

利用知識關聯圖達到外顯知識分享之研究

謝盛文

陳道輝

陳年興

國立中山大學資訊管理學系

國立中山大學資訊管理學系

國立中山大學資訊管理學系

onyxhsw@cc.nsysu.edu.tw

dhchen@elearning.edu.tw

nschen@cc.nsysu.edu.tw

摘要

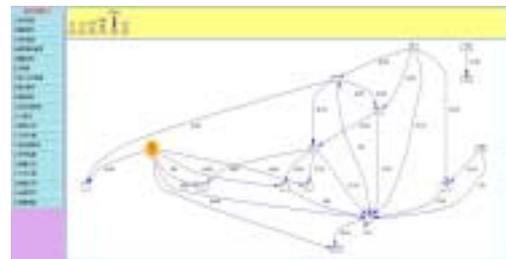
近年來由於網際網路的蓬勃發展，使得資訊的來源不但越來越豐富且更加的多樣化，但正因資料過多導致資訊整合不易，這樣不僅不能進行整體而有系統的知識分享，而且容易造成學習者更大的負擔。因此近年來透過知識關聯圖提供相關的知識字詞與知識字詞間的關聯，使其瞭解相關知識字詞的關聯性，將可以有效的降低學習迷失現象與避免資訊超載以達成知識分享的相關研究越來越多，然而可以發現在建立知識關聯圖的過程中，存在兩個問題，一個是分析的內容不足且關聯程度缺乏統一的衡量標準；另外是研究人員往往只能依靠人工或者半自動的方法對於研究主題進行編碼與分類。因此本研究針對上述二點提出新的解決方式，本研究將修改之前研究者的作法，除了以利用論文之標題、關鍵字與關鍵字出現之頻率外，本研究將利用論文中的摘要內容建立知識關鍵字之間的關聯強度及相關知識關鍵字的關聯相似度。最後依據關聯強度與關聯相似度，實作自動化建立知識關聯圖的系統。

關鍵詞：知識關聯圖、外顯知識、資料挖掘、知識分享、資訊管理

一、研究背景與動機

近年來由於網際網路的蓬勃發展，使得資訊的來源不但越來越豐富且更加的多樣化。使用者在現今網路環境中，往往可以透過各種線上資料庫或搜尋引擎的協助，取得大量的資訊以獲取知識。然而，大量的資訊往往會造成資訊超載，資訊超載是指在網際網路的學習環境中，有著大量且豐富的資訊，能夠提供學習者豐富資料，但資料過多將導致資訊整合不易。因為學習者在網際網路上蒐集資料時，得到的往往是片段而零散的資料，所接受的資訊遠超過學習者所能吸收的程度，這樣不僅不能取得整體而有系統的知識，而且容易造成學習者更大的負擔。且過多片段而零散的資訊，更往往會造成使用者的迷失，導致使用者要花更多的心力進行閱讀與整合。所謂迷失是指學習

者在網際網路或多媒體等學習環境中，學習者對其學習的位置與方向失去概念的一種傾向，大致可分為「不知道下一步要去哪裡」、「知道要去哪裡，但不知怎麼到達」、「不知道現在自己的位置」三種（陳明溥、莊良寶，1999）。迷失對學習效果有相當重要的影響，因為學習者在學習的過程中會在繁多的資訊中尋找答案，但對其所學習的知識字詞沒有一個整體的概念，往往會迷失在探索的過程中，而偏離原先的主題，自然會對學習造成嚴重的阻礙。此外，迷失會造成學習者產生不安的感覺，容易造成學習者對學習歷程的沮喪，進而對學習內容產生恐懼感、降低學習意願，甚至不願意學習（陳明溥、嚴榮泉，2001）。所以如何運用系統化的方式，協助使用者更快、更有效率的從大量網路資源中獲得真正所需的資訊與知識，與及時的提供學習者相關的知識字詞與知識字詞間的關聯，使其瞭解相關知識字詞的關聯性，將可以有效的降低學習迷失現象與避免資訊超載，以達到有效的知識分享。



圖一、以資料探勘為研究主題之知識關聯圖

然而為了更明確的界定本研究知識分享的範圍，我們必須對於知識的本質有更深入的了解，學者 Alavi&Leidner (2001) 建議透過四種知識的特性來了解知識，包含知識能顯現的清楚程度、儲存的單位、現象的瞭解以及實務的運用。由於本研究的目的在於運用系統化的方式，有效率的從大量網路資源中獲得真正所需的知識，並提供學習者相關的知識字詞與知識字詞間的關聯，以有效的降低學習迷失現象與避免資訊超載。因此以知識能顯現的清楚程度而言，本研究是屬於『外顯知識(Explicit knowledge)』（Nonaka, 1994）。對於儲存的單位而言，本研究是屬於『可編碼的知識』。對於現象的瞭解則著重於『關係性的知識』。實務的運用則可以用來做為『設計藍圖』（林東

清, 2002)。對於此種利用可編碼的外顯知識並強調關係性所形成的設計圖, 稱之為知識關聯圖, 圖一是以資料探勘為主之研究主題的知識關聯圖。由於知識關聯圖的產生可以有效的降低學習迷失現象與避免資訊超載, 以達到有效的知識分享。因此近年來學術界有相當多針對知識關聯圖的相關研究(Schultze、Leidner, 2002; He、Hui, 2002; 謝武星, 2000; 楊存一, 2001; 張家華, 2001; 陳年興、孫振凱, 2002; 王惠嘉, 張毓倫, 2002; 李律品, 2002)。其中國內大部分的研究都將全國博碩士論文網(<http://datas.ncl.edu.tw>)的相關資料視為可編碼的知識(例如: 論文之標題、關鍵字、引用文獻內容等), 來探討相關知識也就是關鍵字的關係性。然而這樣的研究僅限於分析論文的標題、引用文獻內容(謝武星, 2000), 或是僅分析論文關鍵字共同出現之頻率(陳年興、孫振凱, 2002), 可能稍嫌不足; 而且這些分析活動往往需要藉由大量的研究人員或是利用半自動的方法對於研究主題進行編碼與分類(王惠嘉, 張毓倫, 2002; 楊存一, 2001)。且進行編碼的過程中由於缺乏各項研究主題之間關聯程度的統一衡量標準, 研究人員往往只能依靠自身的判斷, 或是透過閱讀非常大量的文獻來了解研究主題之關聯程度。在這樣受限制的情況下, 使得研究人員在建置知識關聯圖的過程中, 除了造成非常大的不便外, 更可能無法建立正確的知識關聯圖。

