到95%信心水準下的顯著性差異，亦即選用 SLP 作為趨勢預測指標會比選用 API 為預測指標佳，此情況亦與 Tseng et al.（2008）所提出以 SLP 作為趨勢預測指標之效果優於以 API 作為預測指標之效果相同。

(四) 選用之預測指標於大部分情況下並未改變預測結果

另外根據本研究多次以兩種指標下進行之主題趨勢預測觀察後發現，在會議預測會議和期刊預測會議時，預測指標採用 SLP 會比 API 的預測效果佳，而在選用預測指標上並不會造成影響的是會議預測會議、期刊預測會議、會議預測期刊，也就是說在這三種情況下，預測指標採用 SLP 和 API 所進行的預測效果都是一致的，惟獨當期刊預測期刊時，選用的預測指標會改變其預測結果。

第二節 建議

本研究採用驗證之方式，比較會議文獻與期刊文獻的主題分佈與相對性預測效果，然而因受研究範圍與時間之限制，僅以 SIGIR 會議文獻與 Web of Science 此跨領域資料庫蒐集 1990 年至 2007 年資訊檢索領域的核心期刊文獻進行分析，並以人工彙整主題詞彙作為研究樣本，藉以探討資訊檢索領域之主題分佈與趨勢預測之相對性效果。根據本研究於彙整主題詞彙時所發現的問題，做出以下幾點建議：

一、 主題趨勢相關之建議

(一) 主題詞彙的建議

本研究採用期刊主題詞進行文獻歸類分析時發現，其控制詞彙的部份尚未完善，許多單複數詞彙和縮寫詞彙皆尚需統整，以提供未來進行趨勢分析的研究者更精確的主題詞彙。

在本研究所採用的會議文獻主題上，因各年度的會議都由不同的主辦單位負責，因此其主題詞的變化性也就更多，此外，各年度的會議文獻也受限於每段議程的發表數量，所以會出現會議文獻的主題涵蓋兩種主題，若可於會議主題上給予控制詞彙，會更易於會議文獻的收錄與後續相關研究的進行 。

(二) 主題預測之建議

由本研究所進行的實驗與研究假設檢定可知，未來若要瞭解大領域的趨勢以期刊文獻預測期刊主題的效果較佳，而要了解各自領域的趨勢則是以相同的文獻類型進行預測較佳，亦即欲知道會議文獻的未來趨勢，要以會議文獻做為預測基礎，而欲瞭解期刊文獻的未來趨勢，則是要以期刊文獻做為預測基礎。

此外，在自動化歸類的結果可看出，會議文獻與期刊文獻在主題分佈的數量較為一致，表示自動化歸類後所取得之結果比主題整併之結果較適合進行後續的實驗研究，而且自動化歸類的結果也可以排除人為因素的干擾。

二、 對未來後續研究之建議

由於本研究之研究範圍、時間之限制，因此僅能於有限的時間內採集研究樣本，故研究仍有不足之處，茲列出下列幾點建議以供進一步之相關研究作為參考。

(一) 研究樣本方面

本研究在研究樣本的選取上，受限於 SIGIR 會議文獻以及資訊檢索領域的核心期刊文獻，而實際上關於資訊檢索議題尚有其他的會議文獻可供研究，因此後續可針對資訊檢索領域的其他會議文獻進行探討，或是利用其他灰色文獻進行研究，像是專利文獻或部落格的文章等 。

另外，在資訊檢索領域的期刊文獻樣本的選取上，本研究是透過資料庫限定主題為「資訊檢索」領域的核心期刊文獻進行探討，因資訊檢索領域的期刊文獻過多，故本研究僅限定於核心期刊為範圍，而後續研究也可針對資訊檢索領域的非核心期刊文獻進行全面性的探討。而會議文獻與期刊文獻的作者群亦有差異存在，本研究尚未針對作者群進行分析與探討，因此無法得知是否為作者群之差異而造成預測結果之影響，故後續研究亦可針對作者群部份進行更深入的探究 。

(二) 研究方法方面

本研究為實驗之量化研究，並未考量其他外在因素，如各年代的社會環境與背景狀況等，因此後續研究可針對相關背景進行探討。

在主題趨勢預測前的文獻分類，因受限於資料庫的主題詞以及會議文獻的會議主題，僅能針對字面上或會議主題上可判斷之分類方式進行主題分類，後續研究可採用專家多人比對等方式以達到更加嚴謹的主題分類。

而本研究所進行預測時所採用的驗證集是以預測集之後兩年，此為較簡便之方法，後續研究也可採用其他的更為嚴謹之方法所找出的熱門主題作為驗證對象，可以更加強在預測效果上的正確性。

本研究的主題趨勢預測方面，僅針對可對應的主題進行預測相對性效果探討，而無法相對應之主題並未再針對其趨勢預測效果進行更深入的研究，後續研究亦可從此方面再加強，以補足無法對應的主題之間預測效果之探討。

(三) 後續的實驗研究

由於本研究是採用預測集為 10 年、驗證集為 2 年的情況下進行實驗，但經過預測集與驗證集年代的改變所得之預測結果可知，尚有其他預測集與驗證集年代區間可得到預測之最佳結果(如表 4-95、表 4-96)，故後續實驗亦可以其他的年代範圍進行主題趨勢預測。

參考文獻

英文部分：

Allen, R. S. (1995). The magnitude of conference proceedings published in physics journals. *Special Libraries 86*(2), 136-144.

Åström, F. (2007). Changes in the LIS research front: Time -sliced cocitation analyses of LIS journal articles, 1990-2004. *Journal of the American Society for Information Science and Technology, 58*(7), 947-957.

Baker, D. R. (1990). Citation Analysis: A methodological review. *Social Work Research & Abstracts, 26*, 3-10.

Borgman, C. L., & Furner, J. (1990). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology, 36*, 1-45.

Cai, K.-Y., & Card, D. (2007). An analysis of research topics in software engineering – 2006 *Journal of Systems and Software, 81*(6), 1051-1058.

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics, 22*(1), 155-205.

Chen, C. (2005). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359-377.

Courtial, J. P. (1994). A cowaord analysis of scientometrics. *SCIENTOMETRICS, 31*(3), 251-260

C.W.Hanson, & Janes, M. (1960). Lack of index in reports of conference. *Journal of the Document, 16*(2), 65-70.

C.W.Hanson, J., M. (1961). Coverage by abstract journals of conference papers. *Journal of the Document, 17*(3), 143-149.

Ding, Y., Chowdhury, G., & Foo, S. (1999a). *Mapping intellectual struture of information retrieval: An author cocitation analysis, 1987-1997.* Paper presented at the 7th International Conference of the International Society for Scientometric and Informetrics, Colima, Mexico.

Ding, Y., Chowdhury, G., & Foo, S. (1999b). *Mapping the development in information retrieval speciality: A bibliometric analysis via journals.* Paper presented at the 7th International Conference of the International Society for Scientometric and Informetrics, Colima, Mexico.

Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management, 37*(6), 817-842

Drott, M. C. (1995). Reexamining the role of conference papers in scholarly communication. *Journal of the American Society for Information Science, 6*(4), 301.

Funk, M. E. (1988). The usefulness of monographic proceedings. *Buletin of Medical Library Association, 6*(1), 14-21.

Gardner, J. (1980). *The conference as an integral component in the science and technology information dissemination network- with some thought on the role of librarianship as a facilitator between the scientist / Engineer and the Printed Work.* Paper presented at the Conference Literature in Science and Technology: Its Role in the Distribution in Information.

Garvey, W. D., Lin, N., Nelson, C. E., & Tomita, K. (1972). Research studies in patterns of scientific communication Ⅱ: The role of nationa meeting in scientific and technical communication. *Information Storage and Retrieval, 8*(4), 164.

Garvey, W. D., Lin, N., Nelson, C. E., & Tomita, K. (1972). Research studies in patterns of scientific communication Ⅲ: Information – exchange processes associated with the production of journal articles. *Information Storage and Retrieval, 8*(5), 207-221.

Garvey, W. D., Lin, N., Nelson, C. E., & Tomita, K. (1972). Research studies in patterns of scientific communication Ⅱ: The role of national meeting in scientific and technical communication. *Information Storage and Retrieval, 8*(4), 159-169.

HE, Q. (1999). Knowledge discovery through co-word analysis *LIBRARY TRENDS, 48*(1), 133-159.

Hjerland, B. (2003). Fundamentals of knowledge organization. *Knowledge organization, 30*(2), 87-111.

Kean, P., & Ronayne, J. (1972). Preliminary communications in chemistry. *Journal of Chemical Documentation, 12*(4), 218-220.

Mills, P. R. (1973). Characteristics of unpblished conference proceedings. *Journal of Documentation, 29*(1), 36-50.

Montesi, M., & Owen, J. M. (2008). From conference to journal publication:How conference papers in software engineering are extended for publication in journals. *Journal of the American Society for Information Science and Technology, 59*(5), 816-829.

Mooers, C. N. (1951). Zatocoding applied to mechanical organization of knowledge. *American Documentation, 2*(1), 20-32.

Pelzer, N. L., & Wiese, W. H. (2003). Bibliometric study of grey literature in core veterinary medical journals. *Journal of the Medical Library Association 91*(4), 434-441.

Perry, B. (1995). *Grey literature in the internatonal monetary fund*. Paper presented at the Second International Conference on Grey Literature.

Persson, O. (1994). The intellectual base and research fronts of JASIS 1968 -1990. *Journal of American Society for Information Science, 45* (1), 31-38.

Rokaya, M., Atlam, E., Fuketa, M., Dorji, T. C., & Aoe, J.-i. (2008). Ranking of field association terms using Co-word analysis *INFORMATION PROCESSING & MANAGEMENT 44*(2), 738-755

Smeaton, A. F., Keogh, G., Gurrin, C., McDonald, K., & Sødring, T. (2002). Analysis of papers from twenty-five years of SIGIR conferences: What Have We Been Doing for the Last Quarter of a Century ? *ACM SIGIR Forum 36*(2), 39 - 43

Subramanyam, K. (1981). *Scientific and technical information resources*. New York: Marcel Dekker.

Tseng, Y.-H., Lin, Y.-I., Kuo, C.-H., & Lee, Y.-Y. (2008). *Which kinds of trend metrics are more effective for emerging trend detection?* Paper presented at the Proceedings of the International Symposium on Webometrics and Web Mining, 國立政治大學公企中心.

Tseng, Y.-H., Lin, C.-J., & Lin, Y.-I. (2007). Text mining techniques for patent analysis. *Information Processing and Management, 43* (5), 1216-1247

Tsay, M.-Y., Jou, S.-J., & Ma, S.-S. (2000). A bibliometric study of semiconductor Literature, 1978-1997. *Scientometrics, 49*(3), 491-509.

van Raan, A. F. J., & Tijssen, R. J. W. (1993). The neural net of neural network research. *Scientometrics, 26*(1), 169-192.

White, H. D., & Griffith, B. C. (1981). Author cocitation : A literature measure of intellectual structure. *Journal of the American Society for Information Science,*

32, 163-171.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co -citation analysis of information science, 1972 -1995. *Journal of the American Society for Information Science, 49*, 327-355.

Woibel. (1995). Metadatea the foundations of resources description. *D-LIB Magazine*.

Zhao, D., & Strotmann, A. (2007). Can citation analysis of web publications better detect research fronts? *Journal of the American Society for Information Science and Technology, 58* (9), 1285 - 1302.

