

# 本文章已註冊DOI數位物件識別碼

## ► 基於知識本體之文獻多維度關聯視覺化導覽平台

An Ontology-based Navigation Platform for Browsing Multi-dimensional Relationships of Academic Papers

doi:10.6188/JEB.2011.13(1).03

電子商務學報, 13(1), 2011

Journal of E-Business, 13(1), 2011

作者/Author: 楊亨利(Heng-Li Yang);趙逢毅(August F.Y. Chao)

頁數/Page: 27-54

出版日期/Publication Date: 2011/03

引用本篇文獻時，請提供DOI資訊，並透過DOI永久網址取得最正確的書目資訊。

To cite this Article, please include the DOI name in your reference data.

請使用本篇文獻DOI永久網址進行連結:

To link to this Article:

[http://dx.doi.org/10.6188/JEB.2011.13\(1\).03](http://dx.doi.org/10.6188/JEB.2011.13(1).03)



*DOI Enhanced*

DOI是數位物件識別碼 (Digital Object Identifier, DOI) 的簡稱，是這篇文章在網路上的唯一識別碼，用於永久連結及引用該篇文章。

若想得知更多DOI使用資訊，

請參考 <http://doi.airiti.com>

For more information,

Please see: <http://doi.airiti.com>

請往下捲動至下一頁，開始閱讀本篇文獻

PLEASE SCROLL DOWN FOR ARTICLE



## 基於知識本體之文獻多維度關聯視覺化導覽平台<sup>1</sup>

楊亨利\* 趙逢毅

國立政治大學資訊管理學系

### 摘要

學術研究必須建立於過去已完成的研究，因此文獻探討非常重要。數位時代下，改變了過去研究者必須東奔西跑的尋找文獻方式，但卻也會造成研究者被許多數位文獻淹沒。若奠基於事先定義好的領域本體論，人類的視覺化資訊處理能力應可協助瞭解文獻中多維度的複雜資料關聯。本研究提出一個文獻導覽平台系統架構，協助使用者依預訂的知識本體導覽領域文獻，並能依循「知識本體與關鍵字」、「引文網路」及「人員關係網路」三種關聯來導覽論文，並與呈現的網路圖形資料互動。最後實作視覺化導覽雛型平台，並以國家圖書館之博、碩士論文網所提供的部分論文為雛形測試資料。

關鍵詞：引文網路分析、社會網路分析、知識本體、視覺化資料挖掘

## An Ontology-based Navigation Platform for Browsing Multi-dimensional Relationships of Academic Papers

Heng-Li Yang August F.Y. Chao

Department of Management Information System, National Cheng-Chi University

### Abstract

Paper survey is the most important task for conducting research. Nowadays, this work might return huge amount of information owing to digitalized-documentation and high-speed Internet. Based on predefined domain ontology, human visual data-processing capability might help understand multiple dimensions of complicated information relationships. This study proposes a framework for document exploration to assist users to browse domain literature based on predefined ontology. It also guides users to explore the literature based on

<sup>1</sup> 本研究受行政院國科會專案計畫 (NSC 96-2416-H-004-015-MY2) 補助，特此致謝。

\* 通訊作者

電子郵件：yanh@nccu.edu.tw

three different relationships: network of knowledge ontology and keyword, citation network, and people social network. Finally, a prototype was implemented and part of data from Electronic Theses and Dissertations System of National Library were fed to test the feasibility.

*Key word: Citation Network Analysis, Social Network Analysis, Ontology, Visual Data Mining*

## 1. 導論

論文在經過數位化儲存與網際網路的傳遞，方便了研究學者進行不同領域之間研究資源的參考與查閱。但對一個研究者，面對這麼多過去的研究資料，如何才能找到符合其研究方向的有用資料、瞭解在有興趣的學門裡所探討的相關議題，以及有效地找出以往的參考文獻，最後建立新的研究方向，實為難事。近年來許多新興的網路服務公司，使用網頁搜尋引擎跨越不同資料庫進行文獻資料搜尋，並使用基於被引用次數做為辨定權重原則的— Paper Rank 演算法 (Gori and Pucci, 2006)，能在很短的時間內提供使用者在所搜尋的文獻集中權重較高的文獻。這適合用在使用者已經清楚進行搜尋的論文目標物為何的情境，如某個關鍵字詞、某篇論文主題、某位學者。但是對於初學者，通常所需的並非特定搜尋，而必須廣泛嘗試各種相關搜尋。此外文獻尋找的過程，其實也是學習領域知識與建立概念研究問題的步驟。依引用權重順序性的排序並無法提供研究人員在特定知識領域中，文獻間關聯與概念間的關聯，反而扼殺研究者在尋找文獻時可能意外獲得相關觀念啟發。

在台灣的全國博、碩士論文資訊網，長年來集中保存了許多台灣研究者的心得與精華，包括了論文全文與摘要，分門別類地儲存。雖然系統中設有全文、欄位與關鍵字搜尋，但對一個研究入門者，面對這麼多過去的研究資料，如何才能找到符合其研究方向的有用資料，瞭解其有興趣的學門裡所探討的相關議題，以及有效地找出可供參考文獻，以建立新的研究方向，並不容易。遑論再從單篇論文中的參考文獻列表，延伸閱讀以找到該領域的經典文獻做為參考。

