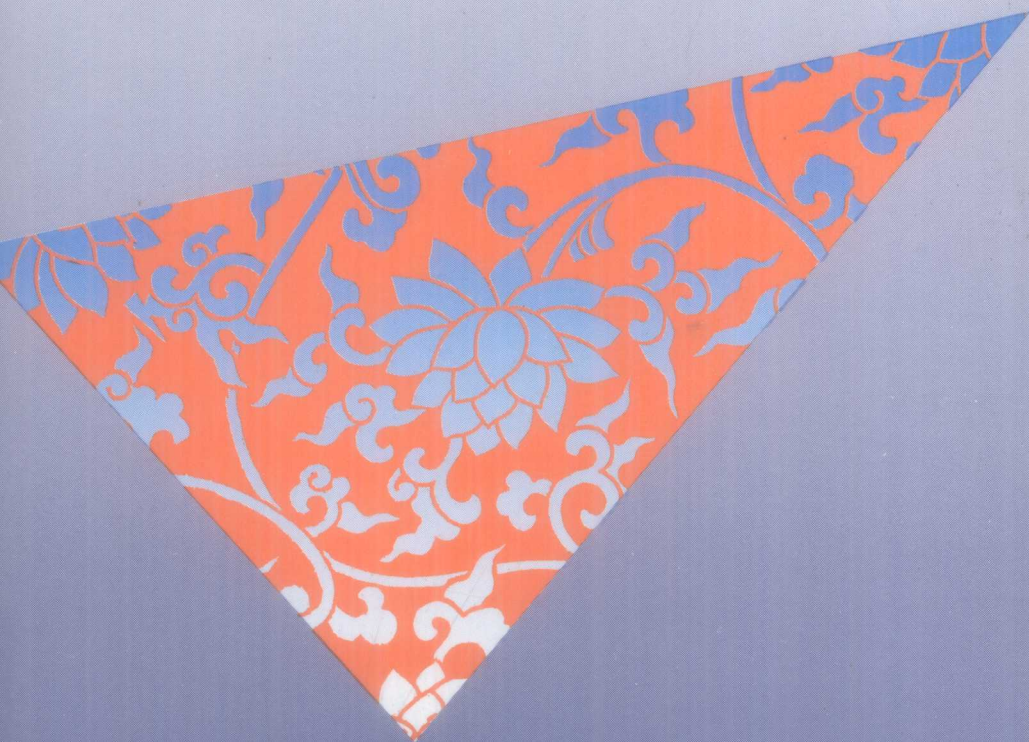


訊中「相關」概念之研究

圖書館學與資訊科學叢書

黃慕萱 / 著

台灣學生書局印行



「相同」概念之研究

ISSN 1120-4180

創刊號 第一卷

民國九十一年五月



資訊「相關」概念之研究 檢索中

黃 慕 萱

臺灣 學 生 書 局 印行

資訊檢索中『相關』概念之研究/黃慕萱著，--初版--

臺北市：臺灣學生，民85

面；公分

參考書目：面

含索引

ISBN 957-15-0748-2(精裝)

ISBN 957-15-0749-0(平裝)

1. 資訊儲存與檢索系統

028

85003360

資訊檢索中『相關』概念之研究

著 作 者：黃 慕 萱

出 版 者：臺 灣 學 生 書 局

發 行 人：丁 文 治

發 行 所：臺 灣 學 生 書 局

臺北市和平東路一段一九八號

郵政劃撥帳號〇〇〇二四六六八號

電 話：三 六 三 四 一 五 六

傳 真：三 六 三 六 三 三 四

本書局登記證字號：行政院新聞局局版台業字第一一〇〇號

印 刷 所：常 新 印 刷 有 限 公 司

地址：板橋市翠華街8巷13號

電話：九 五 二 四 二 一 九

定價 精裝新臺幣三五〇元
平裝新臺幣二八〇元

中 華 民 國 八 十 五 年 四 月 初 版

02323

版權所有·翻印必究

ISBN 957-15-0748-2 (精裝)

ISBN 957-15-0749-0 (平裝)