中文部份：

昌炎新(2006)。核心期刊的淵源以及功效分析。武漢科技大學學報，*8*(5)，21-23。

周靜怡、孫坦、陳濤(2007)。共詞可視化：以人類基因組領域為例。情報學報，*26*(4)，532~537。

邱炯友(2006)。學術傳播與期刊出版。臺北市：遠流。

許雅婷(2006)。資訊檢索文獻老化現象之研究-兼論同時法與歷時法之特質。台北市：國立政治大學圖書資訊與檔案學研究所 。

郭華(2006)。論重要的科技信息資源-會議文獻。圖書館工作與研究(1)，25-27。

張正宏(2008)。運用關鍵詞建構知識管理研究網絡。台北市：國防管理學院資訊管理學系。

曾元顯(2007)。農業研究前沿探勘模式與系統之開發(研究計畫，STPI-C-960906)。台北：財團法人國家實驗研究院科技政策研究與資訊中心。

曾元顯、林瑜一(2006)。文字探勘技術在教育評鑑研究發展趨勢分析之應用 。在教育評鑑國際學術研討會，國立台灣師範大學。

黃淑娟、蔣嘉寧、黃擎天、宋雪芳(1998)。由文獻分析檢視會議文獻的傳播 。教育資料科與圖書館學，*35*(4)，340。原文引自 Felix Liebensy, “Lost Information: Unpublished Conference Papers,”in *International Conference on Scientific Information*(Washington: National Academy of Science -National Research Council, 1959), pp.475-495.

蔡佳玲(1990)。我國社會科學學術期刊論文引用灰色文獻之研究 。國立政治大學圖書資訊研究所碩士論文。

蔡明月(1997)。學術傳播與書目計量學。教育資料與圖書館學，*35*(1)，39。

蔡明月(2003)。資訊計量學與文獻特性。臺北市：編譯館。

蔡明月、劉瓊芳(2007)。1992-2005 資訊計量學研究及其發展演變。圖書與資訊學刊，61，42-56。

蔡明月、賴芊卉(2007)。資訊科學引用與被引用文獻分散現象與主題變化研究：1985-2005。圖書資訊學研究，1(2)，1-31。

吳嘉雯(1998)。半導體文獻雙被引現象分析。台北縣：淡江大學教育資料科學學系。

傅雅秀(1996)。從科學傳播的觀點探討中央研究院生命科學專家的資訊尋求行為。圖書館學刊，*11*，133-163。

黃秀琴(1998)。會議文獻作者生產力與其延續出版品之研究:以國防科技學術研討會為例 。台北縣：淡江大學教育資料科學研究所碩士論文 。

黃淑娟、蔣嘉寧、黃擎天(1981)。會議文獻之探討。教育資料科學，*18*(4)，88。

黃惠美(2007)。臺灣地區圖書資訊學文獻高生產作者的引用圖像—以作者雙被引分析為例。中華民國圖書館學會電子報，*5*。

楊世瑩(2005)。*SPSS* 統計分析實務。臺北市：旗標。

鄭琚媛(2004)。臺灣地區生命科學國際會議文獻生產力與延續出版之研究 。台北縣：淡江大學資訊與圖書館學研究所碩士論文 。

羅思嘉(1990)。引用文獻分析與學術傳播研究。中國圖書館學會會報，*66*，73-85。

羅思嘉、陳光華、林純如(1990)。圖書資訊學學術文獻主題分類體系之研究 。臺灣大學圖書資訊學刊，16，185-208。

資料庫與網路資源：

Science Citation Index®. Retrieved 6/1, 2008, from http://scientific.thomson.com/products/sci/

The 30th Annual International ACM SIGIR Conference 23 -27 July 2007, Amsterdam. Retrieved 6/3, 2007, from http://www.sigir2007.org/history.html

中國大百科全書智慧藏。智慧藏學習科技股份有限公司編著。檢索日期：2008，4/26。檢自：http://dblink.ncl.edu.tw/web/Content.asp?ID=377&Query=1

附錄

附錄 一　第一種和第二種主題整併下主題年代序列表

表 1 全部可對應之主題類別年代序列

主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Link analysis	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Web(IR,search,structure,www)	43	0	0	0	0	0	0	0	0	0	4	3	4	3	3	10	10	6
WORLD-WIDE-WEB(WWW,WEB)	63	0	0	0	0	0	0	0	0	0	4	5	7	12	6	8	9	12
Hypertext	4	2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
HYPERMEDIA	5	0	0	1	0	0	0	1	2	0	0	0	0	0	0	0	0	1
HYPERTEXT(HYPERTEXT IR)	15	1	0	0	0	2	1	3	0	1	0	0	1	1	3	0	0	2
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Semantic Models	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Semantics	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
SEMANTICS	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
LSI(Latent semantic indexing)	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
LSI&Theory	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
LATENT SEMANTIC ANALYSIS	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
Probabilistic(IR)	3	0	2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Probabilistic Models	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
PROBABILISTIC MODEL(S) (IR)	17	0	0	0	1	0	0	0	0	0	1	2	3	2	2	0	5	1
PROBABILISTIC INFERENCE	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0

PROBABILITY	3	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
Theory and Logic	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
Retrieval Logic	3	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
Logic	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
LOGIC	9	1	1	1	0	0	1	0	0	1	1	0	0	1	2	0	0	0
Passage Retrieval	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
Mathematical Models(retrieval Model)	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Language Models	9	0	0	0	0	0	0	0	0	0	0	5	0	0	4	0	0	0
Information Retrieval Models	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Retrieval Model(S)	32	6	1	0	3	0	4	0	3	3	1	4	0	4	0	0	0	3
Formal Models	16	0	0	0	0	0	0	3	0	0	0	0	0	0	6	0	3	4
MODEL(S)	105	2	7	7	5	5	6	5	11	2	7	11	6	3	6	5	9	8
Inference Networks	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
INFERENCE(S)	9	0	0	0	0	2	1	0	0	0	4	1	0	0	0	0	1	0
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Classification&Categorisation(Classifiers)	28	0	0	0	0	3	3	2	4	0	0	3	1	1	0	5	3	3
Text Categorisation	7	0	3	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
Text classification	3	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
Text Categorization	14	0	0	0	0	3	0	0	0	1	0	0	4	6	0	0	0	0
CLASSIFICATION	20	0	1	1	0	0	0	2	0	0	2	0	1	4	0	1	6	2
CATEGORI(S)ZATION	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3
TEXT CATEGORIZATION	4	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	2	0
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007

Evaluation	28	0	0	1	0	0	0	0	0	0	4	3	2	0	0	3	6	9
Performance Evaluation	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
SYSTEM PERFORMANCE	3	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
PERFORMANCE (STANDARDS)	35	0	2	0	0	0	1	2	3	1	2	1	2	5	4	2	5	6
high-precision	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
PRECISION	5	1	0	0	0	0	1	0	1	0	1	0	0	1	0	0	0	0
RECALL	7	1	0	0	0	0	1	0	1	0	1	2	0	1	0	0	0	0
Test collections	3	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
IR TEST COLLECTIONS	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
Data Structures	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dimensionality reduction	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
Efficiency	25	0	0	0	0	2	3	0	3	0	2	0	3	0	3	4	5	0
Efficiency&Scalability	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
RETRIEVAL EFFECTIVENESS	7	0	1	1	0	0	0	0	0	0	0	2	0	0	2	0	1	0
EFFICIENT(EFFICIENCY)	3	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0
Signatures	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Indexing(structures)	5	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	3
INDEX TERMS	2	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Compression	7	3	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COMPRESSION	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Linguistic Analysis	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Language Processing	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Language Analysis	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
Domain specific NLP	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Natural Language Processing	21	2	0	0	3	3	0	3	0	0	4	0	0	0	3	3	0	0
Language	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
LANGUAGE(S)(LANGUAGE IR)	10	0	0	1	0	0	0	0	0	0	1	0	0	0	1	1	4	2
NATURAL-LANGUAGE	5	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Video and image	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
Speech and music	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
Spoken document retrieval	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Multimedia(IR)	18	0	0	0	0	0	2	0	0	3	3	0	2	3	0	3	0	2
Image Retrieval	12	0	0	0	0	0	0	3	1	0	0	0	0	0	5	0	0	3
IMAGE DATABASE	2	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
IMAGES(RETRIEVAL)	2	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
MULTIMEDIA	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
User Modelling	4	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
User behavior and modeling	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
BEHAVIOR(S)	50	0	0	2	2	0	2	2	3	2	4	1	6	4	5	2	7	8
Personalization	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
User Inferfaces	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
User Interaction	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
Users and the web	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3

Users: clarification, feedback, and browsing	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
User Studies	17	0	0	0	0	3	4	0	0	0	0	3	3	0	0	8	0	0
User Issues	4	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
USER(S)	65	2	5	5	3	1	2	2	6	3	3	4	3	2	8	2	7	7
Interactive Retrieval(Use of Phrases)(Interaction)	10	0	1	0	0	0	0	0	3	3	0	0	0	0	0	0	0	3
INTERACTIVE INFORMATION-RETRIEVAL	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
Interfaces(Issues)	7	3	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Search Interfaces	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
Interface Design and Display	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SEARCH INTERFACE	8	0	0	0	1	0	2	0	0	1	1	0	1	0	1	0	1	0
IR INTERFACE(S)	27	0	0	0	2	2	2	2	0	2	2	0	4	5	2	3	1	0
Human	6	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0
HUMANITIES(SCHOLARS)	12	0	1	0	1	0	1	0	0	0	1	1	1	3	1	0	2	0
Visualization	5	0	1	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0
VISUALIZATION	6	0	0	0	0	0	0	0	0	0	1	0	2	1	0	1	0	1
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Relevance Feedback	11	0	0	0	1	0	0	3	0	0	0	0	0	0	0	4	3	0
Relevance	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Feedback	5	0	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
AUTOMATIC RELEVANCE FEEDBACK	4	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0

FEEDBACK	23	2	1	0	1	0	0	1	2	4	1	3	2	2	1	2	1	0
RELEVANCE	85	2	3	1	2	3	4	6	7	7	10	6	5	4	4	3	13	5
RELEVANCE JUDGMENT(S)	4	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1
RELEVANCE FEEDBACK	13	0	0	0	1	0	0	1	1	0	0	2	0	1	2	1	3	1
RELEVANCE INFORMATION	2	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Query and Profile Modification	3	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
Query Refinement	3	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
Queries	8	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	3	0
Query Processing(strategies)	8	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	3
QUERY(S)	31	1	0	1	0	0	0	0	0	1	0	4	6	3	6	2	3	4
Query Expansion	7	0	0	3	1	0	0	3	0	0	0	0	0	0	0	0	0	0
Query Expand	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
QUERY EXPANSION	18	0	0	0	0	3	1	1	1	0	1	1	2	1	0	0	5	2
QUERY LANGUAGE	2	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
QUERY LANGUAGE APPROACH	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
QUERY TRANSLATION	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Learning to rank	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
RANKING	7	0	0	0	1	2	0	1	0	0	0	0	0	0	0	1	0	2
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
DB&IR(Integration)(Structures)	6	0	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DATABASE(S)	34	0	1	0	2	3	2	3	4	0	0	1	2	1	3	6	2	4