為了能在有限的時間內勾勒特定領域中的文獻分佈的概況，以取得有較具代表性之文獻資料，本研究試以知識本體與關鍵字網路、引文網路與人員關係網路，以視覺化的平台揭露各層級網路中的結構關係。此外透過特定領域的知識本體做為參考指標，配合視覺化平台中的互動機制進行三種網路層級間觀點的切換，供三種使用者類型為探索文獻與引用文獻的參考。



## 2. 文獻探討

### 2.1 知識本體

Bunge (1977) 指出本體論是關於真實世界的基本特性哲學理論，用以探討真實世界中存在的實體，並進行有系統的說明。知識本體是一種被認同的概念性分享，並以定義字辭的型式表現 (Uschold and Gruninger, 1996)。分享觀念包括了特定領域建立模式的觀念架構、在內部交互運作的具體內容溝通協定，及溝通特定領域理論的規則。由於知識本體主要的目的是在分享已知領域的知識，而領域專家所建構出的知識本體，一般認為是足以代表特定領域的知識。因此在許多學者的研究中使用已經建構好的專家知識本體，做進一步的排序、探勘、或萃取特定知識，以協助使用者在搜尋結果、網路行為研究、喜好分析等 (曾信誠, 2004; 蔡旺典, 2006)。亦有在情境感知的環境中，透過已經建置好的文物知識本體，依個人的興趣，推薦導覽符合其偏好的文物供參觀 (許明章, 2006)。

知識本體建構到目前為止尚未有明確的方法 (Noy and McGuinness, 2001)，但大體上可以分為七個步驟：(1) 定義知識本體的領域與範圍、(2) 參考現有的知識本體、(3) 列舉知識本體中重要的概念、(4) 定義知識的本體與架構、(5) 定義知識本體中的屬性 (Slots)、(6) 定義知識本體屬性中的事實 (Facets)、(7) 建立實例。而對知識本體的塑造，Uschold 與 Gruninger (1996) 指出除可用由下而上 (Bottom-up)、由上而下 (Top-down) 與由內而外 (Middle-out) 的三種方式外，亦指出所建立的知識本體需要與現存的知識本體整合。此外知識體亦需要視實際情況進行知識本體的改善與演進 (Uschold and Gruninger, 1996; Sure et al., 2004)。

### 2.2 文獻計量

人類的知識是經過時間累積，而且學科的演進是在原先的知識及架構上將學術領域向外延伸推進。在嚴謹學術研究體制下，學理也並非是憑空而來，所有的文獻著作並非是孤立的。學理的建構會透過觀念的延用、引用、借用發生某些關聯，這些關聯反映了科學思想交流與借鏡的軌跡。文獻計量學是對文獻中所引用到的文獻特質、引用文獻間關係、引用文獻與樣本文獻間關係所進行的研究，包含書目對 (Bibliographic)、共被引 (Co-citation) 及共字分析 (Co-word Analysis) 與叢集分析 (Cluster Analysis)、誌謝分析 (Acknowledgement Analysis) 與引用文獻內容分析等 (蔡明月, 2003)。

書目對分析是指任兩篇文獻所引用相同文獻的數量分析；共被引則類似於書目對分析，但是指任兩篇文獻被後來的文獻同時引用的次數，此兩種分析是在瞭解文獻之

間的關連性強度。透過文獻共被引、期刊共被引與作者共被引的叢集圖形，可以大體上看出研究核心 (Braam et al., 1991) 與作者之間的專業領域關聯 (Chen and Paul, 2001)。而共字分析是指對文獻內容所用文字進行統計分析，以瞭解領域間關係的密切度 (Braam, et. al, 1991)。誌謝分析則是將誌謝詞中作者對某(些)人或某(些)組織進行分析 (蔡明月, 2003)，以瞭解其「隱形合作」關係 (Cronin et. al., 1992)。上述分析的方式不僅能用來排除不重要的引用文獻，協助研究人員在有限的時間內查閱較為重要的文獻，同時也供研究人員能夠分析現階段該領域之研究邊界，以找出停滯問題、方向，及最近發表的文獻中該知識領域所擴展的研究趨勢。

此外透過文獻在引文網路結構中的關聯方式，可瞭解該文獻被參考的價值。在 Lawrence 與 Bollacker (1999) 研究學術上文獻引文的分析中，認為連接到共被引的文獻 (即 Hubs) 可以被視為進入該領域的敲門石，其可用作回顧 (Review)、考查研究 (Survey) 或是入門導引 (Tutorial) 的文章。這類回顧類型文獻會整理相關領域中重要的觀念及文獻，而共被引的文獻排名也是非常有用，因為這些文獻內也包括了在特定領域的簡介及先前的研究內容。藉由這樣的文獻在引文網路上的關聯方式，可經由分析網路後提供我們一個尋找較易入門的文獻方法。

### 2.3 社會網路

社會網路是指社會組織中，個人與個人、個人與組織間相互連結的關係集合。因此研究社會網路主要目的是在發掘人與人間相互交集的模式，而這些模式的發掘可以顯示出一個人際關係的特徵 (陳俊彰, 2001)。一般社會網路分析方法，是以節點、線及密度等網路圖形的運算來區分不同的社會網路，透過中心點及中心性瞭解那些節點是在群體中最多數連結到的，並區分在不同社會網路間的中心群聚性；藉由成分 (Component)、核心 (Core)、群集 (Clique)、位置、角色、分組情況 (Cluster) 等分析，以瞭解社會網路的組成概括，並探討對組成的社會網路整體影響 (Scott, 2000)。

