

資訊檢索之五大基本概念探討

Five Fundamental Concepts in Information Retrieval

黃基萱

Mu-hsuan Huang

國立台灣大學圖書館學系

Dept. of Library Science, National Taiwan University

【摘要 Abstract】

本文探討Patrick Wilson所提出之資訊檢索五大基本概念——「資訊」、「關於」、「相關」、「需求」和「用途」。因為無法找出公認一致的定義描述上述概念，很難完全解決其所產生的諸多不便與混亂，然而對上述概念的部分釐清，還是能使學者和從業人員站在較為一致的立足點討論有關資訊檢索的研究。

This article discusses the five fundamental concepts of information retrieval addressed by Patrick Wilson—information, about, relevance, need and use. Although it is impossible to reach consensus definitions for the above concepts, it can be quite helpful for future discussions in information retrieval researches.

關鍵詞 Keyword

資訊檢索、關於、相關、資訊需求、資訊使用

Information retrieval, About, Relevance, Information need, Information use

限實·原



壹、前言

在探討資訊檢索之基本概念之前，對資訊檢索領域之學科屬性及許多重要概念皆應有所認識。首先探討資訊檢索的學科屬性問題。McGrath曾在1978年發表一篇論及軟硬學科（soft/hard disciplines）的文章，文中指出在一些尚未形成研究典範（research paradigm）的學科中，很多名詞無法取得學者們一致公認的定義，而且這些名詞很可能永遠無法獲得其一致公認的定義①。在McGrath所列之學科軟硬排名表中（見表一），數學是最硬的學科，其次是化學，而圖書資訊學的硬度居中間偏上，比經濟學、管理學、教

表一 學科軟硬排列表

數學（最硬）
化學
物理
電腦科學
地理學
法文
天文學
圖書館學
公共行政
經濟學
管理學
歷史
音樂
演講
建築
舞蹈
教育
心理學
社會學
藝術（最軟）

育學都高；而表中硬度最低的三個學科則分別為心理學、社會學和藝術。事實上，部分人士可能會對表一中學科軟硬度的排名提出質疑，認為其在排列順序上有失客觀，圖書資訊學的硬度竟然

在眾多社會科學之上。然而，這畢竟已是十多年前的排名，且學科軟硬度的排名並非本文的重點，筆者只是想藉此說明圖書資訊學仍是一典範尚未建立的學科，因此很多溝通討論的名詞都缺乏其操作型定義，也無法形成較為一致的一般性定義。

事實上，在一些尚未建立典範的實用學科中，其從業人員往往不知其所以然地應用及討論很多實用的技能，Patrick Wilson對這種現象非常不能苟同，他認為這種似是而非的作法如果真的存在，也只能存在於一些較沒有價值（cheap）的學科中②。很不幸地，資訊科學或資訊檢索就是上述狀況之典型實例，此領域中很多名詞都無法獲得一致公認的定義，連最基本的概念——「資訊」都缺乏公眾認同的定義，更遑論其他名詞了。Wilson在其有關資訊檢索基本概念之名著中，論及資訊檢索中最為核心、最需公認一致定義的五個名詞，分別是「資訊」（information）、「關於」（about）、「相關」（relevance）、「需求」（need）和「用途」（use）③。這五大概念的定義問題至今尚未解決，使學者在討論上及從業人員在應用上產生諸多不便與混亂，一代宗師Wilson雖嘗試釐清上述五大概念，但大部分原已存在的問題仍然無法解決。本文擬根據Wilson所提出之資訊檢索五大基本概念逐一進行討論，雖然無法建立公認一致的操作型定義，但相信對釐清五大概念有其一定的貢獻。

貳、資訊

「資訊」在傳播學、資訊科學和圖書館學上都是尚待定義的核心概念，而這些學科也一直在期待此一基本且理論性概念的誕生。嚴格說起來，資訊只存在於人類的心靈之中，所有書籍文獻中的文字都只是記載心路歷程的符號，這些文字符號常被用來表達資訊，但絕非資訊本身，它們

充其量只能被視為心靈資訊之外在儲存體。至於資訊本身，雖然它無法被完整地描述、解釋和了解，但藉由觀察其特質和效果，卻可以增加人際溝通或人物溝通之效益④。事實上，人腦每天都自覺或不自覺地完成無數次資訊處理，這些資訊處理早已完全融入人類生活之中。比如說每天早上6點半聽到鬧鐘準時起床、出門是否需要帶把雨傘、在繁忙的交通中應該選擇那條路可以最早到達目的地、先吃午飯還是先去郵局寄信等，各種資訊與決策，在日常生活中可謂無所不在。

在人類溝通和傳播的歷史上，語言的發明是第一個里程碑，從此人類可以口耳相傳傳播資訊；文字的發明則是第二個里程碑，至此人們在傳播資訊時可以超越時間和空間上的限制。但由於資訊只存在於人類心靈之中，因此其特質應該是絕對主觀的，一旦將其轉化為語言或文字，如果沒有客觀的認知存在，就不可能完成溝通的使命，所以知識或資訊是主觀或客觀的爭議也一直存在。試想，如果資訊不具備客觀的特性，那人與人之間就不可能憑藉信件溝通，更不可能藉著閱讀書本文字產生共鳴。然而，資訊也絕對有其主觀性存在，否則同樣的文字就不可能在不同人身上產生截然不同的詮釋和感受。資訊這種主觀與客觀兼具的特性，尤其是其主觀性，不但使其更難獲得一致公認的定義，也會帶給研究者和從業人員無所適從的感覺。

一般而言，物理科學被歸屬於硬科學是不容置疑的，因此以其二個非常普遍的名詞——「重力」和「電」來比較其和「資訊」之間的不同。「重力」、「電」和「資訊」一樣，都無法以四肢觸摸或心靈感覺體驗其存在（接受電擊時或可感受電的特質，但一般並不存在此狀況），但「重力」和「電」都有具體的計算公式和放諸四海皆準的操作型定義，而「資訊」卻沒有估算公式，更沒有操作型定義。事實上，人文社會科學仍