主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Systems	5	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	0	0
SYSTEM-DESIGN	8	1	2	1	0	0	0	0	0	1	0	0	3	0	0	0	0	0
INFORMATION-RETRIEVAL SYSTEMS	9	0	0	1	0	0	1	0	0	1	1	1	0	1	0	1	1	1
RETRIEVAL SYSTEMS	7	0	0	0	0	0	0	0	0	1	0	1	2	0	0	1	0	2
SYSTEM(S)	146	8	8	10	5	6	4	12	11	8	14	7	10	7	7	11	12	6
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Knowledge Based Information Retrieval	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KNOWLEDGE	34	1	0	3	2	1	2	1	5	1	4	1	1	2	3	3	3	1
KNOWLEDGE ORGANIZATION	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Term Relationships	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
Term Statistics	3	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
TERM(S)	12	0	0	1	2	0	0	0	0	1	2	1	0	0	3	0	1	1
TERM DEPENDENCE	4	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0	0
VOCABULARY PROBLEM	6	0	0	0	0	0	0	1	3	0	0	1	0	1	0	0	0	0
CONTROLLED VOCABULARY	6	0	0	0	1	1	0	2	0	0	0	0	1	0	0	0	1	0
VOCABULARY	5	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	2	0
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Text representation	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
REPRESENTATIONS	5	0	0	0	0	0	1	0	0	0	0	2	1	0	1	0	0	0
DOCUMENT REPRESENTATIONS	4	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2	0

SIGIR 會議文獻　期刊文獻

表 2 補足對應之主題類別

主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Question and answering	20	0	0	1	0	0	0	0	0	0	3	3	0	3	0	3	3	4
Question and answering	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Routing	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
Routing and filtering	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Collaborative Filtering	3	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
Filtering	14	0	0	0	0	0	3	0	0	0	3	0	3	2	0	3	0	0
filtering&collaborative	4	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
Filter	12	0	0	0	0	1	0	0	2	0	2	2	3	0	0	0	2	0
	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Fusion	11	0	0	0	0	3	0	1	2	0	0	0	0	0	0	0	3	2
Fusion	8	0	0	0	0	0	0	0	1	1	0	0	0	0	1	2	3	0
	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Cross language(IR)	25	0	0	0	0	0	0	3	3	0	1	3	4	2	3	3	3	0
Cross language(IR)	23	0	0	0	0	0	0	0	0	2	2	2	0	2	6	3	3	3
	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Multi-lingual IR	4	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
Asian	11	0	0	0	0	0	2	5	1	0	0	0	3	0	0	0	0	0
Processing for Japanese Text	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MIR	2	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Multilingual	14	0	0	0	0	0	1	0	0	0	5	0	0	0	4	0	4	0

	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Text Summarization	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
Summarization&Sentence Selection(multidocuments and new applications)	24	0	0	0	0	0	0	0	0	3	0	7	3	0	0	5	3	3
Summarization	4	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	1
	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Clustering	22	0	1	3	0	0	0	1	3	0	3	0	2	2	4	0	3	0
Cluster	23	0	0	0	2	0	0	0	0	0	3	3	2	2	0	3	3	5
	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
(IR) Theory	22	0	3	0	0	0	0	0	0	0	1	0	3	3	0	12	0	0
(IR) Theory	13	0	0	1	1	0	2	0	0	0	2	2	0	0	0	3	0	2
	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Distributed Systems	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Distributed(IR)	25	0	0	0	0	3	0	1	3	3	1	0	0	3	0	4	4	3
Distributed	23	0	0	1	2	0	1	1	1	0	1	3	1	0	0	4	3	5

SIGIR 會議文獻　期刊文獻

表 3 第一種主題整併之文獻年代序列

組別	總計	主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
1	43	Web(IR,search,structure,www)	43	0	0	0	0	0	0	0	0	0	4	3	4	3	3	10	10	6
	63	WORLD-WIDE-WEB(WWW,WEB)	63	0	0	0	0	0	0	0	0	0	4	5	7	12	6	8	9	12
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
2	57	Language Models	9	0	0	0	0	0	0	0	0	0	0	5	0	0	4	0	0	0
		Retrieval Model(S)	32	6	1	0	3	0	4	0	3	3	1	4	0	4	0	0	0	3
		Formal Models	16	0	0	0	0	0	0	3	0	0	0	0	0	0	6	0	3	4
	105	MODEL(S)	105	2	7	7	5	5	6	5	11	2	7	11	6	3	6	5	9	8
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
3	52	Classification&Categorisation(Classifiers)	28	0	0	0	0	3	3	2	4	0	0	3	1	1	0	5	3	3
		Text Categorisation	7	0	3	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
		Text classification	3	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
		Text Categorization	14	0	0	0	0	3	0	0	0	1	0	0	4	6	0	0	0	0
	29	CLASSIFICATION	20	0	1	1	0	0	0	2	0	0	2	0	1	4	0	1	6	2
		CATEGORI(S)ZATION	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3
		TEXT CATEGORIZATION	4	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	2	0
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
4	34	User Interaction	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
		Users and the web	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
		Users: clarification, feedback, and browsing	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
		User Studies	21	0	0	0	0	3	4	0	0	0	0	3	3	0	0	8	0	0
		User Issues	4	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
	65	USER(S)	65	2	5	5	3	1	2	2	6	3	3	4	3	2	8	2	7	7
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007

5	12	Interfaces(Issues)	7	3	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0
		Search Interfaces	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
		Interface Design and Display	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	35	SEARCH INTERFACE	8	0	0	0	1	0	2	0	0	1	1	0	1	0	1	0	1	0
		IR INTERFACE(S)	27	0	0	0	2	2	2	2	0	2	2	0	4	5	2	3	1	0
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
6	31	Evaluation	28	0	0	1	0	0	0	0	0	0	4	3	2	0	0	3	6	9
		Performance Evaluation	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	39	SYSTEM PERFORMANCE	3	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
		PERFORMANCE (STANDARDS)	36	0	2	0	0	0	1	2	3	1	2	1	2	5	4	2	5	6
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
7	28	Efficiency	25	0	0	0	0	2	3	0	3	0	2	0	3	0	3	4	5	0
		Efficiency&Scalability	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
	10	RETRIEVAL EFFECTIVENESS	7	0	1	1	0	0	0	0	0	0	0	2	0	0	2	0	1	0
		EFFICIENT(EFFICIENCY)	3	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
8	13	Queries	8	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	3	0
		Query Processing(strategies)	5	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	31	QUERY(S)	31	1	0	1	0	0	0	0	0	1	0	4	6	3	6	2	3	4
9	8	Query Expan(d)sion	8	0	0	3	1	0	1	3	0	0	0	0	0	0	0	0	0	0
	24	QUERY EXPANSION	18	0	0	0	0	3	1	1	1	0	1	1	2	1	0	0	5	2
		QUERY LANGUAGE	2	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
		QUERY LANGUAGE APPROACH	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
		QUERY TRANSLATION	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007

10	29	Language Processing	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Language Analysis	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
		Domain specific NLP	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
		Natural Language Processing	21	2	0	0	3	3	0	3	0	0	4	0	0	0	3	3	0	0
		Language	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	15	LANGUAGE(S)(LANGUAGE IR)	10	0	0	1	0	0	0	0	0	0	1	0	0	0	1	1	4	2
		NATURAL-LANGUAGE	5	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
11	18	Relevance Feedback	11	0	0	0	1	0	0	3	0	0	0	0	0	0	0	4	3	0
		Relevance	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
		Feedback	5	0	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
	131	AUTOMATIC RELEVANCE FEEDBACK	4	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0
		FEEDBACK	23	2	1	0	1	0	0	1	2	4	1	3	2	2	1	2	1	0
		RELEVANCE	85	2	3	1	2	3	4	6	7	7	10	6	5	4	4	3	13	5
		RELEVANCE JUDGMENT(S)	4	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1
		RELEVANCE FEEDBACK	13	0	0	0	1	0	0	1	1	0	0	2	0	1	2	1	3	1
		RELEVANCE INFORMATION	2	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
12	26	Routing	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
		Routing and filtering	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
		Collaborative Filtering	3	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
		Filtering	14	0	0	0	0	0	3	0	0	0	3	0	3	2	0	3	0	0
		filtering&collaborative	4	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
	12	_Filter_	12	0	0	0	0	1	0	0	2	0	2	2	3	0	0	0	2	0
13	11	Fusion	11	0	0	0	0	3	0	1	2	0	0	0	0	0	0	0	3	2

	8	_Fusion_	8	0	0	0	0	0	0	0	1	1	0	0	0	0	1	2	3	0
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
14	29	Distributed Systems	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Distributed(IR)	25	0	0	0	0	3	0	1	3	3	1	0	0	3	0	4	4	3
	25	_Distributed_	25	0	0	1	2	0	1	1	1	0	1	3	1	0	0	4	3	5
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
15	25	Cross language(IR)	25	0	0	0	0	0	0	3	3	0	1	3	4	2	3	3	3	0
	23	_Cross language(IR)_	23	0	0	0	0	0	0	0	0	2	2	2	0	2	6	3	3	3
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
16	22	Clustering	22	0	1	3	0	0	0	1	3	0	3	0	2	2	4	0	3	0
	23	_Cluster_	23	0	0	0	2	0	0	0	0	0	3	3	2	2	0	3	3	5
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
17	22	(IR) Theory	22	0	3	0	0	0	0	0	0	0	1	0	3	3	0	12	0	0
	13	_(IR) Theory_	13	0	0	1	1	0	2	0	0	0	2	2	0	0	0	3	0	2
		主題詞彙	小計	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
18	19	Multi-lingual IR	4	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
		Asian	11	0	0	0	0	0	2	5	1	0	0	0	3	0	0	0	0	0
		Processing for Japanese Text	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		MIR	2	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	14	_Multilingual_	14	0	0	0	0	0	1	0	0	0	5	0	0	0	4	0	4	0

SIGIR 會議文獻　期刊文獻

表 4 第二種主題整併年代序列

主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Web；Hypertext；Link analysis	53	2	0	0	0	0	0	0	0	0	6	3	4	3	3	10	10	9
WEB；HYPERMEDIA；HYPERTEXT	83	1	0	1	0	2	1	4	2	1	4	5	8	13	9	8	9	15
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Semantics；LSI；Probabilistic Models；(IR) Theory；Logic；Inference Networks	118	9	6	6	11	3	6	3	3	6	2	9	3	8	11	13	6	7
LATENT SEMANTIC (SEMANTICS)；PROBABILISTIC MODEL；(IR) Theory；INFERENCE(S)	163	3	9	9	7	8	10	5	11	4	16	16	9	6	11	10	17	12
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Classification&Categorisation(Classifiers) ；Clustering ； Filtering ； Fusion	111	0	4	3	6	9	6	4	9	1	6	3	13	11	11	8	9	8
CATEGORI(S)ZATION ； Cluster ； Filter ； Fusion	72	0	1	1	2	1	0	2	3	1	8	5	6	7	1	6	18	10
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Performance Evaluation；Test collections	37	0	0	1	3	0	0	0	0	0	4	3	2	0	3	3	9	9
PERFORMANCE；TEST COLLECTIONS	52	2	2	0	1	1	4	2	5	2	4	3	2	7	4	2	6	6
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Efficiency；Scalability	32	0	2	0	0	2	3	0	3	3	2	0	3	0	5	4	5	0
EFFICIENT(EFFICIENCY)	10	0	1	1	0	0	0	1	0	0	0	2	1	1	2	0	1	0
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Indexing(structures)；Term Relationships ； Statistics ；Question and answering	33	0	0	1	2	3	0	0	0	3	3	3	0	5	0	3	3	7

