自動化建構具時間向度之知識概念圖
--以資訊管理領域之知識及其演進為例

陳年興　　謝盛文　　鄭百勝
國立中山大學資訊管理研究所
nschen@cc.nsysu.edu.tw　　onyxhsw@swh.idv.tw　　m934020039@student.nsysu.edu.tw

摘要

二十一世紀是知識經濟的時代，而知識經濟中最重要的工作就是知識管理，組織必須要有系統性、分析性的利用知識，以提升組織本身的競爭能力。學術界與業界亦相同，學術界有著大量的資訊，本研究希望透過系統性、分析性的方式來針對學位論文進行資料探勘，並根據探勘的結果建構知識概念圖，以減輕使用者在瀏覽學位論文資訊時，資訊過載與使用者迷失的現象。本研究收集全國博碩士論文網 79 到 93 學年度，其中屬於資訊管理的碩士論文共有 5476 筆。並利用論文資訊的「關鍵字」及「摘要」中的位置資訊，分析關鍵字配對在摘要之中的距離，建立研究主題之間的關聯相強度，最後自動畫出研究主題之間的知識概念圖，之後並選定資訊管理領域作為本研究的應用對象，且加以分析不同時段的知識內涵及其演進。研究發現 79 到 83 年度的資訊管理領域，分成電子商務以及專家系統兩個部份。84 到 88 年度發現兩個部份逐漸整合成一個有系統性、整體性的學門。84 到 88 年度中，服務品質開始受到了重視。89 到 93 年網際網路與電子商務的關聯強度在此達到最高點。

關鍵詞：知識概念圖、資訊管理知識演進、資料探勘、知識管理

1. 緒論

由於資訊科技的快速發展，學術界相關資訊的來源也越來越豐富，其中以線上資料庫或搜尋引擎的大量資訊是最為重要的。然而，大量的資訊容易造成資訊過載（Information Overloading）(陳年興、孫振凱，2002)，資訊過載是指在網際網路的環境中，有著大量且豐富的資訊，但往往因為資料過多而導致資訊整合不易，加上使用者在網際網路上蒐集資料時，得到的往往是片段且零散之資料，遠遠超過學習者所能接收的程度，這樣不僅無法取得完整且有系統的知識，亦容易造成學習者更大的負擔，造成使用者迷失（Disorientation）。

因以本研究希望利用資料採礦的方式自動化建構具時間向度之知識概念圖，協助使用者更快、更有效率的從大量資料中獲得真正所需的資訊與知識，與即時的提供學習者相關的知識字詞與知識字詞之間的關聯。透過概念圖的提供將可以有效的降低學習迷失現象與避免資訊過載，以達到有效的知識分享與利用。本研究之「知識概念圖」是由概念圖（Concept Map）延伸而來的，然而，從文獻中可以發現，在建立概念圖的過程中，存在兩個問題；一是未考慮時間的因素，因為知識間的關係與整個領域的概念是會隨著時間而改變的；二是研究人員往往只能依靠人工或者半自動的方法對研究主題進行編碼與分類。因此本研究針對上述二點，並參考黃琬婷(2004) 對於知識概念圖的相關處理方式，且以全國博碩士論文網(http://datas.ncl.edu.tw)做為資料來源，包含有民國 79 到 93 學年度的資料，配合主成份分析與關聯強度，來自動化的建構知識概念圖。為了實證本方法的可行性，本研究選定資訊管理領域為應用的對象，並以實作出來的知識概念圖，分析資管領域之知識內涵及其變遷與演進。本研究主要的目的是建構知識概念圖，所以我們首先介紹概念圖與知識概念圖的相關研究，最後是本研究之相關篩選技術與計算公式。