在繼續致力於其典範的建立與操作型定義的產生（尤其是社會科學），但也有部分學者開始質疑這項努力的必要性，Donald B. Cleveland和Ana D. Cleveland即為其中二位。他們認為即使是在硬學科中，公認一致的操作型定義對了解大自然或心靈現象並非絕對必要，對軟學科則更非必要，因此資訊缺乏公認的定義不會造成太大的問題⑤。

如果從Wilson的角度來探討資訊概念，首先必須觀察承載資訊的載體——文件（document）。一般常說，文件包含（contain）資訊，而這句話可以從二個不同的角度來解釋：其一至少表示文件包含某一主題之資訊，其二更進一步地表示文章包含某一主題之正確資訊⑥。也就是說，資訊載體內可能記載各種不同主題之資訊，但無法保證載體上所記載資訊之正確性。事實上，如果用非常嚴格的態度討論「文件包含資訊」這句話，可以發現這句話有很重的語病，因為書籍或文件實際上只包含文字（texts）——成串的文字或符號，並沒有所謂的資訊存在。倘若更進一步探討文字和資訊間的關係，也很難發現此二者間存在任何絕對關係。換言之，完全相同的文字並不保證其包含完全相同的資訊，例如文中出現「太漂亮了」一詞，可以真的表示此人或此物的漂亮程度，也可能是「太好了」、「太棒了」、或是「就這麼辦」之意；反之，完全不同的文字也可能包含完全相同的資訊，否則就不可能進行任何翻譯工作。正因如此，如果想找出文件中承載的資訊，純粹利用文字的外觀來判斷並不可行，必須設法了解字裡行間及其背後隱含的意義，只有文意才能決定資訊內容，文字本身並無法達到此功能。因此，嘗試將「文件包含資訊」一語改為「文件擁有（have）資訊」，或是改為「文字包含意義」，都無法真正釐清資訊的定義⑦。



由於資訊僅存於人類心靈之中，因此個人心中所有信念（beliefs）的集合可以視為其個人資訊之代名詞，而這些個人資訊或信念，往往是正確與不正確雜陳，因此可以根據其正確與否，將個人心中之資訊或信念做更進一步之區分。一般而言，在資訊傳播的過程中，資訊接受者必須先了解資訊輸出者所欲輸出之資訊內容並相信之，才有可能真正將其轉化為資訊接受者心中之資訊。因此，根據資訊之正確性及其被接受與否，資訊輸出者和資訊接受者因資訊傳播而建立的關係可分為四種：當無法確定資訊輸出者的資訊是否正確時，資訊接受者可能僅止於了解資訊內容，並未更進一步接受資訊成為心中的信念（第一種關係）；當然，資訊接受者也可能自認為了解該資訊，而接受此不一定正確之資訊為心中信念（第二種關係）；而當資訊輸出者的資訊內容確定無誤時，資訊接受者同樣也會產生二個階段的反應，其有可能停留於了解的階段（第三種關係），也可能更進一步接受此正確資訊為心中的信念（第四種關係）^⑧。然而不管資訊本身的正確與否，資訊輸出者所描述的資訊都有被資訊接受者誤解的可能，如果考慮到這一層關係，那麼資訊輸出者和資訊接受者因為資訊傳播而產生的關係將更形複雜。

將資訊區分為「不知其真偽」和「確定其為真」二種不同類型的資訊，在資訊檢索上具有相當重要的意義。因為當所檢索到的資訊真假難分時，不應該稱之為「資訊檢索」（information retrieval），而應該正其名為「內容檢索」（content retrieval）^⑨。也就是說，目前資訊系統（不論其為文件檢索系統或是事實檢索系統）所提供之資訊檢索事實上皆屬內容檢索，因此檢索所獲通常是和某一主題有關之所有資訊，包含正確的資訊及誤訊（misinformation）。但若站在讀者的立場來看，假設讀者想知道唐朝建立之

西元年代（正確的答案應為西元618年），而檢索系統所得到的年代卻為西元589年或是西元630年，讀者的確檢索到一些資訊（應該說是誤訊），但這些不正確的資訊可能比沒有檢索到任何資訊所帶來的傷害更大。因此，一個負責任的資訊系統應該提供的是資訊檢索而非內容檢索，亦即應提供資訊而非誤訊。事實上，這是一個相當嚴肅且影響深遠的問題，例如在評估參考服務時，參考館員正確回答問題的比率通常只有55%^⑩，這固然和館員的素質及訓練有關，但文件中常包含誤訊也是造成此一現象的主因。

Wilson將目前的資訊檢索定位為內容檢索，同時指出資訊檢索應達到的目標為：提供正確的資訊供大眾使用。儘管資訊經常真偽難辨，且過去為正確之資訊在今日不一定為正確，而過去不正確之資訊今日可能變成正確，然而資訊系統還是必須考慮資訊的正確性，惟有提供正確的資訊，才能讓讀者做出正確而無悔的決策。

參、關於

「關於」是一個鮮為國人探討的資訊檢索之基本概念。「關於」在分類編目或索引摘要上都有其舉足輕重的地位，因為不管是分編或索摘，其目的都在於設法描述文件內容，而描述文件內容可視為說明此文件是關於某主題最常使用的手段之一。嚴格說起來，「關於」是探討文字和真實世界間的關係^⑪。舉例來說，如果某文件是在描寫奧地利之維也納市，那這篇文章就是「關於」維也納。但文件涵蓋的主題通常不只一個，一篇文章有可能「關於」數個主題甚至數十個主題，例如孫中山先生的傳記，雖然提及許多其他人物和地名，但其主要主題應是「關於」孫中山。然而有些「關於」數個主題的文件，其主要主題卻不是這麼明顯，此現象在分類上通常會造成一些困擾，因為一份文件只能給予一個分類號，但