整合上述問題, 可以發現在建立知識關聯圖的過程中, 存在兩個問題, 第一個是分析的內容不足, 且關聯程度缺乏統一的衡量標準; 第二是研究人員往往只能依靠人工或者半自動的方法對於研究主題進行編碼與分類。因此本研究針對上述二點提出新的解決方式, 本研究將修改之前研究者的作法, 除了以利用論文之標題、關鍵字與關鍵字出現之頻率外, 本研究將利用論文中的摘要內容, 這是因為學術論文的撰寫, 除了論文本文之外, 尚包括其他的論文資訊, 如摘要、關鍵字等內容。其中最能代表整篇論文的研究主題的便是論文的關鍵字與摘要。在選擇關鍵字時, 作者往往會根據論文之題目與研究範圍, 使用最精簡的詞彙來描述研究的主题; 同樣的, 在撰寫摘要時, 作者也會根據論文的內容與成果, 在最短的篇幅內寫出研究的成果為何。因此, 研究人員在閱讀學術論文時, 必定會先檢視關鍵字與摘要的部分, 以期迅速了解該學術論文的研究主題與大致內容, 並作為接下來閱讀論文本文的輔助; 而由於論文摘要與關鍵字的重要性極高, 又可代表整篇論文的研究內容, 取得也較容易, 因此, 欲針對學術論文進行各項知識管理活動時, 若從論文的摘要與關鍵字著手, 將可收事半功倍的效果。所以本研究將透過計算知

識關鍵字出現於摘要位置與兩兩知識關鍵字於摘要的相對距離, 從中擷取出知識與知識之間的相關距離, 並藉由相關距離的分析, 建立知識關鍵字之間的關聯強度及相關知識關鍵字的關聯相似度。最後由於關聯強度與關聯相似度的建立, 將可使得系統依據關聯強度與關聯相似度, 自動化建立知識關聯圖。

二、相關文獻

(一) 知識的分類

學者 Alavi&Leidner (2001) 將知識作以下幾種不同的分類, 並建議以此來了解知識的特性, 包含知識能顯現的清楚程度、儲存的單位、現象的瞭解以及實務的運用。知識能顯現的清楚程度, Nonaka(1994)首先將知識分為內隱知識與外顯知識。外顯知識其定義是指可用文字、數字、圖形或其他象徵物(Symbol)來清楚表達(如手冊、書本、程式)的知識, 是可定義、可擷取的知識, 而且溝通容易。內隱知識其定義是指高度個人化, 難以正式化只可意會, 而且深植在個人的經驗、判斷、價值、潛意識與心智模式。

儲存的單位, 根據 Blackler 將知識分成五個層次(林東清, 2002), 分別是內含於員工的觀念型知識、內含於員工的經驗型知識、內含於組織文化的知識、內含於組織運作的知識、可編碼的知識。內含於員工的觀念型知識是指內含於員工的腦海裡和潛意識裡, 例如一些觀念化、抽象化、直覺式的知識。內含於員工的經驗型知識是屬於員工的, 例如透過邊做邊學的訓練及練習。內含於組織文化的知識是指組織內含有優良文化、共識和價值觀, 這些東西是員工帶不走而內含在組織裡, 例如優良的團隊精神、忠誠度高等等。內含於組織運作的知識是指內含於例行的作業流程或人際社會間的互動關係, 例如有效率的生產線、有效率的顧客服務等。可編碼的知識, 這是知識的最外層的部分, 是由文字、數字等資訊來傳達、顯示的知識, 例如書、系統手冊、視聽教材等等。由於本研究的來源是全國博碩士論文網其本是就非常類似手冊或者書本, 因此全國博碩士論文網的知識能顯現的清楚程度是屬於外顯知識, 且儲存的單位是屬於可編碼的知識。現象的瞭解, 由於 knowledge 是由 know 所發展出來, 因此可將知識視為一種對於問題的知曉能力, 對於現象的瞭解與問題的知曉能力一般可以從 5 個 W 的思考(林東清, 2002)。分別是描述性知識(Know-what)、程序性知識(Know-how)、因果性知識(Know-why)、情境性知識(Know-when)與關係性知識(Know-with)。由於本研究的目的在於瞭解知

識與知識之間關聯的程度，所以本研究對於現象的瞭解是屬於關係性知識。實務運用方面的分類，這種分類主要是針對組織經營有價值的知識，以其不同存在方式與相關性分類，這種分類法可將企業內的知識歸內為企業流程、經營模式、科技架構、分析報告、專案經驗、企業架構、專利權、軟體程式、經驗、最佳實務、工作計劃、專案與設計藍圖等(林東清, 2002)。由於本研究最後會將研究主題之間的關聯強度與關聯相似度呈現給使用者，提供新進研究人員或是初次接觸學術領域的學生一個參考的指標與引導方向，並透過知識關聯圖的呈現，將學術領域相關的知識以概念節點及連結關係來加以呈現，因此實務運用方面是屬於設計藍圖。對於此種利用可編碼的外顯知識並強調關係性所形成的設計圖，稱之為知識關聯圖，下面將介紹知識關聯圖的相關研究。