2 文獻探討

2.1 知識概念圖

概念圖可表現畫圖者的概念、概念間的關係、及圖的結構(Novak，1979)。本研究認為知識概念圖是一種知識表示的方法，以圖形化的方式來說明知識與知識間的關係、知識結構；McAleese 對概念圖下的基本定義是『一個包含 m 個概念元素的集合{C1 … Cm}和 r 個非空集合的關聯或連線{R1… Rr}的有向性、非環式的多維圖形』McAleese (1994，1999)，如圖 1 McAleese (1994，1999)；事實上就是『概念—關聯—概念』的形式，然而，本研究從學術論文中萃取出來的知識概念圖是以『研究主題—關聯—研究主題』的形式呈現，而研究主題是指學術論文中經過處理的關鍵字，關聯則是代表各研究主題間的關係程度。

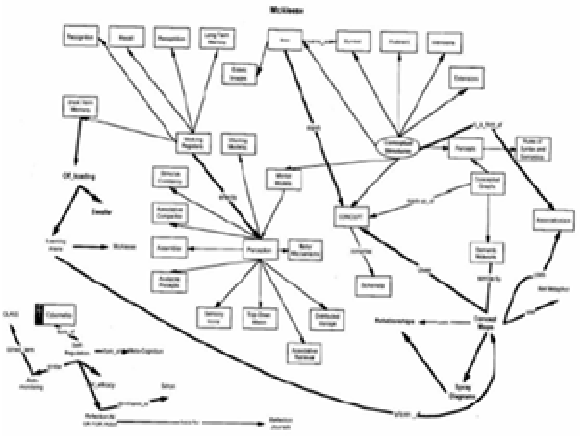


圖 1、McAleese 研究中之概念圖

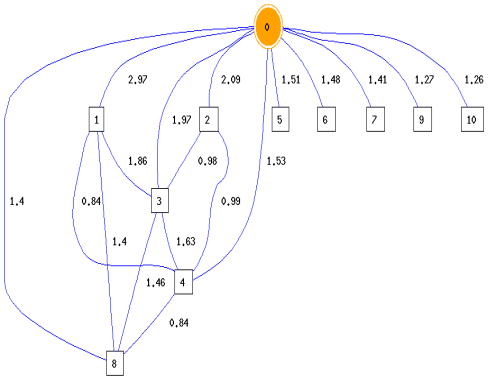


圖 2、知識概念圖

知識概念圖可以有效的降低學習迷失與避免資訊超載且減少傳統建構方法所需的成本及時間，也讓學習者在尋找知識時能更有效率，因此近年來學術界有相當多針對知識概念圖的相關研究(Schultze & Leidner，2002)。圖 2 為知識關聯圖的範例。

2.2 關鍵字詞篩選

由於本研究必須擷取分析資料中的關鍵字，而分析資料為博碩士論文。關鍵字詞篩選觀念，依據不同對象給予字詞不同的權重，其方法有頻率(Ng，Goh & Low，1997)、TFIDF (Term Frequency Inverse Document Frequency)、相關係數等，本研究將利用字詞出現頻率來進行關鍵字詞篩選，同時，從郭祥昊、鍾義信、楊麗(1998)，及張翠英、亢臨生(1998)等學者的研究中得知，漢語中有意義的兩字詞佔了其中的 70%～75%，而許多詞也都是從兩字詞延伸，因此本研究僅選取大於三個字的中文關鍵字做為研究主題。

2.3 關聯強度

針對兩研究主題間之關聯強度，陳道輝、謝盛文、陳年興(2002)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的關聯性。其中，聯繫強度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，而關係發生頻率為此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，例如學位論文或期刊(公式 1)。

$$RS(k1,k2)=\frac{1}{(norm(avg\_dist))^2}\times norm(count)\times type$$

RS＝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

公式 1、陳道輝提出之研究主題關聯強度公式

3.系統設計與實作

本研究提出自動建構知識概念圖的方法，並依此方式實作系統；本研究所開發的系統主要是先以自動化的方式擷取網路上之學術論文，並存入自己的資料庫中；接著將處理論文資訊中的關鍵字，原則上處理完的每個關鍵字都是該領域的研究主題；下一步將針對這些不同的研究主題，計算它們的屬性，存入研究主題資料庫；最後，以主成份分析找出在該時間區段中具有代表性之研究主題，計算其關聯強度，自動產生在不同時間點下之知識概念圖。本研究選定資訊管理領域為應用的對象，系統流程圖如圖 3 所示，之後將先介紹本研究之基本假設，再依序介紹各個步驟的詳細內容。

