

「相關」與「模糊」在資訊檢索領域中關係驗證與分析
Verification and Analysis of the Notions of “Relevant” and “Fuzzy”
in Information Retrieval

李煜基

Yu-Ji Li

私立致遠管理學院資訊管理學系專任講師
Lecturer, Department of Information management,
Diwan University
E-mail : catiris@dwu.edu.tw

洪一梅

I-Mei Hung

中央研究院歷史語言研究所研究助理
Research assistant, Institute of History and Philology, Academia Sinica
國立臺灣大學圖書資訊學系博士生
PhD. Student, Department of Library & Information Science,
National Taiwan University
E-mail : yimay@asihp.net

【摘要 Abstract】

在資訊檢索中，「相關理論」是重要的基礎概念；而在資訊檢索的技術發展中，「模糊理論」亦是重要的議題。本文除針對此二種理論進行深入介紹與概念內涵比較解析外，並以文獻計量方法驗證二者的應用核心領域，及進行二者的關聯研究。文末嘗試依據上述結果，推測「相關」及「模糊」在「資訊檢索」領域中可能的應用發展趨勢。

The “Relevant Theory” and the “Fuzzy Theory” are two important topics in Information Retrieval (IR). In addition to thoroughly explicating the two theories, this paper adopts the method of bibliometrics to verify and analyze the respective core fields of applications and the connection between the two notions of “relevant” and “fuzzy.” Furthermore, this study also suggests probable extensive applications of the two theories in the future trends of information retrieval.

關鍵詞 Keyword

書目計量 相關理論 模糊理論 資訊檢索

Bibliometrics ; Relevant theory ; Fuzzy theory ; Information retrieval

壹、前言

Wilson (1978, p.10) 曾在文獻中提到，技術的發展總是走在科技了解之前，且在資訊技術領域中，往往有能力做很複雜的事，卻說不出核心的概念及理論，資訊檢索領域就是一個明顯的例子。而在這特殊領域中，技術的創新反成為理論建構的基礎，透過抽絲剝繭、歸納與分析，方得為領域中的核心概念形塑，這看似矛盾的理論建構方法，卻也是資訊檢索領域，在實務與技術進展的過程中，許多基礎概念與基礎理論不斷被發現被提出的原因。「實務與技術」、「概念與理論」二大區塊的交互激盪，方促成今日資訊檢索領域的蓬勃發展。解析資訊檢索領域此發展特質，做為一個資訊研究者，看待技術的創新，應有反向為概念與理論形塑、堅實資訊檢索學術基礎的構思。此即本研究的動機，嘗試以這樣的方法，讓個別資訊技術的發展有更清晰的概念基礎，也讓理論得以從技術應用上獲得驗證。

本研究初始，是由於對近來資訊檢索領域中「模糊技術」發展的好奇，在許多文獻中發現，模糊技術是發展於模糊理論，但依據 Patrick Wilson 建構的資訊檢索概念框架，模糊理論並非是資訊檢索基本概念之一，是否有可能，模糊理論已是資訊檢索五大基本概念發展後的應用理論？而模糊理論與五大基本概念間有何關聯？本文將僅針對「模糊理論」與五大基本概念之一的「相關理論」進行關聯研究。

本文將針對「相關理論」及「模糊理論」進行結構式介紹與比較，並以 LISA 中收錄之相關主題文獻進行計量分析，以驗證二者的應用核心領域為「資訊檢索」，並解析二者在資訊檢索領域中的關聯，以此推測二者的未來應用發展。

貳、「相關理論」與「模糊理論」

「相關」與「模糊」理論在資訊檢索研究中的關係，首先以定義、研究歷史、概念內涵與應用等面向，分別探析二者的發展；接著進行資訊檢索中「相關」與「模糊」的關係之比較分析。

一、資訊檢索中的相關理論

所謂相關在資訊檢索中的定義、研究歷史發展、內涵及應用等面向，茲分析如下：

(一) 定義

「相關」在資訊檢索中是公認的核心概念，有時我們可能覺得它熟悉到不需要定義，但有時卻又覺得它複雜到無法以任何定義加以釐清。(黃慕萱，民 85，頁 3)在此，我們嘗試給「相關」一個較非應用性的定義：三個角色分別為 A、B 及 C，相關即為 C 與「B 對 A 的看法」的關係程度或影響程度。因此，「相關」事實上是 B 主觀地去判定 C 與 A 的關係程度或 B 主觀地受影響程度。以這樣的定義來申論，其實社會中無處不在發生相關現象，人們在生活中亦隨時皆在進行相關判定。

圖書資訊學界對「相關」的實證型研究，較多都是界定在某些主題領域上的應用，也多是引自資訊科學中發展的相關概念，如「資訊檢索」主題領域中，Wilson 提出資訊檢索五大基本概念：「資訊」、「關於」、「相關」、「需求」、及「用途」。「相關」不僅為其一，Harter 亦曾說：「資訊即相關」，意即：此五大概念實以相關為核心。「相關」操控著資訊檢索的所有角色與行為，包含使用者(需求者、檢索者及資訊管理者)、資訊、系統，並且環環相扣。(黃慕萱，民 85，頁 4-5)

(二)研究發展歷史

「相關」的研究一直是資訊科學的重點研究之一，從 1950 到 1970 年代，無論是實證型研究或是理論型研究，皆著重在「相關」本質的釐清及理論的建立。1958 年，ICSI(International Conference for Scientific Information)的二個重要爭辯，何種相關在哲學上和經驗上是最好的？何種相關可以被測量？其爭辯主要環繞在 Vickery 在會中所提出的二個觀念上：系統導向的主題相關；使用者導向的使用相關。而在 1967 年時，Rees and Schultz 曾為文指出，ICSI 的辯論促成了四點共識：

- (1) 「相關」不只是在系統內部所進行的作業；
- (2) 「相關」不只是文件內容中僅有的特質；
- (3) 「相關」不是二分式的抉擇；
- (4) 存在著一種可以被判定的「使用者的相關」。

同年，Caudra and Katter 在其研究中指出，「相關」可表示：(1)詞彙或文件與使用者興趣領域的關係；(2)系統輸出與使用者資訊需求的關係。並發展出「相關評估模式」。其對「相關」的突破性研究成果在於「找出使用者變數」，但其研究限制中亦提到其難以直接測量與使用者內心狀態相關的變項，因此，其建議，未來研究應以發展使用者與情境的研究模式為優先。此外，Rees and Schultz 著重在「相關判定」的過程與方法，其研究成果發現：(1)個人差異的確會影響相關判定；(2)愈多資訊可作為判定依據時，其相關排序愈低。由結果指出，相關判定是取決於內在差異及/或個人知識狀態的動態改變，因此，其建議未來研究可採取認知途徑(Cognitive approach)。1973 年，Cooper 的研究將邏輯相關的定義侷限在主題方面的考量，將此部份導引回到傳統的系統導向；而在使用者面向，其認為“Utility”一詞可涵蓋所有使用者認為有價值的