INDEX TERMS；TERM(S) DEPENDENCE；VOCABULARY；Question and answering	46	0	1	1	3	2	3	4	4	1	3	5	3	2	4	1	7	2
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Natural Language Processing；Cross language；MIR	75	2	2	4	3	3	6	11	4	3	5	3	7	2	6	6	3	5
LANGUAGE；Cross language；Multilingual	52	0	0	1	0	0	1	0	1	2	9	2	0	2	12	5	12	5
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
User behavior；Human；Interface Design and Display ；Visualization	73	3	5	3	6	5	7	4	3	6	0	3	3	6	0	8	6	6
USER(S)；BEHAVIOR(S)；HUMAN；SEARCH INTERFACE；VISUALIZATION	170	2	6	7	9	3	9	6	9	8	12	6	17	16	17	9	18	16
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Relevance Feedback；Query；Query Expansion	48	0	2	5	2	3	4	6	3	0	2	0	5	0	0	4	6	3
RELEVANCE FEEDBACK；QUERY EXPANSION	186	6	5	3	4	7	6	10	11	12	14	17	15	12	14	9	27	14
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Distributed(IR)(Systems)	29	4	0	0	0	3	0	1	3	3	1	0	0	3	0	4	4	3
Distributed	25	0	0	1	2	0	1	1	1	0	1	3	1	0	0	4	3	5
主題詞彙	總次數	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
DB&IR(Integration)(Structures)；Systems；Knowledge Based Information Retrieval	14	0	4	2	0	0	0	0	0	0	1	4	0	0	0	0	0	0
DATABASE(S)；SYSTEM(S)；KNOWLEDGE	232	9	9	14	9	10	9	17	20	11	19	11	15	11	13	22	18	15

SIGIR 會議文獻　期刊文獻

附錄 二　主題萃取系統歸類結果

1(37):

- 17354 : 300筆 : 0.0011 (**retrieval:** 73.2, **document:** 48.1, **information retrieval:** 37.8, **rank:** 36.0, **algorithm:** 25.7)
  - 14656 : 104筆 : 0.0121 (**information retrieval:** 11.2, **research:** 10.7, **term:** 8.2, **structure:** 7.8, **document:** 6.0)
    - 9750 : 43筆 : 0.0382 (**citation:** 19.0, **map:** 15.6, **science:** 7.8, **identify:** 4.0, **relevance:** 3.2)
      - 1531 : 19筆 : 0.1312 (**identify:** 2.7, **citation:** 2.6, **relevance:** 2.6, **document:** 2.1, **judgment:** 1.5)
        - 403 : 8筆 : 0.1929 (**sensitivity:** 8.5, **dimension:** 5.3, **lpi:** 4.9, **manifestation:** 4.2, **orthogonal:** 3.0)
          - 138 : 134 : 5筆,0.10(dimension:14.14)
          - 140 : 173 : 3筆,0.26(sensitivity:9.52, precision:0.80, retriev:0.66, search:0.35)
        - 25 : 55 : 11筆,0.20(e. g.:31.00, information:0.27)
      - (1) 8247 : 24筆 : 0.0475 (**map:** 14.3, **citation:** 11.6, **information science:** 9.3, **research:** 5.1, **journal:** 4.7)
        - 172 : 10筆 : 0.2345 (**map:** 11.8, **co-citation:** 6.3, **aca:** 4.9, **author:** 4.5, **co-citation analysi:** 4.2)
          - 150 : 150 : 5筆,0.09(map:8.98, information:0.13, retrieval:0.08)
          - 28 : 143 : 5筆,0.20(author:12.83, co-citation:6.97, map:2.30, information retrieval:0.50)
        - 4 : 14筆 : 0.5116 (**cit:** 17.6, **information science:** 13.4, **citation:** 9.3, **journal:** 6.9, **field:** 6.3)
          - 37 : 156 : 4筆,0.19(cit:22.03, reference:3.29, scientific:2.00, citation:1.48, relationship:0.77, document:0.72 ...
          - 74 : 55 : 10筆,0.04(science:17.54, information:1.94)
    - 13999 : 61筆 : 0.0051 (**term:** 11.3, **retrieval:** 9.2, **document:** 6.5, **information retrieval:** 5.5, **system:** 5.5)
      - 1198 : 17筆 : 0.1427 (**identification:** 7.8, **community:** 6.7, **structure:** 5.4, **domain:** 5.0, **scientific:** 3.8)
        - 253 : 13筆 : 0.2166 (**sublanguage:** 14.1, **abstract:** 7.0, **worm:** 6.3, **thesauru:** 5.6, **scientific:** 4.7)
          - 158 : 109 : 6筆,0.09(abstract:15.55, sublanguage:12.23, structure:1.20)
          - 70 : 82 : 7筆,0.15(thesauru:18.01, term:1.43, approach:1.37)
        - 112 : 171 : 4筆,0.12(identification:7.21, identify:0.53, information retrieval:0.17)
      - 9375 : 44筆 : 0.0403 (**visualization:** 29.4, **document:** 8.4, **retrieval:** 6.3, **search:** 5.4, **interface:** 4.9)
        - (2) 339 : 30筆 : 0.2025 (**visualization:** 42.4, **display:** 10.2, **visualize:** 4.9, **document:** 4.5, **interface:** 4.4)
          - 95 : 21筆 : 0.2740 (**visual:** 16.9, **display:** 12.8, **vibe:** 12.0, **visualization:** 10.0, **subtopic:** 7.8)
            - 151 : 53 : 11筆,0.01(interface:3.97, information retrieval:3.36, system:2.40, document:0.98)
            - 187 : 62 : 10筆,0.00(visual:9.75, retrieval:1.29, document:1.08, information:0.70)
          - 117 : 74 : 9筆,0.12(visualization:27.42, space:2.32, information:1.39)
        - 4428 : 14筆 : 0.0796 (**ontology:** 6.3, **search:** 4.8, **searche:** 2.4, **recall:** 1.8, **document:** 1.4)
      - (3) 638 : 21筆 : 0.1722 (**graph:** 18.3, **page:** 14.0, **link:** 13.0, **topic:** 6.7, **algorithm:** 6.6)
        - 154 : 15筆 : 0.2406 (**link analysi:** 9.9, **page:** 9.7, **topic:** 9.6, **topic distillation:** 7.8, **graph:** 7.4)
          - 31 : 107 : 6筆,0.07(link analysi:13.05, page:5.62, pagerank:4.23, algorithm:3.44, graph:2.83)
          - 59 : 66 : 9筆,0.05(topic:16.37)
        - 174 : 113 : 6筆,0.01(graph:2.35, search:0.51, information:0.26, document:0.17)
      - 3199 : 29筆 : 0.0958 (**model:** 8.2, **document:** 4.6, **measure:** 4.2, **family:** 4.0, **system:** 3.8)
        - 400 : 21筆 : 0.1938 (**performance measure:** 7.1, **model:** 5.7, **hypergeometric:** 5.6, **song:** 4.9, **family:** 4.9)
          - 15 : 69 : 9筆,0.00(measure:4.08, document:0.26, retrieval:0.84, information:0.45)
          - 5 : 40 : 12筆,0.00(model:6.36, document:1.77, retrieval:0.84, information:0.33)
        - 7 : 85 : 8筆,0.07(message:23.99, information:1.15, system:0.79)
    - 12704 : 146筆 : 0.0220 (**collection:** 33.9, **retrieval:** 31.9, **rank:** 31.2, **relevance:** 30.1, **test:** 22.9)
      - 9819 : 34筆 : 0.0378 (**genetic:** 34.5, **function:** 8.4, **rank:** 6.2, **order:** 4.4, **collection:** 4.1)
        - 397 : 24筆 : 0.1940 (**genetic:** 44.9, **genetic algorithm:** 22.8, **function:** 12.3, **program:** 5.5, **collection:** 4.1)
          - (4) 81 : 19筆 : 0.2829 (**genetic:** 30.2, **genetic algorithm:** 12.4, **fitne:** 9.9, **function:** 9.1, **rank function:** 7.4)
            - 27 : 31 : 14筆,0.00(genetic:30.37, algorithm:11.66, relevance:2.77, retrieval:0.65, information:0.42)
            - 52 : 139 : 5筆,0.05(rank function:10.05)
          - 114 : 147 : 5筆,0.02(query evaluation:5.70, text:0.94, technique:0.71, documen:0.43)
        - (5) 37 : 10筆 : 0.3359 (**markov:** 11.4, **walk:** 9.2, **random:** 8.9, **hmm:** 7.1, **website:** 6.9)
          - 132 : 100 : 6筆,0.10(random:16.25, walk:9.16, field:2.27, model:0.98)
          - 59 : 154 : 4筆,0.15(markov:6.62, hmm:3.99, hidden:2.56, chain:2.23, model:0.95, rank:0.62, system:0.10)

      - 11508 : 112筆 : 0.0283 (**test:** 30.1, **retrieval:** 21.7, **relevance:** 21.2, **evaluation:** 20.6, **collection:** 20.3)
        - 7846 : 79筆 : 0.0101 (**relevance:** 48.2, **test collection:** 29.6, **judgment:** 28.9, **document:** 20.5, **evaluation:** 15.0)
          - 248 : 67筆 : 0.2175 (**relevance:** 37.8, **judgment:** 37.0, **test collection:** 36.7, **highly relevant:** 22.5, **relevanc judgment:** 20.8)
            - 83 : 58筆 : 0.2804 (**relevance:** 32.1, **judgment:** 30.7, **highly relevant:** 26.1, **relevance judgment:** 24. **relevant document:** 19.5)
              - (6) 6 : 47筆 : 0.4966 (**highly relevant:** 32.0, **relevance judgment:** 30.3, **relevant document:** 24.6 **highly relevant document:** 23.0, **criteria:** 20.2)
                - 143 : 7 : 29筆,0.01(relevance:57.58, document:4.15)
                - 6 : 18 : 18筆,0.00(relevan:33.87, relevant document:17.62, relevance:9.05, system:1.45, retrieval:0.68, information ...
              - 18 : 50 : 11筆,0.00(precision:7.60, average:7.21, information:1.10, retrieval:0.07)
            - 80 : 72 : 9筆,0.04(test collection:16.26, retrieval:1.02)
          - 1133 : 12筆 : 0.1453 (**rank list:** 7.8, **topology:** 5.4, **topological:** 4.2, **bootstrap:** 4.2, **position:** 3.5)
            - 0 : 93 : 7筆,0.00(list:5.50, rank:2.65, collection:1.74, system:0.55, information:0.14, retrieval:0.10)
            - 172 : 129 : 5筆,0.08(metric:8.17)
        - 3576 : 33筆 : 0.0900 (**stem:** 42.8, **arabic:** 13.2, **morphological:** 9.3, **stemmer:** 7.8, **word:** 7.2)
          - (7) 835 : 28筆 : 0.1603 (**stem:** 49.8, **arabic:** 15.4, **morphological:** 10.8, **word:** 9.3, **stemmer:** 9.1)
            - 23 : 19筆 : 0.3585 (**stem:** 36.0, **morphological:** 13.4, **stemmer:** 11.3, **spoken:** 10.0, **finnish:** 9.9)
              - 14 : 33 : 15筆,0.00(word:6.66, language:4.44, retrieval:1.14)
              - 149 : 114 : 6筆,0.09(stem:21.28, stem algorithm:7.77, stemmer:4.98)
            - 9 : 67 : 9筆,0.00(arabic:14.94, retrieval:2.08, query:1.31)
          - 83 : 125 : 5筆,0.04(h:t:6.01, better:2.46, retrieval:0.09)

