圖書與資訊學刊
47(Nov '03)15-35
ISSN 1023-2125

基於詞語抽取的圖書與資訊學刊研究主題分析

Term Extracting-based Analysis of Research Subjects in the Bulletin of Library and Information Science

林 頌 堅
Sung-chien Lin
世新大學資訊傳播學系助理教授
Assistant Professor, Department of Information and Communication Studies
Shih-Hsin University
E-mail：scl@cc.shu.edu.tw

【摘要 Abstract】

本論文持續《圖書與資訊學刊》第 42 期「圖書與資訊學刊論文的高頻詞語抽取與分析」的研究，提出一系列基於自然語言處理與資訊檢索技術的研究主題分析方法。在這個方法中，首先應用關鍵詞語抽取技術，從中英文雙語的論文資料中，抽取出具有意義並且完整的詞語。接著利用隱含語義索引和叢集分析等技術進行主題辨識，將相關的詞語叢集成概念集合，辨識出期刊內重要的研究主題。最後則利用統計上的多維量尺法與電腦繪圖工具將主題繪製在二維的圖形上，便於觀察主題間的結構。我們對《圖書與資訊學刊》第 16 到 45 期的論文資料進行分析，結果發現本期刊的 6 個重要研究主題為(1)圖書館服務與館員訓練、(2)學校圖書館利用教育與資訊素養、(3)圖書資訊學教育與課程、(4)資訊檢索與使用者研究、(5)檔案學和(6)網路與電子技術。

In line with Lin (2003)'s study, this paper proposes a method to analyze research subjects using natural language processing and information retrieval techniques. Keywords are extracted from bilingual corpus, and latent semantic indexing and cluster analysis help identifying research subjects that are co-occurring with keywords. The multidimensional scaling technique is applied to determine thematic structures among research subjects. This study examines 209 articles in the 《Bulletin of Library and Information Science》 and finds the following six areas of research subjects: (1) library services and librarian training, (2) library user education and information literacy, (3) curriculum in library and information science, (4) studies on information retrieval and users, (5) archive studies, (6) internet and electronic technologies.

關鍵詞 Keyword

領域分析 關鍵詞語抽取 圖書與資訊學刊

Domain analysis；Keyword extraction；Bulletin of Library and Information Science

壹、緒論

領域分析(Domain analysis)是藉由分析某一個學術領域內進行的學術活動，諸如論文發表、研討會舉行等等，探索這個領域所關心的研究主題，了解研究人員所重視的資訊與各種資訊生產、傳播與使用的情形，因此是圖書資訊學界相當重視的一個研究方向。(Hjørland & Albrechtsen, 1995)領域分析可以從巨觀的角度，了解某一學科內普遍共識的研究主題及知識架構，提供研究人員了解學科發展的現況。若是單就一份期刊的論文發表情形來進行領域分析，則可以了解期刊重要的作者及研究機構，也可以辨識出期刊內重要的研究主題，提供編輯對收錄的論文主題加以檢視，了解學科發展的趨勢，調整編輯的方向；此外，還可提供作者與讀者全面了解期刊所涵蓋的研究範疇以及提供在檢索上的參考。比方說，對一系列 Journal of American Society of Information Science 及其前身的 American Documentation 所進行的研究中，我們可以發現這份期刊在 1972 到 1990 年期間，主要收錄的論文來自資訊科學基礎理論的探討，這些論文多爲相關科系研究人員獨立完成的研究，較少受到各種機構的補助；(Harter & Hooten, 1992)但是這種現象近年已經有所轉變，逐漸走向由多位跨國、跨機構的作者共同合作，而且有部分論文接受來自於政府、大學和基金會的補助。(Koehler, 2001)另外就 1986 到 1990 年出版的 Journal of American Society of Information Science 論文的引用文獻加以分析，結果顯示「文獻計量學」(Bibliometrics)與「資訊檢索」(Information retrieval)是這份期刊的兩個最大而重要的研究主題，但是這兩個研究主題間有明顯的分離現象，缺乏整合的理論與方法。(Persson, 1994)

在領域分析方法中，文獻計量學利用客觀的量化方法統計出版的期刊論文、研討會論文及專利等文獻，分析領域中重要的研究人員、相關文獻、研究主題等等，已經獲得了相當不錯的研究成果。(Borgman, 1990)例如對相關期刊論文以作者的國籍或所屬機構進行統計，可以得知不同國家或機構在這個領域的生產力。而對於研究主題的分析，文獻計量學常利用「作者共被引分析」(Author co-citation analysis)技術，發掘領域中重要作者共同被引用的群聚，再推測這個領域中重要的研究主題，並且利用統計方法配合電腦繪圖技術呈現分析出來的資訊，更容易解釋分析的結果。(White & McCain, 1998)然而在中文的學術環境中，由於現成的論文資料庫大多以書目資料的呈現爲主要考量，較少具備統計分析的功能，在進行文獻計量的研究上有相當的困難。尤其目前較少資料庫提供國內期刊的論文引用資訊，無法利用共被引分析技術來分析國內期刊的研究主題。

對於上述的困難，我們曾利用分析論文題名與摘要的關鍵詞語，對領域中的研究主題分析提出一個較可行的量化研究方法。由於論文中出現的詞語是作者用來指涉想要傳達的概念，這些詞語代表了研究的問題、理論、方法與解釋等研究相關的主題，如果可以將一篇論文的關鍵詞語抽取出來，便大致可以揭露出這篇論文的研究主題。進一步來說，對一份期刊中所有的論文或是對一個領域中相關的論文所成的集合，進行詞語抽取分析，找出在期刊或領域的論文中出現多次的關鍵詞語，便可以獲得該期刊或這個領域的重要研究主題。相較於過去利用論文或是作者作爲統計的對象，以詞語進行分析在資料量小的環境中更可以得到較可信的結果，而且利用詞語本身具有語義，在解釋上更爲容易而有說服力。這個技術已經應用在分析圖書資訊學的數位圖書館研究(林頌堅，2003)以及國內圖書資訊學領域的重要期刊。(林頌堅，2002a)在《圖書與資訊學刊》42 期的論文(林頌堅，2002b)中，我們已對本學刊 16 期到 39 期的 164 篇論文，利用

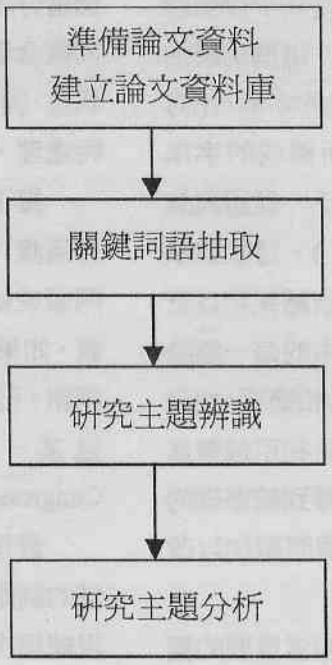
詞語抽取方法對論文中常出現的主題進行分析。在這項初步研究上，所抽取出來的 233 個中文詞語，在與論文關鍵詞比較之後，已經初步驗證這個方法的可行性，同時發現《圖書與資訊學刊》中出現的詞語以圖書館學的相關概念最多，代表多數論文所探討的研究主題以圖書館的管理、服務與技術爲主，其次資訊使用研究、網路與電子文件等資訊科技、圖書資訊學教育、圖書館利用教育和檔案學也是這份學刊常見的研究主題。

本論文是從上一篇論文的基礎繼續發展，除了將期刊收錄的範圍擴增到第 45 期，共計 209 篇論文外，由於《圖書與資訊學刊》的論文題名與摘要部分兼具有中英文兩種語文，爲了充分利用論文資訊中的語義訊息，我們嘗試發展一套可以同時適用於中文與英文的多語詞語抽取方法。在分析上，除了從詞語本身字面所具有的語義來加以詮釋該期刊常見的研究主題之外，我們進一步利用「隱含語義索引」(Latent semantic indexing)與「叢集分析」(Cluster analysis)技術，辨識期刊中的重要研究主題，進而發現相關的論文與重要作者，並以圖形呈現研究主題之間的結構關係。本論文即是報告以上的研究過程與結果。