因素，包含：Usefulness、Accuracy、Informativeness、Credibility、Clarity、Qualities of system、Possible negative factors、Possible positive Factors 等，但 Cooper 並未實際檢測這些因素。1975 年，Saracevic 認為「相關」即是用來測量來源與目的之間的有效接觸程度，其將研究變項分為：文件與文件表徵、檢索問題、判定情境、判定尺度、判定者。Saracevic 認為「判定者」為主核心變項，而對判定者影響最大的變項為其主題知識，情境因素亦會有一定的影響力。研究主要結果：由使用者判定的相關雖然複雜，但可被測量。另外亦歸納出三個研究假設：(1)相關判定是主觀的；(2)認知叢集的改變被用來解釋相關判定的過程；(3)由於對資訊的需求目的不同，因此存在著不同類型的判定。其結論指出，認知因素應是所有觀點中的「相關」的基本特性。(Schamber, Eisenberg & Nilan, 1990, pp.755-776)

上述的研究對「相關」的內涵及應用發掘愈來愈多，但對「相關」的定義，從簡單到複雜，再從複雜到簡單，卻仍難以獲得一典範型的定義內容。也因此，到了 1980 年代，「相關」研究的進展有限。而在研究發展沿革中，雖然相關的定義一直存在著爭議，但其中最一致的觀點為：「相關」與人的關係終究多過與系統功能的關係。但在上述的研究中，也發現系統功能卻總在「相關」判定上位居主控的角色。一直至 1990 年代，科技的發展帶來了「相關」新的研究目標。(黃慕萱，民 85，頁 53-62)研究者開始運用過去對「相關」的看法去發展新的科技，如類神經網絡、人工智慧及模糊系統等，同時，也嘗試以「相關」去評估系統功能，進而改善。「相關」成為不僅只是系統的發展基礎原理，也成為實際可檢測系統的工具指標，由此也發現「相關」本質的多樣化。

(三)概念內涵

「相關」與「模糊」在資訊檢索領域中關係驗證與分析

在資訊科學領域，學者對相關的研究結果，反映在定義的「相關」概念上，而這些概念的成長變化，正是人類認識「相關」的發展歷程，而這些概念經過許多研究者的實證研究，正確地說，僅可認定皆是「相關」的內涵之一。亦即研究結果不斷地彌補著人類對「相關」的認識，而「相關」實際的完整的內涵為何？目前仍未有結論。

以下以 1970 年代前後及 1990 年代，學者所提出的「相關」概念，來陳述我們現今所認識的「相關」的內涵：

1. 主題相關

Caudra and Katter 在 1967 年，以系統觀點，認為所謂相關是資訊條件敘述(即輸入系統之檢索問題)和文章內容間之一致性，亦即文章所涵蓋的內容對資訊條件敘述的適合程度。對所謂的「適合程度」仍為抽象概念。

2. 邏輯相關

Cooper 在 1971 年，以使用者的資訊需求觀點，提出所謂相關是一句子和資訊需求邏輯相關的必要條件是其所屬文件(以儲存的句子表達之)必須包含構成資訊需求之最小前提組(Minimal premise set)。(黃慕萱，1996，頁 13)對所謂最小前提組，其哲學定義為「能推論出所需結果之最小前提集合，在此集合中，如果刪除任一前提，就無

法以邏輯推理得到所需結論」。試圖找出客觀的可測量因素，以主題相關為基礎。其判定相關的程序是採用「演譯法」。(Cooper, 1971, pp.19-37)

3. 情境相關

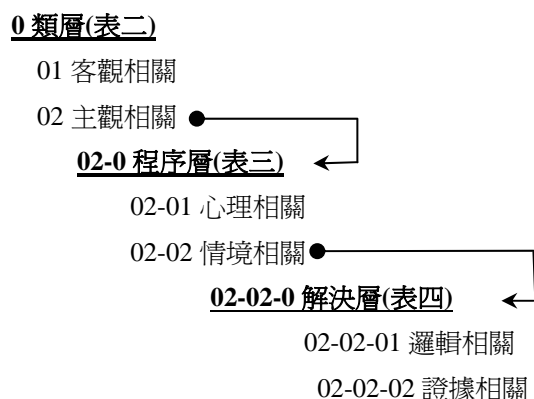
Wilson 在 1973 年，以使用者的個別化資訊需求情境觀點，以邏輯相關為基礎，但考慮到個人的認知狀態、興趣和喜好，以及上述個人情境隨時間不同而產生的變化。相關是動態且不斷變化的判定。以邏輯相關為基礎，並以證據相關彌補之。判定程序併同歸納法。

4. 心理相關

到了 Harter (1992, pp.602-615)以心理及認知觀點，認為相關的資訊就是能改變人類認知狀態(或知識狀態)的資訊，亦即，能產生文字關聯效果的資訊。包含主題、邏輯及情境相關。

前述 Harter 的「心理相關」、Cooper 的「邏輯相關」及 Wilson 闡述的「證據相關」與「情境相關」等相關之概念內容，以下嘗試以模式結構重組「相關」的概念發展框架(如表一)，將相關區分為三層：類層、程序層及解決層。類層是主客觀的相關類別；程序層主要在「主觀相關」中發生，以影響目標來區分；解決層則是在「情境相關」中發生，以解決方法加以區分。

表一：「相關」概念發展框架



第一層類層是以主客觀面向區分，所謂的客觀相關，是資訊本身的主題相關，在相關的判定上屬單一式程序，而這樣的系統運用，通常屬專家導向或系統導向；而主觀相關，則是指具改變認知狀態

的元素，在相關判定上屬互動式程序，在系統運用上屬需求者導向或實踐者導向。其內涵、程序與導向如表二。

表二：類層

類 層			
類	內涵	程序	導向
客觀相關	資訊本身的主題相關	單一式程序	專家導向 系統導向
主觀相關	改變認知狀態的元素	互動式程序	需求者導向 實踐者導向

第二層程序層是發生在「主觀相關」中，分為「心理相關」與「情境相關」。心理相關主要內涵是在「相關判定」上，系統實務則反應於人機間的檢索行為變動，其主要影響在個人，執行判定的程序是以系統答案比對檢索者的需求；情境相關，主

要內涵是在「認知影響」上，系統實務是反應在整體環境設計的變動及實際相關上，其影響目標為整體環境，執行程序以整體的需求環境來決定答案的相關。程序層中二種相關的內涵、系統實務、影響目標與程序之比較如表三。

表三：程序層

程 序 層				
相關	內涵	系統實務	影響目標	程序
心理相關	相關判定	檢索變動	影響個人	以系統答案比對需求
情境相關	認知影響	設計變動 - 實際相關	影響環境	以需求環境來收納答案