2(28):

- 17172 : 405筆 : 0.3017 (**model**:189.3, **query**:160.7, **retrieval**:119.9, **document**: 98.2, **term**: 90.4)
  - 2688 : 8筆 : 0.1048 (**preference**: 5.2, **user preference**: 4.9, **clas**: 4.2, **negative**: 4.2, **dbm**: 3.5)
    - 11 : 131 : 5筆,0.07(preference:13.46, model:0.67, information:0.11, retrieval:0.07)
    - 123 : 182 : 3筆,0.1 : (multiuser:3.69, information-retrieval:0.50, system:0.20)
  - 16166 : 397筆 : 0.3052 (**model**:194.9, **query**:147.5, **document**:140.9, **retrieval**:112.0, **term**: 92.1)
    - 13744 : 62筆 : 0.0164 (**inference**: 10.8, **information**: 10.5, **model**: 9.2, **query**: 8.3, **information retrieval**: 8.0)
      - 5857 : 41筆 : 0.0658 (**agent**: 16.1, **network**: 14.5, **concept**: 12.2, **conceptual**: 10.5, **inference**: 8.7)
        - 214 : 28筆 : 0.2244 (**network**: 24.1, **agent**: 23.2, **plausible**: 7.8, **neural network**: 7.4, **kohonen**: 6.3)
          - 105 : 14 : 20筆,0.03(network:53.89, information:0.05)
          - 72 : 75 : 8筆,0.00(agent:12.28, information:2.16, information retrieval:0.80, system:0.61)
        - 775 : 13筆 : 0.1634 (**concept**: 10.5, **conceptual**: 8.8, **alteration**: 7.1, **data model**: 6.3, **article title**: 4.9)
          - 129 : 115 : 6筆,0.02(data model:6.36, integration:3.03, level:2.00, database:1.30, concept:1.20, experiment:0.75)
          - 382 : 99 : 7筆,0.08(concept:11.77, article:2.21, information:1.17)
      - (8) 3411 : 21筆 : 0.0924 (**logic**: 31.1, **information retrieval**: 2.5, **represent**: 1.6, **document**: 1.5, **query**: 1.4)
        - 190 : 16筆 : 0.2304 (**logic**: 23.7, **logical**: 5.8, **terminological logic**: 3.5, **nonmonotonic**: 2.8, **nonmonotonic logi** 2.8)
          - 51 : 84 : 3筆,0.17(logic:27.35, logical:10.55, information:1.71, retrieval:1.32, information retrieval:1.04)
          - 77 : 80 : 8筆,0.04(information retrieval:5.03)
        - 68 : 132 : 5筆,0.15(belief:16.50, belief revision:9.20, model:1.64, information:0.21, retrieval:0.08)

- 13897 : 335筆 : 0.0156 (**model**:153.9, **document**:100.3, **query**:100.3, **retrieval**: 82.2, **term**: 77.1)
  - 10300 : 324筆 : 0.0352 (**model**:140.0, **document**: 89.5, **query**: 86.1, **retrieval**: 74.7, **language**: 71.7)
    - 6372 : 311筆 : 0.0611 (**model**:151.2, **query**: 95.5, **term**: 86.0, **document**: 77.8, **language**: 71.2)
      - (9) 25 : 24筆 : 0.3567 (**space**: 13.4, **vector space model**: 13.0, **document and query**: 9.9, **concept space**: 6.3, **metasearch**: 5.6)
        - 144 : 24 : 15筆,0.01(space:29.63, vector:9.47, model:2.41, information:0.04)
        - 159 : 82 : 8筆,0.01(document and query:7.77, model:3.96, probabilistic:2.30)
      - 4423 : 287筆 : 0.0796 (**model**:122.9, **term**: 88.5, **query**: 80.9, **cluster**: 69.6, **document**: 63.6)
        - 2245 : 235筆 : 0.1132 (**model**: 87.5, **query**: 78.5, **language**: 71.9, **expansion**: 67.5, **document**: 66.5)
          - 342 : 56筆 : 0.2024 (**retrieval model**: 35.1, **probabilistic**: 21.1, **probabilistic retrieval**: 13.8, **dependence**: 13.0, **term**: 11.7)
            - (10) 57 : 43筆 : 0.3151 (**retrieval model**: 25.7, **probabilistic retrieval**: 17.0, **operator**: 13.1, **mental model**: 9.9, **probabilistic model**: 7.9)
              - 188 : 22 : 17筆,0.00(probabilistic:30.34, model:8.59, retrieval:2.80, information:0.16)
              - 189 : 9 : 26筆,0.00(model:31.48, retrieval:10.03, information:2.98)
            - 153 : 34 : 13筆,0.01(term:11.78, information:0.71, retrieval:0.57)
          - 462 : 179筆 : 0.1874 (**cluster**: 59.0, **query**: 54.2, **query expansion**: 52.9, **language**: 52.0, **question**: 42.9)
            - 177 : 121筆 : 0.2337 (**cluster**: 54.3, **query expansion**: 41.2, **measure**: 32.1, **document**: 28.3, **document cluster**: 25.0)
              - (11) 20 : 65筆 : 0.3699 (**query expansion**: 61.3, **measure**: 17.1, **interactive query expansion**: 15.5, **term**: 15.3, **mismatch**: 10.6)
                - 165 : 4 : 45筆,0.01(expansion:122.32, query expansion:84.92)
                - 191 : 15 : 20筆,0.00(measure:30.87, similarity:10.80, term:5.97, information:3.20), document:2.88, retrieval:1.24 ...
              - (12) 56 : 56筆 : 0.3154 (**cluster**: 53.9, **document cluster**: 35.9, **cluster algorithm**: 10.6, **document collection**: 7.4, **scatter**: 7.1)
                - 181 : 10 : 25筆,0.00(document:20.93, retrieval:1.03)
                - 91 : 6 : 31筆,0.03(cluster:138.44, document:13.75)
            - 67 : 58筆 : 0.3025 (**language model**: 44.0, **question answer**: 19.8, **definitional**: 12.7, **natural language**: 11.4, **smooth**: 9.2)
              - (13)(14) 146 : 5 : 37筆,0.01(language:78.65, model:20.32, retrieval:2.89)
              - 99 : 12 : 21筆,0.12(answer:90.31, question:65.75, system:1.91)
        - 2991 : 52筆 : 0.0992 (**weight**: 29.0, **lsi**: 26.4, **latent**: 22.3, **latent semantic index**: 21.1, **weight scheme**: 16.5)
          - (15) 8 : 27筆 : 0.4729 (**lsi**: 38.0, **latent semantic index**: 30.1, **svd**: 9.9, **matrix**: 8.4, **term frequency**: 6.3)
            - 135 : 48 : 11筆,0.03(term:7.08, frequency:4.78, model:3.32, approach:1.84, retrieval:0.93)
            - 73 : 23 : 16筆,0.04(latent semantic index:38.45, lsi:26.96, document:1.51)
          - (16) 273 : 25筆 : 0.2128 (**collaborative filter**: 28.2, **weight scheme**: 23.6, **term weight**: 9.2, **recommendation**: 9.2, **prediction**: 6.8)
            - 175 : 29 : 14筆,0.08(collaborative:51.79, collaborative filter:33.77, information:0.03)
            - 50 : 47 : 21筆,0.06(weight:29.32, term:5.80, query:1.34, retrieval:0.60, information:0.15)
      - 610 : 13筆 : 0.1736 (**information system**: 9.8, **prediction**: 7.2, **prr**: 5.6, **subject literature**: 4.2, **performance prediction**: 4.2)
        - 167 : 87 : 8筆,0.01(prediction:7.29, performance:1.81, query:1.46, system:0.85, retrieval:0.32)
        - 55 : 138 : 5筆,0.05(interpretation:3.62, theory:2.66, knowledge:2.25, development:2.25, science:1.92, information ...
    - 250 : 11筆 : 0.2172 (**object**: 19.0, **proximity**: 9.2, **condition**: 9.0, **textual object**: 5.6, **data object**: 4.5)
      - 190 : 83 : 8筆,0.08(object:23.47, data:1.80, text:0.80, retrieval:0.29)
      - 62 : 180 : 3筆,0.16(proximity:14.68, number:0.82, algorithm:0.63, query:0.37)

3(25):

- 16780 : 424筆 : 0.0029 (**information**:112.8, **query**: 63.7, **information retrieval**: 61.9, **document**: 58.5, **model**: 45.5)
  - 13843 : 355筆 : 0.0158 (**information**: 82.9, **search**: 79.7, **document**: 59.3, **query**: 46.0, **information retrieval**: 45.2)
    - 10213 : 116筆 : 0.0356 (**text**: 24.3, **approach**: 19.1. **word**: 17.0, **information**: 14.9. **feature**: 14.1)
      - 297 : 29筆 : 0.2093 (**digital library**: 32.8, **resource**: 9.9, **project**: 5.2, **information**: 3.8, **database**: 3.7)
        - 76 : 19筆 : 0.2878 (**digital library**: 24.6, **music information retrieval**: 15.5, **graphic**: 7.8, **musical**: 7.1, 4.2)
          - 143 : 70 : 9筆,0.10(digital:26.86, digital library:20.97, information:1.23)
          - (17) 24 : 2[illegible] : 10筆,0.07(music:40.53, music information retrieval:22.35)
          - 16 : 57 : 10筆,0.00(resource:8.62, algorithm:2.64, information:1.00, system:0.78)
        - 5027 : 87筆 : 0.0734 (**categorization**: 33.5, **text**: 31.2, **text categorization**: 29.3, **feature**: 19.6, **approach**: 17.[illegible]
          - 2488 : 64筆 : 0.1084 (**categorization**: 40.3, **text categorization**: 35.0, **approach**: 12.9. **feature**: 11.9, **alg**
            - 125 : 30筆 : 0.2597 (**web page**: 35.4, **extract**: 13.7, **place**: 7.0, **title**: 6.6, **html**: 6.4)
              - 155 : 32 : 13筆,0.01(extract:17.81, extraction:14.65, documen:2.11, retrieval:0.09)
              - 96 : 21 : 17筆,0.05(page:46.41, web page:33.05)
            - 180 : 34筆 : 0.2328 (**categorization**: 33.8, **text categorization**: 30.8, **classifier**: 15.1, **approach**: 5
              - (18) 17 : 25筆 : 0.3329 (**categorization**: 24.1, **text categorization**: 20.7, **classifier**: 19.1, **linear**: **classifier**: 6.3)
                - 154 : 39 : 12筆,0.01(text:12.16, classifier:9.56, approach:1.52)
                - 60 : 35 : 13筆,0.05(categorization:31.07, text categorization:18.61, category:10.47)
              - 88 : 73 : 9筆,0.03(selection:24.42, information:0.13)
          - 294 : 23筆 : 0.2096 (**window**: 12.0, **sense**: 11.9. **word**: 11.2. **prun**: 7.6, **word sense**: 7.4)
            - 162 : 54 : 11筆,0.01(term:5.52, query:4.79)
            - 82 : 45 : 12筆,0.04(word:20.47, retrieval:1.89, term:1.38, information:0.35)