在標示主題標目或敘述語時，由於一份文件可以多個敘述語索引之，所以比較不會構成問題。

一般而言，「關於」可以視為找出文件主題的處理動作，因此研究「關於」的學者，經常將研究重點擺在應該或如何找出該文件所「關於」之主題。不可否認地，圖書資訊學者也希望能找出許多名詞的操作型定義，就如同數學家和物理學家一般，學者們可以站在公認的定義上進行討論，同時，任何人只要根據操作型定義操作，都可以得到相同的結果。因此，「關於」的操作型定義之所以重要，除其有助於了解「關於」之本質外，更是確保索引一致性的最有效方法，因為索引者與檢索者若能根據操作型定義選擇詞彙，應該會得到完全一致的結果，如此一來則存在已久的索引一致性問題即可迎刃而解。Maron就曾致力於「關於」操作型定義之研究，他認為文件「關於」某一主題（假設此主題以詞彙 i 表示），係指當檢索者在尋找該文件時，會以詞彙 i 作為檢索詞彙來找尋資料^⑫。

事實上，Maron的定義和真正的操作型定義間還有一段相當遙遠的距離，與其稱之為「操作型定義」，不如將其正名為「規則」。具體而言，Maron的規則僅指出「關於」之處理過程，但其他人在進行相同的處理時，卻未必見得會選擇相同的詞彙，因此無法成為放諸四海皆準的操作型定義。一般而言，在使用 Maron 之規則決定文章「關於」之主題時，並非經由思索此篇文章涵蓋之主題，而是思索當讀者在找尋此類資訊時，他們會使用那些詞彙進行檢索？如果確定大多數讀者都會以詞彙 i 來查詢此一文件，那這篇文章就應該以詞彙 i 作為其索引詞彙^⑬。換言之，Maron的規則是從使用者導向之觀點來看「關於」的定義，頗為合乎目前使用者導向的系統設理念。

Wilson批評 Maron 的規則無法區別主題索引

和非主題索引，所以不能將其專用於主題索引上^⑭。但從資訊組織的觀點看來，這應該是優點而非缺點，因為文件不僅可依其主題進行分編和索摘，也可以根據其他非主題層面進行分編和索摘，例如以文件之研究方法組織資訊即為可行的方法之一。以上述孫中山先生的傳記為例，如果以主題進行索引，那索引詞彙自然非「孫中山」莫屬，但坊間存在的孫中山先生傳記相當多，每本傳記可能各有其特色，對讀者而言，以這些特色進行索引，再配合原先即已存在的主題索引，所提供的檢索效益應該大為增加。因此同為孫中山先生之傳記，倘若該本傳記的特色是以非常新穎之傳記調查方法撰寫完成，那麼此種傳記調查方法應該是非主題索引上非常好的選擇之一。

「關於」和整理組織資訊的方法間有不可分離之關係。事實上，目前一直有一股浪潮在質疑圖書館員整理組織資訊之方式。以線上檢索為例，檢索所得的書目，讀者之使用率只有25%左右^⑮；而不做任何分編或索摘所得到的檢索結果，居然和經過分編索摘所得到的結果不相上下^⑯。上述警訊在在提醒圖書館界重新思索其整理組織資訊的方法。一般而言，如果傳統的分編和索摘系統成功地達成整理組織資訊的使命，那檢索所得書目的利用率應該會大為提高，也不可能發生自動索引和人工索引檢索效率相似的情況。雖說此種呼籲在圖書館界的迴響十分微弱，大部分從業人員都安於過去組織整理資訊的方式而不願尋求改變，然而值此圖書資訊科學正面臨轉型的關鍵時刻，學界與業界絕對得認真思索此一問題，否則很可能面臨專業領域遭到蠶食鯨吞的厄運。正因如此，Wilson提出其心目中理想之整理組織資訊的方式，雖然此種方法仍具有相當大的爭議，但不失為拋磚引玉的作法。

Wilson整理組織資訊的方法架構在資訊本身的結構上，他認為任何文件內容均有其規律的結

構，而資訊規律的結構往往是該文件言談結構的外顯^⑪。舉例而言，任何人在撰寫文章時，一定有其中心思想，文章的內容應依照一定的順序環繞該中心思想展開論述。事實上，一篇文章就像一棵大樹一樣，大樹的樹幹就好比文章的中心思想，枝幹和樹葉就像文章內容一樣環繞著樹幹生長，通常愈靠近樹幹的部分愈接近中心思想，因此整篇文章就如同大樹一樣有其一定的架構。正因為文章（或文件）結構有其邏輯上的順序，因此Wilson認為可以根據文章內容之言談結構來組織整理資訊，這些可用來組織資訊的層面包括該文件之歷史背景、問題陳述、解決問題的方法及其遭遇之難題等^⑫。當然，不同作者可能會對文件結構有其不同的詮釋，比如說可以在上述四層面中加上研究方法及研究結果等。筆者以為，利用文件本身的結構來整理組織資訊是一個可行的方向，但至於何種結構最為自然且最有效率，還有待無數實證研究繼續努力。

不管是Maron或Wilson所提出之「關於」，其用來描述文件之詞彙通常是主題標目或敘述語。但在電腦化資訊檢索已極其普遍的今日，題名及摘要上很多自然語言都可以用來描述文件所「關於」之主題。除此之外，文章中之目次表、前言和結論都蘊含相當豐富之主題資訊，在思考一個新的組織整理資訊的方式時，這些項目都應該列入考慮。

肆、相關

一個名詞的定義通常反映出學者對此現象之具體研究成果，因此從相關一詞在定義上的變化，可以很明顯地看出相關概念之成長。一般而言，早期的相關多屬主題相關，爾後慢慢擴展至其他非主題相關，因此本文擬根據1970年以後一些最具代表性的相關定義來探討相關，希望藉此說明相關概念的發展。這些較具知名度的定義，分

別是Cooper在1971年所提出之「邏輯相關」（logical relevance）^⑬，Wilson在1973年所提出之「情境相關」（situational relevance）^⑭及Harter在1992年所提出之心理相關（psychological relevance）^⑮。從這些論述中，不僅可看出相關概念發展的歷史及方向，還可以了解大師們論述的方法及其精闢的見解。