自 序

「相關」概念在資訊檢索中之所以位居核心概念，主要是由於資訊系統的目的在檢索相關資訊，及系統評估所使用的測量值大多是相關導向的回收率和精確率等因素所造成。因此，惟有加強對相關概念的了解，才有可能從本質上改進資訊系統，讓資訊系統真正成為知識溝通的橋樑。尤其是處於資訊時代的今日，掌握相關或有用的資訊已成為提高國力及競爭力的不二法門，所有有關相關之理論型及實證型研究也愈發凸顯其重要性。

本書從理論與實證二個不同的觀點研究相關概念。在理論方面，首先介紹以相關為核心的資訊檢索五大基本概念，其次介紹相關概念之歷史、發展、及其實證型研究結果，而後探討相關概念在系統評估上的應用、爭議、及其所帶來的省思。在實證方面，本書分為二大部分探討，第一部分的重點在描述資訊需求者相關判斷的結果，除實際分析訪問所得之每一次相關判斷評級、瀏覽欄位、判斷依據、和所耗費時間外，並企圖探討可能影響相關判斷結果之因素。第二部分則從引用文獻的角度探討心理相關，除設法了解檢索所得書目和實際引用書目間的差異外，並系統化地研究相關書目不被使用的原因及其他非線上檢索所得之參考書目來源，企圖深入探討相關概念和讀者使用間微妙而複雜的關係。

本書為作者研究及教學之心得結晶。返國數年來，承蒙

師長的提攜及同儕的關愛，深感在心！本書雖力求完妥，然受限於時間及個人才力，疏漏自所難免，尚祈學界與業界先進不吝賜教為禱！

本書承台大圖書館學研究所王慧玉及曾繁絹二位同學協助打字、排版、校對和編製索引等事宜，特此致謝！最後，感謝父母親的愛護與幫忙，以及外子的支持與鼓勵，使筆者在教書、研究之餘，能專心撰稿，而無後顧之憂，在此致上最深的謝意！

黃慕萱 謹識

民國 85 年 4 月

於國立台灣大學

資訊檢索中「相關」概念之研究

目 次

自序	I
第一章 緒論	1
第一節 研究緣起與目的	1
第二節 相關概念之理論探討	4
第三節 相關概念之實證研究	9
第二章 研究方法	15
第一節 研究設計	16
第二節 抽樣方法與研究方法	20
第三節 資料蒐集、編碼與分析	25
第三章 資訊檢索之基本概念	31
第一節 資訊	33
第二節 關於	38
第三節 需求	42
第四節 用途	46

第四章 「相關」概念之探討	53
第一節 相關研究的歷史	53
第二節 相關的定義與發展	62
第三節 影響相關判斷之因素及相關判斷之依據	71
第四節 其他相關研究的成果	78
第五章 「相關」概念與檢索系統評估	89
第一節 測量值的歷史	90
第二節 效用派理論	99
第三節 檢索評估	107
第四節 Cranfield 研究	117
第五節 Salton 和其他研究結果	125
第六章 資訊需求者相關判斷實證分析	135
第一節 檢索之基本特性分析	135
第二節 相關判斷之結果分析	146
第三節 相關判斷所花費之時間分析	152
第四節 影響相關判斷之因素探討	158
第七章 心理相關之實證探討	173
第一節 參考書目之特性分析	173
第二節 檢索所得書目和被引用書目之差異探討	181
第三節 相關書目不被引用的原因	188
第四節 非線上檢索書目之來源	201
第八章 結論	213

第一節 研究相關概念的意義	214
第二節 相關判斷之特色	218
第三節 心理相關之應用極限	221
參考書目	231
附錄	253
中文索引	281
英文索引	291

圖 表 目 次

圖 1-1	資訊檢索之五大概念與資訊系統關係圖	6
圖 4-1	影響相關判斷之因素組及其變數說明圖	72
表 2-1	檢索者教育程度統計表	23
表 2-2	檢索者所屬學院統計表	23
表 2-3	檢索者所屬系別統計表	24
表 3-1	學科軟硬排列表	32
表 4-1	影響讀者選擇書目之因素表	74
表 4-2	相關判斷因素出現次數表	76
表 5-1	回收率、精確率和雜訊比之 2 乘 2 表格	110
表 5-2	影響檢索結果之檢索者特質表	112
表 5-3	系統評估之標準	115
表 5-4	33 種索引語言之檢索效益排名表	121
表 6-1	不同教育程度檢索者之檢索次數統計表	137
表 6-2	各學院檢索者之檢索次數統計表	137
表 6-3	各學系檢索者之檢索次數統計表	138