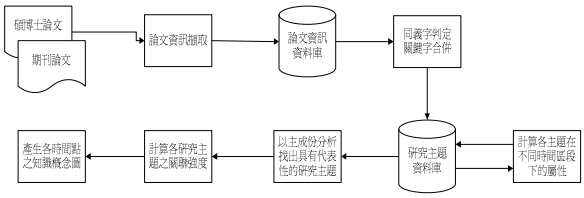


圖 3、系統流程圖

3.1 本研究之基本假設

若要從所獲得的論文資料中擷取具有代表性的資訊做為該論文之重要、關鍵的研究主題，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」來進行分析(陳道輝等人，2002)。因此，若要針對學位論文進行知識管理，可以藉由分析「關鍵字」與「摘要」來進行。論文作者在選擇關鍵字時，會選擇與該論文最相關的詞彙。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(謝盛文、陳道輝、陳年興，2003)。是以本研究假設：一篇論文的每一關鍵字均代表著論文的研究主題之一，而且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。若一篇文章中，兩個名詞之間的平均距離越近，則兩者的關聯性也越高(陳道輝，2003)。

3.2 擷取資訊管理領域相關論文資訊

在建置資訊管理領域之知識結構概念圖前，必須先收集相關的資訊管理論文資料，本研究分析之資料為學位論文，全國博碩士論文資訊網為國內收錄博碩士論文資料庫；包括論文標題、作者、指導教授、學位別、年度、摘要、論文目錄、參考文獻等。其中，資訊管理研究所論文資料包含範圍自民國 76 年至民國 93 年 9 月，共有 5476 筆。由於各式線上資料庫基於安全及效率的考量，僅提供 Web 介面讓使用者進行查詢，而本研究後續的計算與分

析會大量使用網頁上論文的相關資訊，故先透過程式擷取論文網頁畫面至本地端，分析網頁內容，擷取論文的關鍵字作為研究主題，分析完畢之後存入自有的資料庫中，往後的分析與計算均使用自建的資料庫。

3.3 論文關鍵字的擷取與處理

由於論文作者在選擇關鍵字時，會選擇與該研究最相關的詞彙做為關鍵字；同時，從(郭祥昊、鍾義信、楊麗，1998)，及(張翠英、亢臨生，1998)等學者的研究中得知，漢語中有意義的兩字詞佔了其中的 70%～75%，而許多詞也都是從兩字詞延伸，因此本研究僅選取大於三個字的中文關鍵字做為研究主題。故根據本研究之第一個基本假設，經過處理之後的每個關鍵字可以代表該領域之一個研究主題。此外，由於國內學界對於英文專有名詞的中文譯名並沒有一定的依循標準，行政院教育部也沒有硬性規定英文專有名詞的中文譯名為何，因此造成了國內學界對於翻譯名詞沒有一定的準則，許多用法僅是約定成俗的結果，如此一來也造成了一個名詞可能會有許多個中文譯名；對於研究者而言，不管是在查詢論文上，或是論文寫作時詞彙的選擇上，均有不便之處。因此，本階段主要是將具有相同意義的關鍵字進行整合，以合併相同的研究主題；此外，有些關鍵字出現的頻率實在太少，表示此研究主題並不重要，因此也予以刪除；經過這樣一連串的處理之後，剩下的每個關鍵字皆可當成一個研究主題。自民國 76 至 93 年 9 月的資訊管理領域來說，關鍵字共有 10248 個，但經過關鍵字處理與合併後產生之研究主題共有 8910 個，確實大幅減少不必要的關鍵字，並合併相同的研究主題。