一、主題相關

在系統觀點盛行的年代，所討論的相關多為主題相關，因此相關通常被定義為檢索詞彙和描述文章詞彙間的一種吻合關係。顧名思義，主題相關單純就主題的觀點探討相關，其背後的假設是主題相關的資訊應能滿足檢索者之資訊需求，因此它是一種客觀相關，也就是說，張三和李四輸入相同的檢索問題，他們就會得到完全相同的輸出資訊，這種純以主題決定相關的作法，仍是目前資訊檢索系統的主流。但由於相關判斷是相當複雜的決策過程，其判斷過程絕對不可能由主題單獨決定，因此主題相關也不斷遭受質疑，它之所以能在目前資訊系統上大行其道，只因為它是目前人們惟一有能力處理的相關層次之實際應用。

一般而言，最常使用的主題相關定義是Cuadra和Katter在1967年所提出之定義，這是一個影響相當深遠並廣為使用的定義，其逐字指辭如下：「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。」^⑯將之對應到目前的資訊檢索系統中，資訊條件敘述是以檢索詞彙表示，而文章內容則是以描述文章之索引詞彙表示，可見此定義和早期之相關定義毫無差異。附帶一提，Saracevic曾對相關定義做過整理，這些定義大多為1970年代以前的定義，其中所有定義都沒有超越主題相關的範圍，現將其列舉如下^⑰：



- 文件和問題之間的一致性，可用以測量文件提供資訊的程度。
- 文件根據問題提供解答，而此解答之優劣或其滿足資訊需求的程度。
- 文件和問題間之關係程度，此關係包含相關程度、重複程度和適合程度等。
- 答案對問題的適合性。
- 文件對問題的適合程度。
- 文件和資訊需求（問題）間連結程度之測量值。
- 測量答案及其有用程度之測量值。
- 令人滿意的答案。

二、邏輯相關

邏輯相關是由William Cooper所提出，他認為邏輯相關能正確推論是否型問題（yes-or-no question）之相關性，但對於其他類型的問題（包括事實檢索或書目檢索等），邏輯相關可能無法推論出完全正確的答案，不過它對相關概念在理論層次上的提升卻有相當顯著的功用。Cooper首先批評Cuadra和Katter對相關概念所下的定義，他認為「一致性」（correspondence）和「適合」（appropriate）都是非常模糊的名詞，對定義相關或釐清相關的爭議不可能有很大的貢獻²⁶。事實上，上述Saracevic所整理出之相關概念也都停滯於模糊抽象的層次，距離相關之操作型定義尚有一段遙遠的距離。因此，Cooper從文章中部分資訊即能滿足資訊需求的角度探討相關，逐漸發展出其邏輯相關的概念，企圖提高相關概念化的層次。

如果說一篇文章中的部分資訊能滿足讀者之資訊需求，則此文章應被視為相關文章，因此，假設一篇文章中有一個句子能滿足讀者之資訊需求，則這篇文章應被歸類為相關文章。在此假設下，Cooper將相關的限制型定義以下列三個限制條件加以敘述：

1. 檢索問題必須為是否型問題，

2. 儲存資料之敘述方式必須為正式語言中之句子（sentence），3. 資訊系統必須具有推理的功能²⁷。也就是說，當檢索問題的答案（在是否型問題中只有是與否）可由儲存之句子中推論得知時，這篇文章就是相關文章²⁸。舉例來說，A成分可以用來治療老人痴呆症，B食物中含有A成分，所以B食物應可治療老人痴呆症。將這個問題以Cooper的三個限制條件表示，首先，檢索問題應被敘述為「B食物是否可治療老人痴呆症」（滿足條件一之是否型問題），由儲存的資料配合推理的功能，可以得知這個問題的答案應為「是」（正確）。

以限制型相關為基礎，Cooper將其進一步擴展為一般性相關定義，其逐字措辭為：「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）。」²⁹所謂「最小前提組」，其在哲學上的定義為「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，就無法以邏輯推論得到所需結論」³⁰。因此，根據一般性邏輯相關之定義，只要文件中包含構成資訊需求之最小前提組，此篇文章即被判定為相關文章。筆者以為，邏輯相關是一種主題相關，但其判斷主題相關的方法，並不是將代表資訊需求的詞彙和代表文章內容的詞彙互相配對比較，而是先尋找該文件中是否包含能滿足資訊需求之最小前提組。所以，對任何一個檢索問題，只要能夠找出滿足其資訊需求或是提供解答之最小前提組，相關判斷的難題即可迎刃而解。但由於可以滿足同一資訊需求之最小前提組本身即已包羅萬象，再加上邏輯相關推理的本質，找出最小前提組所有不同的組合及表達方法，將成為資訊檢索上相當高難度的一項挑戰。

Cooper是效用派大師，居然反其道地為相關派撰述文章。Cooper自己提出的解釋有二：一為

相關在資訊檢索的重要性是不容置疑的，二為邏輯相關是目前系統設計者惟一有能力處理之效用因素²⁰。也就是說，在Cooper的眼中，邏輯相關所決定的不僅是文章的相關與否，同時也是決定效用最重要且最容易測量的因素。

三、情境相關

情境相關是由Patrick Wilson所提出，此相關是以Cooper之邏輯相關為基礎，加上歸納邏輯（inductive logic）所推得之證據相關（evidential relevance），再考慮讀者個人之知識狀態及其關心（concerns）的重點衍生而成²¹，也就是說，情境相關的本質是一種邏輯相關，但其不同於邏輯相關，因為邏輯相關屬於客觀相關，所以張三或李四只要檢索相同的主題，他們會得到相同的檢索結果；而情境相關由於考慮到個人的認知狀態、興趣和喜好（preference），因此張三和李四即使檢索相同的主題，他們很有可能得到完全不同的輸出結果。