文獻資料內所提供的社會網路資訊，此關係有共同著作 (Co-author)、論文指導，及在論文網中同在系、所資訊。單一篇論文內的共同著作是指單一篇文獻由一個以上的作者共同著作，一般會出現在非博、碩士論文中的文獻資料中；論文指導則是在國家圖書館的論文資訊網中的資料欄位，除了說明本篇作者的指導教授外，亦透漏出作者與指導教授間專業領域的特殊關聯性，及人員在特定知識領域上有一定程度的關聯。

### 2.4 資訊視覺化與互動介面

視覺化就是將資料、資訊與知識轉化為視覺的形式，使得人類與生俱來的視覺能力得以發揮，以做為人類心智與現代化的電腦系統所具有的強大資訊處理系統之間



的介接管道。資訊的視覺化是使用互動的方式表達抽象資料，以加強感官的認知（Shneiderman and Plaisant, 2005）。透過視覺化的介面，我們能夠與大量的資料進行立即的互動，並且有效發掘資料背後所隱藏的特性、模式（Pattern）或趨勢（Trend）。

隨著視覺化工具與技術的進步，使得知識結構的表現大有可為，並且能夠藉由與互動資料，以直覺的方式探索知識（Chen et al., 2001）。依不同的原始資料，使用者能與資料最終視覺化結果互動的方式也其有所差別（Keim, 2002）。雖將特定領域的知識視覺化工作是難以機械化且消耗時間（Börner et al., 2003），但是透過引文網路將某一特定領域的發展視覺化地呈現出來，則研究學者便能更進一步透過使用者介面的與資料互動，而進行大量文獻資料的分析及導覽。Shneiderman 與 Plaisant（2005）指出使用者與視覺化資料的基本互動功能有七項：概觀、縮放、資料過濾、資料探索、相關性、歷史記錄及資料抽取。透過這些功能，不但能讓我們瞭解其領域內部的關係脈絡、直接與資料互動，視覺化的資料更還有以下兩項優點（Keim, 2002）：(1)能容易地處理大量且不同類型的資料，(2)探索是靠直覺，並且不需要瞭解任何複雜的數學或統計演算法或參數。因此透過視覺化後的知識領域資料互動瀏覽、移動、調整、分類、分群、脈絡剖析，特別能讓一個尚未對該知識領域有十分瞭解的研究者，進行研究探索性的導覽，這點與閱讀搜尋引擎所查詢結果並依特定權重所列資料有很大的不同。

## 2.5 引用文獻分析與搜尋相關研究

使用自動建立引用關係的 CiteSeer（Giles et al., 1998）及其改進版本的 CiteSeerX（Li et al., 2006）系統，是目前網路上最大的文獻 Meta-Data 來源（Petinot et al., 2004）。在 CiteSeer 中，使用 Baseline Simple、Word Matching、Word and Phrase Matching 與 Likelt 進行引用文獻的辨認。雖然 Likelt 法則的辨認效能不理想，但其原則即能讓 CiteSeer 不需經人工辨視文獻，並統計被引用次數及估計文獻的重要程度。在使用介面上，則是依一般的搜尋方式表現，也就是鍵入關鍵字後，系統列出相似的文獻，並依相似程度順序式地將結果列出。此外，使用者可透過網頁中的超連結，進行文獻修正或是引用文獻、Meta-Data 查詢、使用 API 查詢與個人化的興趣比對（Petinot, et. al, 2004）。

HistCite（Garfield et al., 2003）與前者不同的是，將引用文獻資料以視覺化的方式，依文獻發表年代表現出文獻之間的引用關係。CiteSpace（Chen, 2006）使用不同的學術文獻網路服務，以視覺化網路圖形表達，除了呈現文獻網路之間的引用關係，並以顏色區別出文獻的發表年，以達到如同 HistCite 的文獻成長效果。Ke 等人（2004）使用大型網路分析軟體 Pajek，進行 ACM 中的文獻引用網路視覺化。而陳銘翔（2006）與高孔明（2008）則是使用 Prefus 函式庫（Heer et al., 2005），建立引文

網路視覺化平台，讓使用者在操作介面上有較大的彈性。

綜合上述研究，雖然都進行引文網路的建立與視覺化以協助重要文獻或人員的分析與辨視，但無法提供研究入門者特定知識領域的概括導覽，且都只適合對該領域已有認知的研究人員才能進行引文網路間的分析。當研究人員對知識領域還不甚瞭解或是不清楚其領域中的知識概念，便很難區別文獻之間的引用關聯為何。因此對於不同的研究者角色，必需輔以不同的文獻導覽功能協助其進行文獻探討工作。此一導覽功能不僅需要能表現出知識領域所包涵的專屬概念，亦需能夠透過已知的知識概念、文獻與人員的關聯原則以推敲文獻所內涵的知識內容。