首先，先描述本研究所提出之主題分析方法，除了方法的整體架構之說明與詞語抽取方法的改進之外，重點在如何從詞語辨識領域中重要的研究主題以及利用圖形表達這些研究主題間的結構關係。其次，報告對於《圖書與資訊學刊》16 期到 45 期的論文進行研究所得到的結果，包含所抽取出來的詞語、重要研究主題與分析。最後則是本研究的結論與未來可以發展的研究課題。

貳、主題分析研究方法

圖一爲本研究進行主題分析的流程架構，茲將此流程說明如下。



圖一：本論文的研究主題分析方法

首先，準備將要進行分析的論文資料。對於期刊的每一篇論文，我們將論文的題名、摘要、出版年與作者資訊建置在論文資料庫中，便於進行詞語抽取與其後的結果分析。其次，從資料庫中取出論文的題名與摘要，利用字串出現的統計特徵與經驗法則進行詞語抽取。接著，從隱含語義索引估算詞語之間的相關程度，並以叢聚演算法辨識重要的研究主題。最後即可進行研究主題分析，獲知相關的論文與作者等資料，並以圖形呈現期刊中研究主題間的結構關係。

以下分別介紹關鍵詞語抽取方法的改進、研究主題辨認以及研究主題關係之圖形呈現等部分的做法。

一、關鍵詞語抽取方法的改進

之前發表的論文已經提出適用於中文的關鍵詞語抽取方法(林頌堅，2002b)，利用出現總頻次和前後接字複雜度等統計特徵，對中文題名與摘要裡所有曾經出現的字串進行過濾，保留具有意義且完整的詞語，雖然該方法的適用性已經初步得到驗證，但是仍具有以下幾個問題。首先，這個詞語抽取方法無法過濾少數不相關與不完整的字串，比方說，由高頻的停用詞(Stop words)所構成的字串(如：「的檔案」)或是一些概念較廣泛、從語義無法判斷研究主題的詞語(如：「推展」)。這些字串需要進一步的處理，使得接下來的分析結果可以更加精確。其次，《圖書與資訊學刊》中的每一篇論文都具有中文和英文兩種語文的題名和摘要，如果可以同時抽取中英文兩種語言的詞語，利用較豐富的語義資訊，在分析研究主題時可以得到較準確的結果。因此，本論文針對上述的幾個問題加以改進。

在自然語言的文句處理上，一個相當重要的觀念是語言的基本書寫單位是字(Characters)，在文句中，由一個字或一序列的字串構成具有意義的詞(Words)，從詞再建構出更大的意義單位，如詞組(Phrases)、句子(Sentences)等，傳達作者想表示的概念。可見若要進行文句的自動化處理，字是首先可以辨識的單位，但若要處理意義層次的問題，則需要能辨識詞或詞以上的單位。在書寫上，英文的詞由少量的字所組成，詞與詞之間有明顯的界限，區分較容易，但有許多規則或不規則的變化來表示事物或概念的單複數、時式(Tense)與時態(Aspect)，處理上同一詞語以不同形態出現者需要能加以辨識，而且詞組雖然由一序列的詞構成，但指涉的意義往往較爲專殊甚至不同，也需要先行辨識；中文詞雖然少有形式上的變化，但字的總數高達數千以上，且詞間缺乏界限，在區分上相當困難，在中文文句的處理上需要先行進行斷詞(Word identification)，找出具有意義的詞語。(Chen & Liu, 1992)因此，在關鍵詞語抽取方法的改進上，首先對英文的文句中以詞作爲處理的單位，中文文句則以字爲處理單位，利用詞語對領域的重要程度(總頻次)和詞語的完整性(前後接字複雜度)等統計特徵區分出中文的詞以及中英文的詞組，至於具有相同概念但不同形態的英文詞語，辨識上需要更多的訊息，因此這部分將留到下一階段的研究主題辨識時處理。

接下來針對由高頻的停用詞所構成的字串進行過濾，我們觀察到這些停用詞都是出現在字串的開頭或結尾。因此可以建立一個中英文的停用詞表，如果詞語的開頭或結尾是停用詞，則過濾這個詞語，但是如果是停用詞出現在詞語的中間，通常是某一個特定的專有名詞，如“Library of Congress”，這種情形則不加以過濾。

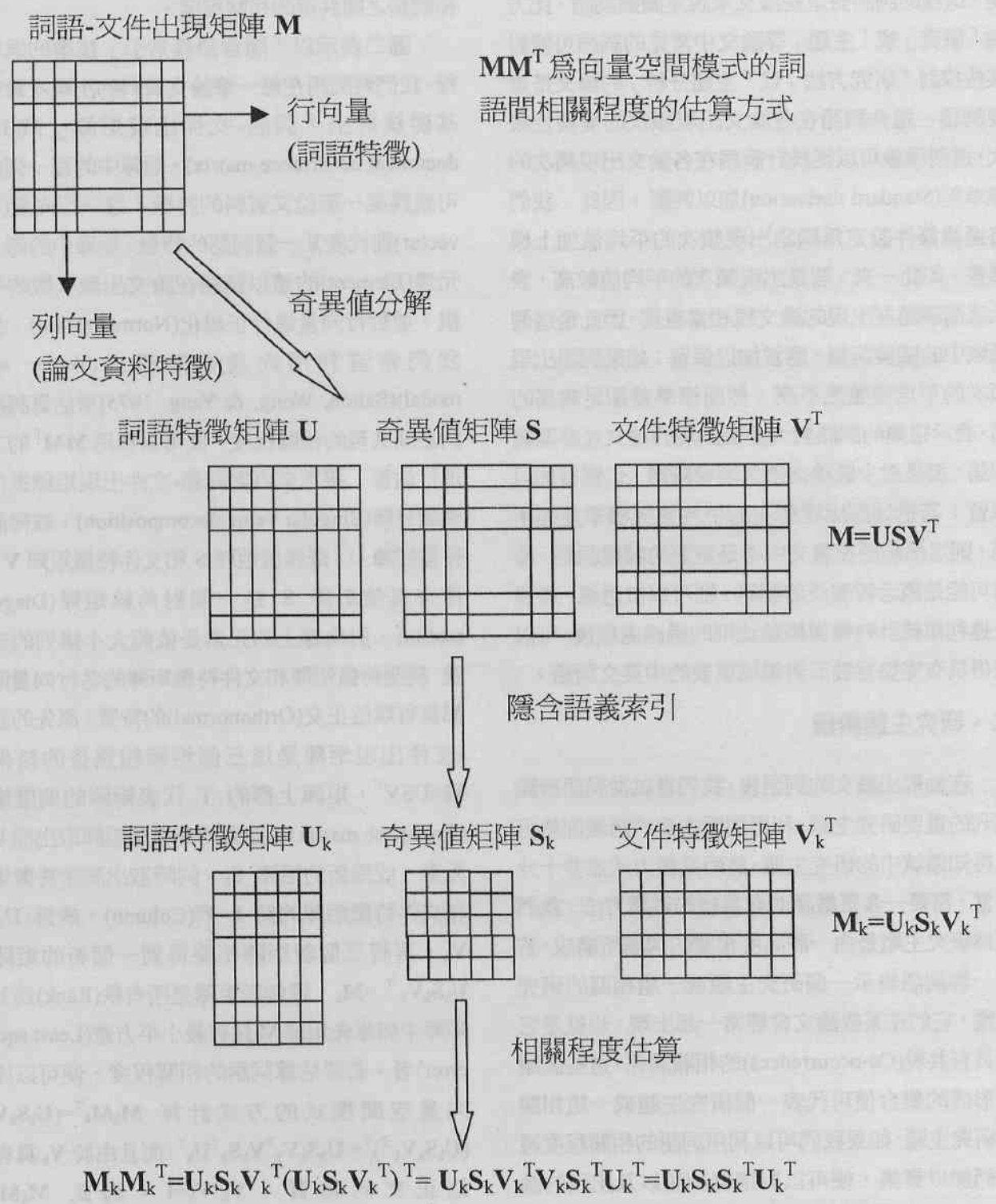
最後針對一些出現在多數的論文而且概念廣泛的詞語加以過濾，這些詞語在統計上的特徵是出現總頻次雖然高，但是在每篇論文出現的次數並不多，利用詞語在論文資料中出現頻次的平均值(Mean)，可以順利去除。然而，有些詞語雖然在多