Cooper的邏輯相關是從演繹法的觀點來看相關，他認為當一篇文章包含構成答案或是可推論出答案之最小前提組時，這篇文章即被判定為邏輯相關，但Wilson認為單從演繹（推論）的觀點來探討相關是不夠的，必須兼由歸納的角度來看相關，也就是說，當文章中之資訊能強化某一前提、假設或概念時，這篇文章也應被視為證據相關之文章²²。事實上，證據相關在資訊檢索上相當重要，因為能強化讀者假設或推論的文章，對讀者而言，應該都是有價值的文章，所以用證據相關來補充邏輯相關的不足，對相關的定義是一種正面的突破。

一般而言，情境相關所指的情境是資訊需求者所看到的情境，而非其他任何人由任意觀點所看到的情境，因此判定情境相關的先決條件，必須先了解並描述資訊需求者個人所處之情境（不管此情境是否為資訊需求者本人所關心）。基本

上，資訊需求者所提出的問題應該是其關心的問題（但此關心的動力有可能是被動的），但決定相關與否的情境卻不一定為其所關心之情境。從定義上來說，直接情境相關（directly relevant situationally）是指相關情境為關心之情境時；而間接情境相關（indirectly relevant situationally）則是指相關情境為非關心情境的狀態²³。舉例而言，張三想知道他還有多少存款，則其存款數目為直接情境相關之資訊；但張三借了1,500元給李四，此事實與張三有多少錢則為間接情境相關。

由於情境相關考慮到個人知識狀態，而個人知識狀態不斷地變化，因此昨日相關的資訊今日不見得相關，而昨日不相關的資訊今日卻可能成為相關資訊。此外，Wilson尚提出顯著資訊（significant information）的概念，其定義為能改變個人知識狀態或認知狀態的資訊²⁴。他認為檢索者在進行線上檢索時，通常並不預期找到所有的相關資訊，新資訊和能夠改變知識狀態的資訊應該更受歡迎。當然，顯著資訊可能為直接情境相關，也可能是間接情境相關。

Wilson的情境相關對相關的研究貢獻甚鉅，他不但提出證據相關以補邏輯相關之不足，還將相關的範圍延伸至考慮個人的知識狀態，這些想法和今日資訊系統設計的理念不謀而合。一般而言，以情境相關作為系統判斷相關的標準，其所面臨的最大難題在描述個人的知識狀態及文字與文字間之推論與歸納關係，而這個困難似乎不可能在短時間內克服與解決。因此，在認定系統必須考慮檢索者個人之知識狀態，同時提供邏輯相關且證據相關之顯著資訊時，研究者對於認知心理學、學習理論、及人類思考等，不得不有較為深入的研究與了解。

四、心理相關

心理相關為Harter在1992年所提出，其概念

源自日常生活中對話交談的靈感，其中心思想是由Sperber和Wilson所提出^⑩。事實上，不管是中文或英文，一般對話中提及相關時都不是指「關於某主題」(on the topic)，反倒是一些互相關聯的主題、加強或減弱個人認知的信念、或是從另外的角度來看事情之觀點比較容易被習稱為相關。也就是說，設計資訊檢索系統時所使用的「關於」(about)，並不是一般人對相關的定義，而是遷就早期以詞彙代表資訊需求及文件主題之相關定義。Harter認為將相關的概念局限在「關於」(主題相關)的層次，對相關概念的發展是致命的傷害，並因而打擊資訊科學理論的發展^⑪。因此，如何從認知或心理的角度來了解相關，尤其是從認知狀態或知識狀態的改變來看相關，已經成為相關概念發展的新方向。

在Sperber和Wilson的心目中，個人的認知環境或知識狀態係指其在某一特定時間內所能明白 (manifest) 的事實 (facts) 和假設 (assumptions)，在言談進行的過程中，個人的認知狀態 (或知識狀態) 會隨著談話內容不斷地改變。一般而言，在對話或交談時，談話者通常會選擇他認為相關值最大之關聯文字 (context) 作為話題，因此對各別交談者而言，所謂相關就是指產生最大文字關聯的情境或是需要最少資訊處理的情境。Harter將Sperber和Wilson的理論應用在圖書資訊學對相關的定義上，他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，換句話說，就是能產生文字關聯效果 (contextual effect) 之資訊。據此，Harter將主題相關定義為潛在相關 (weak relevance)，因為潛在相關之資訊是最有可能讓讀者產生文字關聯效果 (或改變知識狀態) 之資訊，也是目前資訊系統惟一能夠處理的相關定義^⑫。

事實上，從心理學及認知的角度來看相關，就等於承認相關本身變化的特質，因此和Wilson

之情境相關有異曲同工之效。不過，心理相關的涵蓋範圍更廣，在此概念之下，不論是資訊需求、資訊檢索、書目計量學、及相關本身，都被心理相關緊密地扣成一環。在Harter的心目中，資訊需求是一種不斷變化的知識狀態，所以研究的重點應該擺在目前的資訊需求，而不是過去或原始之資訊需求^⑬。同理，由於資訊需求隨時在變化，因此必須重視資訊檢索的過程。Harter甚至以為資訊即為相關，因為資訊檢索的目的在找尋相關資訊，所以資訊本身自然包含相關的意義。此外，Harter尚提出書目計量學和資訊檢索的關聯性，如果將相關資訊視為改變知識狀態或產生文字關聯效果之資訊，引用文獻正是相關資訊最好的來源之一，因為一般作者所引用之書目，必為改變其知識狀態或是產生文字關聯之書目；同時，由於引用書目並不局限於同一主題，因此從引用文獻的角度探討相關，不但可以突破傳統主題相關的限制，也可以讓書目計量學和資訊檢索之概念合流^⑭。