### 3. 建議的文獻導覽平台系統架構

在經過上述的文獻探討之後，本研究建議一個文獻導覽平台系統架構（圖 1）。依據三種不同的使用者（一般者、入門者、釐清研究方向者）角色，並針對其需求提供符合的功能。主要系統功能區分為文獻資料收集與關聯系統（Paper Collection and Relation Building System, PC&RB）與視覺化導覽平台（Visual Browsing Platform, VBP），此二者間的資訊溝通透過應用程式介面（API, Application Interface）交換文獻資料。

#### 3.1 使用者需求角度

為正確在滿足不同研究者文獻搜尋需求，本研究依尋找文獻資料使用者需求，將其區分為三種：一般查詢瀏覽者、知識領域入門者、釐清研究方向者。

- (1)一般查詢瀏覽者：此類使用者不一定有特定的目的，除想要瞭解該領域知識的大體概念外，也會想從已知的知識著手瞭解該領域知識。針對此類使用者，資訊提供應著重在知識概念的瞭解，系統不僅需要提供一般基本的介面操作功能，亦應快速反應使用者查詢、提供宏觀網路層級、網路層級間的切換、由知識本體出發等需求。
- (2)知識領域入門者：此類使用者因為興趣或需要想對其有更深入的瞭解，但透過學術搜尋引擎卻無法瞭解該特定領域的知識，更不知道自己該從何處深入瞭解。這類使用者期待能瞭解這個知識領域中常見的觀念、文獻、人物，或是其本理論及該領域發展現況和大方向，並且能取得這相關文獻進一步研讀、探討。此類使用者所需要是從已知的知識概念去瞭解領域知識在實際研究上所代表的義涵，並且更進一步去推敲目前研究所應用到的領域與在同領域中的研究方法。系統則需協助使用者能夠進行網路資料探索（包括宏觀與微觀切換）、重要資



料的過濾、由瀏覽領域知識出發的引文網路。因此，系統資訊提供則會著重由知識本體與關鍵字層中的知識本體骨架網路所延伸出去的關鍵字、文獻、人員與引用文獻及這些資訊的關聯方式。

- (3)釐清研究方向者：此類使用者對該知識領域已有一定的認識，想釐清可能的研究方向。這類使用者需要能瞭解目前已經完成的領域之各項研究方向、議題及與其他應用知識概念間的關聯與脈絡，以決定研究方向的可行性。當然這類使用者也會想瞭解已經有那些議題曾被完成，而在現今的外在條件則能接續延伸、藉用新觀念產生研究方向，或者從過去已知的領域或人員發掘出可能研究缺口，進一步找出合適的研究方向。系統在協助此種角色，除了提供大量的資料供使用者瞭解該領域的詳細情況外，也應提供更多的過濾與介面操作的相關資訊，以協助釐清研究方向者能進行更多的文獻分析。

### 3.2 系統架構功能說明

本概念性系統架構乃結合了文獻探討中所提及之引文分析、人員關係網路分析、知識本體與關鍵字分析及資訊視覺化表示方法，配合研究人員的使用需求分析，並加以修正、整合而成。因此會將過去文獻資料依引文關係、人員關係網路及知識本體與關鍵字網路整理，再透過視覺化導覽平台（VBP）配合互動機制呈現資料，以供使用者進行文獻探索與延伸查詢。