(二) 知識關聯圖的相關研究

目前國內外學界所對於對學術界的知識管理，均有提出許多作法。其中 Schultze & Leidner (2002) 將六本資訊管理領域最富盛名的期刊中，94 篇關於知識管理的論文或文章選出，並且利用人工編碼與分類，將這些文章分為四個類別，分別為 Normative Discourse、Interpretive Discourse、Critical Discourse 與 Dialogic Discourse，並且對未來知識管理的研究提出建議與看法。He & Hui (2002) 將 1987 至 1997 年關於資訊擷取的 1466 篇期刊論文中的引用文獻取出，進行作者共同引用文獻之分析，並且利用分群技術及多變量分析等方式，建立資訊擷取技術領域的作者群組，並且可以讓使用者快速了解與某作者相近領域之作者與期刊文章。謝武星(2000)應用了特徵擷取、中文字詞分析、統計及演算等方法，針對中文學術論文的參考文獻及標題進行分析，透過專業詞彙資料庫與虛字符號資料庫的建立，擷取最適合描述論文研究主題的詞彙；並以中文文獻及字詞的資訊化處理此研究主題的論文為例，建立系統以驗證其所提出的作法，也達到了良好的效果。張家華(2001)透過文獻調查法與內容分析法，針對資訊管理研究所的學位論文所援用之理論進行分析，以了解台灣資訊管理學界領域近三年的研究重點與各個學校的發展方向；並透過前述資訊，建構國內資訊管理學界的知識地圖，期望對於未來研究者在搜尋文獻資料時有一個良好的方向。這些研究的出現，顯示目前針對知識關聯圖的相關研究有日益增多的趨勢。李律品(2002)將國外著名之資訊管理領域期刊之內容，針對作者、學校、標題、關鍵字等欄位加以編碼，將資訊管理領域之期刊論文加以分類，並以系統化的方式描繪資訊管理領域的研究發展歷程，並且

依研究者研究議題的差異，進行資訊管理研究生態演化之分析。陳年興與孫振凱(2002)透過網際網路上搜尋引擎的功能，針對全國碩博士資訊網進行關鍵字詞的擷取，並分析關鍵字詞出現在網頁上的關聯，建構網際網路上的知識分佈圖，以輔助線上教學，讓學習者探索網際網路上的資料時能夠更有效率，避免學習者發生認知迷失與資訊超載的情況。楊存一(2002)利用類神經網路之自適應共振理論網路，對資訊管理的碩博士論文進行分類，並進行關鍵議題的找尋及排名，以了解國內資訊管理學界關鍵議題之發展趨勢，同時讓研究者了解過去該議題被研究的情況。但目前關於博碩士論文資料的研究，僅限於分析論文之標題、引用文獻內容，或是僅分析論文關鍵字共同出現之頻率，而不考慮關鍵字出現位置；或是需透過人力對論文內容或研究方法進行編碼，無法達成自動化擷取、分類的功能。由於目前的相關研究，均有其不足之處，本研究希望能改進之前的作法，分析論文關鍵字在摘要之中出現的位置，以建立關鍵字之間的關聯強度與關聯相似度。

三、系統設計與實作

首先，為了分析研究主題之間的關聯性，本研究提出了兩個衡量指標：關聯強度與關聯相似度，這兩個指標均是從論文內容所產生。接下來本研究將實作系統，首先透過自動擷取與分析程式，將網際網路上之資訊相關領域(包含資訊管理、資訊工程、資訊科學等)的學位論文資料進行分析與擷取。之後運用這兩個指標，根據論文內容中的摘要與關鍵字，計算研究主題與其他研究主題之關聯強度與關聯相似度，讓使用者了解兩研究主題之間的關聯性，並且以網頁的方式將最後計算結果呈現給使用者。最後，本研究將評估本研究所提出之關聯強度與其他關聯計算方法之不同，同時，根據計算出來的關聯強度與關聯相似度，依照不同年度、不同來源、不同學校等分類結果，探討研究主題之間的關聯性為何，並對結果提出分析以及評估，以描述主題之關聯性。

(一) 研究主題關聯衡量指標的三個基本原則

由於本研究所提出之衡量指標是根據論文內容計算而來，因此本節將先簡介學術論文的內容格式，並且根據論文內容與格式，提出本研究之基本原則。最後再根據基本原則，提出本研究之衡量指標：關聯強度與關聯相似度。由於學術論文在撰寫上均具有一定之格式，以符合學術界的要求。以博碩士學位論文為例，通常包含以下的內容，在將論文送交國家圖書館典藏時，也必須附上以下的資訊，供

建檔及往後查詢使用：論文題目、作者、指導教授、學年度、學校、系所名稱、中英文摘要、關鍵字、目錄、論文本文、參考文獻、附錄等。而在這些內容與資訊中，除了論文本文與論文題目之外，以「摘要」與「關鍵字」最能代表一篇論文之研究主題。相較於論文本文，「摘要」與「關鍵字」內容均非常簡短，在極短的時間內即可閱讀完畢；而研究者在閱讀論文時，也經常藉由閱讀論文之摘要與關鍵字，快速的了解論文的大致內容，如方法、流程、結果等，以決定是否要對論文本文詳加閱讀。因此，若要針對學位論文進行知識管理，可以藉由分析論文內容中最能代表論文研究主題的「關鍵字」與「摘要」來進行。所以，本研究也將針對論文中之「關鍵字」與「摘要」兩個必備內容進行分析與處理，論文中提出之衡量指標也由這兩個內容發展而來。由於論文作者在選擇關鍵字時，會選擇與該論文研究最相關的詞彙，如研究的方向、使用到的理論基礎、進行研究的方法等；而作者會選擇這些關鍵字，代表在這些關鍵字之間，必定存在著研究上的關係，作者才會將這些關鍵字同時放在一篇論文的關鍵字中。因此本研究提出以下兩項原則：一篇論文的每個關鍵字均代表著論文的研究主題之一、一篇論文的每個研究主題，兩兩之間必定有研究上的意涵存在。根據這兩項原則，本研究便可以依照這些假設了解哪些研究主題之間具有關聯性存在，並依此計算兩者之關聯強度與關聯相似度。此外，由於中文文法的特性，通常在寫作或是日常對話中，兩個名詞若在文字中的距離越接近，且如果此種關係出現的次數相當頻繁的話，則兩者的關係也會越強烈。因此，本研究提出第三項原則：整體而言若兩個名詞之間的平均距離越近，則兩者的關聯性也越高。以圖二為例，網際網路與資訊科技兩名詞之間的距離為 72 個字，而網際網路與再仲介化之距離為 92 個字。由於網際網路與資訊科技的距離較再仲介化為近，且如果說存在這樣關係的文章越多，則本研究認為網際網路與資訊科技之關聯性，比網際網路與再仲介化來的高。根據前述三項原則，本研究提出衡量兩研究主題關聯性之指標，分別為「關聯強度」與「關聯相似度」，前者可衡量研究主題之直接關聯性，後者則可衡量研究主題之間接關聯性。