第三層之解決層在「情境相關」中發生，其主要在實際進行推論、判定及解決問題。邏輯相關採用演繹法，其判定元素主要為「關心」及「興趣」，主要著重在解決問題的方法上；證據相關是採歸納

法，其判定元素為「程度」及「可能性」，主要著重在解決問題的程度。解決層中邏輯相關與證據相關之比較如表四。

表四：解決層

解 決 層			
相關	推論方法	判定元素	解決問題
邏輯相關	演繹法	關心與興趣	解決問題的方法
證據相關	歸納法	程度與可能性	解決問題的程度

(四)相關在資訊檢索的應用

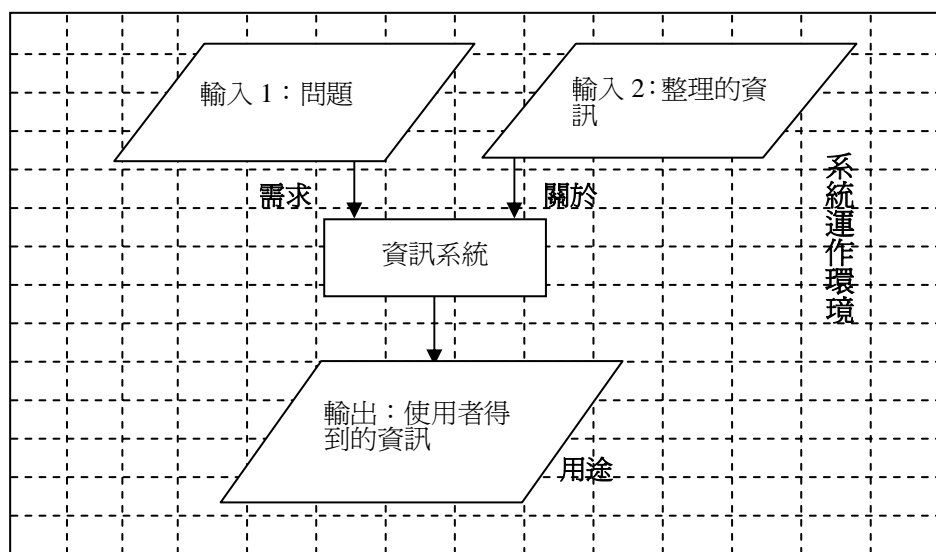
由相關之概念發展框架可知，「相關」可以是一個「結果」，也可以是一個「判定的程序」，同時也可以是一個「解決問題的工具(方法與程度)」。相關理論在資訊檢索的應用，主要可分區分為系統概念與實務應用二方面加以陳述：

1.系統概念

以 Harter 整合「相關」概念的所提的看法說明之：依據 Wilson 提出的資訊檢索五大基本概念，「資訊」、「關於」、「相關」、「需求」、及「用途」，Harter 認為「資訊」即「相關」，因為「資訊」檢索的目的是為檢索「相關」資訊，資訊的本身即包含了「相關」的意義。而，相關資訊是為滿足使用者之資訊「需求」，相關的判定方法之一，是以符合「需求」程度來判定。而整理資訊的重點在於分

析、描述及組織文獻所「關於」之主題，而所謂「關於」，亦即在分析、描述及組織文獻彼此間或與知識結構間的「相關」。以系統程序來說，系統必須蒐集、對映使用者之資訊需求，及文獻所關於之主題，使用者才有可能檢索出其所需之相關文獻，因此，「需求」和「關於」已在系統程序中與「相關」概念互相結合。至於「用途」，雖然效用派學者認為系統應提供有用的(Useful)資訊，但因效用派理論已逐漸融入「相關」理論中，因此「用途」可視為某種「相關」。綜合言之，資訊檢索的五大概念以「相關」為核心，環環相扣。(黃慕萱，民 85，頁 4-6)

本文嘗試將黃慕萱教授根據 Soergel 的「資訊系統結構圖」所繪製的「資訊檢索之五大概念與資訊檢索系統關係圖」提出調整性的看法如圖一。



虛線縱軸：資訊

虛線橫軸：相關

圖一：資訊檢索之五大概念與資訊檢索系統關係圖

依圖一，我們可以說，資訊檢索系統是在運作「資訊」與「相關」。而「資訊」與「相關」是遍佈在整個系統運作環境中的。

2. 實務應用面

「相關」在資訊檢索系統上的實務應用包含：
 (1)以系統運算「相關」：存在於系統中的資訊，藉由系統的運作予以整理，整理包含了組織、分析、描述，而這些步驟都參雜了「相關」的運算。
 (2)以「相關」評估系統：許多系統檢索的評估，所依據指標為：回收率、精確率、雜訊比等，而這些指標皆是以「相關」運算而來。(黃慕萱，民 85，頁 109)
 (3)交互相關(Interrelation)，資訊檢索中有一重要影響變數，即「使用者」(使用者包含需求者、檢索者及資訊管理者)，且在系統運算與評估系統的過程中，使用者的內在需求及使用外操作行為，包含將內在轉化為外在的情境程序，皆是「相關」的最大決定因素。Saracevic (1976, pp.84-85)

認為使用者的知識與使用者的問題，經過表示程序，皆影響了「相關」的系統運作，形成「主題相關」(Topical relevances)、「動機性相關」(Motivational relevances)及「解釋性相關」(Interpretational relevances)的交互相關。

因此，我們可以說，「相關」的應用是遍佈在基本資訊檢索系統模式中的。

二、資訊檢索中的模糊理論

以下從定義、研究歷史發展、內涵及應用等面向探討文獻中所謂的「模糊理論」：

(一) 定義

所謂模糊理論基本上是為了處理模糊概念所發展出來的。它實際上是模糊集合、模糊關係、模糊邏輯、模糊控制、模糊量測…等理論之泛稱。是一門將模糊概念量化的學問。模糊理論以模糊集合為基礎，以研究不確定事物為目標，接受模糊現象

「相關」與「模糊」在資訊檢索領域中關係驗證與分析

存在的事實，根據不清晰的訊息，透過近似推理但實有根據的過程而得到較符合正確的結果。在傳統明確集合(Crisp set)的特徵函數是採用非 0 即 1 的二分法，而模糊集合則是將其擴展為由 0 至 1 的任何值，稱為歸屬函數(Membership function)，當一個元素屬於某集合的程度愈大時，其歸屬函數就愈接近 1，否則愈接近 0。藉由歸屬函數對模糊概念進行量化之後，便可以利用精確的數學方法進行模糊資訊的分析和處理。(林信成、蕭勝文，2003，頁 63-64)