#### 3.2.1 文獻收集與關聯子系統（PC&RB）

在此 PC&RB 中主要區分成四個部分，分別為知識本體維護模組、網頁服務溝通模組、引文網路分析與網路計算模組與 PC&RB API（圖 1），依序分述如下。

##### 3.2.1.1 知識本體維護模組

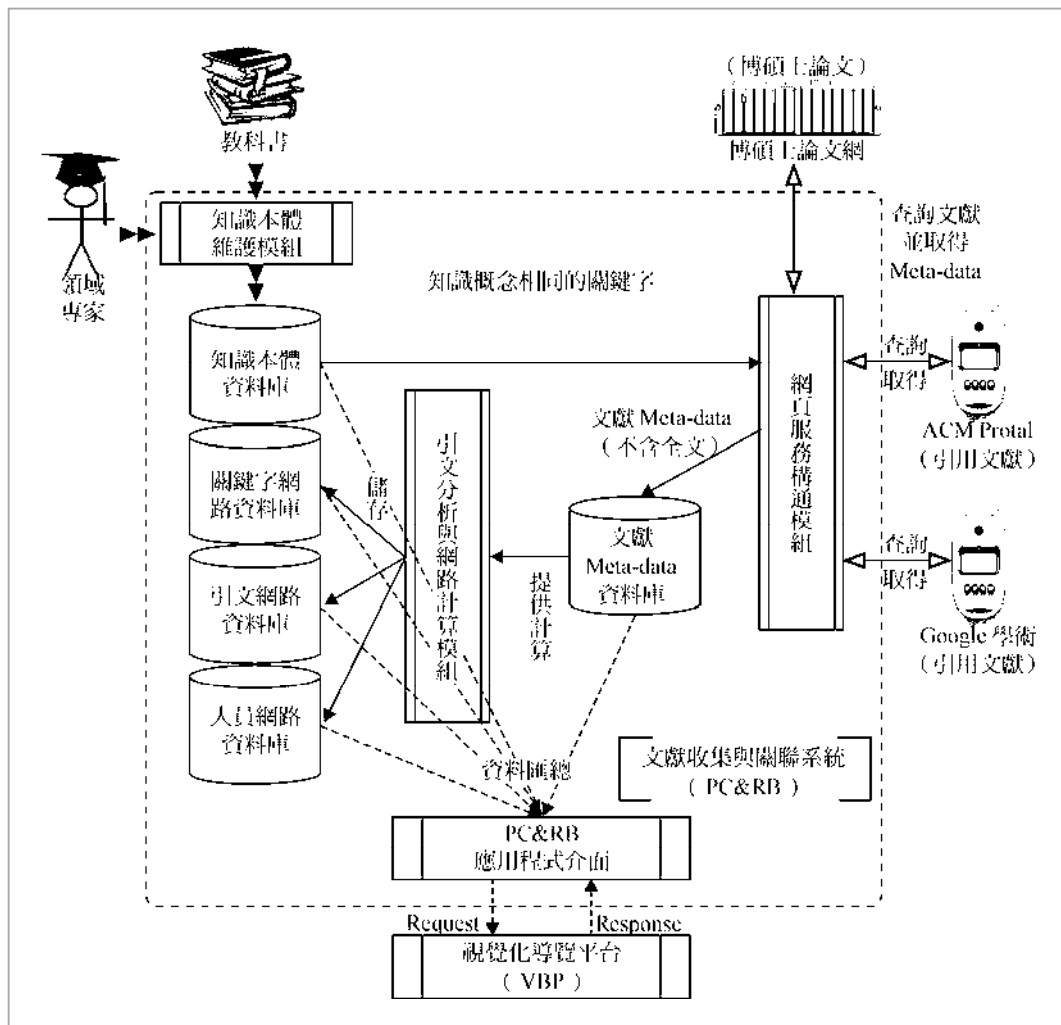
本研究參考 Noy 和 McGuinness（2001）、Uschold 和 Gruninger（1996）及王奕涵（2006）知識本體建構方式，並請領域專家來協助建立本系統架構所需的導覽用領域知識本體，因此沒有細部定義到屬性、事實與實例。此一模組提供知識本體維護介面。

##### 3.2.1.2 網頁服務溝通模組

此模組主要為與不同的網路文獻查詢服務溝通，以取得文獻的 Meta-Data。假設我們以國家圖書館碩博論文網為主要論文資料來源，透過文獻查詢取得的 Meta-Data 包括了作者、指導教授、學位別、校名、系所名、關鍵詞、摘要、引用文獻等資料欄位。由於該論文網中的各論文引用文獻編寫的格式不統一，無法自動正確辨別引用文



獻的資料欄位，而需經過資料清理之後才能進行引文網路的擴展搜尋。本研究利用向其他的文獻入口（如 ACM、Google）搜尋服務查詢，不僅未來可向外擴展多層查詢引用文獻外，並可藉助該等入口網之自動濾掉常用字詞與錯字糾正功能，協助辨別原論文網中不同引文格式，校正並補足該引文的作者、篇名、發表期刊、卷期等資訊，以協助資料清理。