表 6-4	各次檢索之目的分析表	140
表 6-5	檢索時或相關判斷時之研究階段分析表	141
表 6-6	檢索前所掌握之相關資料筆數分析表	142
表 6-7	檢索前之準備時間分析表	143
表 6-8	資訊需求者認為所需之相關文章篇數統計表	144
表 6-9	各次檢索所得書目筆數分析表	145
表 6-10	各次檢索之精確率分析表	146
表 6-11	各次檢索檢出之相關書目筆數表	148
表 6-12	檢索書目之相關判斷結果分布表	148
表 6-13	相關判斷瀏覽欄位分析表	150
表 6-14	相關判斷依據欄位分析表	150
表 6-15	相關判斷所費時間分析表	153
表 6-16	判斷相關文章與不相關文章所需時間比較表	156
表 6-17	判斷「非常相關」、「相關」、「部分相關」 及「不相關」文章所需時間比較表	157
表 6-18	相關判斷結果與情境變數間之相關係數表	162
表 6-19	三組變數之相關係數表	165
表 6-20	研究目的對相關判斷結果及精確率之 變異數分析表	167
表 6-21	研究階段對相關判斷結果及精確率之	

t 檢定表.....	167
表 7-1 報告之參考書目數量分析表	175
表 7-2 文章類型對參考書目篇數之影響	175
表 7-3 報告之中文參考書目數量分析表	176
表 7-4 文章類型對中文參考書目篇數之影響	177
表 7-5 報告之英文參考書目數量分析表	178
表 7-6 文章類型對英文參考書目篇數之影響	180
表 7-7 報告所引線上檢出書目之相關判斷結果	182
表 7-8 引用書目分布於「非常相關」判斷結果之 篇數統計表.....	183
表 7-9 引用書目分布於「相關」判斷結果之篇數 統計表.....	184
表 7-10 引用書目分布於「部分相關」判斷結果之 篇數統計表.....	185
表 7-11 引用書目分布於「不相關」判斷結果之篇數 統計表.....	186
表 7-12 不同相關判斷結果之被引用率比較表	187
表 7-13 低引用率成因表.....	190
表 7-14 檢索書目被判斷為相關卻不引用的原因	193
表 7-15 Wilson 之相關文獻不被使用理論與本研究	

所發現之相關文獻不被使用原因對照表	198
表 7-16 非檢索所得之參考書目來源	202
表 7-17 正式管道與非正式管道之參考書目來源表	207

第一章 緒論

第一節 研究緣起與目的

資訊系統的主要目的是在提高溝通（communication）的效果，而「相關」（relevance）正是提昇溝通效果的關鍵，惟有不斷地提供「相關的」（relevant）資訊，溝通才能發揮最大的效果。（註1）因此，惟有加強對「相關」概念的了解，才有可能從本質上改進資訊系統，讓資訊系統真正成為知識溝通的橋樑。尤其是處於資訊時代的今日，掌握相關或有用的資訊已成為提高國力及競爭力的不二法門，所以關於相關的理論型及實證型研究也愈發凸顯其重要性。

在資訊科學中，相關是一個不斷在討論中成長的核心概念，將其稱之為資訊檢索之心臟一點也不為過。一般而言，相關之所以在資訊科學中佔有如此重要的地位，除因資訊檢索的目的在檢索相關資訊外，系統評估所使用的測量值也大多與相關有關，例如回收率和精確率都必須求得相關資料之筆數，因此不管在資訊檢索或是系統評估中，相關都扮演舉足輕重的角色。而除資訊科學外，在傳播學中，相關也是非常重要的概念，不過就此二領域而言，相關都仍是模糊且尚