(二)研究發展歷史

早在 1960 年，Maron and Kuhns 就曾利用概率論的方法，研究資訊檢索系統的索引和檢索特性，並提出了「相關數」(Relevant number)的概念，而這已經隱含了某種模糊處理的因素。Zadeh 在 1965 年提出模糊集合(Fuzzy sets)理論，為模糊概念的數學描述和處理提供了有力的工具。1976 年，Tahani 提出了模糊檢索模型，接著 Radechi 進行了更深入的研究，給出檢索語言的 λ (水平語義)。 λ 事實上是索引詞與資訊內容在語言意義上比較的一個測度。但缺點是檢索輸出的資訊相關度量值隨著 λ 的變化而變化。1982 年，Zenner, Decaluwe and Kerre，對 Radechi 關於 λ 的定義做了修改，解決了穩定性的問題。但 λ 在資訊檢索系統上的應用研究，並沒有實質性的進展。此外，戎行等人於 1980 年，在 SJTU(上海交通大學)的系統上首先實現了一個簡單的 Fuzzy 模型，稱為相似度檢索模型；黃祥喜則探討了資訊檢索的模糊數學描述；顧其鈞、李書宏則分別探討了模糊檢索中歸屬函數的分布問題；都小鍵、康耀紅則建立了多值相關下的檢索理論。(王玫，1995，頁 8-9)

(三)概念內涵

模糊理論的內涵具有數學及哲學的基礎，源自

於數學，是在進行量化的測度；源自於哲學則是其邏輯的分析與運算，曾經模糊邏輯被稱為多重價值邏輯(Many-valued logic)。模糊的概念內涵，可以其數學基礎、模糊邏輯、相似、非常真實與機率加以描述。(Jacquette, 2002, pp.595-605)

1. 數學基礎：以符號表示： \approx (相似度)、 λ (水平語義)、 \leq 、 \geq 。
2. 模糊邏輯：模糊邏輯可區分為概觀(Broad sense)與縮觀(Narrow sense)二種，概觀等同於模糊集合理論及其應用；縮觀是建立在多重價值(Many-valued)邏輯上的近似推理(Approximate reasoning)理論。其基本主張為：
(1)以 1 代表真實(Truth)，以 0 代表虛假(Falsity)，以比較值來區分的真實 - 價值(Truth-values)的集合。
(2)在真實 - 價值(Truth-values)區分之後，再以真實 - 函數(Truth-functional)來測量。
(3)連結可能的真實 - 函數(Truth-functional)及持續性 - 真實標準(Continuous t-norms)。
(4)真實 - 函數(Truth-functional)的應用是相對於真實標準(T-norms)的餘值(Residuum)。
(5)真實 - 函數(Truth-functional)的否定，即為虛假(Falsity)。
3. 相似(Similarity)：相似是一種模糊等式，此概念出現在多數已知的模糊社群(Fuzzy community)中，設定相似的判定標準後，運用邏輯推理，即使完全不同的值，仍有可能相似。
4. 非常真實(Vary true)：若以縮觀的多元價值邏輯(Many-valued logic)來看，則模糊真實價值(Fuzzy truth-values)即為非常真實(Very true)的判定，若是以模糊集合及特徵函數來進行多元價值判定，亦即真實函數的模糊邏輯(Truth-functional fuzzy logic)。
5. 機率(Probability)：模糊的方法提供答案不是精確的二分結構，而是機率的高低。

(四)模糊理論在資訊檢索的應用

從許多文獻中解析，模糊理論促進了資訊檢索系統的多樣化及多功能發展，包括：(1)概念面，即權重概念，例如文獻模糊權重、權重式模糊問題、權重式詞彙-問題；(2)系統面，如模糊知識導向系統、動態系統、模糊網絡、程序控制系統、模糊認知地圖、模糊分類系統、智能控制系統、人工智慧等；(3)技術面，包含模糊搜尋、權重式布林邏輯、模糊相似模式及模糊分類等技術發展。(Mohammadian, 1999, 176-194)

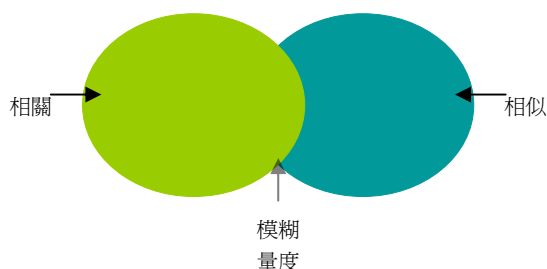
三、資訊檢索中「相關」與「模糊」的關係分析

依據相關文獻中所論述，資訊檢索中「相關」與「模糊」的關係可區分三個面向：「概念內涵」、「哲學與數理基礎」及「應用與發展面向」，進行

二者間關係之比較。

(一)在概念內涵的關係

相關與模糊的概念關係，以其內容觀點，模糊量度可藉由「相似」來判定「相關」，亦可藉由「相關」來判定「相似」。因為二者間有部份意義的重疊(如圖二)。舉例來說，定義 x, y ，當 $|x-y| \leq 1$ 時，則 x 與 y 為「相似」，而這也是一種簡單明瞭的「相關」，3 和 4 是「相似」，4 和 5 也是，而 3 和 5 則不是，當不是絕對地僅能是 3 和 4，也可以推論到 4 和 5 時，這就是一種模糊量度，而這個模糊量度透過 $|x-y| \leq 1$ 這個關係來判定，亦即以「相關」來判定「相似」。反之，3 和 4 具有 1 個差的相似，4 和 5 也具有 1 個差的相似，則 (3, 4) 與 (4, 5) 這二組數具有都是 1 個差的「相關」，這則是以「相似」來判定「相關」。(Jacquette, 2002, pp.600-601)



圖二：模糊量度內涵概念圖

而以其背景(Context)而言，相關應是多元價值的判定，具有程度上的差異，而非單純是與非的二元判定。(Saracevic, 1976, p.97)模糊則是提供了明確多元價值判定的方法。

(二)在哲學與數理基礎的關係

在哲學基礎上，從相關邏輯(Relevant logic)來看系統運作的「相關」、從模糊邏輯(Fuzzy logic)來看系統運作的「模糊」，二者皆運用了「推論邏輯」，無論是模糊理論中的「相似」或是相關理論中的「相關」，在某些情境下是一致的，亦意指相同的內容。再清楚地說明，相關的二分判定，可視

「相關」與「模糊」在資訊檢索領域中關係驗證與分析

為一種二元價值(Two-valued)邏輯的分類，而模糊的 Many-valued 邏輯即包含了二元價值邏輯，更進一步，相關的判定也可應用模糊理論來改善二分法所帶來的誤差或錯誤的精確判定。在二者的哲學基礎上，亦分別發展了源自集合理論(Sets theory)的 "Relevant set theory" 及 "Fuzzy set theory"，而這都是資訊系統在運算時的重要理論結構。(Mares, 2004, 195-199)