數論文的出現不多，可是在少數的論文出現頻次很高，這樣的詞語對這些論文來說是關鍵詞語，比方說「研究」或「主題」等論文中常見的詞語可能對某些探討「研究方法」或「主題分析」的論文是重要詞語，這些詞語在各論文出現頻次的變異性較大，這個現象可以從統計詞語在各論文出現頻次的標準差(Standard derivation)加以判斷。因此，我們將過濾條件設定爲詞語出現頻次的平均值加上標準差。如此一來，若是出現頻次的平均值較高，表示這個詞語在出現的論文裡相當重要，因此是這個領域中的關鍵詞語，應當加以保留；如果詞語出現頻次的平均值雖然不高，然而標準差卻足夠高的話，表示這類的詞語對大多數出現的論文並非關鍵詞語，但是對少數論文則是關鍵詞語，也應當加以保留；若是詞語出現頻次的平均值與標準差都不高，則這個詞語在論文中不是重要的關鍵詞語，極有可能是概念較廣泛的詞語，則可以被過濾。經過上述利用統計特徵與經驗法則的過濾處理後，可以獲得具有完整意義而對領域重要的中英文詞語。

二、研究主題辨識

在抽取出論文的詞語後，我們嘗試從詞語辨識期刊的重要研究主題。利用詞語本身的語義固然可以得知領域中的研究主題，然而這種方式並非十分妥當，需要一套更嚴謹而有系統的辨識方法。我們認爲研究主題是由一群高度相關的詞語所構成，若有一群詞語表示一個研究主題或一組相關的研究主題，它們在某些論文會經常一起出現，也就是它們具有共現(Co-occurrences)的相關關係，這些詞語所形成的集合便可代表一個研究主題或一組相關的研究主題，如果我們可以利用詞語的相關程度將它們加以叢集，便可以對這群詞語以及出現的論文、論文的作者等資料加以分析，了解研究主題的情況。在本研究中以兩種叢集演算法先後對詞語進行分群，並且利用「隱含語義索引」(Latent semantic indexing)技術(Deerwester, et. al., 1990)來估算詞語和詞語之間共現的相關程度。

圖二表示以「隱含語義索引」技術的處理過程。我們將詞語在每一筆論文資料的出現次數做爲基礎統計出「詞語-文件出現矩陣」**M**(Term-document occurrence matrix)，矩陣中的每一列向量可視爲某一筆論文資料的特徵，每一行向量(Row vector)則代表某一個詞語的特徵，矩陣中的每一個元素(Element)的值以詞語在論文出現次數的平方根，並對行向量進行正規化(Normalization)。如果我們希望利用向量空間模式(Vector space model)(Salton, Wong, & Yang, 1975)來估算詞語與詞語間共現的相關程度，便可以利用 $\mathbf{MM}^T$ 的方式進行估算。接下來再對詞語-文件出現矩陣進行奇異值分解(Singular value decomposition)，取得詞語特徵矩陣 **U**、奇異值矩陣 **S** 和文件特徵矩陣 **V**，使得奇異值矩陣 **S** 是一個對角線矩陣(Diagonal matrix)，對角線上的元素是依照大小排列的奇異值，詞語特徵矩陣和文件特徵矩陣的各行向量間則都具有單位正交(Orthonormal)的特質，原先的詞語-文件出現矩陣是這三個矩陣相乘後的結果，$\mathbf{M}=\mathbf{USV}^T$，矩陣上標的 T 代表矩陣的倒置矩陣(Transpose matrix)。如果將奇異值矩陣取出前 k 個元素，成爲新的矩陣 $\mathbf{S}_k$，同時取出詞語特徵矩陣和文件特徵矩陣的前 k 列(Column)，成爲 $\mathbf{U}_k$ 和 $\mathbf{V}_k$，再將三個新矩陣相乘得到一個新的矩陣，$\mathbf{U}_k\mathbf{S}_k\mathbf{V}_k^T=\mathbf{M}_k$。這個新矩陣是所有秩(Rank)爲 k 的矩陣中與原先矩陣 M 具有最小平方差(Least square error)者。若要估算詞語的相關程度，便可以比照向量空間模式的方式計算 $\mathbf{M}_k\mathbf{M}_k^T=(\mathbf{U}_k\mathbf{S}_k\mathbf{V}_k^T)(\mathbf{U}_k\mathbf{S}_k\mathbf{V}_k^T)^T=\mathbf{U}_k\mathbf{S}_k\mathbf{V}_k^T\mathbf{V}_k\mathbf{S}_k^T\mathbf{U}_k^T$，而且由於 $\mathbf{V}_k$ 具有單位正交的特質，$\mathbf{V}_k^T\mathbf{V}_k=\mathbf{I}$，因此 $\mathbf{M}_k\mathbf{M}_k^T=\mathbf{U}_k\mathbf{S}_k\mathbf{S}_k^T\mathbf{U}_k^T$。利用奇異值分解，將詞語特徵及論文資料特徵的重要資訊保留在奇異值較大的維度中，也就是前幾個維度，可以取得詞語對論文的



圖二：利用隱含語義索引技術估算詞語間相關程度

隱含語義結構。換句話說，隱含語義索引認爲基本上兩個相關詞語間必定存在較大的共現關係，但可能由於統計上的雜訊(Noise)，造成相關詞語在論文資料上不存在共現關係或是關係較弱，因此可以透過奇異值分解，以較前面的維度的統計訊息爲主，忽略雜訊的影響，使得這些詞語的相關程度可以獲得較大的估算值。因此一般認爲隱含語義索引較單純的向量空間模式在相關程度的估算上有效，在本研究便利用隱含語義索引估算詞語間的相關程度。

其次說明詞語的叢集，本研究對詞語的叢集分爲兩個階段，第一階段將相關詞語叢集起來，形成概念集合，第二階段則針對核心的概念集合，再度進行叢集，找出領域中重要的研究主題。概念集合的目的是希望將具有相近概念的詞語叢集起來。如前面所述，英文的詞語由於單複數、時式或時態的不同，而有多種的型態，或者中英文具有相同概念的詞語，需要能夠加以辨識。其次，在不同論文中，不同作者可能使用多種不同的詞語來代表相近概念，比方說對於未來圖書館(Libraries in the future)的形式，許多論文使用「數位圖書館」來強調館藏資源的形式，有些則以「電子圖書館」來強調未來圖書館中電子出版與電子傳遞的網路資源特色，也有些論文則將「虛擬圖書館」作爲未來圖書館著重在共享的虛擬館藏上，但研究人員多認爲這三個詞語具有相近的概念(Bawden & Rowlands, 1999)，出現的論文大多探討相同的研究主題。再者，在論文中同一詞語所指涉的概念不一定相同，以圖書資訊學中相當重要的「相關」概念而言，雖然在某些論文中「相關」特別指涉於資訊對問題情境上的合適性，然而由於這個詞語也是一個相當一般的詞語，常會出現在許多句子中，若是論文中出現該詞語，需要進一步分析此時所代表的是何種概念。若只以詞語在論文中的出現情形判斷論文的研究主題在統計上無法適用於以上兩種狀況，對研究主題分析上較爲不足。