3.4 主成份分析

由上述處理步驟，經過處理後具有代表性的資訊管理領域共有 8910 個研究主題，哪些研究主題是具有代表性、能代表該時間區段的研究主題呢？本研究採用主成份分析來找出在該時間區間裡，最具有代表性的研究主題。主成份分析是種統計方法，首由 K. Pearson 於 1901 年提出，再由 Hotelling (1993) 加以發展而成的一種統計方法。主要的目的就是希望以較少的變數去解釋原來資料中大部分的變異，更期望能將這些相關性的資料轉換成彼此互相獨立的變數。因此，本研究以三個變數做為主成分的綜和性指標(每個研究主題與其他主題的關聯個數、出現的頻率、和持續的時間)。研究主題與其他主題的關聯個數愈高，表示與越多領域有關，所以越重要；研究主題出現的頻率越高，表示越多人進行研究，也就越有代表性；研究主題持續的時間越久，表示不是一個曇花一現的研究，而是能被大家所接受，且並值得持續的研究；因此，各項變數若越高，則該研究主題在該時間區段將越具代表性，利用此三項變數之綜合性指標，來判斷那些研究主題可以做為該時間區段下的代表性主題。表 1 為主成份分析後的結果，由表中可以發現只有第一個因素（即由三個選定之主成份綜合的指標）的特徵值大於 1，且其解釋變異量達到 61%；在實際研究裡，研究者如果未使用超過五或六個成份，就能解釋變異之 61%，已是令人滿意的結果(Hotelling，1993)。所以本研究以此組係數大小(表 2)來做為加權的依據是可行的，故本研究根據此係數加權計算後，值愈大代表該研究主題愈具有代表性。

表 1、主成分解釋變異表

	特徵值	解釋變異%	累積解釋變異%
1	1.847	61.553	61.553
2	0.791	26.364	81.917
3	0.363	12.083	100.000

表 2、主成之係數

主成份	關聯個數	出現頻率	持續時間
係數	0.835	0.877	0.617

3.5 關聯強度

如果兩研究主題彼此之間有關聯，則以線段連接兩個研究主題，並標示其關聯強度。然而，要經過兩個步驟才能完成，第一步驟，對每篇論文研究主題配對，第二步驟，計算該配對在論文摘要中的平均距離。

第一步驟，每篇論文研究主題兩兩建立配對，研究主題配對必須兩個關鍵字同時出現在該篇論文摘要中，計算後共有 33966 組配對，但基於本研究之基本假設，扣除關鍵字在摘要中沒有同時出現，其距離為 0 者，兩兩研究主題之間有關聯的即減少為 15515 個。第二步驟，計算這些研究主題配對在論文摘要中的平均距離，且兩個研究主題必須同時出現在單一句子中，如此才當作研究主題配對之距離計算，在摘要部份取一個平均距離來代表該研究主題配對之關係，計算完畢後即可依公式計算各研究主題中彼此的關聯強度。計算主題間的關聯強度，為知識概念圖的前置作業。依照擷取資訊管理領域相關論文資訊、論文關鍵字的擷取與處理、主成份分析、關聯強度，本研究將可產生在不同時間點下之知識概念圖。