圖二、距離越近的兩名詞，其關聯性也越高

(二) 關聯強度(R)

陳俊彰(2001)曾提出聯繫強度衡量方法，計算兩位老師之間的關係強度，以建構教師社會網路圖，如公式 1。本研究乃根據此衡

量方法，經過修正之後提出兩研究主題之間的關聯強度計算公式，以衡量兩研究主題之間直接關聯性。修正之後的公式如下(2)，接下來將對本研究所提出之公式做詳細的說明。

$$強度 = \left[\sum (\text{關係強度分數} \times \text{關係發生頻率}) \right] \times \text{關係發生類別} \quad (1)$$

$$R(k1, k2) = \frac{1}{(\text{norm}(\text{avg_dist}))^2} \times \text{norm}(\text{count}) \times \text{type} \quad ..(2)$$

R = 兩研究主題之關聯強度, norm = 標準化函數

avg_dist = 兩研究主題在論文摘要中之平均距離

count = 出現頻率, type = 出現類型

本研究之關聯強度計算公式乃由聯繫強度衡量方法修正而來，原公式中之關係強度分數即為兩研究主題在摘要中之平均距離，而關係發生頻率為本關聯在摘要中的出現頻率，關係發生類別則為本關聯出現類型。根據本研究之第二項原則，一篇論文的每個研究主題，兩兩之間必定有研究上的意涵存在，因此若兩研究主題不曾同時出現在一篇論文中，則表示兩研究主題沒有研究上的意義存在，因而兩研究主題將沒有關聯強度值；而根據本研究之第三項假設，兩研究主題在摘要中的距離越近，則兩者之關聯強度應越高，同時強度的變化程度也應是成指數倍數變化，而非線性變化，故本研究先將兩研究主題之間的平均距離轉換為 Z 分數；而 Z 分數有可能為負值，所以本研究再將 Z 分數轉換為平均數為 5，標準差為 1 之標準化分數，以確保標準化分數為正值。最後，再以標準化分數平方之倒數為關係強度分數。舉例來說，假設「網際網路」與「電子商務」曾經出現於 32 篇博碩士學位論文中，而平均距離為 86.7；轉換為標準化分數之後的出現次數為 7.5，距離為 5，則此兩研究主題之關聯強度為 $1/5 \times 7.5 \times 1 = 1.5$ 。關聯強度越高，表示兩研究主題在摘要中的平均距離越短，同時也表示兩者同時出現在摘要中的次數也越多，則兩研究主題之間的直接聯繫程度越強，兩研究主題的研究程度越密切；反之，關聯強度越低，表示兩研究主題在摘要中的平均距離越長，同時也表示兩者同時出現在摘要中的次數也越少，則兩研究主題之間的直接聯繫程度越弱。

(三) 關聯相似度(S)

在過去的研究中，Jaccard 係數是在衡量交易資料集時最為廣泛使用的相似度量測標準 (Guha, Rastogi & Shim, 1998)。Jaccard 係數亦稱為 IOU 指標 (Intersection Over Union measure)，假設兩物件 a, b 所屬交易之集合分別為 X 與 Y，則 Jaccard 係數分子為 X 與 Y 交集之大小，分子為 X 與 Y 聯集之大小。公式如下。舉例來說，假設物件 a 出現於交易 T1, T2, T3, T4 中，物件 b 出現於交易 T2, T3, T5, T6 中，則 a 與 b 之 Jaccard 係數為 $2/6=0.33$ 。Jaccard

係數介於 0 與 1 之間,0 表示兩者完全不相同, 1 表示兩者完全相同。Jaccard 係數主要的計算方式是觀察物件共同出現的情形,而忽略物件沒有出現的情形。由於 Jaccard 係數乃是計算兩物件所屬之交易資料集之間的相似度,所以若將論文之關鍵字視為一個物件,每個與該關鍵字具有關聯存在的關鍵字視為一次交易,那麼便可利用 Jaccard 係數的概念,計算兩研究主題之間的關聯相似度。本研究乃根據 Jaccard 係數公式,來做為衡量兩研究主題之間的關聯相似度的公式。因此關聯相似度越大,表示研究主題之間重疊的研究主題越多,則兩者的間接聯繫強度越強,兩研究主題之關聯研究主題越相似;反之,若關聯相似度越小,則兩者之間接聯繫強度則越弱,代表兩研究主題之關聯研究主題越不相似。