本研究利用 Cliques 叢集演算法(Kowalski & Maybury, 2000)，將詞語分類成概念集合。Cliques 是一個多重分類的叢集演算法，只要詞語與集合中其他各個詞語間的相關程度都超過某一個預設值，一個詞語可以被分類到多個集合中，避免單一分類中單一連結(Single link)演算法過鬆的條件，或者完整連結(Complete link) 演算法的過於嚴格。以概念集合取代詞語，一來可以利用集合代表一組具有相同概念的詞語，使得在論文中所出現的不同詞語，如將「圖書館員」、「librarian」和「librarians」等不同形態的詞語或「數位圖書館」、「電子圖書館」與「虛擬圖書館」等相近概念詞語叢聚起來，表達同一個概念，出現這些詞語的論文便可視爲具有相同的研究主題。二來對於可以表示多種概念的詞語，在經過叢聚後可以從集合中其他相關的詞語辨識這個詞語相關的某一特定研究主題，以上述的詞語「相關」爲例，在指涉資訊對問題情境的合適性時，可能一起出現的詞語有「資訊需求」、「資訊行爲」、「搜尋」等等，當將這些詞語與「相關」叢聚成詞語集合時，若是論文中出現「相關」，便可以由集合中其他的詞語是否出現多次，而辨識此論文是否包含「資訊對問題情境的合適性」的研究主題。因此，本研究根據前述之隱含語義索引估算詞語間的相關程度，並利用 Cliques 叢集演算法，產生期刊中曾出現的各種概念集合。而在概念集合之間的相關程度估算上，概念集合的特徵可以視爲是集合內各個詞語特徵(也就是詞語在 M_k 中的相對行向量)的總和，再利用隱含語義索引的方式進行估算。

由於期刊中曾出現的各種概念集合範圍包羅萬象，在研究主題辨識上不易分析，因此在目前的研究中先選擇若干核心的概念集合進行分析，找出期刊中最明顯而重要的主題。核心集合的選擇是依據概念集合與其他概念集合之間的連結程度來判斷，在本研究中概念集合與其他概念集合間相關程

度是否超過預設的閾值，來決定它們是否連結。如果概念集合與越多的概念集合連結，則這個概念集合可能構成領域中重要的研究主題；反之，如果概念集合與其他概念集合之間的相關程度都不強，也就是這個概念集合連結到其他概念集合的個數相當少，則這個概念集合可推斷是在領域的邊陲部分。依據以上的原則，核心集合的選擇方法如下：1.估算所有概念集合之間的相關程度。2.統計每一概念集合連結其他概念集合的個數。3.設定一個最小的連結個數，如果概念集合的連結個數大於或等於這個最小的連結個數，則將這個概念集合視爲是核心集合。最後仍然依據隱含語義索引的方式估算概念集合間的相關程度，對概念集合進行 Ward 叢集分析(Kowalski & Maybury, 2000)，分成若干大類，將分類結果做爲期刊的研究主題。

尋找出研究主題後，可以再次將研究主題的特徵視爲是概念集合特徵的組合，便可以再度利用隱含語義索引找到期刊中與主題相關的論文，進而檢索出這個主題的重要作者與發表時間的分佈，提供進行分析的參考。

三、研究主題的關係之圖形呈現

最後是將研究主題的結構關係加以呈現的問題。在本研究中，我們希望能參考作者共被引分析的做法，將重要的研究主題之間的關係呈現在一個二維(Two-dimensional)的圖形中。作者共被引分析(White & McCain, 1998)的主要步驟是先選取出領域中重要的作者形成一個集合，在集合中的作者除了具備論文多次被引用的條件之外，還需與集合中的某一群作者共同被引用多次，這個集合內的作者便是接下來進行分析的對象。對集合裡的每一位作者統計他與其他集合內的作者在論文資料庫中共同被引用的論文數目，做爲該作者的特徵(Profile)。接下來便可以利用特徵間的相關係數(Correlation coefficients)計算集合中每兩位作者之間的相關程度。若是兩位作者的特徵有相似的現象，意味著在與其他作者共同被引用的程度上，這兩位作者具有相似的關係，換言之，在這兩位作者中的一位 A 如果和集合中的其他某一位作者有很高的共被引次數，另一位作者 B 便也具有很高的共被引次數，但反之其他作者若與 A 的共被引次數低，則與 B 的共被引次數也不高，我們可以說這兩位作者間在論文的被引用上具有較大的相關程度。再依據作者間的相關程度利用叢集分析和多維量尺法(Multi-dimensional scaling)等統計技術進行分析。叢集分析可以使相關程度大的作者叢集在一起，因爲他們的研究被認爲具有相關性，被叢集所形成的集合便可代表領域中某一個研究主題。使用多維量尺法則可以將集合中每一作者投射到二維空間上的某一點，點與點的距離代表了作者間的相關程度，被引用的相關程度越高的作者，在空間上所對應點之間的距離越小，反之，越不相關的作者，他們在空間上所對應的點距離越大。如此一來可以形成幾個板塊狀的研究主題。配合叢集分析，我們便可以在利用多維量尺法所形成的二維圖形中，分析作者叢集所代表的研究主題之間的關係。在圖形上兩個相鄰的作者叢集表示對應的研究主題之間有某種程度的相關性，而與分離的作者叢集則代表研究主題之間少有交集。例如，White 和 McCain(1998)利用作者共被引分析資訊科學的研究現象，在多維量尺法所形成的二維圖形中明顯的有兩群作者，這說明了在資訊科學的學術領域主要包含兩個次領域，一是資訊檢索研究，一是領域分析研究。資訊檢索主要研究的範疇是人-電腦-文獻之間的介面(Human-computer-literature interface)，包含了實驗性和實作性檢索(Experimental and practical retrieval)、一般圖書館系統理論(General library systems theory)、使用者理論(User theory)、OPAC、索引理論(Indexing theory)等研究主題；領域分析是對學術文獻與其社會脈絡(Learned

literatures and their social contexts)進行分析，研究範疇中則有引用分析和引用理論(Citation analysis and citation theory)、文獻計量學(Bibliometrics)、科學傳播(Scientific communication)等主題。再進一步從圖形上的分佈可以觀察到這兩個次領域間少有交集，換言之，雖然同屬於資訊科學的領域中，資訊檢索研究與領域分析研究雙方面的研究人員甚少利用對方所發展出來的研究方法和理論來充實本身的研究成果。從上面的研究可以看出若能將統計出來的訊息進行圖形化，在解釋上將更加容易而有說服力。

本研究參考了上述作者共被引分析的方法，同樣利用多維量尺法將研究主題顯示在圖形上。不同於作者共被引分析利用領域中重要的作者做爲分析對象，本研究以詞語的概念集合爲分析對象。因此，本研究中多維量尺法的輸入是詞語的概念集合在論文資料裡共現的相關程度，而如前所述，概念集合間的相關程度是由隱含語義索引估算得到的。輸出的結果則是在圖形上概念集合所對應的座標，如果兩個概念集合間共現的相關程度愈大，則這兩個概念集合在圖形上相對應的點距離愈近。在前面已經將彼此間相關程度較大的概念集合叢集起來，形成若干研究主題，因此圖形上屬於同一研究主題的概念會形成一個聚落，研究主題間的關係便可以從相對應的聚落之間的位置觀察得到。

參、詞語抽取及主題分析結果

在本節中將依序呈現研究所得到的結果，包括從論文資料中抽取出來的詞語、利用詞語的共現現象和隱含語義索引對詞語進行叢集所辨識得到的研究主題、各研究主題相關的論文及重要作者，以及利用多維量尺法所得到的研究主題間關係的圖形。