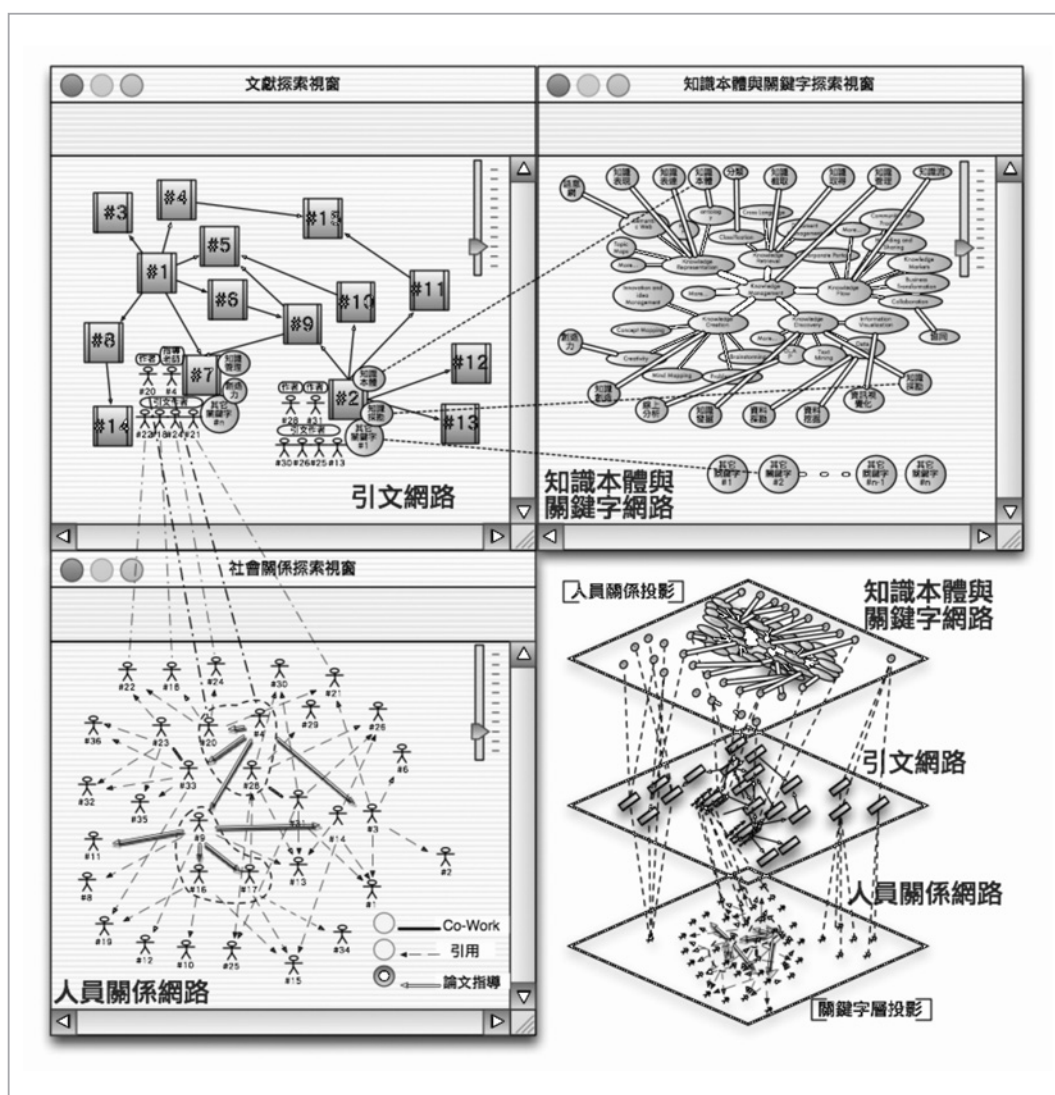
▲ 圖 1 文獻資料收集與關聯計算系統架構圖

### 3.2.1.3 引文分析與網路計算模組

為了建立三層網路關聯，必需分析所取得的資料文獻、關鍵字、相互引用、人員關聯方式，以建立系統中必需的「知識本體與關鍵字網路」、「引文網路」與「人員



關係網路」等三層網路（如圖 2）。「知識本體與關鍵字網路」是由知識本體與其同義之關鍵字所建立而成的網路關聯，其中知識本體為該網路層骨架，可表現知識概念間關聯方式與知識概念的同義英、中翻譯字詞。「引文網路」依循文獻間的實際引用情況，揭露了文獻之間的引用關聯，透過此關聯可瞭解前述所介紹之文獻的叢集、書目對、及共被引關係。「人員關係網路」則是透過文獻及引文中的 Meta-Data 取得的作者資訊，將人員共引用（Author Co-citation）關係與共同作者（Co-author）關係揭露出來。



▲ 圖 2 文獻中 Meta-Data 對應圖

透過上述的三種網路關聯建立，則可將文獻的 Meta-Data 從基本的文獻引用關係，延伸到人員網路層與知識概念層中。更進一步透過不同層級網路之間的投射，即可以瞭解特定知識概念能經過引文網路而投射到人員網路層的範圍；反之，亦可得到特定人員已經完成了那些知識領域中的研究（如圖 2）。

### 3.2.2 應用程式介面（PC&RB-API）

前述的 PC&RB 會透過此一介面傳輸給視覺化導覽平台（VBP）需要的資料，以滿足因應不同使用需求，所提供的不同導覽互動方式，以及三種網路層級間的圖層切換。為了讓所收集且處理好的資料能被有效地再利用，建議透過 Web Service 來傳遞資料。而所包括的資料內容則依所要呈現的內容規劃，視情況外擴展網路範圍。此外為減少視覺化導覽平台的不必要的計算，所傳送資料使用 GraphML 格式來傳遞。

### 3.2.3 視覺化導覽平台（VBP）

本研究參考陳銘翔（2006）與高孔明（2008），並結合 Shneiderman 與 Plaisant（2005）所建議互動功能，配合前述之使用者需求分析，設計視覺化導覽平台以協助瞭解文獻間的義涵。VBP 平台中設計各節點間的推力與節點關聯的拉力，使節點能呈現自動群聚的效果。而互動機制除了能提供使用者導覽更多文獻資料與網路層級切換外，也能透過平台中視覺化參數的微調（如放大、縮小等）達到最好的視角。系統包涵之互動功能有：(1)概觀功能：概觀三個網路圖層之間的交互關係，以瞭解全局脈絡及趨勢；(2)縮放功能：使用資料過濾功能，減少網絡圖型之圖層、節點、關聯，以瞭解局部細節；(3)資料過濾功能：依使用者設定的觀察焦點，過濾關聯距離較高的資料；(4)資料探索：依循平台中的資料間關聯方式，探索平台內之資料；(5)相關性：平台中透過不同顏色的節點與單 / 雙向線以揭露資料間相關內容；(6)資料抽取：提供資料匯出之功能，圖形匯出與 GraphML 檔案格式匯出。此外本研究使用顏色來辨別節點與關聯，在不同層級中，同樣的節點性質與關聯性質採用統一相同的顏色，以便使用者直覺瞭解在複雜網路裡連結圖形所代表的涵義。另一方面參考文獻中的引文網路文獻推薦之 PaperRank 演算法（Gori and Pucci, 2006），以計算所得之權重來顯示節點大小。在知識本體與關鍵字層中，除了可以明白知識本體之間概念上下層級的關係外，亦可知道與知識本體同義的論文關鍵字是如何設定的，供研究人員進行關鍵字間的共字分析。在引文網路層中，不同於一般網頁論文搜尋服務的呈現方式，VBP 是使用網路圖形供使用點選查詢與瀏覽，讓研究人員能透過視覺化的書目對與共被引連結圖形，呈現較佳的參考文獻。而人員關係網路層則除了表現出作者間的師徒與師門之社會關係外，亦可知道作者在撰寫論文時所引用的文獻作者，藉由社會關係與作者共引關聯，呈現研究人員之間的引用偏好。