$$S = \frac{|X \cap Y|}{|X \cup Y|}, S = \text{關聯相似度} \dots\dots\dots(3)$$

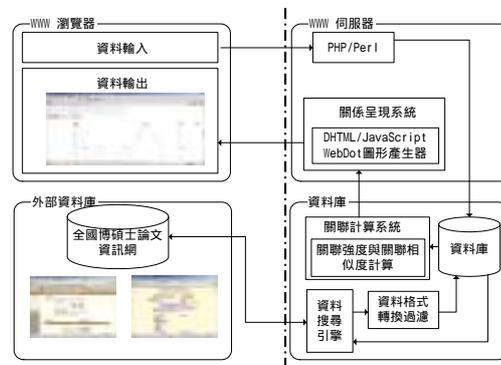
四、系統建置

為了評估本研究所提出之研究架構及衡量指標,本研究透過實際建置系統以實作關聯強度與關聯相似度兩指標。圖三為本研究提出知識關聯圖系統架構圖。系統架構圖,本系統之主要流程可分為四個階段:論文資料擷取階段、關鍵字處理階段、衡量指標計算階段及結果呈現階段。由於篇幅關係,簡要各階段之內容分述如下。階段一為論文資料擷取階段,自動化擷取國內資訊相關系所的博碩士畢業論文,並且自動擷取論文的各種屬性,如標題、校系、作者、關鍵字、摘要等。階段二關鍵字處理階段,由於國內並未強制規定博碩士論文可以使用的關鍵字為何,造成關鍵字數量過於龐大,因此本階段將對論文所使用的關鍵字進行處理,將具有相同意義與意義相似的關鍵字進行合併,並且設定門檻值過濾使用率較低的關鍵字,最後的結果將作為衡量指標計算階段與結果呈現階段中,相同或相似研究主題合併之用。階段三衡量指標計算階段,針對每一篇論文,先行計算衡量指標所需的數值資料,如關鍵字配對個數、關鍵字之間的距離等,以提升最後結果呈現階段的指標計算速度。階段四結果呈現階段,根據使用者所輸入之各種參數,如年份、學校、系所等,實際計算關聯強度與關聯相似度,並且將最後結果透過網頁呈現給使用者。

五、評估與分析

在本章中,首先將針對本研究所提出之關聯相似度,以資料挖掘中之關聯法則進行評估,並與本研究所提出之關聯強度與關聯法則

產生之規則進行比較。接下來則將針對單一研究主題,以本研究所提出之關聯強度與關聯相似度分析其他相關研究主題之關聯性強弱。同時,並以不同的面向,如年份、學校、科系等參數,分析在各種面向下研究主題關聯性強弱之變化,並藉此看出研究主題之主流與發展趨勢。



圖三、知識關聯圖系統架構圖

(一) 關聯強度結果評估

本研究所提出之關聯強度,是衡量兩個研究主題直接強度之指標,使用者可以根據衡量指標之大小了解兩者關係的強弱;而根據本研究之假設:「一篇論文的每個研究主題,兩兩之間必定有研究上的意涵存在」,若兩研究主題之間有關聯強度存在,則兩者必定同時出現於某一篇文章的關鍵字欄位中。因此,若要對關聯強度進行評估,可以利用其他演算法檢驗兩研究主題之間是否有關聯存在,與關聯的強弱大小。一篇論文通常具有多個關鍵字描述論文主題。若將一篇論文視為一次交易,而將論文中的關鍵字視為一次交易所購買的項目,則便可藉由資料挖掘中之購物籃分析,利用關聯法則之方法,分析是否有研究主題出現關聯法則,若有研究主題之間有關聯法則存在,則表示兩者的關聯性極高。此外,若有關聯法則出現,亦可分析關聯法則之有用性指標值與確定性指標值,以評估關聯強度之正確性。本研究乃根據購物籃分析之概念,透過關聯法則評估關聯強度。評估流程如下,首先將論文與其關鍵字之資訊轉換為相似於交易記錄之格式。透過資料挖掘軟體,尋找其中是否有規則存在。分析執行結果,並將執行結果與關聯強度進行比較。本研究透過 DBMiner 公司所推出之 DBMiner AX 2002 尋找資料中是否有關聯法則的存在。DBMiner AX 2002 具有高效能的關聯法則挖掘能力,同時可針對單一維度或是多維度資料進行關聯法則挖掘。針對挖掘時的參數值,在設定上如下, Minimum Support (有用性指標): 1%、Minimum Confidence (確定性指標): 25%、Minimum Lift: 1。根據前述之參數設定,最後執行結果發現:即使在所有參數值均是最低的情況下,

DBMiner AX 2002 依然沒有找出任何一條規則。發現此結果之後，本研究嘗試將論文資料依照系所別分開處理，希望藉由系所的分別，找出該戲所中具有高強度的研究主題關聯性。將不同系所的論文分開處理之後，於資訊管理系所中僅產生了一條規則，其他的系所則無規則出現。產生的規則如表一。

表一、由 DBMiner AX2002 所產生之關聯法則

編號	IF	Then	有用性指標	確定性指標	Lift
1	網際網路	tokent1?	1.59	25.34	4.23

因此將論文視為交易，將關鍵字視為交易中的項目，並以此設定透過資料挖掘軟體進行關聯法則挖掘是可行的，所產生的結果也與本研究所提出之關聯強度相類似。但是，即使將有用性指標、確定性指標等參數值調到最低，產生出的規則仍然非常非常之少，僅有在資訊管理系所產生出一條規則，剩下之系所均無任何規則存在。相較於本研究所提出之關聯強度，可以將所有有關聯強度存在之研究主題取出，使用者可藉由已排序之輸出資料，了解兩研究主題之間的關聯程度，以及以多個面向，如不同的時間、學校、系所等，了解在這些面向下哪些主題具有高度的關聯性。因此，相較於一般傳統的資料挖掘方法，本研究所提出之關聯強度更適用於衡量學術論文中研究主題之關聯性。