首先從《圖書與資訊學刊》第 16 期到第 45 期，共 30 期 209 篇論文的題名和摘要等資料中抽取關鍵詞語。在本研究中，關鍵詞語的前後接字複雜度必須在 0.5 以上，詞語在論文資料中出現頻次的平均值與標準差總和的閾值則設爲 2.5，另外，對於出現總頻次的閾值，較短的詞語(中文的二字詞或三字詞)設爲 15 次，較長的詞語則設爲 10 次，抽取的結果共得到 278 個詞語，中文與英文詞語分別有 138 個和 140 個。出現總頻次較高的前 75 個詞語以及它們在論文出現的總頻次、出現頻次的平均值與標準差，如表一所示。

表一：出現總頻次排序前 75 個的關鍵詞語

詞語名稱	F_t	f_t	σ_t	詞語名稱	F_t	f_t	σ_t	詞語名稱	F_t	f_t	σ_t
使 用	145	2.16	1.73	服 務	140	2.37	1.78	網 路	127	2.44	2.21
資 料	110	2.29	2.09	系 統	104	3.06	2.41	分 析	102	1.62	1.12
system	101	2.46	2.30	taiwan	92	1.92	1.08	檢 索	87	2.49	3.74
research	85	1.89	1.44	service	82	1.91	1.44	教 育	75	2.21	1.94
應 用	74	1.54	1.00	檔 案	73	3.84	3.18	工 作	72	1.95	1.82
education	70	2.50	2.31	文 獻	70	1.75	1.51	services	69	1.92	1.36
librarians	67	2.48	2.49	讀 者	67	2.31	1.86	課 程	65	3.61	2.81
技 術	63	1.85	1.22	大 學	62	2.38	1. 50	electronic	61	2.77	2.13
管 理	60	1.88	1.39	analysis	59	1.74	1.82	archives	57	3.56	3.00
resources	57	1.90	1.40	internet	56	2.24	2.08	technology	56	1.65	1.00
期 刊	**52**	**4.33**	**3.09**	network	51	1.96	1.58	management	50	1.79	1.15
web	50	2.50	2.20	大學圖書館	50	2.63	1.60	組 織	50	2.00	2.04
users	49	1.69	1.09	美 國	48	1.71	1.10	圖 書 館 學	48	2.82	2.87
科 技	46	1.53	1.41	data	45	2.14	1.73	university	45	1.67	1.12
subject	44	3.38	2.53	功 能	43	1.72	1.40	access	41	1.78	1.59
電 子	41	1.64	0.97	chinese	40	2.22	1.08	knowledge	39	2.79	3.30
我 國	39	1.77	1.12	知 識	39	3.25	3.22	專 業	39	1.60	1.68
設 計	38	1.81	1.18	history	37	1.68	1.94	systems	37	1.76	1.38
館 員	37	2.64	2.12	**資 訊 素 養**	**36**	**5.14**	**3.04**	行 爲	35	2.50	1.30
調 查	35	1.94	1.13	館 藏	35	1.84	1.60	content	34	2.27	2.35
中 心	34	1.70	1.00	比 較	34	1.79	1.10	格 式	34	2.83	1.52
實 施	34	1.89	1.24	**閱 讀**	**34**	**6.80**	**3.25**	環 境	34	1.62	1.00
collection	33	1.65	1.24	search	33	2.75	3.19	資 料 庫	33	1.94	1.16
資訊科學	33	3.00	1.60	模 式	33	2.06	2.01	論 文	33	2.54	2.06
students	32	2.29	1.48	user	32	1.88	1.74	社 會	32	1.60	0.97

F_t：詞語在論文資料中出現的總頻次

f_t：詞語在論文資料中出現頻次的平均值

σ_t：詞語在論文資料中出現頻次的標準差

從表一可以發現本研究所抽取出來的詞語，不但可以抽取中英文兩種語言的詞語，適用於多語環境下的研究主題分析，而且可以有效過濾許多不完整的詞語。就抽取出來的關鍵詞語，很明顯地可以看出大多與圖書資訊學或檔案學的重要概念相關。尤其在表一的高頻詞語中，出現頻次平均值相當高的詞語，比方說「閱讀」(f_t=6.80)、「資訊素養」(f_t=5.14)和「期刊」(f_t=4.33)等，在圖書與資訊學領域中都是非常重要的詞語。而且應用這方法，對於概念廣泛、對研究主題較沒有區別能力的詞語，在審慎選擇詞語篩選的閾值後，也可以過濾。

在詞語抽取之後，我們依據它們在論文資料中出現的情形建立詞語-文件出現矩陣，利用奇異值分解進行隱含語義索引，估算詞語間的相關程度，將相關程度較大的詞語叢集成概念集合。在本研究中，對所有抽取出來的詞語，以詞語-文件出現矩陣的前 50 個奇異值進行隱含語義索引，換言之，k 值的設定值爲 50，集合中詞語間的最小相關程度則設爲 0.4，結果共得到 212 個概念集合。其次計算這 212 個概念集合與其他概念集合的連結程度，在這裡兩個概念集合互相連結的條件是它們之間的相關程度必須超過 0.4，而且連結個數超過 10 或以上的概念集合，才會保留爲核心的概念集合，結果共分析出 67 個核心集合。接下來再對核心集合進行分類，找出《圖書與資訊學刊》的研究主題，經過分析共找到 6 個研究主題，我們依據構成詞語的語義，爲 6 個研究主題命名爲(1)圖書館服務與館員訓練、(2)學校圖書館利用教育與資訊素養、(3)圖書資訊學教育與課程、(4)資訊檢索與使用者研究、(5)檔案學和(6)網路與電子技術。表二便是這 6 個主題的名稱及它們所包含的概念集合。從表二，可以觀察到各個概念集合已經將具有接近概念的中英文詞語叢集起來，概念集合中詞語的語義也可以明顯地表示所表述的研究主題，以下分別介紹各主題的命名和相關論文、重要作者。

表二：《圖書與資訊學刊》分析所得到的研究主題與核心集合

主題	核心集合	主題	核心集合
(1) 圖書館服務與館員訓練	5. librarians, training, 館員	(2) 圖書館利用教育與資訊素養	14. graduate, students, university
	16. 大學圖書館, 網路教學		21. school, 利用教育
	37. reference, service		22. learning, medical, 學習
	55. 大學圖書館, 服務		44. students, university, 大學
	97. 調查，館員		57. 論文, 學院
	141. service, 服務		59. 網路教學, 學習
	152. services, 服務		60. university, 大學, 大學圖書館
	161. 問卷, 調查		63. learning, students, 學生, 學習
	163. university libraries, 大學圖書館		68. school, students, 學生
			73. graduate, 學院
			101. information literacy, 能力, 資訊素養
			144. information literacy, learning, 資訊素養, 學習

（續下表）

（接上表）

(3) 圖書資訊學教育與課程	27. education, library and information science, 教育, 圖書館學	(4) 資訊檢索與使用者研究	23. databases, 資料
	43. courses, information science, schools, 資訊科學, 圖書資訊學		24. behavior, 因素, 行爲
	46. 教育, 課程		26. access, opac
	84. library and information science, 資訊科學, 圖書館學		30. retrieval, 檢索
	87. lis, programs, 階段		48. users, 讀者
	131. education, 專業, 教育		56. 分析, 使用
	186. taiwan		62. mining, 知識
			66. 系統, 使用
			88. 使用, 檢索
			89. dissertations, 引用
			119. data, 資料
			128. readers, 讀者
			148. users, 使用
			158. analysis, 分析
			190. 文獻
			194. 功能
(5) 檔案學	41. archival, 檔案, 檔案描述	(6) 網路與電子技術	4. wide, 網路
	90. archival, archives, 檔案, 檔案館		10. network, 資源共享
	98. archival, ead, 檔案描述		13. web, 電子
	103. archives, 行政, 機關, 檔案		36. technology, 技術, 網路
	113. 國際, 標準		45. bibliographic, control, 書目, 控制, 網路資源
			51. resources, 環境
			64. document, 傳遞, 電子
			72. web, wide
			77. 技術, 電子
			102. network, technology, 網路
			106. 網路, 應用
			116. document, 文件, 電子
			122. technology, 科技
			126. resources, 資源
			143. internet, 網路
			154. electronic, 電子
			156. center, 中心
			193. 出版