  - 13107 : 69筆 : 0.0197 (**machine**: 15.7, **dictionary**: 15.3, **language**: 13.0, **bilingual**: 11.3, **test**: 8.9)
    - 185 : 10筆 : 0.2315 (**opinion**: 24.1, **comparative**: 13.8, **comparative sentence**: 4.9, **sentiment**: 3.5, **subjective**: 3.5)
      - 161 : 108 : 6筆,0.01(comparative:6.51, part:3.38, test:2.37, experimen:1.09, data:0.62, retrieval:0.30, system:0 ...
      - 43 : 163 : 4筆,0.18(opinion:21.00, sentence:3.07, sentiment:2.88, subjective:2.49, product:1.66, exper:0.77, combination ...
    - 12382 : 59筆 : 0.0237 (**translation**: 68.3, **cross-language**: 32.5, **cross-language information retrieval**: 29.9, **clir**: 19.3, **machine**: 19.3)
      - (19) 9 : 42筆 : 0.4719 (**translation**: 91.7, **cross-language**: 43.2, **cross-language information retrieval**: 39.3, **query translation**: 26.2, **clir**: 25.6)
        - 1 : 37筆 : 0.6499 (**translation**: 69.9, **query translation**: 32.4, **cross language**: 31.0, **cross language information retrieval**: 28.6, **translate**: 15.1)
          - 120 : 17 : 19筆,0.02(cross-language:26.62, translation:35.69, cross-language information retrieval:35.59, query:10 ...
          - 46 : 20 : 18筆,0.05(translation:54.14, query translation:15.79, document:2.24, retrieval:2.10, information:0.32 ...
        - 109 : 143 : 5筆,0.12(clir:8.47, cross-language:1.75, translation:1.50, corpu:1.35, performance:1.02, information ...
      - 298 : 17筆 : 0.2092 (**transliteration**: 19.0, **japanese**: 13.4, **automatic index**: 10.4, **consistency**: 8.5, **transliteration model**: 5.6)
        - 139 : 38 : 13筆,0.02(index:25.37)
        - 86 : 155 : 4筆,0.13(transliteration:19.08, dictionary:3.14, proper:2.63, alphabet:2.31, japanese:2.11, english:1 ...

4(13):

- 16723 : 104筆 : 0.0030 (**classification**: 64.9, **algorithm**: 9.3, **performance**: 6.6, **data**: 5.7, **retrieval**: 5.1)
  - 11854 : 48筆 : 0.0265 (**image**: 95.8, **query**: 7.6, **retrieval**: 7.2, **content**: 6.5, **automatic**: 5.9)
    - 652 : 13筆 : 0.1715 (**busine**: 14.1. **boe**: 6.3, **busine intelligence**: 6.3, **evidence**: 5.6, **evidence combination**: 4.9)
      - 103 : 104 : 6筆,0.12(busine:20.56, intelligence:8.01, information:0.70, experiment:0.67)
      - 184 : 96 : 7筆,0.08(evidence:17.90, combination:6.82, different:2.07, retrieval:1.13)
    - 2259 : 35筆 : 0.1129 (**image**: 91.9, **image retrieval**: 28.4. **visual**: 10.1, **content**: 9.8, **content-based**: 8.0)
      - (20) 15 : 3[illegible] 筆 : 0.3924 (**image**: 72.8, **image retrieval**: 35.0, **annotation**: 20.4, **visual**: 12.5, **image annotation**: 12.0)
        - 131 : 52 : 11筆,0.11(annotation:50.48, image:26.93, automatic:5.77, model:2.10, retrieval:0.10)
        - 19 : 13 : 20筆,0.07(image:94.96, image retrieval:24.18)
      - 110 : 164 : 4筆,0.11(archive:9.52, speech:2.30, request:2.20, material:1.66)
  - 11495 : 56筆 : 0.0283 (**classification**: 54.1, **classifier**: 17.8, **algorithm**: 9.5, **learn**: 9.5, **category**: 6.9)
    - 8180 : 52筆 : 0.0479 (**classification**: 49.2, **classifier**: 13.6, **algorithm**: 11.5, **learn**: 11.2, **data**: 8.2)
      - 843 : 2[illegible] 筆 : 0.1599 (**category**: 5.7, **domain**: 6.2, **classification**: 5.6, **distribution**: 4.4, **data**: 2.6)
        - 179 : 16筆 : 0.2335 (**domain**: 8.1, **chemoinformatic**: 7.1, **user query**: 5.6, **molecule**: 4.9, **belief function**: 4.2)
          - 145 : 46 : 11筆,0.01(query:4.60)
          - 72 : 137 : 5筆,0.04(domain:7.09, knowledge:1.21, information:0.11)
        - 40 : 123 : 5筆,0.18(maximum:15.64, entropy:7.10, data:0.89, document:0.53, retrieval:0.11)
      - 2636 : 31筆 : 0.1056 (**classification**: 30.3, **learn**: 18.7, **algorithm**: 9.7, **machine**: 9.5, **svm**: 9.4)
        - (21) 92 : 19筆 : 0.2762 (**classification**: 18.8, **text classification**: 13.0, **machine**: 7.3, **classification algorithm**: 7.1, **support vector machine**: 7.1)
          - 121 : 141 : 5筆,0.11(machine:7.46, support vector machine:5.03, learn:2.58)
          - 52 : 28 : 14筆,0.05(classification:36.73)
        - 59 : 12筆 : 0.3101 (**learn algorithm**: 10.6, **instruction**: 6.3, **linear**: 5.2, **scaffold**: 4.9, **learn environment**: 4.9)
          - 55 : 118 : 5筆,0.15(linea:11.39, classifier:1.80, algorithm:1.67, model:0.72)
          - 75 : 94 : 7筆,0.04(learn:13.51, information:0.12, retrieval:0.10)
    - 87 : 153 : 4筆,0.03(regularize:2.99, score:2.50, grad:2.44, regression:1.32, text:0.78, categorization:0.69, document ...

5(12):

- 14122 : 124筆 : 0.0146 (**relevance:** 38.5, **feedback:** 37.2, **performance:** 29.1, **trec:** 25.3, **retrieval:** 24.8)
  - 325 : 6筆 : 0.0047 (**interpret:** 8.1, **lesson:** 3.5, **usability:** 3.4, **analytic:** 3.0, **ppp:** 2.8)
    - 100 : 192 : 2筆,0.32(usability:3.75, cognitive abilities and information system usability:1.41, user characteristic ...
    - 43 : 157 : 4筆,0.06(analytic:2.49, interpret:2.03, performance:1.93, information retrieval:0.97, measure:0.64, filter ...
  - 12281 : 118筆 : 0.0242 (**relevance:** 39.7, **feedback:** 34.3, **document:** 31.9, **performance:** 27.1, **trec:** 22.6)
    - 1123 : 39筆 : 0.1456 (**retrieval performance:** 22.2, **document representation:** 15.5, **query:** 9.2, **exhaustivity:** 6.8, **measure:** 5.
      - 528 : 17筆 : 0.1308 (**document representation:** 9.4, **extend boolean:** 9.2, **optimal:** 7.0, **extend boolean model:** 6.5, **age** 6.3)
        - 4 : 97 : 7筆,0.08(representation:20.87, document representation:15.41, retrieval:0.46)
        - (22) 47 : 59 : 10筆,0.06(boolean:21.32, extend:5.64, query:3.15)
      - (23) 30 : 22筆 : 0.3439 (**recall and precision:** 13.4, **retrieval performance:** 12.6, **variation:** 5.4, **measure:** 4.9, **stair:** 4.9)
        - 142 : 44 : 2筆,0.01(retrieval performance:13.92, query:1.53)
        - 78 : 50 : 10筆,0.04(recall:25.45, recall and precision:15.63, retrieval:0.86, information:0.23)
      - 5773 : 79筆 : 0.0966 (**filter:** 43.1, **feedback:** 28.0, **relevance feedback:** 26.7, **trec:** 25.5, **data:** 17.2)
        - 1751 : 63筆 : 0.1251 (**filter:** 55.3, **feedback:** 25.6, **relevance feedback:** 23.8, **data fusion:** 22.8, **profile:** 17.2)
          - (24) 62 : 39筆 : 0.3058 (**data fusion:** 32.6, **relevance:** 10.6, **relevance feedback:** 9.5, **feedback:** 9.5, **hoc:** 7.5)
            - 115 : 8 : 2筆,0.02(feedback:66.06, relevance feedback:48.72, document:23.01, retrieval:0.27)
            - 143 : 51 : 11筆,0.02(data fusion:32.20, system:1.95, retrieval:0.88)
          - (25) 94 : 24筆 : 0.2746 (**filter:** 35.1, **information filter:** 12.7, **adaptive:** 10.2, **profile:** 8.3, **rocchio:** 7.0)
            - 64 : 19 : 18筆,0.05(filter:52.76, information:0.62)
            - 75 : 112 : 6筆,0.14(adaptive:10.23, adaptive filter:7.07, performance:0.48, system:0.33)
        - 186 : 16筆 : 0.2314 (**okapi:** 9.8, **trec:** 9.5, **rank svm:** 8.4, **scal:** 5.7, **bmr:** 4.9)
          - 134 : 64 : 9筆,0.10(trec:18.23, interactive:3.62, retrieval:0.07)
          - 90 : 88 : 7筆,0.03(rank:12.50, document:2.24, term:1.35, information:0.05)

7(9):

- 14490 : 55筆 : 0.0129 (**file:** 33.9, **compression:** 21.3, **fast:** 14.4, **time:** 12.8, **text:** 10.3)
  - 227 : 13筆 : 0.2216 (**compression:** 8.5, **pattern:** 3.2, **algorithm:** 1.3, **text:** 1.1, **system:** 0.8)
    - 50 : 8筆 : 0.3225 (**bitmap:** 9.9, **data stream:** 6.3, **compression:** 5.6, **decompression:** 4.9, **stream data:** 4.2)
      - 84 : 169 : 4筆,0.33(bitmap:15.00, compression:2.99, technique:0.32)
      - 93 : 151 : 4筆,0.03(data stream:5.65, decompression:3.46, pattern:0.36, error:0.86, algorithm:0.64, text:0.51, search ...
    - 39 : 149 : 5筆,0.06(code:3.66)
  - 3461 : 42筆 : 0.0916 (**file:** 42.9, **invert:** 27.6, **fast:** 12.1, **time:** 12.0, **update:** 10.7)
    - 1582 : 30筆 : 0.1298 (**invert file:** 20.5, **fast:** 15.1, **time:** 11.5, **identifier:** 9.1, **compression:** 7.8)
      - (26) 3 : 14筆 : 0.5376 (**invert:** 29.7, **invert file:** 29.3, **identifier:** 13.0, **compression:** 11.6, **indexe:** 8.5)
        - 219 : 136 : 5筆,0.11(compression:9.96, index:7.33, invert:4.49, time:1.32)
        - 33 : 71 : 9筆,0.37(invert:28.14, invert file:20.40, index:1.70, document:0.96)
      - (27) 148 : 16筆 : 0.2462 (**tree:** 12.4, **trie:** 8.5, **key:** 6.7, **historical:** 6.3, **space efficiency:** 6.0)
        - 21 : 43 : 12筆,0.00(text:5.05, retrieval:1.34, search:1.80)
        - 45 : 170 : 4筆,0.17(fast:4.93, time:1.97, space:1.15, structure:0.59, query:0.57, rank:0.54, search:0.35)
    - (28) 55 : 12筆 : 0.3156 (**signature:** 28.9, **signature file:** 16.2, **partition:** 11.4, **directory:** 9.9, **directory structure:** 7.8)
      - 156 : 117 : 6筆,0.01(directory:8.30, directory structure:4.95, communication:2.93, retrieval:0.22, system:0.20)
      - 54 : 116 : 6筆,0.17(signature:29.84, signature file:12.01, partition:5.07, query:0.94, analysis:0.77, performance ...