待定義之基本概念。

事實上，相關在日常生活中是一個相當普遍然缺乏定義的名詞，談話過程需要相關話題才能繼續進行，檢索時也需要相關資訊讀者才會覺得滿意。一般而言，二人對話的意義是建立在彼此對相關事物的共識上，如果雙方無法取得共識，一定會覺得話不投機，尤其是在具有目的的談話中。在言談進行時，話題的相關與否不一定是由主題決定，而每個人對相關的範圍也可能有不同的認定。就如同上課時學生提出之問題一般，學生一定是覺得該問題與課堂上的討論相關才會發問，但老師卻不一定認為此問題相關，其他同學對此問題之相關程度也可能會有不同的判斷。再以石門水庫乾旱缺水為例，由於每個人所處的情境不同，有人可能會馬上想到自己家裡的用水問題，但也有人會想到環保問題或是旅遊問題。資訊檢索上之相關亦然，主題相關並不是決定文章相關與否的惟一因素，很多情境因素都會影響相關判斷的結果，所以相關在資訊檢索上始終是個相當具爭議性及複雜性的概念。

線上資訊檢索系統可視為一部相關判斷的機器，檢索者通常以詞彙和布林邏輯運算元結合來表達其資訊需求，系統即依據檢索者輸入之檢索條件進行相關判斷，將其（系統）所判斷為相關之資訊呈現給讀者，由於系統通常只是根據檢索詞彙和描述文件詞彙之間的配對（match）判斷相關，因此不管是張三或李四，只要輸入相同的詞彙，他們一定會得

到相同的檢索結果。這種以詞彙配對的檢索方式，完全無法超越主題相關的層次，很可能是資訊檢索結果不能滿足讀者需求的主要原因之一。事實上，即使是針對相同主題的檢索問題，資訊需求者對相關資訊之要求還是各有不同，所以理想的系統應該根據讀者的知識狀態給予不同的檢索結果，絕非一成不變地對相同的問題提供相同的檢索結果。換言之，資訊檢索系統相關判斷的結果，應該和資訊需求者相關判斷之結果一致（即電腦判斷相關的結果等於資訊需求者個人判斷相關的結果），因此，惟有先了解人類相關判斷的過程與決策，才有可能設計出此一理想之資訊檢索系統。

總而言之，相關在資訊檢索中是一公認的核心概念，但人們有時可能認為它熟悉到不需要定義，有時又可能覺得它複雜到無法以任何定義加以釐清，因此，圖書資訊學界有關相關的實證型研究，實驗對象都是以個人心目中之相關概念進行相關判斷，難免在研究結果的解釋上帶來一些爭議。正因如此，在缺乏典範及操作型定義的資訊科學研究中，如何了解相關的本質並成功地定義相關，絕對是一個值得努力的方向。相關概念的重要性雖然無庸置疑，然而卻是國內學者鮮少涉獵的研究領域，所以更凸顯此方面研究之重要性。因此本書的目的在針對相關概念進行理論性探討和實證性分析，除剖析相關概念之本質、發展、和應用外，並嘗試實際蒐集資料，以實證的角度深入地了解國人相關判斷的特性，並測試心理相關的應用層面。本書共分八章，第一章和第二章

章分別為緒論和研究方法，第三章至第五章的重點是相關概念之理論探討，第六章和第七章為相關概念之實證型研究，而第八章則為結論。希望藉此喚起國人對相關研究的重視，因為語言文化的不同，一定會造成國人和歐美人士相關判斷上的差異，我們不可能完全沿用國外相關研究的結果，非得建立本土化的相關研究，才有可能設計出適合國人使用之理想資訊檢索系統。

第二節 相關概念之理論探討

由於理論為實證研究之基礎，故本書第三至第五章將致力於相關概念之理論探討，並對各家重要研究加以闡述與評析。一般而言，相關概念之所以重要，在於資訊檢索的目的是為找尋相關資訊，及系統評估的測量值大多與相關有關等因素，因此本研究將以資訊檢索和系統評估二種不同角度來審視相關概念。在資訊檢索方面，本書將從相關概念本身及資訊檢索的其他重要概念與相關概念的關係來探討此核心概念。至於系統評估方面，本書除探討相關導向之測量值外，並將對效用派理論詳細地加以介紹，然後再討論系統評估之研究結果帶給圖書館界的省思。

首先探討資訊檢索之基本概念。Patrick Wilson 認為資訊檢索領域包含五大基本概念，分別是「資訊」、「關於」(about)、「相關」、「需求」(need)、和「用途」(use)。