伍、需求

本文所謂之需求係指資訊需求。一般而言，張三擁有資訊需求意謂張三需要一些資訊來滿足其需求。事實上，在資訊科學或資訊檢索中，需求是一極須定義的名詞，它和許多類似的名詞經常被混為一談，而這些名詞雖然有其實質上的差異，卻經常被替代使用，如「需求」(need)、「想要」(want)、和「缺乏」(lack) 等。Derr在其討論資訊需求的名著中指出：必須擁有資訊目的 (information purpose) 才可能產生資訊需求^⑮。他並嘗試釐清許多類似概念和需求之間的不同。下文試以一簡單例子說明「需求」和「想要」之間的異同。假設張三年過四十，他對體檢的資訊應有「需求」，不管他是否「想要」做健康檢查；同樣地，李四「想要」知道隔壁張

太太的花邊新聞，但他實在不「需要」此類資訊。換句話說，人們可能想要他所不需要的資訊，也可能需要他所不想要的資訊。至於「缺乏」，則和「想要」一樣，與需求之間並不存在絕對的關係。例如張三可能知道很多有關李四的糗事（表示他不缺乏此種資訊），但他還是想知道更多；當然，張三也有可能完全不知道李四的糗事（表示他缺乏此種資訊），但他卻一點也不想知道任何有關李四的事件。綜上所述，缺乏並不一定要導致需求，而在不缺乏的狀況下，需求同樣可能產生。

Wilson對資訊需求也有其獨到的見解，他認為一定要有目的（goal），才可能產生需求，這和Derr所提出之「須具有資訊目的才可能產生資訊需求」的說法，實有異曲同工之妙^⑩。舉例來說，張三需要錢買一台冷氣機，意謂張三如果沒有足夠的錢，他就無法達成購買一台冷氣機的目的，因此滿足張三需求的最簡單方式就是張三用錢去買一台冷氣機。Wilson特將此種方式稱為積極滿足需求的方式。但如果張三沒有足夠的錢，是否象徵著他沒有辦法滿足其需求呢？由於需求和目的息息相關，因此張三可以改用其他方法或降低其目的層次來滿足其需求，例如去租一台冷氣機使用（用其他方式達成同樣的目的），或是改買一台電風扇（更改或降低目的層次，使其易於達成）。在Wilson看來，這些則是屬於消極達成目的的方法。換句話說，可以藉由某種途徑直接且完全滿足的需求屬於積極的需求，而必須改用其他方法或是設法降低目的層次才能達成的需求則稱為消極的需求^⑪。在上述定義下，圖書資訊學中消極的資訊需求，其數量遠較積極的資訊需求為大。

進一步闡述消極之資訊需求在資訊檢索上所衍生的意義。由於同一資訊目的通常可透過各種不同方法分別達成，比如李四想使用計算機求數

學四則問題之解，除非此四則問題過於複雜，否則當計算機不在手邊的時候，他通常會以手算解決問題，儘管計算機具有增加效率和確保品質等優點。資訊需求亦然，由於資訊需求大多屬於消極性質，再加上資訊和資訊之間又具有替代性，因此只有在少數情境下，才會發生非某資訊無法滿足其需求的狀況，因之造成資訊科學的不確定性，間接影響到資訊從業人員的專業性。

事實上，人類的需求相當多，資訊需求只是其中的一種，通常經由提供資訊而達到滿足的需求皆可稱之為資訊需求。不過，資訊需求的滿足感相當主觀，即使檢索者已經提供大量相關資訊，讀者還是有可能覺得不滿足；反之，檢索者僅提供一、二篇相關文章，對某些讀者而言卻已十分滿足。尤有甚者，提供誤訊照樣能滿足某些讀者的資訊需求，特別是在讀者將誤訊信以為真時。^⑫Cooper一直主張以讀者對系統的滿意程度來評估系統^⑬，卻無法獲得多數學者的認同，其原因不外乎資訊需求的滿足感過於主觀及無法排除誤訊所帶來的傷害。假設讀者對一經常提供誤訊的系統高度滿意，此系統還是不夠資格被稱為良好之資訊檢索系統。

在有關資訊需求的研究中，學者不斷地呼籲必須找出讀者原始、真正的資訊需求，因此找出讀者真正之資訊需求已成為參考唔談最主要的目的之一。不過，由於資訊需求不斷變化的動態特質，回到原始資訊需求的理論也逐漸受到挑戰。Kuhlthau的研究證實檢索問題的文字性敘述，可以有效地反應出檢索者在某一特定時間內思考的重點及思路的變化，充分顯示出資訊需求不斷變化的特色^⑭。以撰寫報告為例，作者最初報告的構想和最後報告的成果往往有些差距，有時甚至連題目都可能完全更換，因此對讀者而言，回到原始的資訊需求並沒有太大意義，重要的反在於掌握目前的資訊需求。Harter在其心理相關一



文中更明白指出，資訊需求可以修正為讀者目前對某一特定問題的認知狀態，這種狀態基本上是動態的，因此研究的重點不應擺在原始或最初的資訊需求，所須了解掌握的應該是目前的知識狀態⁴⁴。換句話說，在整個研究過程中，資訊需求是動態的、複雜的且不斷成長的，因此，從任何時間點切入，都會產生一資訊需求，代表讀者在此一剎那對某問題之認知狀態。雖說可藉由了解讀者目前的知識狀態解決資訊需求不斷變化所產生的問題，而圖書資訊從業人員也可以認定自己的目標為滿足讀者目前的資訊需求，但在資訊科學的研究上，絕不可就此劃地自限。在書目控制或資訊的組織整理上，如果對資訊需求沒有更深入的研究，就無法決定應該以何階段之知識狀態為基礎進行分編或索摘，也不能確定當原始的資訊需求在資訊檢索上已不具意義時，是否要反過頭來追求最後的資訊需求（但所謂「最後的資訊需求」，究竟是取開始撰寫報告時之需求，亦或是完成報告時之需求，仍是一亟待斟酌之問題）。很多時候，設法釐清讀者問題和推向較後階段的資訊需求之間很難區分，但在沒有大量研究投入有關資訊需求的研究前，這些說法都只是一些缺乏證據的推測。