談到數理基礎，算術(Arithmetic)是很重要的人類活動，而算術亦是人類得以進行許多推論的理論基礎，而在相關邏輯(Relevant logic)與模糊邏輯(Fuzzy logic)的內涵中，除邏輯上的哲學基礎一致外，二者的數理基礎都成為其得以進行邏輯推論的前提，而事實，數理基礎亦為資訊系統真正在進行運算時的重要方法結構。在許多資訊檢索系統中，運作「相關」最普遍的作法是布林邏輯，而模糊檢索實際上即是布林邏輯的擴展，對於被檢索的資訊，模糊檢索會給一個相對於檢索需求(或檢索問題)的函數值，這個數值在程度上即反映了資訊對檢索需求的相關度，再與「基本設定值」(王玫稱閥值)比較，相關度量值 \geq 基本設定值的資訊才會被作為相關資訊，予以輸出，同時也可按相關度量值的大小排序輸出，以顯示各資訊對檢索需求的量化相關程度。因此，以數理基礎來說，模糊實際上是對相關性的定量程度描述。(王玫，1995，頁9)

(三)在應用與發展面向的關係

模糊現象源自於「相關」的判定，而模糊理論則適時地解決了「相關」判定中二分法的缺陷。模糊理論並非提供明確的解答，而是符合需求的機率高低，因此，可解決「相關」判定一些盲點，以避免各種不同程度「相關資訊」的遺漏。二者在應用與實務發展上的關係解析：

1. 互動系統及行為系統

傳統的資訊檢索建立在系統導向的運算上，但當使用者的角色愈來愈突顯後，我們發現數理的運算逐漸整合了認知模式，將使用者安裝在現代資訊系統的環節上，導致愈來愈多的互動系統或稱行為系統的出現。(Saracevic, 1996, p.3)

2. 專題選粹服務

模糊理論在資訊檢索上的改善，包含了資訊組織(如：模糊分類)及檢索結果的評估(如：相關的多重價值判定)，而這提供了專題選粹服務的絕佳基礎，近來，有些研究開始導向服務品質的提昇及主動性的資訊服務，導入「讀者模糊分類」(林信成、蕭勝文，2003，頁74)，期望以模糊技術來更細緻地描述讀者的需求，依此建立模糊專題選粹服務的模式，即模糊讀者、模糊資訊組織，以模糊系統與技術建立二者的相關判定之模式。

3. 專家系統

一般的資訊檢索系統在多語言的檢索問題上總是遭遇困境。而在此的多語言，分為二個部份：一為自然語言，可能因國別或種族不同而有的原始語言差異；一為專業語言，可能因專業背景的不同，在語法及用詞上的差異。可運用多語關鍵詞中語意關連的知識(內含模糊概念)，分析使用者的資訊需求，並建立文獻與問題之間的相關，以提供適合的結果。(Mohammadian, 1999, 488-489)

4. 網路資訊檢索系統

網路空間的動態模式，整併「相關」的判定與檢測及「模糊」的演算技術，促進了網路資訊檢索系統的發展與熱絡。(MacCall, 1998, p.13)

綜合上述，二者在資訊檢索領域中，從基礎的概念內涵到應用發展，「相關」與「模糊」實具有密切的關係。在定義界定的層次與程度，在哲學與

數理的演化基礎，在應用上的立基與發展，皆可看到「相關」與「模糊」理論的相輔相成。

參、「相關」與「模糊」在資訊檢索中應用發展的驗證

針對「相關」及「模糊」理論在資訊檢索中的應用發展，以文獻計量方法，分別驗證確認二者的發展核心，及解析其應用關係。

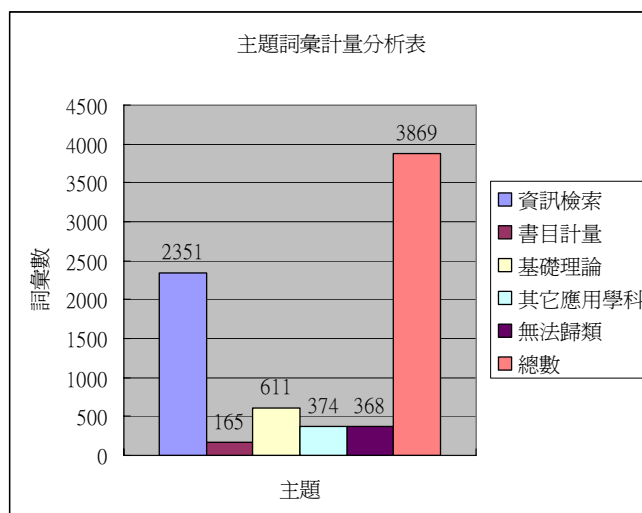
一、發展核心的驗證

雖然我們常可以在資訊檢索領域的文獻中看到「相關」及「模糊」理論，許多學者在討論「相關」與「模糊」理論時，均脫離不了資訊檢索的應用。以下嘗試以文獻計量分析的方法，驗證這樣的現象，確認與「相關」及「模糊」理論有關的文獻，最大的應用核心即為「資訊檢索」。

此驗證採取的資料來源為 Library and Information Science Abstracts，LISA 資料庫，資料蒐集日期從 1969 年至 2005 年 12 月 31 日，資料選取的方式(查尋指令)為：DE=(relevance or relevant or fuzzy)，獲取資料量共計 675 筆，合計 3851 個主題詞彙。所採取的計量分析程序，首先以 White

and McCain (1998, pp.327-355)以作者共同被引的方式，實證研究資訊科學的領域分析，將資訊科學的領域，大致分為兩大區塊，即資訊檢索與書目計量。本研究據此再將主題區分為：(1)資訊檢索；(2)書目計量；(3)基本理論(即為「相關」與「模糊」的基本理論與概念)；(4)其它應用學科；(5)無法歸類等五大類；在歸類過程中，有 18 個詞彙重複歸類，因此計量詞彙總數為 3,869。其次，再以各主題所得詞彙數，做為「相關」與「模糊」的應用強度，各主題詞彙計量結果顯示如下，並以圖三表示：

1. 「資訊檢索」主題詞彙，共計 2351 個，佔總詞彙數中的 60.7%。
2. 「書目計量」主題詞彙，共計 165 個，佔總詞彙數中的 4.2%。
3. 「相關」與「模糊」基礎理論主題詞彙，共計 611 個，佔總詞彙數中的 15.8%。
4. 其它應用學科的主題詞彙，共計 374 個，佔總詞彙數中的 10%。
5. 無法歸類之主題詞彙，共計 368 個，佔總詞彙數中的 9.5%。



圖三：應用強度之主題詞彙計量分析表

從 LISA 所獲得的「相關」與「模糊」理論的文獻計量結果顯示，「資訊檢索」確為二者發展的應用核心，佔了約 60%的強度。而「書目計量」領域，這二個理論應用較少。同時，在歸類過程中，我們發現，其它應用學科中，應用最多的是「醫學」領域，再則是「圖書資訊學」領域(「資訊檢索」及「書目計量」之外的主題)。