(二) 單一研究主題之結果分析

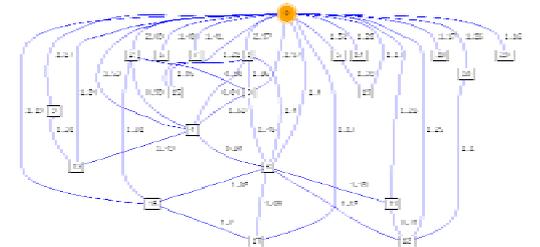
由於本系統提供使用者兩種不同的查詢功能，因本研究也將針對兩種不同的執行結果進行分析。本節將先針對「單一研究主題查詢」之查詢功能，由於篇幅有限僅以資訊管理最熱門的「知識管理」研究主題為例進行查詢，並且針對執行結果進行分析，在下一節則將針對「整體性研究主題查詢」之查詢功能，分別以不同之年份、系所、學校等面向進行查詢，並且針對執行結果進行分析。以「知識管理」為研究主題，查詢資訊管理學系學位論文中與「知識管理」相關之研究主題關聯強度前 20 名，可以得到如表二之結果，其中 R 為關聯強度、S 為關聯相似度、A 為平均距離、N 為文章總數，而這 20 個研究主題之間的交互關聯強度圖如圖四。

從圖四中可以看出此 20 個研究主題之關聯程度，其中，有 6 個研究主題與編號 4 資料倉儲有關聯強度存在，有 7 個研究主題與編號 8 組織學習有關聯強度存在，表示這 20 個與知識管理相關的研究主題，許多均與資料倉儲與組織學習有關。然而資料倉儲與組織學習之間雖然有關聯強度存在，兩者的卻關聯強度僅有 0.84，代表兩者之間的直接關係不高，因此由圖四的內容，本研究推論：與知識管理相關的研究，也與組織學習與資料倉儲有著密切的

關係。而從表二中可以發現，與知識管理關聯強度最強之研究主題為知識管理系統等 4 個同義或相似關鍵字，出現在論文中的頻率也最高，共有 33 篇論文。而兩者的關聯相似度也是這 20 個中最高的，為 10.58%。這兩者之間詳細的關聯強度、關聯相似度之值與論文資料，則如表三。若使用者想深入了解「知識管理」與「知識管理系統架構」之研究，本系統也包含了詳細的學術論文資料與指導教授資料，使用者可以藉此尋找學位論文，或是尋找該領域之專家，如表四。

表二、與知識管理關聯強度前 20 之研究主題

編號	研究主題	R	S	A	N
1	知識管理系統,知識管理系統,知識管理系統,知識管理系統	2.9	10.58%	111.6	3
2	資訊科技	2.8	4.17%	236.4	1
3	網路管理,網路管理,網路管理,網路管理,網路管理,網路管理	1.8	3.95%	82.1	10
4	網路管理,網路管理,網路管理,網路管理,網路管理,網路管理	1.5	3.76%	363.1	4
5	資訊管理	1.5	3.76%	464.5	1
6	企業學習,企業學習,企業學習,企業學習,企業學習,企業學習	1.4	3.71%	44.5	1
7	知識管理	1.4	3.14%	94.5	1
8	知識管理,知識管理,知識管理,知識管理,知識管理,知識管理	1.4	3.14%	147.2	6
9	知識管理,知識管理,知識管理,知識管理,知識管理,知識管理	1.2	4%	411.9	2
10	企業學習	1.2	3.88%	199.9	1
11	資訊管理	1.2	3.88%	159.4	2
12	知識管理,知識管理,知識管理	1.2	3.88%	177.6	6
13	知識管理	1.2	3.17%	159.4	2
14	知識管理	1.2	3.0%	185.9	2
15	知識管理	1.2	3.52%	115.2	2
16	知識管理,知識管理,知識管理	1.2	3.49%	143.2	3
17	資訊科技應用	1.1	3.17%	143.2	2
18	知識管理	1.1	3.51%	195.4	4
19	資訊管理	1.1	3.76%	143.2	1
20	知識管理	1.1	3.76%	189.7	1



圖四、以知識管理為研究主題之知識關聯圖表三、「知識管理」與「知識管理系統」及同義與相似關鍵字之詳細關聯資料

關鍵字 1	關鍵字 2	關聯強度	關聯相似度	論文篇數	平均引文次數
知識管理	知識管理系統	2.97	6.8%	33	88 學年度: 5 篇 89 學年度: 5 篇 90 學年度: 18 篇
知識管理	知識管理系統,知識管理	1.19	1.9%	1	90 學年度: 1 篇
知識管理	知識管理系統,知識管理	1.63	2.8%	4	88 學年度: 1 篇

表四、「知識管理」與「知識管理系統架構」之詳細論文與指導教授資料

關鍵字 1	關鍵字 2	論文名稱	指導教授
知識管理	知識管理系統架構	網路資訊管理與多媒體之研究 (88)	張清河 (1)
		企業知識管理系統之建置 (88)	張清河 (1)
		以網路導向式資料倉儲為基礎的知識管理系統 (88)	曾中全 (1)
		知識管理系統架構之比較研究 (88)	張清河 (1)

(三) 整體性研究主題之結果分析

查詢「整體性研究主題」時，可以依據不同的年份、不同的學校與不同的科系等三個面向進行查詢。本節中也將依據這三個面向分別進行結果分析與探討，依順序分別為：根據不同年份進行查詢、根據不同學校進行查詢、根據不同系所進行查詢。透過系統查詢 81 學年度至 90 學年度之全國資訊管理學系、所研究主題之間的關聯強度與關聯相似度，可以得到如近十年的結果。由於篇幅有限僅列出 88-90 年研究主題之間關聯強度之前五名。從表中可以看出，88 學年度中網際網路與電子商務的研究到達最高峰，兩者的關係強度高達 5.77。