以主題(1)來說，概念集合 5 包含的詞語有「librarians」、「training」以及「館員」，這個集合即可以表示館員訓練的概念。另外在這個研究主題中還包括了分別具有詞語「服務」和「service」及「services 」的概念集合 141 和 152 等與服務相關的概念，因此將這個研究主題視爲是有關「圖書館服務與館員訓練」。《圖書與資訊學刊》中與這個主題較相關的論文如表三所示，從論文題名我們可以再次驗證這個主題與圖書館的讀者服務以及館員訓練相關。從表三所列出的這些論文查詢相對應的作者資料，可以發現論文未集中在少數的作者，所有的作者只有一或二篇的論文，另一點值得注意的是這 20 篇論文出版的年份集中在 1996 到 1999 年，共有 15 篇，2001 年和 2002 年各只有 1 篇。

表三：研究主題(1)「圖書館服務與館員訓練」所包含的論文及相關程度

論　文　題　名	相關程度
國內北區大學圖書館參考館員接受在職訓練之課程內容需求探討	0.33
國立政治大學圖書館參考服務使用研究	0.28
網路教學帶給大學圖書館的挑戰與機會	0.27
台灣地區大學圖書館館員工作滿意度影響因素探討	0.26
台灣地區大學院校圖書館教授指定參考書服務之調查研究	0.25
大學圖書館網站內容分析	0.20
淺論參考服務的組織和人員	0.19
公共圖書館效能評量指標與面向之研究：以臺北市立圖書館爲例	0.18
網路科技環伺下省思大學圖書館兩個問題—圖書館與電腦中心之關係及讀者政策之制定	0.17
臺灣地區大學暨獨立學院圖書館館藏淘汰之調查研究	0.17
圖書館資訊素養之培養方針與評量指標	0.16
國內北區大學圖書館實施閱選訂購之現況探討	0.16
我國大學圖書館館員工作輪調之研究	0.16
館際互借資源共享的新契機-台灣地區期刊聯合目錄暨館際合作系統	0.15
圖書館在網際網路提供的讀者服務	0.13
圖書館學碩士的生涯發展研究	0.13
美國資料中心之參考諮詢與推廣服務	0.12
政大學生圖書館焦慮之探討	0.12
探討臺灣地區學術圖書館的收費服務之現況	0.12
大學圖書館資訊素養網站之研究	0.11

研究主題(2)則以概念集合 21 的「school」和「利用教育」，以及概念集合 101 和 144 的「information literacy」和「資訊素養」等，命名爲學校圖書館利用教育與資訊素養，表四列出相關的論文與相關程度，從論文出版的年份，可以觀察到自 1999 年後，由於推展資訊素養的觀念之後，在這份期刊裡這類主題的論文有增加的現象，但與研究主題(1)同樣地分散於多位作者。

表四：研究主題(2)「學校圖書館利用教育與資訊素養」所包含的論文及相關程度

論　文　題　名	相關程度
問題導向學習與醫學生資訊素養之探討	0.43
網路教學帶給大學圖書館的挑戰與機會	0.35
由資訊素養提昇「知」的能力	0.32
我國研究生對圖書館利用教育態度之探討：以淡江大學國際研究學院和工學院研究生爲例	0.28
政大學生圖書館焦慮之探討	0.24
以新聞學博碩士論文評鑑政治大學傳播學院圖書分館館藏	0.24
技術學院圖書館利用教育之現況調查研究	0.20
以學生爲中心的資訊素養教育課程之探討	0.19
光碟檢索者文獻搜尋行爲之研究：以國立政治大學研究生爲例	0.18
Favorite Books of Middle Students：Why They Are Favorite	0.18
資訊素養的意義、內涵與演變	0.17
大學圖書館資訊素養網站之研究	0.13
圖書館資訊素養之培養方針與評量指標	0.13
台灣地區大學院校圖書館教授指定參考書服務之調查研究	0.13
臺灣地區大學暨獨立學院圖書館館藏淘汰之調查研究	0.12
國立政治大學法學外文期刊評鑑	0.11
技專校院圖書館使用者行爲與滿意度因素分析－以景文技術學院爲例	0.11
大學圖書館的中心地位：一個老隱喻與新概念	0.11

在研究主題(3)中包含概念集合 27、43、46、84 和 131，從它們所包含的詞語「library and information science」、「information science」、「圖書館學」、「圖書資訊學」、「資訊科學」、「education」、「教育」、「course」和「課程」等，表示這個研究主題和圖書資訊學中教育理念和課程設計相關，因此命名爲圖書資訊學教育與課程。表五列出的論文中，可以看到大部分與圖書資訊學的專業教育相關，而且許多研究採用比較分析做爲研究方法。這個主題的論文在作者方面也沒有集中的現象，發表的年份也相當分散。

研究主題(4)與電子資料庫、OPAC 等資訊系統的使用行爲有關，概念集合包括了「databases」、「behavior」、「access」、「opac」、「retrieval」、

「users」、「readers」、「行爲」、「檢索」、「讀者」、「使用」等中英文相關的詞語，可見得這個主題是資訊檢索中與使用者理論和一般圖書館系統理論相關的研究，從表六論文的題名中，也可以歸納出相同的看法，因此被命名爲資訊檢索與使用者研究。在這主題上，發表較多的學者是黃慕萱教授，有 3 篇論文在《圖書與資訊學刊》中發表。由於這方面的研究是圖書資訊學重要的主題，因此在過去的幾年中都有許多相關論文在《圖書與資訊學刊》發表。

表五：研究主題(3)「圖書資訊學教育與課程」所包含的論文及相關程度

論　文　題　名	相關程度
80 年代以來中國內地圖書館學信息學教育之發展與展望	0.44
談圖書館學與資訊科學教育 大學部學程及學科專長問題	0.31
台灣與美加地區圖書資訊學資訊科學課程之研究	0.29
圖書資訊學研究方法課程的現況與問題	0.27
圖書館與資訊科學專業教育之探討	0.25
圖書館學之理論基礎	0.22
技術學院圖書館利用教育之現況調查研究	0.19
從學校相關系所名稱的改變探究圖書館的轉型	0.16
以學生爲中心的資訊素養教育課程之探討	0.15
圖書與資訊學刊論文的高頻詞語抽取和分析	0.15
媒體出版與圖書館學整合課程模組芻議：圖書與資訊研究	0.14
澳洲圖書資訊學教育	0.14
圖書館資訊素養之培養方針與評量指標	0.14
紐西蘭圖書資訊學教育	0.14
美國圖書館暨資訊科學委員會與我國教育部圖書館事業委員會之比較	0.13
國內北區大學圖書館參考館員接受在職訓練之課程內容需求探討	0.13
圖書館在「研究方法與論文寫作」課程中的角色	0.13
從九年一貫課程改革看國內中小學圖書館之發展	0.12
資訊資源管理：融合圖書館學、檔案學本科教育的平臺	0.12
圖書館學碩士的生涯發展研究	0.12