陸、用途

文件中所蘊含的資訊通常具有多種用途，Wilson將其分為主要用途（primary use）和進一步用途（further use）二大類。一般而言，描述文件之主要資訊內容即為說明該文件之主要用途的手段⁴⁵，但主要用途絕非該文件唯一的用途，甚至可能也不是該文件最顯著的用途。而進一步用途通常可以包括該文件所能應用之計畫類型、該文件所能加速之決策、該文件所能支持之論點、及該文件所能證實之預測等⁴⁶。事實上，當前整理組織資訊的方法只考慮到該文件的主要

用途，如果能以文件的進一步用途進行資訊之組織整理，應該是可行性相當高之組織整理資訊的新方法。也就是說，理想的整理組織資訊的方式絕不能單從主題分析著手，考慮的層面絕對要超越主題層面，才有可能真正提高讀者對檢索的滿意程度。目前坊間絕大多數的資訊檢索系統都僅提供主題檢索，所作的資訊分析也僅止於主題分析，這可能是檢索所得書目引用率偏低的主要原因之一。舉例而言，某讀者想蒐集資訊證明其論點無誤，或是尋求他人的數據來證明自己的研究假設，若由系統依主題（主要用途）來判斷相關，找到的相關文獻可能會相當多，但由於讀者是以能否支持其論點或假設（非主要用途）為判斷相關的標準，因此真正判斷成相關的文獻在數量上可能會相當少，甚至其他主題的類似論點或數據都較主題相關之資訊來得實用。必須說明的是，此處將用途和相關混為一談，乃因目前相關派已充分吸收許多效用派（用途派）學說之精華，且效用派精神多已全然融入相關派學說所致。

如果更詳細說明文件的用途，在最極端的情況下，書本可以當做鎖匙、禦寒或擋雨的工具，但這些都不屬於本文討論的範圍。一般而言，即使在探討文件的主要用途時，仍然必須對資訊和誤訊作出區分。也就是說，如果資訊檢索系統不舌資訊之真偽，只是蒐集有關此問題或主題之訊息（包含資訊及誤訊），則充其量只能稱之為內容檢索系統，而不能稱為資訊檢索系統。在內容檢索系統中，讀者可以找到有關某主題之訊息，更具體的說，應該是某作者對某主題的看法或意見（不嘗試去分辨其真偽）。換言之，目前的資訊檢索系統不但只能根據主要用途（主題上的用途）提供資訊，且其所提供之資訊真偽不分，因此只能稱之為內容檢索系統，無法提供正確且超越主題之相關（或有用）資訊。

在資訊檢索上，對於檢索應該提供相關的（

relevant) 資訊或是有用的 (useful) 資訊，在過去一直是個爭論的重點。認為應該提供相關資訊的學者被稱為相關派，而主張提供有用資訊的學者則被稱為效用派 (utility)。由於相關的資訊並不一定有用，而有用的資訊也不一定相關，因此這二個學派曾經各自獨霸一方。但由於相關派不斷地成長，所考慮之觀點逐漸增多，再加上其不斷吸收效用派的理論，相關派顯然已成為系統評估的主流。事實上，如果完全就用途的觀點來看，理想的線上檢索系統應該檢索出讀者最後使用的書目，而該書目可能是根據主要用途而來，也可能根據進一步用途而來。上述文字雖然是利用效用派的觀點陳述用途的重要性，但由於「用途」早已融入「相關」之中（例如Harter之心理相關即完全包含這些多元化的進一步用途），因此也等於說明相關的重要性。

七、結語

本文根據Wilson所提出之資訊檢索五大基本概念—「資訊」、「關於」、「相關」、「需求」和「用途」，逐一進行深入的探討。事實上，資訊科學的進步一日萬里，再加上從事研究的人逐漸增多，對重要概念或基本概念的認知，各家學者可能會有不同的看法。一般而言，將「資訊」、「相關」和「需求」列為主要概念的爭議可能較少，但「關於」和「用途」則可能被視為較為式微的概念。筆者之所以仍將這二個概念列入討論，是想經由「關於」概念闡述整理組織資訊的重要性，並藉「用途」概念再度思索相關概念的意義，以求對資訊檢索之基本概念進行更深入且完整地探討。

在Wilson所提出之資訊檢索五大基本概念中，若根據Harter所提出之「資訊即相關」的看法，此五大概念實以相關為核心互相貫串。事實上，Harter之所以會提出資訊即相關的說法，是因為資訊檢索目的是為檢索相關資訊，所以資訊本身自然包含相關的意義在內^⑦。再者，相關資訊是為滿足讀者之資訊「需求」，而整理組織資訊的重點則在探討文件所「關於」之主題，也就是說，系統設計者必須蒐集檢索者之資訊需求及文件所關於之主題，使用者才有可能檢索出其所需之相關文獻，所以需求和關於二概念已與相關概念互相结合。至於用途概念，雖說相關派學者認為系統應提供相關的資訊，而效用派學者認為系統應提供有用的 (useful) 資訊，但由於效用派的理論已逐漸融入相關派中，因此用途可視為相關的類同義詞。換言之，此五大概念以相關為核心環環相扣，彼此產生密不可分的關係。

總之，理想的資訊檢索系統應該提供資訊而非訊息（此處之訊息指資訊和誤訊難陳之意），就如同一般人選擇諮詢對象一樣，通常是根據其知識豐富與否，不太可能因為此人舌燦蓮花而決定委任於他^⑧。同時，理想的資訊系統不應只根據主題來分析資訊，必須根據進一步用途更深一層地組織整理資訊，例如分析文件對解決某類型問題的貢獻，或是文件所能證明之假設，或是文件所支持之論點等。也就是說，資訊檢索系統不能再停留於內容檢索的層次，必須檢索出所有有用（包含主要用途和進一步用途）的資訊（不包含誤訊），才能真正滿足讀者的需求^⑨。