8(8):

- 16530 : 52筆 : 0.0037 (**catalog:** 27.9, **search:** 5.5, **library:** 5.0, **design:** 4.2, **framework:** 3.5)
  - 9717 : 40筆 : 0.0384 (**catalog:** 32.7, **online:** 15.8, **session:** 9.1, **search:** 7.4, **library:** 6.3)
    - 2621 : 22筆 : 0.1059 (**catalog:** 23.3, **online:** 14.4, **catalogue:** 8.6, **design:** 7.5, **system:** 4.0)
      - (29) 70 : 17筆 : 0.3010 (**online:** 18.5, **catalog:** 17.6, **online catalog:** 15.5, **opac:** 10.9, **catalogue:** 10.6)
        - 106 : 106 : 6筆,0.12(opac:9.91, catalog:3.33, subject:2.07, library:1.72, project:1.36, online:1.33, search:1.29 ...
        - 92 : 49 : 11筆,0.03(online:25.06, search:1.92, system:1.72)
      - 126 : 121 : 5筆,0.02(requirement:3.62, design:2.86, system:0.48)
    - 1665 : 18筆 : 0.1275 (**session:** 13.3, **site:** 12.6, **navigational:** 6.5, **search:** 4.1, **number:** 3.8)
      - 875 : 13筆 : 0.1578 (**web site:** 13.4, **session:** 9.3, **thesaurus-based:** 4.9, **decision:** 3.9, **homepage:** 3.8)
        - 125 : 127 : 5筆,0.02(predict:3.18, relevance:1.16, search:0.73)
        - 89 : 35 : 8筆,0.03(site:16.39, web site:12.98, page:5.02, information:0.28)
      - 166 : 133 : 5筆,0.06(behaviour:10.77, journal:6.84, view:3.43, analysi:1.36, number:1.25)
  - 621 : 12筆 : 0.1729 (**sampl:** 17.9, **rough:** 7.8, **fuzzy:** 5.6, **citation profile:** 4.2, **approximation:** 4.2)
    - 186 : 65 : 9筆,0.00(fuzzy:10.21, set:4.64, retrieval:1.66, system:1.47, information:0.14)
    - 95 : 177 : 3筆,0.32(sampl:11.77, distribut:0.21)

9(7):

- 6749 : 59筆 : 0.0579 (**student:** 38.4, **information:** 12.1, **research:** 9.8, **retrieval:** 7.0, **task:** 5.1)
  - 5560 : 39筆 : 0.0686 (**student:** 26.6, **search:** 19.9, **tactic:** 19.3, **search tactic:** 14.6, **experience:** 11.4)
    - 176 : 30筆 : 0.2341 (**tactic:** 25.0, **search tactic:** 19.0, **experience:** 15.1, **search experience:** 13.0, **student:** 10.2)
    - (30) 35 : 17筆 : 0.3375 (**tactic:** 16.3, **search tactic:** 13.0, **animation:** 8.5, **experience:** 5.4, **crosslingual:** 4.9)
      - 108 : 53 : 10筆,0.03(knowledge:12.09)
      - 98 : 91 : 7筆,0.03(task:8.60, search:1.94, information:1.06)
    - 22 : 37 : 13筆,0.00(search:10.44, information:1.17, retrieval:0.36)
    - 980 : 9筆 : 0.1519 (**student:** 8.3, **print:** 4.5, **internet:** 3.5, **children:** 2.8, **permuterm:** 2.8)
      - 111 : 103 : 6筆,0.03(internet:7.07, problem:3.16, information:0.47)
      - 113 : 176 : 3筆,0.12(student:5.27, print:4.61, information:0.14, retrieval:0.06)
  - 91 : 20筆 : 0.2763 (**device:** 12.0, **intelligent:** 7.0, **uncertainty:** 6.9, **intelligent information:** 6.3, **undergraduate:** 6.3)
    - (31) 26 : 25 : 15筆,0.00(intelligent:14.90, information:3.32, system:2.56, retrieval:1.10)
    - 44 : 124 : 5筆,0.06(development:4.95, information:0.70, retrieval:0.44)

10(7):

- 15327 : 41筆 : 0.0090 (**text:** 8.5, **information:** 4.5, **article:** 2.4, **system:** 1.9, **different:** 1.9)
  - 216 : 12筆 : 0.2241 (**cognitive:** 22.3, **physical:** 4.9, **polyrepresentation:** 4.2, **view:** 4.0, **paradigm:** 3.8)
    - 141 : 75 : 8筆,0.10(cognitive:26.13, model:1.07, search:1.05, system:0.32, information:0.02)
    - 49 : 167 : 4筆,0.06(paradigm:4.12, situational:2.68, nature:1.32, cognitive:1.31, relevance:1.12, view:1.06, methodology ...
  - 12940 : 29筆 : 0.0206 (**summary:** 26.1, **sentence:** 11.3, **text:** 9.2, **system:** 2.9, **article:** 1.9)
    - 284 : 21筆 : 0.2107 (**summary:** 34.3, **summarization:** 18.5, **sentence:** 15.0, **text:** 6.2, **evaluation:** 2.2)
    - (32) 21 : 15筆 : 0.3604 (**summary:** 25.8, **summarization:** 23.1, **text summarization:** 15.5, **sentence:** 10.1, **summarizer:** 7.1)
      - 147 : 76 : 8筆,0.10(summarization:27.32, text summarization:21.64, sentence:4.06, summary:2.97, document:0.22)
      - 71 : 90 : 7筆,0.17(summary:21.35, new:2.14)
    - 41 : 102 : 6筆,0.06(standard:11.44)
  - 27 : 8筆 : 0.3505 (**surrogate:** 14.7, **medicine:** 7.1, **clinical:** 7.1, **evidence-based:** 5.6, **guideline:** 2.5)
    - 280 : 140 : 5筆,0.00(clinical:7.74, medicine:6.03, problem:0.60, information:0.35, retrieval:0.10)
    - 42 : 181 : 3筆,0.18(surrogate:10.60, search:0.53)

11(7):

- 13640 : 45筆 : 0.0170 (**database:** 12.3, **subject:** 7.1, **language:** 6.7, **retrieval:** 6.6, **support:** 6.3)
  - 1042 : 37筆 : 0.0345 (**database:** 18.2, **relational:** 16.5, **language:** 6.3, **support:** 5.6, **keyword:** 5.2)
    - (33) 1174 : 28筆 : 0.1437 (**relational:** 21.6, **database:** 11.7, **query language:** 6.6, **normalization:** 5.1, **support:** 5.1)
      - 156 : 22筆 : 0.2402 (**xml:** 29.7, **xml retrieval:** 12.7, **database:** 8.9, **query language:** 8.1, **xml document:** 7.8)
        - 110 : 68 : 9筆,0.03(database:16.58, information:0.14)
        - 116 : 36 : 13筆,0.12(xml:78.36, retrieval:2.28, document:1.77, information:0.57)
      - 97 : 105 : 6筆,0.03(relational:5.91, application:2.59, support:1.57, data:0.86, query:0.31, system:0.28)
    - 39 : 9筆 : 0.3393 (**controll:** 12.7, **record:** 6.4, **subject:** 6.1, **subject descriptor:** 5.6, **vocabulary:** 5.5)
      - 29 : 120 : 5筆,0.07(descriptor:6.93, bibliographic:5.87, subject:5.10, record:4.16)
      - 30 : 159 : 4筆,0.20(controll:11.30, vocabulary:1.54, natural:1.44, versu:1.20, language:0.75, information:0.55, index ...
  - 4406 : 8筆 : 0.0790 (**clarification:** 8.5, **negation:** 3.5, **implicit:** 3.4, **display:** 3... **display time:** 2.8)
    - 163 : 178 : 3筆,0.23(clarification:11.00, ambiguity:2.18, query:0.48, system:0.26)
    - 75 : 142 : 5筆,0.15(implicit:5.06, seaches:4.48, display:3.11, feedback:1.16, information:0.18, retrieval:0.06 ...

12(6):

- 12715 : 30筆 : 0.0166 (**interactive:** 23.0, **behavior:** 13.3, **system:** 7.5, **activity:** 7.0, **dynamic:** 4.3)
  - (34) 7416 : 24筆 : 0.0571 (**interactive:** 30.0, **information-seeking:** 12.2, **iir:** 9.1, **activity:** 8.8, **behavior:** 6.7)
    - 1180 : 22筆 : 0.1435 (**interactive:** 24.5, **information-seeking:** 14.3, **iir:** 10.6, **interactive information retrieval:** 7.4, **dynamic** 6.4)
      - 2 : 18筆 : 0.5543 (**interactive:** 18.3, **information-seeking:** 17.7, **information-seeking strategy:** 9.9, **interactive information retrieval:** 9.1, **intention:** 7.1)
        - 185 : 26 : 14筆,0.00(interactive:24.55, system:4.20, retrieval:2.67, information:2.14)
        - 32 : 165 : 4筆,0.15(information-seeking:12.96, information-seeking strategy:9.89, intention:4.99, interactive:2 ...
      - 157 : 161 : 4筆,0.24(simulate:12.60, real:2.52, evaluation:1.56, interactive:0.72, test:0.58, system:0.55, query:0 ...
    - 160 : 151 : 2筆,0.22(navigation:3.70, search and navigation:2.83, autococ:2.12, search and navigation behavior:2.12 ...
  - 808 : 6筆 : 0.1620 (**forum:** 5.7, **mention:** 5.4, **behavior:** 2.8, **unstructur:** 2.8, **cross-language evaluation forum:** 2.8)
    - 136 : 155 : 4筆,0.02(mention:6.34, metadata:2.39, different:1.46)
    - 53 : 187 : 2筆,0.17(objective:2.14, cross-language evaluation forum:2.12, s objective function:1.41, optimal use of an ...