此時，知識管理與知識管理系統、供應鏈與供應鏈管理之研究也開始增加。89 與 90 年兩者並無太大的不同，網際網路與電子商務之關聯強度仍然名列前茅，但到了 90 學年度已開始下降；而 89 學年度開始資料倉儲與「資料挖掘」之研究開始大量增加，這種現象代表關於「資料倉儲」與「資料挖掘」之相關研究已成為現今資訊管理學界的熱門主題之一。

表五、88-90 年研究主題之間關聯強度前五名

學系	排名	關聯字 1	關聯字 2	數	%
網路	1	資訊科技	網際網路	3.73	37%
	2	電子商務	網際網路	3.47	34%
	3	資料庫	資料探勘,資料挖掘,網際網路,資料倉儲	3.28	32%
	4	資訊系統	資料庫	3.00	30%
網路	1	網際網路	網際網路	3.73	37%
	2	電子商務	網際網路	3.08	30%
	3	資訊科技	網際網路	4.1	41%
	4	網際網路	網際網路	3.79	37%
網路	1	資料倉儲,資料倉儲系統	資料探勘,資料挖掘,資料倉儲	3.71	37%
	2	資料倉儲	網上資料,網上資料處理	3.70	37%
	3	電子商務	網際網路	3.77	37%
	4	網際網路	網際網路	3.44	34%
網路	1	知識管理	知識管理系統,知識管理系統,知識管理	3.02	30%
	2	網際網路	電子商務	3.42	34%
	3	網際網路	網際網路	3.42	34%
	4	網際網路	網際網路	3.42	34%

透過本系統，分別查詢中山大學、中央大學、台灣大學、政治大學等國內四所歷史悠久且具有高知名度之資訊管理研究所，學位論文之研究主題關聯強度與關聯相似度，可以得到如表六的研究主題關聯列表，表中僅列出研究主題前五名。從表中可以發現，在四所學校的資管系中，電子商務與網際網路之研究仍然佔最大宗。除台大資管系外，在其他的學校均位居關聯強度第一名。中山大學資管系中具有高關聯強度之研究主題頗為平均，同時也是四所學校中有關組織變革與資訊科技關聯強度排名最高者。中央資管系中具有高關聯強度之研究主題，前三名均與網際網路有密切關係。同時，也是四所學校中關於資料挖掘與關聯規則強度最強者，顯示中央大學進行相當多關於資料挖掘之研究。政大資管系中具有高關聯強度之研究主題中，供應鏈與供應鏈管理排名為第二，僅次於電子商務與網際網路之後；同時，排名第三與第四者均與資料庫有關，可見政大資管研究著重於供應鏈與資料庫的相關研究。台大資管系中具有高關聯強度之研究主題與其他三所學校頗為不同。排名第一的為服務品質與網際網路，而其他學校均為電子商務與網際網路，可見台大資管系之研究著重於網路服務品質的部分；此外，排名第四的拉格蘭氏鬆弛法與最佳化，此關聯幾乎不存在於其他三校中，同時其他三校的研究主題與台大資管系的重複率甚低，可見台大資管系的走向與其他學校的不同。從同一學校不同系所的查詢，更可看出同為一間學校內之資訊相關領域系所，其研究方向之不同。本研究分別以中山大學之資訊工程與台灣大學之資訊工程兩系所為例進行查詢，則可得到之結果如表七，中山資管所之結果請參考表六。從表六和表七中，可以很明顯的看出中山資管所與中山資工所研究主題上的差異。中山資管所之主要研究主

題與整體資管界並無太大不同；而由表七中，可以看出中山資工所與全國所有資訊工程系所研究主題的差異，網際網路在中山資工所並不是一個熱門的主題，研究主題關聯強度前五名中僅有第二名包含網際網路；相對的分散式與相關之研究在中山資工所中是較為熱門之研究主題。由表六和表七中亦可看出台大資工所與台大資管所之研究方向差異，台大資管所與整體資訊管理學界的研究方向有很大的不同，而台大資工所之研究方向與整體資工系所研究雖然仍有不同，但差異不如台大資管來的大，網際網路在台大資工所之研究方向上仍佔有一定之比率。而從表七的內容中，可以看出台大資工所中排名第三之「資料探勘」與「頻繁型樣」之相關研究在其他資訊工程系所中甚少出現，同時兩者之關聯強度與關聯相似度都很高，因此可見台大資工所之主要研究方向包含了資料探勘中發掘頻繁型樣之方法。

表六、中山資管、中央資管、台大資管、政大資管研究主題之間關聯強度之前五名

學系	排名	關聯字 1	關聯字 2	數	%
中山資管	1	電子商務	網際網路	3.50	37%
	2	資訊科技	網際網路	3.73	38%
	3	商業國際化學習法	網際網路	2.83	29%
	4	網際網路	網際網路	3.24	34%
	5	知識管理	資料倉儲	3.24	34%
中央資管	1	電子商務	網際網路	3.83	33.3%
	2	知識管理	網際網路	3.21	27.5%
	3	網際網路	網際網路	2.89	24.7%
	4	應用資訊系統,應用資訊系統	網際網路	2.46	21.1%
	5	資料探勘,資料探勘,資料探勘	網際網路	2.88	24.7%
政大資管	1	電子商務	網際網路	3.93	33.7%
	2	網際網路	網際網路	2.84	23.7%
	3	資料庫	資料庫	2.47	20.6%
	4	資料庫	網際網路	2.23	18.6%
	5	知識管理,知識管理,知識管理	網際網路	2.23	18.6%
台大資管	1	服務品質,服務品質,服務品質	網際網路	3.50	33.3%
	2	電子商務	網際網路	3.47	32.6%
	3	資訊科技	網際網路	3.61	33.8%
	4	知識管理,知識管理,知識管理	網際網路	3.30	30.9%
	5	資料庫	資料庫	2.30	21.8%