4.成果分析

本研究利用資訊管理領域為實作的對象，並且呈現該領之知識概念圖，並說明有哪些重要研究主題，且描述資訊管理領域在時間上的演進為何。因

此分兩部分，第一部分將介紹 76 年到 93 年資訊管理領域的概念；第二部份會分為不同的時間區段以概略描述國內資訊管理領域的變遷與演進。

4.1 資訊管理領域整體之知識概念圖

此部分主要是以不同大小及個數的研究主題，來呈現有哪些重要的知識概念與各知識概念之間的關係，由知識概念圖可以讓我們更快及更清楚知道資訊管理領域內知識的關聯。圖 4 與圖 5 分別為 5 個及 10 個最重要之研究主題所實作現出來的知識概念圖，使用者可以獲知目前資訊管理領域的研究主題，及這些重要主題之間的關聯強度來判斷兩個不同知識概念之關聯程度。表 3 為研究主題之中英對照表。由圖 4 可以發現，國內 76 年到 93 年資訊管理領域主要 5 個研究主題是電子商務、網際網路、知識管理、網頁探勘及服務品質等研究領域，本研究發現電子商務和網際網路是資訊管理領域研究的重心和主軸，本研究去比對碩博士論文網資料發現，資訊管理領域從 76 年到 93 年論文共 5476 筆，其中包含電子商務與網際網路為關鍵字研究主題共 1459 筆；另外，本研究發現在近幾年，知識管理這新興的研究越來越受到重視，已在整個資訊管理領域中佔有重要的一席之地。

表 3、研究主題中英對照表

網際網路	Internet	類神經網路	NN
電子商務	EC	顧客關係	CRM
網頁探勘	WM	應用軟體服務	ASP
知識管理	KM	資料倉儲	DW
關聯法則	MAR	服務品質	Quality Service
資訊系統	IS	專家系統	ES
系統動力學	SD	決策支援系統	DSS
全球資訊網	WWW	模糊理論	Fuzzy

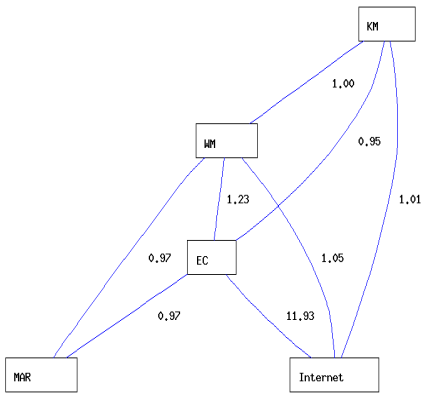
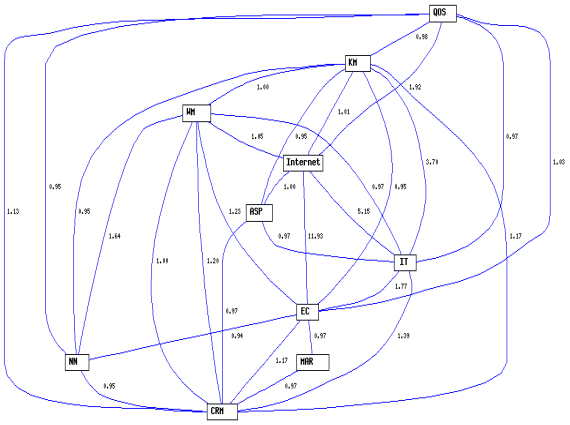


圖 4、前 5 個重要概念之知識概念圖

由圖 5 亦可看出除了上述的研究主題之外，顧客關係管理、類神經網路、應用軟體服務供應商、資訊科技議題，亦在資訊管理領域中佔有重要的地位。此外，亦可由圖看出國內的資訊管理領域，以電子商務和網際網路的關係最強，網際網路與資訊科技次之；再以資訊科技來說，使用者可以了解其



與電子商務、知識管理、網頁探勘、應用軟體服務供應商、服務品質及顧客關係管理的研究領域有關，及彼此之間的相關程度為何；如此，知識概念圖可以做為使用者知識概念的建立、領域關係的強弱或知識管理的指標。