表六：研究主題(4)「資訊檢索與使用者研究」所包含的論文及相關程度

論　文　題　名	相關程度
從線上公用目錄的功能探討圖書館期刊編目政策	0.24
資料探勘於圖書館行銷及顧客關係管理之應用	0.22
台灣電子期刊使用者行爲分析：以 Elsevier SDOS 電子期刊系統爲例	0.21
探討讀者使用線上公用目錄檢索點及主題檢索之情形	0.20
光碟檢索者文獻搜尋行爲之研究：以國立政治大學研究生爲例	0.19
修改行爲探討：以國立台灣大學之終端使用者爲例	0.18
檢索背景對檢索技巧及檢索結果之影響	0.18
中華民國政府文獻使用研究：以政治大學社會科學學院博士論文引用文獻爲例	0.17
技專校院圖書館使用者行爲與滿意度因素分析：以景文技術學院爲例	0.16
古文書檢索系統功能之研究：以台大電子圖書館與博物館系統爲例	0.16
數位化典藏內涵知識獲取模式之探討：以多媒體典藏爲例	0.15
主題複分效用分析：以 NBINET 爲例	0.13
以新聞學博碩士論文評鑑政治大學傳播學院圖書分館館藏	0.13
國立政治大學法學外文期刊評鑑	0.13
圖書館檢索區電腦數量之配置：排隊理論之應用	0.13
報紙資料庫在圖書館所扮演的角色	0.12
Metadata 管理系統之分析與設計	0.12
相關概念在資訊檢索中之發展與趨勢	0.12
WAPOPAC 系統設計與行動圖書館通訊技術之探討	0.12
終端使用者之檢索類型變化研究	0.12
數位圖書館中權威控制系統的設計	0.12
兒童閱讀行爲之探討	0.11
科學家資訊搜尋行爲的探討	0.11
應用 XML 技術之電子文獻傳遞系統架構	0.11
圖書館自動化與讀者焦慮類型	0.10
館際互借資源共享的新契機：台灣地區期刊聯合目錄暨館際合作系統	0.10
論資訊流通中摘要著作權	0.10
TREC 現況及其對資訊檢索研究之影響	0.10
資訊尋求的理論與實證研究	0.10
成人閱讀之研究：以台北市立圖書館永春分館讀者爲例	0.10

從研究主題(5)概念集合中的詞語可以很明顯的看出，這個主題包含與檔案學相關的研究，如檔案描述、檔案館等詞語，表七是這個主題的相關論文，而發表與這個主題相關論文最多的作者是薛理桂教授，共計發表 5 篇論文，另外莊樹華女士和劉佳琳女士也各有 3 篇論文在此期刊發表。

表七：研究主題(5)「檔案學」所包含的論文及相關程度

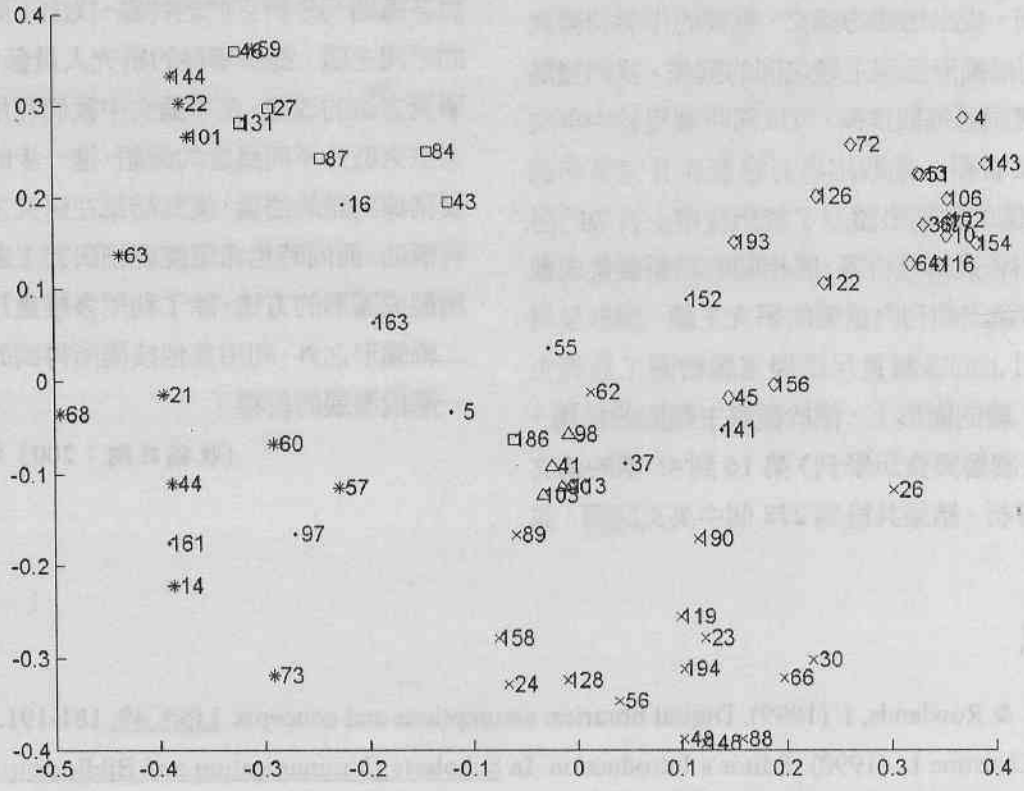
論　文　題　名	相關程度
檔案描述編碼格式在中文檔案之應用：以國史館及中研院近史所檔案館爲例	0.44
中文檔案名稱權威檔之實作：以戰後臺灣經濟發展相關人物及機關爲例	0.36
中文檔案描述規則之擬訂：基於國際檔案描述標準（ISAD(G)）	0.35
中央研究院近代史研究所檔案館管理概述	0.28
臺灣地區檔案名稱權威檔建置之現況分析	0.27
檔案描述編碼格式(EAD)之發展與實施	0.27
檔案鑒定模式：「黑盒子鑒定模式」淺探	0.23
淺論我國檔案中央主管機關應有之功能與發展	0.22
從館藏發展政策探討檔案鑑定工作之擬定	0.19
簡介大陸地區的檔案事業和與民國史編纂的關係	0.16
檔案館的功能與角色	0.16
談地區代碼系統與工商資訊檢索	0.10
大陸地區現存明清檔案的分佈及其史料價值	0.10
UNIMARC 的發展與應用	0.10

圖書資訊學相當重視資訊科技的應用，近年來由於電腦與網路技術發達，更是直接衝擊圖書資訊學相關研究。因此研究主題(6)的詞語許多與資訊科技相關，如「network」、「web」、「technology」、「internet」、「electronic」、「網路」、「電子」、「技術」、「科技」等。另外，如何利用網路技術傳遞電子文件，改善目前資訊使用的環境，這是圖書館學相當重要的研究，因此「resources」、「bibliographic」、「documents」、「control」、「資源」、「書目」、「文件」、「網路資源」、「控制」、「傳遞」等詞語表達了這樣的意涵。因此，我們將研究主題(6)命名爲網路與電子技術相關的研究，相關論文如表八所示。《圖書與資訊學刊》在這個主題的重要作者爲陳亞寧先生，共 5 篇相關論文。