在文獻中，提及的「相關」與「模糊」的基礎理論有權重(Weighted)、關於(Aboutness)、關連(Relation)、模糊叢集理論 (Fuzzy set theory)、模糊邏輯(Fuzzy logic)、演算法(Algorithms)、模糊理由(Fuzzy reasoning)、模糊控制(Fuzzy control)、認知(Precognitive)、模糊關連(Fuzzy relation)等。從主題詞彙中看到這些基礎理論的概觀，或可做為深入分析「『相關』與『模糊』在資訊檢索應用發展」的應用方法與理論基礎。

而在「資訊檢索」領域中，我們發現「相關」與「模糊」理論在有些資訊檢索次領域的應用發展非常熱絡，如查尋、索引、資訊儲存與資訊檢索、資訊作業、技術服務、人工智慧、系統評估、專家系統、網際網路技術應用等等。

二、應用關係的解析

表五：共同主題詞計量分析表

主 題 焦 點	筆 數
Searching	4
Relevance	4
Fuzzy set theory	3
Technical services	3
Information storage and retrieval	3
Information work	3
Subject indexing	3
Boolean strategies	2

關於「相關理論」及「模糊理論」在資訊檢索領域中的應用關係，在文獻計量方法，希望藉由以「相關」及「模糊」理論為共同主題詞或為共同關鍵詞的文獻來進行主題詞彙分析，以找出「相關」與「模糊」在資訊檢索中應用發展的關係。研究程序分為三階段，第一階段為共同主題詞；第二階段為共同關鍵詞；第三階段為歸類分析。分析的資料來源皆為 LISA 資料庫，資料時間為 1969 年至 2005 年。

研究分析方法，採用共同主題詞分析、共同關鍵字分析與歸類分析等三階段，茲分述如下：

第一階段：共同主題詞

本階段所採用的資料蒐集方法是以查尋指令“DE=fuzzy and (DE=relevance or DE=relevant)”進行查尋，結果共獲得 4 筆，再將每一筆文獻書目資料中的 “Descriptors(DE)” 予以統計，排序每一個 DE 在 4 篇文章中所含的次數(必小於 4)。而此方法的基本設定為：LISA 的 Descriptors 為真。計量分析如表五。

Terms	2
Weighting	2
Ranking	1

(接上表)

(續下表)

Fuzzy logic	1
Feedback	1
User feedback	1
Modification	1
Online information retrieval	1
Computerised information retrieval	1
Computerized information storage and retrieval	1
Strategies	1
Text retrieval systems	1

由表五得知，4 篇文獻皆出現的主題詞彙，有 "Searching" 及 "Relevance"，依此我們可以看出，「相關」及「模糊」最主要的共同應用在「資訊查尋」上，而二者的根本理論基礎都是「相關」。而主要在資訊檢索領域中屬「資訊與技術服務」，多運用「主題索引法」、「布林邏輯」、「詞彙」及「權重」等方法來進行。

本階段所採用的資料蒐集方法是以查尋指令 "KW= fuzzy and (KW= relevance or KW= relevant)" 進行查尋，結果獲得 42 筆文獻，將每一筆文獻書目資料中的 "Descriptors(DE)" 予以統計，排序每一個 DE 在 42 篇文章中所出現的次數(必小於 42)。而此方法的基本設定為：LISA 的 Descriptors 為真。計量分析如表六：

第二階段：共同關鍵詞

表六：共同關鍵詞計量分析表(表列中排除出現次數為一次者)

主 題 焦 點	筆 數
Searching	21
Technical services	17
Subject indexing	17
Information work	16
Information storage and retrieval	16
Fuzzy set theory	14
Online information retrieval	11
Fuzzy logic	7
Terms	6
Computer applications	6
Relevance	6

Computerized information storage and retrieval	6
Weighting	5
Strategies	5

(續下表)

(接上表)

Mathematical models	5
Computerised information retrieval	5
Internet	4
Set theory	4
Models	4
Expert systems	4
Artificial intelligence	4
Retrieval performance measures	3
Medicine	3
Evaluation	3
Boolean strategies	3
Ranking	2
World Wide Web	2
User feedback	2
Tourism	2
Performance measures	2
Decision making	2

以關鍵詞計量所進行的分析，與第一階段主題詞計量所進行的分析，情況大致無二，其中較大的不同，因為詞量較多，明顯看出「線上資訊檢索」的應用熱絡，並逐漸出現與網際網路技術的整合；同時「專家系統」與「人工智慧」亦有發展。

第三階段：歸類分析

為避免單獨詞彙分析所可能造成的關係缺

失，本階段依據前二階段所獲取的文獻書目資料，清查所有的主題詞彙“Descriptors (DE)”，忽略字詞差異，如有同義異詞者，依實際主題內容進行歸類，重新排序。歸類方式仍以「資訊檢索」為核心，依表七區分為主類、相關類及理論基礎類，非資訊檢索則納入其它主題，各類之計量分析表如表八至表十一：

表七：歸類方式

類
資訊檢索主類
資訊檢索相關類
資訊檢索理論基礎類

其它主題類

表八：資訊檢索主類主題詞計量分析表

資 訊 檢 索 主 題	
Searching 查尋	21
Subject indexing 標題索引	17
Information storage and retrieval 資訊儲存與檢索	16
Online information retrieval 線上資訊檢索	11
Computerised information storage and retrieval 自動化資訊儲存與檢索	6
Terms 詞彙	6
Weighting 權重	5
Strategies 策略	5
Computerized information retrieval 自動化資訊檢索	5
Expert systems 專家系統	4
Artificial intelligence 人工智慧	4
Retrieval performance measures 檢索執行測量	3
Evaluation 評估	3
Boolean strategies 布林策略	3
Ranking 排序	2
User feedback 使用者回饋	2
Performance measures 執行測量	2
Decision making 抉擇	2
Text retrieval systems 全文檢索系統	1
Knowledge based systems 知識導向系統	2
Data processing 資料處置	1
User interface 使用者界面	1
Systems design 系統設計	1
Search strategies 查尋策略	1
Search engines 搜尋引擎	1
Binding 連結	1
Classes 類別	1
Computerized subject indexing 自動化標題索引	1
Clustering 過濾	1
Classification 分類	1
Dolphin Search 浮標式查尋	1
Forecasting 預測	1

(續下表)

(接上表)

FIRST (Fuzzy Information Retrieval System) 模糊資訊檢索系統	1
Feedback 回饋	1
Fuzzy systems 模糊系統	1
Intelligent systems 智慧系統	1
INSTRUCT 指示	1
Modification 修正	1
Output 輸出	1
Online catalogues 線上目錄	1
Precision 精確性	1
Probabilistic retrieval 概率檢索	1
Recall 回現率	1
Query formulation 問題表示	1
Retrieval languages 檢索語言	1
Relevance feedback 相關回饋	1
合 計	146