圖 5、前 10 個重要概念之知識概念圖

4.2 資訊管理領域之變遷與演進

除了能以知識概念圖看出整體資訊管理領域的知識與概念外，尚可由不同年度區間的知識概念圖來分析，即可了解國內資訊管理領域內知識概念的變遷與演進，因此，本研究將收集的資料以五年為一個單位，共分為三個階段，並將此三個階段分別繪製知識概念圖，圖 6 至 8 分別為 79 到 83 年度、84 到 88 年度、89 到 93 年度的知識概念圖。由圖 6 可以看出 79 到 83 年度的資訊管理領域，分成兩大領域，期中之一是電子商務為主，另一則是以專家系統為主；除此之外，我們發現網際網路在這階段已經開始興起。由圖 7，84 到 88 年度，與圖 6 比較可以發現 79 到 83 年度資訊管理領域的兩大部份在進行整合成一個有系統性、整體性學門。另外也出現了許多新興的研究主題，例如:知識管理、網頁探勘等，更重要的是，網際網路之議題，已越來越受到重視；84 到 88 年度中，服務品質開始受到了重視。最近五年(89 到 93 年)，國內資訊管理領域最主要的就是網際網路的相關議題，網際網路與電子商務的關聯強度在此區段達到最高點，而且由圖 8 發現，各研究主題間關係變的多元，例如: 在 84 到 88 年度，服務品質與網際網路等議題相關，但到了 89 到 93 年度則多了知識管理與電子商務之關係。此外，本研究發現多了資料探勘及資料倉儲這

兩個研究主題，且進一步發現這兩個新興的研究主題並非獨立而是過去研究主題的再延伸，例如:知識管理與資料探勘、資料倉儲與網際網路、服務品質與知識管理。

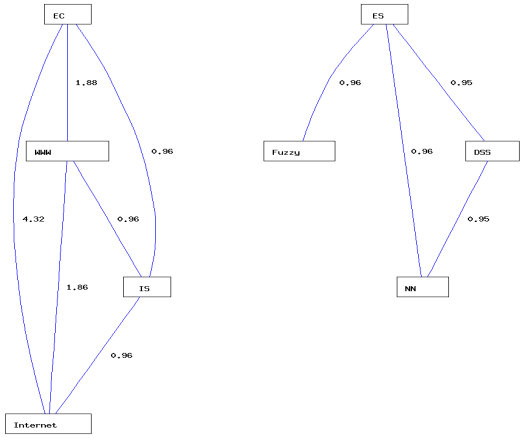


圖 6、79 到 83 年度之知識概念圖

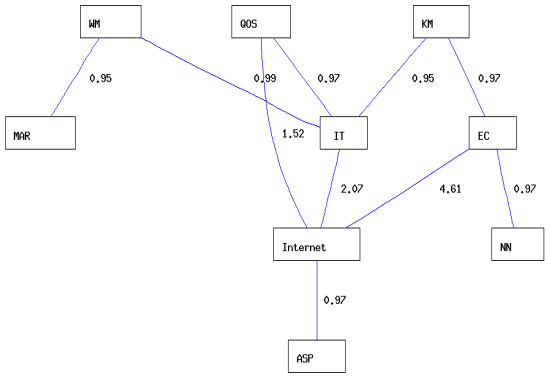


圖 7、84 年到 88 年度之知識概念圖

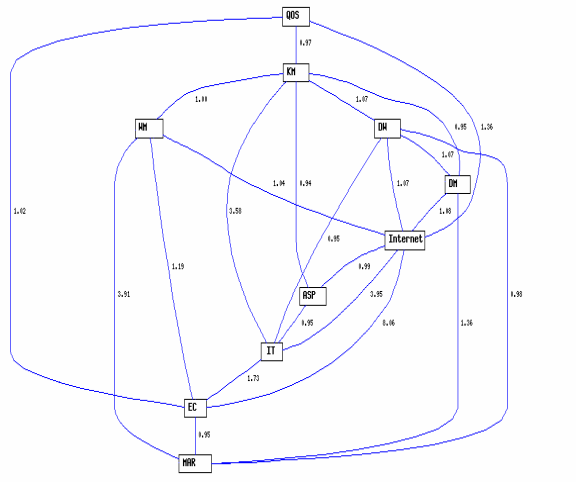


圖 8、89 到 93 年度之知識概念圖

5.結論

由於資訊科技的普遍與網路通訊的發達，使得數位文件的產生與累積的速度極快，而有過多的文件資料，容易導致迷失與資訊過載的問題。學術界亦然，學術界已經產生與累積了大量的數位文件，為解決這些問題，本研究使用文字探勘(Text Mining)中之關聯分析及區分不同時段等方法，使得，數位文件變得系統性、分析性。