表七、中山資工、台大資工研究主題之間關聯強度之前五名

學系	排名	關聯字 1	關聯字 2	數	%
中山資工	1	資料庫	資料庫	3.87	40%
	2	知識管理	網際網路	3.62	37%
	3	分散式系統	遠端操作,遠端操作	3.86	40.0%
	4	分散式	分散式	3.80	39%
	5	修正結構,修正結構,修正結構	修正結構,修正結構,修正結構	3.72	38.5%
台大資工	1	分散式	分散式	3.88	73.6%
	2	資料庫	資料庫	3.76	71.3%
	3	資料探勘	頻繁型樣,頻繁型樣,頻繁型樣,頻繁型樣,頻繁型樣,頻繁型樣	3.4	64.7%
	4	服務品質	網際網路	3.32	62.7%
	5	資料庫	網際網路	2.20	41%

(四) 實務上之應用

對於剛接觸資訊領域的研究人員，本系統提供了一個便利而快速的方法查詢資訊領域中研究主題的關聯性。例如，若一個新接觸資訊管理領域的研究人員，便可透過本系統查詢資訊管理領域中具有高關聯強度之研究主題，也可以探討各個學校之間具有高關聯強度之研究主題的不同，進而了解研究人員所在學校之研究走向；此外，若想針對某個研究主題進行更深入的了解。也可藉由「單一研究主題查詢」的功能，查詢研究人員有興趣的研究主題與其他研究主題的關聯程度，進而快速的取得相關學位論文列表，以閱讀相關的研究文

獻。而對於一個在資訊領域多年的研究人員，利用本系統也可了解其主要研究方向與其他主題之間的關係，並且可藉此發現新的研究主題，進行新的研究方向；同時，若研究者想進行跨領域研究時，也可藉由本系統之查詢結果了解其他領域之研究概況，進而快速取得其他研究主題之研究文獻列表與該領域之研究人員，讓研究者能夠快速的了解跨領域之研究概況。

六、結論

本研究之研究成果包含了以下數點，第一點為提出衡量兩研究主題之間關聯性之衡量指標：關聯強度與關聯相似度，可用來衡量研究主題之間的直接關係與間接關係。第二點為以自動化的方式，擷取網際網路上關於資訊相關領域的學位論文資源，並計算資訊相關領域研究主題之間的關聯強度與關聯相似度，分析現今資訊相關領域中各個研究主題之間的關聯性，並且提供相關研究主題與相關學位論文給研究人員，作為對領域知識之基本認識與往後研究之用。第三點為建立同義關鍵字與相似關鍵字資料庫，以整合國內學術論文關鍵字用法分歧的現象，並且實際運用於系統中，讓使用者在進行查詢時可以一併了解相關研究主題。第四點為針對單一研究主題，分析與其具有高度直接關係之研究主題，以及這些研究主題之間的關聯性。第五點為根據不同的時間點，分析研究主題之間的關聯性，以了解在不同的時間下，研究主題的關聯強度與關聯相似度的變化，以及具有高關聯強度的研究主題變化。論文中以 81 學年度至 90 學年度之資訊管理領域之論文為例，分析十年內資訊管理學界研究方向之變化。第六點為分析不同的學校、系所，在相同的研究主題上關聯強度與關聯相似度的差異，以及各學校與系所的研究主流。論文中分別分析了國內資訊工程、資訊管理、資訊科學三個領域研究方向之差異，以及中山大學、中央大學、政治大學、台灣大學四所學校之資訊管理學系研究方向，與中山大學資管系與資工系、台灣大學資管系與資工系之研究方向。

致謝

本研究由國科會計畫獎助。名稱為知識分享：顯性知識採用知識關聯圖及隱性知識採用知識地圖。計畫編號：92-2213-E-110-041-。

七、參考文獻

[1] 陳明溥、莊良寶(1999)。全球資訊網學習

環境中學習活動型態與學習成效之研究。TANET 99 研討會論文，高雄。

- [2] 陳明溥、嚴榮泉(2001)。網路化問題導向學習系統建構模式之研究。第十屆電腦輔助教學國際研討會論文，台中。
- [3] 林東清(2002)。資訊管理-e 化企業的核心競爭力。台北智勝。
- [4] 謝武星(2000)。針對「學術論文」的知識管理技術研究，政治大學，資訊管理研究所碩士論文。
- [5] 楊存一(2001)。利用自適應共振理論網路探討 MIS 學術論文關鍵議題的發展趨勢，雲林科技大學，資訊管理研究所碩士論文。
- [6] 張家華(2001)。資訊管理學科文獻內援用理論之探討 - 以台灣碩、博士論文為例，政治大學，資訊管理研究所碩士論文。
- [7] 陳年興、孫振凱(2002)。透過網頁分析建構知識分布圖以輔助線上教學，第十三屆國際資訊管理學術研討會，675-682。
- [8] 王惠嘉，張毓倫(2002)。研討會之知識管理及其系統之建構與發展，第十三屆國際資訊管理學術研討會，659-666。
- [9] 李律品(2002)。資訊管理研究之生態演化分析，中山大學資管研究所碩士論文。
- [10] 陳俊彰(2001)。從網頁中發掘教師知識分佈圖，中山大學資管研究所碩士論文。
- [11] Alavi, M., and Leidner D.E., "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues," *MIS Quarterly* (25:1), Mar. 2001.
- [12] Schultze, U. & Leidner, D.E. (2002), "Studying Knowledge Management in Information System Research: Discourses and Theoretical Assumptions", *MIS Quarterly*, vol.26, no.3.
- [13] He, Y. & Hui, S.C. (2002), "Mining a web citation database for author co-citation analysis", *Information Processing and Management* vol.38, 2002.
- [14] Guha, S. R. Rastogi & Shim, K. (1998), "Cure: An efficient clustering algorithm for large databases", *Proceeding of 1998 ACM SIGMOD International Conference on Management of Data*.