表九：資訊檢索相關類主題詞計量分析表

圖書資訊學 (圖書資訊學相關，但無法明確定義 入「資訊檢索類」者)		資 訊 科 學		網 際 網 路	
Technical services 技術服務	17	Information work 資訊工作	16	Internet 網路	4
Bibliometrics 書目計量	1	Computer applications 電腦應用	6	World Wide Web 網際網路	2
University libraries 大學圖書館	1	Computers 電腦科學	1	合 計	6
Saint Mary's University (Nova Scotia) Library	1	DBGuide 資料庫指南	1		
British Library Research and Development Department BLRD report 5651	1	Databases 資料庫	1		
Information services 資訊服務	1	Recycling 迴圈	1		
Library and Information Science Abstracts 圖書資訊科學摘要	1	合 計	26		
Periodicals 定期期刊	1				
合 計	24				

表十：資訊檢索理論基礎類主題詞計量分析表

相關及模糊理論基礎	
Fuzzy set theory 模糊集合理論	14
Fuzzy logic 模糊邏輯	7
Relevance 相關	6
Mathematical models 數理模式	5
Set theory 集合理論	4
Models 模式	4
Sets 集合	1
合 計	41

表十一：其它主題類主題詞計量分析表

其 它	
Computer assisted instruction 電腦輔助教學(教育)	1
Curricula 課程(教育)	1
Professional education 專業教育(教育)	1
Medicine 醫藥(醫學)	3
Radiation therapy 放射療法(醫學)	1
Waste products 廢棄物(醫學)	1
Genetic algorithms 遺傳演算(醫學)	1
Tourism 旅行	2
Threshold values 臨界值	1
Canada 加拿大	1
Ventura	1
Thesauri	1
Staff relations 人員關係	1
Conservation 保存	1
Design 設計	1
Electroplating 電鍍	1
Handwriting recognition 手跡識別	1
LexisNexis	1
Law 法律	1
Magnetic resonance imaging 磁波圖	1
Monitoring 顯示器	1
Preservation 維護	1

(續下表)

(接上表)

Push technology 推動技術	1
Robotics 機器人	1
合 計	27

透過主題歸類，我們可以更清楚地發現：

1. 從表八的資訊檢索主類中，可看出「相關」及「模糊」主要的應用仍是著重在查尋、檢索等資訊尋求方法上，而主要的運算元為「詞彙」及「標題索引」，運算模式為「權重」及「布林策略」，在系統上朝向「專家系統」及「人工智慧」進行發展。
2. 而表九的相關類別，說明了在圖書資訊學領域，「相關」及「模糊」應用主要屬技術服務；而在資訊科學中，其為資訊工作及電腦應用，另外，我們也發現網際網路的技術已逐漸被重視及整併應用。
3. 從表十中亦可發現，「相關」及「模糊」的內涵主要為「邏輯」、「集合」及「數理」等基礎。
4. 而除了上述之外，從表十一中可看到「相關」及「模糊」在「教育」及「醫學」上有較多的應用，教育主要是涉及電腦輔助教學；醫學主要是醫藥技術的開發。

三、文獻計量之綜合分析

經由文獻內容解析及主題詞計量分析的結果，發現「相關理論」及「模糊理論」，在概念內涵上可謂系出同源，哲學上的相似與相關及數理基礎上的邏輯判定使得「模糊理論」更深化了「相關理論」；而在應用發展上，由「模糊理論」基礎所發展出來的模糊技術，更成就了「相關理論」的操作型定義。

而此二種理論，在資訊檢索領域上的應用一直被視為必然，透過相關文獻的主題詞彙分析，也確實驗證了從 1965 年以來，資訊檢索領域即為「相關理論」與「模糊理論」的應用發展核心，而主要

的共同應用即為資訊檢索技術的研發與拓展。

「相關」與「模糊」從內涵到應用發展核心，看似密切又相關，二者在應用發展過程的關係，在本章節中亦透過文獻主題詞彙的計量分析，發現在資訊檢索領域中，二者的關係主要展現在資訊技術與服務上，而以二種理論為共同基礎而發展的「權重」、「布林」及「邏輯判定」等技術皆成為發展現今資訊檢索系統的主要應用。若從最原始的資訊檢索概念框架來看，「相關」是最早發展的基本概念之一，「模糊」則是包含了「相關」又整合了「相似」「邏輯」等其它概念的多元理論，而二者在實際的應用發展中，又結合了網路技術等發展了更多的資訊檢索系統。

在整體資訊檢索領域中，對資訊檢索系統而言，「相關」是隨處存在的現象，也是評估的指標，而「模糊」則是可處於系統運作各個階段的工具或方法，模擬系統運作的變數－使用者，用以提昇資訊檢索的效能，延伸系統輸出端的功能，減少使用者自行判定是否符合需求所需耗費的心力。由本研究中，我們充份發現，在資訊檢索領域的發展中，「相關」理論可說是系統建構的基本面，而「模糊」理論所發展的模糊技術，則是提昇系統功能與效益的加值面。二者的相互激盪結合發展中的資訊科技，如網際網路等，將觸發出更為符合使用者需求的現代與未來資訊檢索系統。而資訊檢索系統仍為未來資訊檢索領域發展的主要核心。

肆、相關與模糊在資訊檢索領域發展趨勢之預測

依據上述綜合解析的結果，嘗試針對此二種理論在資訊檢索領域的發展趨勢，進行推測：

一、在理論與方法上

- (一)從「相關」的二元判定，朝向多元價值判定，並以推論邏輯為基礎，運用集合理論，發展「模糊」。
- (二)從傳統資訊檢索系統的檢索過程的布林邏輯應用，延伸到資訊組織及資訊使用者的「布林策略」，再因應多元價值判定，發展「權重」方法，予以判定標準的權重值。
- (三)系統的運算元發展，從標題索引及詞彙的運算，逐步發展至全文內容(Text)及使用者問題與使用者回饋的運算。使文獻與使用者的關係在判定結果上呈現多重價值的模糊現象，但也促使文獻與使用者在多元面向中更精確地相關。
- (四)多元理論基礎的技術發展：相關與模糊最大的貢獻在於資訊檢索技術的發展，其影響在於其非單一理論的發展，而是多元理論的發展，未來，這樣的發展模式將更為明顯，目前許多的資訊檢索技術都已在這樣的趨勢下成長，如資訊過濾、資料探勘(資料採礦)等。