表八：研究主題(6)「網路與電子技術」所包含的論文及相關程度

論文題名	相關程度
World-Wide Web 文獻安排之探討	0.27
應用 XML 技術之電子文獻傳遞系統架構	0.25
全國學術電子資訊資源共享聯盟概況	0.22
無線網路：圖書館網路應用之新風貌	0.20
XML 與電子文件展示技術之探討	0.19
電子資訊資源、電子出版、學術傳播	0.19
網路資源之書目控制	0.19
網路科技對公共圖書館教育活動影響之研究	0.19
館際互借資源共享的新契機：台灣地區期刊聯合目錄暨館際合作系統	0.17
由各行業在 Internet 之應用探討未來圖書館服務之主要改變	0.16
從資訊科技的應用發展探討圖書館資訊服務的策略管理：以中央研究院計算中心爲個案研究	0.16
學位論文數位化之可行性研究：問題與展望	0.15
網路發展與電子媒體	0.15
網路時代之館際互借資源共享	0.15
分散式圖書館線上館際合作系統 (DILCS) 開發之研究	0.14
Using Web Resources	0.14
網路教學帶給大學圖書館的挑戰與機會	0.14
澳洲分散式系統技術中心與都柏林核心集	0.13
期刊刪除與館際合作之探討	0.13
中文全文文件群集索引理論研究與實證	0.12
淺談電子訂購的交換格式	0.12
以知識探索爲本之知識組織方法論及研究分析	0.12
網際網路的「接近使用」問題	0.12
美國資料中心之參考諮詢與推廣服務	0.12
英國文獻傳遞服務	0.11
The Library Automation and Network Development and Prospect in Taiwan Area. An Analytical Study of Information Science Curriculum in Taiwan and the United States (including Canada)	0.11
媒體出版與圖書館學整合課程模組芻議：圖書與資訊研究	0.10
建立以使用者爲中心、以網路爲基礎、服務驅動的圖書資訊工作機制	0.10
Ohio LINK 在網路環境裏對資訊獲取和資源共享的成就	0.10

最後利用多維量尺法將各個研究主題的核心集合投影到一個二維的圖形上，如圖三所示，對主題間的關係進行分析。圖三上每一個點代表一個核心集合，距離愈近的點表示核心集合間的相關性愈大，旁邊的數字則是對應核心集合的編號。以不同符號代表核心集合所屬的 6 個研究主題，在圖的最左方以星形(*)所表示的是主題(2)「學校圖書館利用教育與資訊素養」；其次，接下來在主題(2)的右方，以正方形(□)出現在圖形的左上方的是主題(3)「圖書資訊學教育與課程」；在這個主題的下方以黑點(•)表示的主題則是主題(1)「圖書館服務與館員訓練」；圖形中央，以上三角形(△)所表示的主題是「檔案學」；最後圖形右方上下，分別以菱形(◇)和交叉(×)表示主題(6)和(4)的「網路與電子技術」與「資訊檢索與使用者研究」。依據研究主題所在的橫軸位置，從左到右可以看成是《圖書與資訊學刊》的各研究主題在「教育訓練」方面(研究主題(2)、(3))、「專業管理」方面(研究主題(1)、(5))和「技術應用」方面(研究主題(4)、(6))的關係，其意義可能在於本期刊主要爲對圖書館與檔案管理的專業上各項議題與發展的探討。因此，相關研究便包含了提升使用者運用資訊能力的利用教育以及增強工作人員專業能力的專業訓練等教育與訓練相關課題；另一方面也包含各種資訊技術引進並對資訊的使用行爲進行了解等技術與應用相關課題。



圖三：以多維量尺法呈現各研究主題的關係

肆、結論

領域分析是針對學術領域進行各種學術活動的研究與分析，其中研究主題的分析，有助於對這個領域有全面性的了解。然而領域分析需要大量專家花費相當龐大的時間與金錢來進行研究，因此利用學術論文做爲分析的對象，並利用自動化的方式來取得初步的分析，較爲客觀並且可以節省許多的資源。針對這個問題，我們認爲利用論文的關鍵詞語本身所具有的語義以及共現關係，將有助於研究主題的分析。基於這樣的理念，本論文提出一系列基於自然語言處理與資訊檢索技術的研究主題分析方法。在這個方法中，首先從論文資料中抽取關鍵詞語，再依據關鍵詞語在論文資料中的共現關係辨識期刊中重要的研究主題，最後對辨識出來的主題進行分析，找出相關的論文、重要的作者與發表年份，並利用圖形呈現主題之間的關係。我們發展出來的關鍵詞語抽取技術，可以同時適用於中英文雙語的論文資料，抽取出具有意義並且完整的詞語。主題辨識的技術是整合了資訊檢索上有效的隱含語義索引和叢集分析等，將相關的詞語叢集成概念集合，辨識出期刊內重要的研究主題。圖形呈現則利用統計上的多維量尺法與電腦繪圖工具將主題繪製在二維的圖形上，便於觀察主題間的結構。我們利用《圖書與資訊學刊》第 16 到 45 期的論文資料進行分析，結果共得到 278 個中英文詞語，並且分析出本期刊重要的 6 個研究主題：(1)圖書館服務與館員訓練、(2)學校圖書館利用教育與資訊素養、(3)圖書資訊學教育與課程、(4)資訊檢索與使用者研究、(5)檔案學和(6)網路與電子技術。

對於以上的結果，除了可以提供《圖書與資訊學刊》編輯的參考之外，對於作者而言，可以了解這份期刊的意旨，與上述主題相關的論文可以投向這份期刊，而對於讀者來說，如果對圖書館的利用教育與資訊素養或是檔案學感到興趣的話，這份期刊是相當重要的資訊來源。

在進一步的研究中，近期將進行的是對於國內圖書資訊學領域的重要期刊進行全面的分析，一方面對近年來的國內圖書資訊學的研究進行一個全面性了解，找出這個領域內重要的研究主題，並與國外相關研究所得到的結果進行比較；另一方面也對各種期刊分析它們的特點，找出不同期刊所著重的不同主題，提供學科的研究人員參考。此外，在研究方法的改進，在本論文中我們利用簡易的叢集方法來區分不同概念的詞語，進一步的研究中希望更精煉詞語的語義，使其結果在研究主題分析上更有幫助；而同時也希望能詳細研究主題之間的關係繪製成圖形的方法，除了利用多維量尺法所得到的二維圖形之外，利用其他技術所得到的圖形也是下一階段發展的目標。

(收稿日期：2003 年 9 月 24 日)

參考文獻

Bawden, D. & Rowlands, I. (1999). Digital libraries: assumptions and concepts. Libri, 49, 181-191.

Borgman, Christine L. (1990). Editor's Introduction. In Scholarly Communication and Bibliometrics (pp.10-27). Thousand Oaks, CA: Sage Publications.

Chen, Keh-jiann Chen & Liu, Shing-Huan Liu (1992). Word identification for Mandarin Chinese sentences. In Proceedings of the 14th International Conference on Computational Linguistics(COLING 92).

Deerwester, Scott, et. al. (1990). Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41(6), 391-407.

Harter, Stephen P. & Hooten, Patricia A. (1992). Information Science and Scientists: JASIS, 1972-1990. Journal of the American Society for Information Science, 43(9), 583-593.

Hjørland, B. & Albrechtsen, H. (1995). Towards a new horizon in information science: domain-analysis. Journal of the American Society for Information Science, 46(6), 400-425.

Koehler, Wallace (2001). Information science as 'Little Science': the implications of a bibliometric analysis of the journal of the American society for information science. Scientometric, 51(1), 117-132.

Kowalski, G. J. & Maybury, M. T. (2000). Document and term clustering. In Information storage and retrieval systems: theory and implementation, 2nd ed., (Chapter 6, pp.139-163). Boston, MA : Kluwer Academic.

Persson, Olle (1994). The Intellectual base and research fronts of JASIS 1986-1990. Journal of the American Society for Information Science, 45(1), 31-38.

Salton, G., Wong, A. & Yang, C. S. (1975). A Vector Space Model for automatic indexing. Communications of the ACM, 18, 613-620.

White, Howard D. & McCain, Katherine 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.

林頌堅(2002 a)。基於高頻詞語的圖書資訊學研究領域分析之初步探討。中國圖書館學會會報，69，138-154。

林頌堅(2002 b)。圖書與資訊學刊論文的高頻詞語抽取與分析。圖書與資訊學刊，42，15-28.

林頌堅(2003)。從相關詞語探勘數位圖書館的概念與研究取向。在淡江大學資圖系編，2003 年資訊科技與圖書館學術論文研討會論文集(頁 1-22)。台北縣：編者。