本研究以資訊管理領域之論文為分析對象，並實作出資訊管理領域知識概念圖，透過對知識概念圖的演進進行分析。研究發現資訊管理領域之研究重點集中在電子商務、網際網路、網頁探勘、知識管理等主題。而從時間的角度來看，本研究以五年為一階段共分三個階段，從第一階段的主要由兩大部分，電子商務與專家系統的研究所主導，到第二階段這兩個部份則出現相互整合的現象並進而衍伸出現一些新興研究主題，例如:知識管理、網頁探勘、等研究主題。到了第三階段本研究發現，新興的研究主題並非是獨立出現反而是某些研究主題的分支，這表示著資訊管理領域已越來越多元發展。透過知識概念圖可以看出資訊管理領域所著重的概念，進而幫助資管領域內的相關研究人員或新進人員不易迷失方向與資訊過載。且對商業組織而言亦可以導入此方式，建構出該商業領域的知識概念圖，以搶得商機。

在未來研究的部分，可以思考除了本研究所使用的三個主成分之外，其他的方法來提高主成分的解釋變異程度。還有，對關鍵字詞的處理及合併方式加以改進以增加關鍵字詞的正確性。並將資料來源擴大到期刊如:MISQ、ISR、SSCI 等之文章分析差異性。

致謝: 本研究是由國科會研究計畫所贊助，計畫編號：NSC93-2524-S-110-001。

6.參考文獻

[1] 郭祥昊、鍾義信、楊麗，基於兩漢詞簇的漢語快速自動分詞演算法，情報學報，第 17 卷第 5 期，1998

[2] 張翠英、亢臨生，三字歧義錬自動分詞方法，情報學報，第 17 卷第 3 期，1998

[3] 陳道輝、謝盛文、陳年興，以資料挖掘技術分析資訊管理領域相關知識之間的關聯，第八屆資訊管理研究暨實務研討會，2001

[4] 陳年興、孫振凱、黃琬婷，利用網頁建構知識分佈圖，中華民國資訊學會通訊，第 5 卷第 3 期， 2002

[5] 陳道輝，利用學位論文資訊萃取資訊相關領域之研究主題關聯性，國立中山大學資訊管理研究所碩士論文，2003

[6] 謝盛文、陳道輝、陳年興，利用知識關聯圖達到外顯知識分享之研究，全國計算機會議，2003

[7] Alavi, M. & Leidner D.E., Review：Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues, MIS Quarterly, 2001, Vol.25, No.1

[8] Hotelling, H., Analysis of a complex of statistical variables into principal components, Journal of Educational Psychology, 1993, Vol.24, pp.417-441

[9] McAleese, R., Special double issue—Concept mapping—book review, Journal of Interactive Learning Research, 1999, Vol.8

[10] Nonaka, I., A Dynamic Theory of Organizational Knowledge Creation, Organization Science, 1994, Vol.5, No.1, pp.14-37

[11] Novak, J.D (1979).A Theory of Education as a Basis for Enviornmental Education.In T.S. Bakshi (Ed.), Enviornmental Education: principles,Methods,and Applications .Enviornmental Science Research series,18,London:Plenum

[12] Schultze L. & Leidner, D.E., Studying Knowledge Management in Information System Research: Discourses and Theoretical Assumptions, 2002, MIS Quarterly, Vol.26, No.3