13(6):

- 8777 : 72筆 : 0.0441 (**hypertext:** 67.9, **link:** 18.8, **information:** 16.1, **interaction:** 9.2, **search:** 8.3)
  - 1902 : 65筆 : 0.1210 (**hypertext:** 40.6, **link:** 24.5, **information seek:** 22.6, **search:** 11.2, **model:** 8.3)
    - (35) 13 : 27筆 : 0.4205 (**hypertext:** 28.9, **link:** 13.6, **hypertext system:** 10.6, **hypertext link:** 10.6, **world:** 10.5)
      - 168 : 78 : 8筆,0.01(world:10.21, wide:3.98, search:1.09, information:0.84)
      - 36 : 18 : 19筆,0.06(hypertext:61.15, link:12.38, system:2.35, retrieval:0.66)
    - 237 : 38筆 : 0.2151 (**information seek:** 32.6, **elicitation:** 32.4, **patron:** 13.4, **documentation:** 11.3, **aboutne:** 9.2)
      - 17 : 30 : 14筆,0.00(elicitation:32.98, information:1.35, retrieval:2.10, information retrieval:1.79)
      - 23 : 11 : 24筆,0.00(information seek:32.85)
  - 608 : 7筆 : 0.1737 (**e journals:** 6.3, **propie:** 6.3, **interact:** 4.6, **newspaper:** 4.4, **environment:** 4.4)
    - 177 : 184 : 2筆,0.22(architecture:2.73, environment:1.62, framework:1.41, conceptual architecture:1.41, extensible ...
    - 85 : 118 : 5筆,0.03(electronic:6.01, e journals:5.42, database:1.01, interface:0.93, information:0.19)

14(6):

- 16372 : 31筆 : 0.0048 (**story:** 18.7, **new:** 10.4, **topic:** 3.4, **text:** 1.7, **term:** 0.9)
  - 10 : 6筆 : 0.4424 (**variant:** 17.6, **biomedical:** 7.9, **katakana:** 7.1, **tokenization:** 4.9, **finite state:** 4.5)
    - 1 : 186 : 2筆,0.21(variant:5.70, katakana:4.95, finite-state:2.50, finite-state graph:2.12, equivalence:2.12, japanese ...
    - 152 : 190 : 4筆,0.24(biomedical:13.27)
  - 14559 : 25筆 : 0.0126 (**story:** 23.1, **new:** 11.7, **track:** 8.2, **topic:** 5.3, **new story:** 4.9)

(36)
- 421 : 23筆 : 0.1906 (**story:** 26.8, **track:** 9.8, **new:** 8.1, **topic:** 6.8, **new story:** 5.7)
  - 18 : 16筆 : 0.5781 (**story:** 12.7, **track:** 12.3, **event detection:** 10.6, **ned:** 7.8, **new event detection:** 6.3)
    - 173 : 168 : 4筆,0.22(track:8.46, topic track:5.55, event track:4.24, train:0.64, model:0.27)
    - 67 : 42 : 12筆,0.05(detection:26.43, new:5.17, document:0.54)
  - 178 : 98 : 7筆,0.00(new:1.56, text:0.73, document:0.52)
- 102 : 188 : 2筆,0.31(merge:6.92, current news:4.24, multilingual thesauri:2.12, metasearch:2.12, thesauri merge:2.12 ...

15(4):

(37)
- 6002 : 25筆 : 0.0645 (**segmentation:** 26.0, **word segmentation:** 14.1, **chinese:** 13.7, **text:** 9.8, **phrase:** 5.7)
  - 2190 : 23筆 : 0.1145 (**segmentation:** 30.5, **chinese:** 16.4, **word segmentation:** 16.3, **text:** 7.8, **word-based:** 5.1)
    - 136 : 18筆 : 0.2529 (**segmentation:** 38.0, **word segmentation:** 20.0, **korean:** 14.8, **chinese text:** 10.9, **compound noun:** 8.5)
      - 124 : 111 : 6筆,0.11(korean:23.00, compound:12.50, korean information retrieval:5.65, index:2.75, term:0.81, text:0 ...
      - 179 : 41 : 12筆,0.03(chinese:37.77, segmentation:33.85, word:11.21, text:6.08, retrieval:2.14, information:0.32)
    - 3 : 146 : 5筆,0.00(keyphrases:5.2, principles:5.0, retrieval:0.50)
  - 10 : 185 : 2筆,0.21(phrase recognition:3.53, short:2.68, short scientific text:2.12, web query:1.59, storage and retrieval ...

17(3):

(38)
- 10598 : 17筆 : 0.0336 (**structur:** 11.3, **document:** 3.6, **frequency:** 2.9, **retrieval:** 1.8, **theory:** 1.7)
  - 232 : 14筆 : 0.2202 (**structur:** 14.4, **structur document:** 8.5, **structur query:** 4.9, **typical:** 3.0, **influence diagram:** 2.8)
    - 34 : 63 : 9筆,0.07(structur:17.62, document:1.07, retrieval:0.44)
    - 57 : 126 : 5筆,0.05(structur query:6.70, document:0.63, retrieval:0.53, information:0.10)
  - 104 : 179 : 3筆,0.31(dr:6.73, inverse:2.56, argumen::2.31, weight:1.86, frequency:0.96, term:0.46, document:0.24 ...

附錄 三　系統自動萃取主題之文獻數量年代排序

編號	主題		1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
1	citation	C	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3
		J	0	0	0	1	1	1	0	0	0	4	2	2	3	1	2	0	2	2	21
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
2	visualization	C	1	1	1	1	0	2	1	0	0	1	0	0	0	1	0	0	0	1	10
		J	0	1	0	1	1	1	0	1	0	2	1	3	3	2	1	2	0	1	20
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
3	link analysis	C	0	0	0	0	1	0	0	2	1	0	1	1	1	0	1	4	2	1	15
		J	0	1	0	1	0	0	0	0	0	0	1	1	0	0	1	0	1	0	6
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
4	genetic algorithm	C	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	2	1	7
		J	0	0	0	0	0	0	1	1	3	1	0	0	2	1	2	0	1	0	12
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
5	markov	C	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	2	2	1	8
		J	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	2
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
6	highly relevant	C	1	0	0	0	1	0	0	1	2	0	2	5	1	1	1	1	2	4	22
		J	0	1	0	0	1	0	1	0	2	0	2	5	4	1	1	2	2	3	25
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
7	morphological	C	0	0	0	0	1	1	2	0	1	0	1	0	3	0	1	1	1	2	14
		J	0	0	0	0	1	0	3	1	0	0	0	1	0	0	2	1	4	1	14
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計

8	logic	C	0	0	0	1	2	0	2	0	1	1	1	0	0	0	1	0	0	0	9
		J	3	2	2	0	2	0	0	0	3	0	0	0	0	0	0	0	0	0	12
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
9	space	C	0	1	1	0	1	0	0	0	0	0	0	2	1	1	0	0	1	0	8
		J	0	0	0	1	0	1	0	1	1	1	0	0	3	0	1	2	1	4	16
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
10	retrieval model/probabilistic retrieval	C	1	3	3	2	2	2	0	1	0	0	0	0	0	1	2	4	1	0	22
		J	0	1	0	0	1	0	0	0	1	2	1	1	3	1	2	3	2	3	21
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
11	query expansion	C	0	2	1	4	1	1	1	5	3	1	1	1	1	1	0	1	2	3	29
		J	0	1	1	0	0	2	2	0	1	1	3	6	1	4	4	1	5	4	36
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
12	cluster	C	1	1	4	1	0	2	2	0	2	0	2	1	2	2	4	1	2	4	31
		J	0	0	0	0	0	0	0	0	1	2	3	4	2	3	3	3	2	2	25
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
13	language model	C	0	1	1	0	0	1	0	0	1	0	0	4	4	1	3	3	2	1	22
		J	0	0	0	1	0	2	0	0	1	0	0	1	0	1	4	3	1	1	15
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
14	question answer	C	0	0	0	1	0	0	0	0	0	0	3	3	1	1	1	2	2	4	18
		J	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	3
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
15	latent semantic index	C	0	0	1	0	2	0	0	0	0	2	1	1	1	0	2	3	0	0	13
		J	0	0	1	1	0	1	0	0	1	0	0	1	0	2	0	5	2	0	14

			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
16	collaborative filte	C	0	0	0	0	0	0	0	0	0	1	0	0	2	1	2	2	3	1	12
		J	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	2
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
17	music	C	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
		J	0	0	0	0	0	0	0	0	0	0	0	0	2	0	6	0	1	0	9
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
18	text categorization	C	1	0	0	0	2	2	2	0	1	1	0	2	1	4	1	0	0	0	17
		J	0	0	0	0	0	0	1	0	0	0	1	1	2	0	1	2	0	0	8
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
19	cross-language	C	0	0	0	0	0	0	0	1	3	2	1	3	2	0	3	3	2	1	21
		J	0	0	0	0	0	0	0	0	0	2	1	1	1	3	4	3	5	1	21
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
20	image retrieval	C	1	1	0	0	0	1	1	2	1	1	2	0	0	4	1	3	0	2	20
		J	0	0	0	0	2	0	0	1	0	1	0	3	1	0	0	0	2	1	11
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
21	text classification	C	0	0	0	0	0	1	0	0	0	0	1	1	3	2	3	1	0	1	13
		J	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	0	1	5
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
22	boolean	C	1	0	0	0	2	0	0	0	1	0	1	0	0	0	0	0	0	0	5
		J	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	1	0	5
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
23	recall and precision	C	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	4

		J	0	1	1	0	2	1	4	0	2	1	2	1	0	1	0	0	0	2	18
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
24	data fusion	C	0	0	1	1	2	1	1	0	0	2	1	2	0	1	0	4	4	1	21
		J	0	0	0	0	0	1	0	1	0	2	1	1	0	2	0	2	5	3	18
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
25	filter	C	0	0	0	0	1	0	2	1	2	0	3	0	0	0	2	1	1	1	14
		J	0	1	0	0	0	0	0	0	1	1	0	3	3	0	1	0	0	0	10
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
26	invert file	C	1	0	0	1	1	0	0	0	0	1	0	1	2	0	1	0	0	0	8
		J	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	1	2	0	6
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
27	space efficiency	C	2	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	0	9
		J	0	0	0	0	0	0	0	0	0	0	2	1	1	0	1	0	2	0	7
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
28	signature file	C	3	1	2	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	8
		J	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	4
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
29	opac	C	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3
		J	0	0	1	1	1	1	4	0	0	0	0	0	0	1	1	0	1	3	14
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
30	search tactic	C	1	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	6
		J	0	0	0	2	1	0	0	0	0	0	0	2	0	0	2	1	1	2	11
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計

31	intelligent	C	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	0	4
		J	1	2	1	0	0	0	0	1	2	0	0	1	0	1	1	0	1	0	11
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
32	text summarization	C	0	0	0	0	0	1	0	0	0	1	1	3	2	0	0	2	2	0	12
		J	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	3
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
33	xml retrieval / database	C	0	1	0	1	0	0	0	0	0	0	0	1	0	1	3	2	2	0	11
		J	0	2	2	2	0	1	0	1	0	1	1	0	2	0	0	3	2	0	17
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
34	interactive&information-seeking	C	0	0	0	0	1	0	0	0	2	0	0	0	0	1	0	1	0	2	7
		J	1	1	2	0	1	2	2	5	0	5	5	3	4	2	3	2	0	3	41
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
35	hypertext	C	0	1	0	1	0	0	0	2	1	0	1	0	0	0	0	1	0	1	8
		J	0	0	0	2	1	1	3	4	3	1	1	1	1	0	1	0	0	0	19
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
36	track	C	0	0	1	0	0	0	0	1	2	1	2	3	3	2	2	1	1	3	22
		J	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
37	word segmentation	C	0	0	0	0	0	0	2	2	0	3	0	0	1	0	0	0	0	1	9
		J	0	0	0	2	0	0	0	0	0	6	2	1	1	1	1	0	1	1	16
			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	總計
38	structur query	C	0	2	1	2	1	1	0	1	1	0	0	0	0	1	0	1	0	0	11
		J	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0	3