二、資訊檢索系統

在資訊檢索系統的發展上，技術與理論的交互激盪，將使下列資訊檢索系統相關議題更加蓬勃發展：

- (一)語意網：網際網路技術逐漸被重視與開發後，Mares 曾說：運用鄰近語意(Neighbourhood semantic)、模式不相關(Modelling irrelevance)、模式相關(Modelling relevance)及合法慣例(Valid formulae)與模式推論邏輯(Modelling inference)及多重價值判定等理論基礎，並整合了自然語言處理、系統網絡、網際網路及數理基礎與運

算，造就了語意網的發展。(Mares, 2004, pp.19-38)

- (二)專家系統與人工智慧：從「相關」的哲學基礎探討，觸發資訊檢索領域學者對使用者角色的重視，並逐步導引了使用者導向系統的發展，當科技愈發展技術愈純熟時，使用者不再純然是被服務對象，其慢慢成為資訊服務的主導者及主控者，並且融入資訊服務程序中，成為資訊服務的決策者、運作者及改良者。因此，專家系統及人工智慧是以使用者角色為建構核心的系統，其發展仍將持續。
- (三)系統模式：資訊建構者、資訊服務者及資訊使用者三者間的角色不再以階段式模式運作，彼此間的交流與交互影響成為系統的重要模式變化，使系統朝向下列三個方向發展：使用者導向、互動模式、主動服務機制。

三、領域影響因素

從資訊檢索領域的研究中，我們發現下列三個面向的交互變動，影響了「相關理論」及「模糊理論」在此領域中的各種應用發展：

- (一) 資訊需求面向 - 運算方法：涉及環節包括內在需求、需求呈現(表達型態與表達方式)、需求行為(尋找)、需求滿足(回饋)。
- (二) 資訊載體面向 - 運算元：資訊載體的多元格式，如：書面格式、電子格式、數位格式等。
- (三) 傳播模式面向 - 運算模式：分為「本體模式」的資訊組織及「擴散模式」的資訊傳播。

伍、結語

從「相關」及「模糊」理論的應用發展過程，我們發現在學術發展與實務創新銜接上的結構程

「相關」與「模糊」在資訊檢索領域中關係驗證與分析

序，從「概念」的發展到「理論」的創造，再研發出「技術」，這看似容易的演化，事實上是歷經每一階段多元概念、多元理論及多元技術的崩裂及融合，方得以成就下一階段的創新。雖然有時我們已發展了技術，卻未明顯看到「理論」的存在，但這並不代表其不存在，而是需要更多的解析與型塑，

讓概念讓理論彰顯，亦使其具體。而這樣的方法也提供了未來無論是學術發展或實務創新上交互激盪的新思維。

(收稿日期：2006 年 04 月 06 日)

參考書目：

- Bateman, J. (1998). Changes in relevance criteria: A longitudinal study. In American Society for Information Science (Ed.), Proceedings of the 61st annual meeting of the American Society for Information Science (pp.23-32). Medford, NJ: Information Today, Inc.
- Bookstein, A. (1981). A comparison of two systems of weighted Boolean retrieval. Journal of the American society for Information Science, July 1981, 275-279.
- Brooks, T. A. (1998). The semantic distance model of relevance assessment. In American Society for Information Science (Ed.), Proceedings of the 61st annual meeting of the American Society for Information Science (pp.33-44). Medford, NJ : Information Today, Inc.
- Buell, D. A. (1985). A problem in information retrieval with fuzzy sets. Journal of the American Society for Information Science, 36(6), 398-401.
- Cooper, W. S. (1971). A definition of relevance for information retrieval. Information Storage & Retrieval, 7, 19-37.
- Harter, S. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43(9), 602-615.
- Jacquette, D. (2002). A companion to philosophical logic. Malden, Mass.: Blackwell.
- MacCall, S. L. (1998). Relevance reliability in cyberspace: Toward measurement theory for internet information retrieval. In American Society for Information Science(Ed.), Proceedings of the 61st annual meeting of the American Society for Information Science (pp. 13-22), Medford, NJ: Information Today, Inc.
- Mares, D. (2004). Relevant logic: A philosophical interpretation. Cambridge: The Press Syndicate of the University of Cambridge.
- Mohammadian, M. (1999). Computational intelligence for modeling, control & automation: Evolutionary computation & fuzzy logic for intelligent control, knowledge acquisition & information retrieval. Netherlands: IOS Press.
- Petry, F. E., Buckles, B. P. & Prabhu, D. (1993). Fuzzy information retrieval using genetic algorithms and relevance feedback. In American Society for Information Science(Ed.), Proceedings of the 56th annual

- meeting of the American Society for Information Science (pp.122-125). Medford, NJ: Information Today, Inc.
- Saracevic, T. (1970). The concept of "Relevance" in information science: A historical review. In Introduction to Information Science. (pp.111-151). New York: Bowker.
- Saracevic, T. (1976). Relevance: A review of the literature and a framework for thinking on the Notion in information science. In Voigt, M.J. & Harris, M. H.(Ed.), Advances in librarianship, 6, (pp.81-139). New York: Academic Press.
- Saracevic, T. (1996). Modeling interaction in information retrieval (IR): A review and proposal. In American Society for Information Science (Ed.), Proceedings of the 59th annual meeting of the American Society for Information Science (pp.33-44). Medford, NJ: Information Today, Inc.
- Schamber, L., Eisenberg, M. B. & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. Information Processing & Management, 26(6),755-776.
- Singh, S. & Dey, L. (2005). A rough-fuzzy document grading system for customized text information retrieval. Information Processing and Management, 41, 195-216.
- Spink, A., Greisdorf, H. & Bateman, J. (1998). Examining different regions of relevance: From highly relevant to not relevant. In American Society for Information Science(Ed.), Proceedings of the 61st annual meeting of the American Society for Information Science (pp.3-12). Medford, NJ: Information Today, Inc.
- Tseng, Y. H. (1998). Solving vocabulary problems with interactive query expansion. Journal of Library & Information Science, 24(1), 1-18.
- 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(4), 327-355.
- Wilson, P. (1973). Situational relevance. Information Processing & Management, 9, 457-471.
- Wilson, P.(1978). Some fundamental concepts of information retrieval. Drexel Library Quarterly, 14(2), 10-24.
- 王玫 (1995)。現代情報檢索理論特性述略。 圖書與情報 , 4 , 7-9,16。
- 林信成、蕭勝文 (2003, September)。模糊詮釋資料及其在圖書館分類編目之應用。 教育資料與圖書館學 , 41(1) , 61-76。
- 曾元顯、林瑜一 (1998)。模糊搜尋、相關詞提示與相關詞回饋在OPAC系統中的成效評估。 中國圖書館學會會報 , 61 , 103-125。
- 黃慕萱 (1996)。資訊檢索之五大基本概念探討。 圖書與資訊學刊 , 19 , 7-21。
- 黃慕萱 (民 85)。 資訊檢索中「相關」概念之研究。臺北市：臺灣學生。