（收稿日期：1996年10月12日）

^⑦ 這一部分其實文句混用其意，用資訊即「關於」之意，其實應該是「相關」之意，因為「關於」是狹隘之謂，「相關」才是廣泛的。

^⑧ 這一部分其實文句混用其意，用資訊即「關於」之意，其實應該是「相關」之意，因為「關於」是狹隘之謂，「相關」才是廣泛的。

註 釋

- 註①：McGrath William, "Relationships between Hard/Soft, Pure/Applied, and Life/Nonlife Disciplines and Subject Book Use in a University Library," Information Processing and Management, 14 (1978), p.22.
- 註②：Patrick Wilson, "Some Fundamental Concepts of Information Retrieval," Drexel Library Quarterly, (1979), pp.10-24.
- 註③：同上註，pp.17-28.
- 註④：Donald B. Cleveland and Ana D. Cleveland, Introduction to Indexing and Abstracting, 2nd ed. (Englewood: Libraries Unlimited, 1990), p.9.
- 註⑤：同上註。
- 註⑥：同註②，p.10.
- 註⑦：同註②，p.11.
- 註⑧：同註②，pp.11-12.
- 註⑨：同註②，pp.12-13.
- 註⑩：F. W. Lancaster, If You Want to Evaluate Your Library... (Champaign, Ill: Univ. of Illinois, Graduate School of Library and Information Science, 1988), p.109.
- 註⑪：同註②，p.14.
- 註⑫：M. E. Maron, "On Indexing, Retrieval, and the Meaning of About," Journal of the American Society for Information Science, 28 (1977), pp.38-43.
- 註⑬：同上註，p.40.
- 註⑭：同註②，p.15.
- 註⑮：Sara J. Penhale and Nancy Taylor, "Integrating End-User Searching into a Bibliographic Instruction Program," RQ, 26 (Winter 1986), pp.212-220.
- 註⑯：Cyril Cleverdon, "The Crandfield Tests on Index Language Devices," Aslib Proceedings, 196 (1967), pp.173-194.
G. Salton, "Recent Studies in Automatic Text Analysis and Document Retrieval," Journal of the ACM, 20:2 (1973), pp.258-278.
- G. Salton, E. A. Fox, and H. Wu, "Extended Boolean Information Retrieval," Communication of the ACM, 26:11 (1983), pp.1022-1036.
- 註⑰：同註②，pp.15-16.
- 註⑱：同註②，p.16.
- 註⑲：William S. Cooper, "A Definition of Relevance for Information Retrieval," Information Storage & Retrieval, 7 (1971), pp.19-37.
- 註⑳：Patrick Wilson, "Situational Relevance," Information Processing & Management, 9 (1973), pp.457-471.
- 註㉑：Stephen P. Harter, "Psychological Relevance and Information Science," Journal of the American

Society for Information Science, 43:9 (1992), pp.602-615.

註⑩：Carlos A. Cuadra and Robert V. Katter, Experimental Studies of Relevance Judgments: Final Report I: Project Summary (Santa Monica, Calif.: System Development Corporation, 1967), p.51.

註⑪：Tefko Saracevic, "The Concept of 'Relevance' in Information Science: A History Review," in Introduction to Information Science, ed. Tefko Saracevic (New York: Bowker, 1970), p.120.

註⑫：同註⑩，p.20.

註⑬：同註⑩，p.23.

註⑭：同上註。

註⑮：同註⑩，p.24.

註⑯：同上註。

註⑰：同註⑩，p.36.

註⑱：同註⑩，p.458.

註⑲：同註⑩，p.460.

註⑳：同註⑩，p.467.

註㉑：同註㉐，pp.467-468.

註㉒：D. Sperber and D. Wilson, Relevance: Communication and Cognition (Cambridge, N.A.: Harvard University Press, 1986).

註㉓：同註㉐，pp.602-603.

註㉔：同註㉐，pp.602-615.

註㉕：同註㉐，pp.610-611.

註㉖：同註㉐，pp.612-613.

註㉗：Richard L. Derr, "A Conceptual Analysis of Information Need," Information Processing & Management, 19:5 (1983), pp.273-278.

註㉘：同註㉗，p.19.

註㉙：同註㉗，pp.19-20.

註㉚：William S. Cooper, "On Selecting a Measure of Retrieval Effectiveness, Part I. The 'Subjective' Philosophy of Evaluation," Journal of the American Society for Information Science, 24:2 (1973), pp.87-100.

William S. Cooper, "On Selecting a Measure of Retrieval Effectiveness, Part II. Implementation of the Philosophy," Journal of the American Society for Information Science, 24:6 (1973), pp.413-424.

註㉛：Carol Collier Kuhlthau, "A Process Approach to Library Skills Instruction," School Library Media Quarterly, 13 (Winter 1985), pp.35-40.

Carol Collier Kuhlthau, "Longitudinal Case Studies of the Information Search Process of Users in Libraries," Library and Information Science Review, 10 (1988), pp.257-304.

註㉜：同註㉛，pp.602-615.

註⑩：同註②，p.21。

註⑪：同註②，p.22。

註⑫：同註②，pp.611-612。

註⑬：同註②，p.23。

註⑭：同上註。

員人呼號與組織參照表

On the Reference Service Organization and Personnel

附錄三

Reference Staff

參考係列員

Reference-Catalogue Staff

【續表 A：參照】

該組成員們所參與之參考服務工作大致可分為二種：即諮詢或外借諮詢及參考資料之整理、編目與統計工作。其工作內容又可依其工作性質之不同而有若干次級之分工；例如：參考資料員，參照資料員，諮詢員等。各組員主要工作內容大抵不外乎搜集、整理、編目、統計、諮詢與外借等項。

The staffs of this department are mainly engaged in reference service. Their work may be divided into two main categories: consulting or lending consultation, and the arrangement, cataloguing and statistics of reference materials. According to their nature of work, they may be further subdivided into several secondary categories: reference material arranger, reference cataloguer, consultant, etc. The main work of each staff member is roughly the same, but there are some differences in the scope of their work.

【續表 B：問題類】

參照員：諮詢員：統計員：編目員

reference material arranger, statistician, cataloguer, and consultant