## 4. 雛型系統實作、情境說明與驗證

### 4.1 雛型系統實作

依上述建議的系統架構，本研究使用軟體工程的瀑布式模型（Peters and Pedrycz, 2004）來開發雛型系統。我們使用數種程式語言來建置系統雛型，以求快速開發驗證，這包括使用 Ruby 語言對文字處理優點及穩定的執行緒功能來開發各種溝通模組、以 PHP 開發 Web Service 為基礎的 PC&RB-API 平台、使用 Prefuse 函式庫（Java）來設計 VBP，並使用 MySQL 建立所有關聯資料庫。其中 PC&RB 中的溝通模組直接將收集到的文獻資料存入資料庫中；VBP 與 PC&RB-API 則是以 Web Service 的方式傳遞經過計算的文獻網路資訊。

#### 4.1.1 領域知識本體建立



▲ 圖 3 本雛形實際包含的 Data Mining 知識本體

系統雛型所使用的知識本體建置步驟如下：(1)定義知識本體的領域與範圍：在有限的時間資源下，本研究選定一小範圍「Data Mining 概念下的應用技術 (Technique)」的知識概念來制作系統雛型知識本體，並以「Data Mining Technique」一詞做為此知識本體的總合概念 (Aggregate Concept)；(2)參考現有的知識本體：為了讓此知識本體能達到導覽研究入門者之功能，本研究參考資料挖掘的數本教科書 (包含 Williams and Simoff, 2006; Tan et al., 2005; Roiger and Geatz, 2002; Hand et al., 2001; Berry and Linoff, 1999; 丁一賢與陳牧言, 2005) 以建立知識本體雛型；(3)列舉知識本體中重要的概念；(4)定義知識的本體與架構；(5)請 2~3 位曾經教授過資料挖掘之教授，透過知識本體維護模組修改此知識本體雛型本體。最後本雛型領域知識本體如圖 3。

#### 4.1.2 文獻與引文收集

本研究使用上述全部知識本體為關鍵字，以程式從國家圖書館的碩博士論文網中查詢，並限制論文為 90~97 學年間、作者所屬系所字串中必需包括有「資訊管理」四個字，來取得論文資訊。共取得論文 1,071 篇。在論文網中所提供的論文摘要頁面中，包含作者、指導教授、學位別、校名、系所名、關鍵詞、摘要、文獻等資料欄位，且回傳的資料型態為 HTML 語言，透過 HTML 解析器 (Parser) 後，取得各特定 Meta-Data 內容。

在取得論文 Meta-Data 後，還需進行資料清理。如各論文引文資料內容會因為作者編寫引用文獻的格式不同，而無法正確提供欄位資訊內容。此類資料清理工作則分為兩階段，(1)人工修正資料型式；(2)透過外部學術文獻查詢服務，以取得正確之引文資訊。這乃透過網頁服務溝通模組，將英文引文送到 Google Scholar 與 AMC Portal 進行資料查詢以取得 Endnote 資料，而中文引文則用百家姓以取得作者資訊，並以最長字串原則決定該引文文獻主題。經過上述之資料清理之後，取得總計引文文獻量共 22,854 筆，存入文獻 Meta-Data 資料庫中。依原架構，是可將引文的引文再次納入 Meta-Data 資料庫的，由此可無限擴充，但目前雛形只將碩博論文的引文納入，並未再往外擴充。

#### 4.1.3 關聯建立

我們透過引文分析與網路計算模組分析上述所取得的文獻與引文間的關聯，以建立知識本體、關鍵字、文獻與人員之間的關聯。在「知識本體與關鍵字」層中，建立了英文關鍵字、中文關鍵字與知識本體概念間的同義關聯，亦計算文獻間的共字關係次數為知識本體與關鍵字之權重，並分別存入知識本體資料庫與關鍵字網路資料庫中。「文獻網路」除計算文獻之引用關係以定出權重，供視覺化節點大小之參考外，亦計算與人員節點及關鍵字節點間的關聯，一併存入文獻網路資料庫中。「人員關係網路」



則是依論文 Meta-Data 中的作者與指導老師的資料欄位來建立師徒關聯，再將作者與所引用文獻的作者建立引用關聯，並依引用次數訂出權重，最後存入人員關係網路資料庫中。

#### 4.1.4 視覺化平台建立

為了讓使用者能直覺地進行文獻導覽，且能在各網不同層級間切換，本研究使用 Prefuse 資料視覺化函式庫以呈現複雜的網路資料。視覺化平台建立包含兩部分工作：(1)建立 PC&RB-API（文獻收集與關聯子系統之應用程式介面）：提供 Web Service 呼叫，以取得 GraphML 規格的網路資訊。(2)建立 VBP（視覺化導覽平台）核心：VBP 依前述系統規劃方式進行雛形 Java Applet 平台設計，將呈現介面區分為知識本體與關鍵字層（包含本層之知識本體骨架網路，即領域知識本體）、論文網路層與人員關係網路層，並提供互動機制以協助使用進行導覽。

VBP 的互動機制，包含前述 3.2.3 所列之六種互動功能：概觀、縮放、資料過濾、資料探索、相關性及資料抽取，此外，另加上三種互動功能：(1)一般者模式：本系統為了提高表現之引文節點的品質，並減少同時出現在視覺化平台中的節點數量，因此參考梁定澎（2003）對資管領域之期刊排名進行引文節點過濾，保留重點期刊之引文資料。雛型系統中，預設將「過濾模式」開啟來過濾引文節點，對一般查詢瀏覽者與知識領域入門者在進行平台探索與查詢時能提供較好的引用文獻資料以供參考，同時也能有較減少視覺化系統資訊過載。(2)網路觀點切換：在不同的網路圖層間，使用者可依興趣節點雙擊後，進入由興趣節點延伸之對應網路圖層（表 1）進行探索。(3)自動群聚：本研究設計節點間的推力與拉力，並使視覺化平台不斷計算節點之最佳位置，以達到視覺化介面中節點的自動群聚效果。

另一方面，為了區別出各節間之間的屬性與其重要性，在平台中以顏色區分出同屬性節點與關聯（表 1），且透過權重計算結果以節點的大小表現出該節點在網路結構的重要性。權重計算的方式在不同層級中的計算方式不同。在「知識本體與關鍵字」層中，各關鍵字節點的大小透過該字在系統所取得論文使用的次數決定；在「引文網路」層則使用 PaperRank 之權重計算原則，決定文獻與引文節點的權重；在「人員網路層」中，則依人員被引用的次數決定人員節點之權重。

#### 4.2 雛型系統情境說明

以下依前述之三種不同角色情境：一般查詢瀏覽者、知識領域入門者、研究方向尋找者，來說明本雛型系統如何去滿足不同角色之需求。

目前圖層	關聯型態	節點型態	切換圖層至	切換功能說明
知識本體骨架網路	(紫色) 單向箭頭為概念上層，反之為下層。 雙向箭頭為同義詞。	知識概念 (黃色)	知識本體 與關鍵字層	探索以所選之知識本體，尋找與其概念相同的關鍵字，與此相同概念之關鍵字共同出現在同一篇文章中的共用使用關鍵字。
知識本體與關鍵字層	(紫色) 單向箭頭為概念上層，反之為下層。 雙向箭頭為同義關聯。 (皮膚色) 同時出現於特定的論文的共字關聯。	知識概念 (黃色)	知識本體 與關鍵字層	與在知識本體層中雙擊知識概念相同。
		關鍵字 (淡藍色)	引文網路層	查詢所點選關鍵字所出現的論文，並列出其所引用文章與共同引用文章之關聯網路圖。
引文網路層	(灰色) 論文內所標註的關鍵字、作者姓名或指導教授、引用文獻。	關鍵字 (淡藍色)	引文網路層	與在知識本體與關鍵字層中點選關鍵字相同。
		文獻 (亮紅色)	人員網路層	尋找所點選之文獻中作者，並且尋找該作者有共同著作(論文指導關係)之作者、所撰論文被引用之作者(cited in)與所撰論文引用之碩博論文作者(cite out)。
		引用文獻 (淡綠色)	引文網路層	尋找同時引用該文獻之碩博論文。
人員網路層	(灰色) 論文所標註的作者或是指導教授。 (粉紅色) 論文之師徒關係，箭頭方向為老師。 (淡藍色) 論文作者引用關係，箭頭方向為被引用文獻之作者。	人員 (皮膚色)	引文網路層	列出所點選作者所撰之文獻，與此文獻所用之引文、關鍵字。

#### 4.2.1 一般查詢瀏覽者

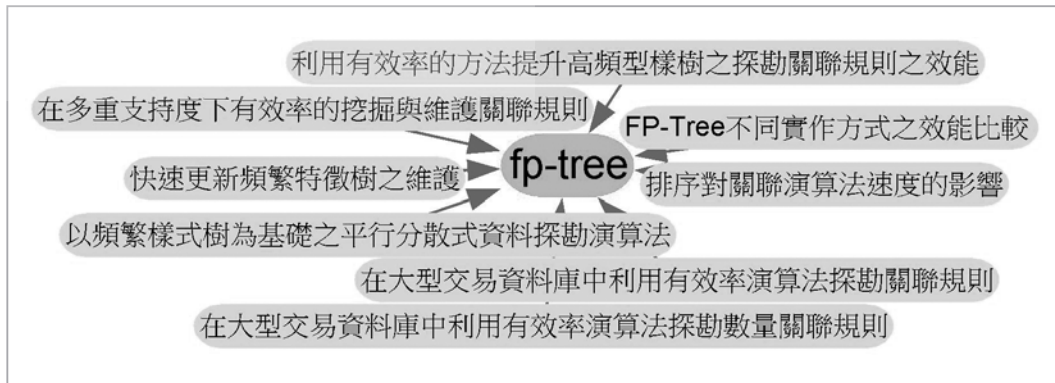
一般查詢瀏覽者對本系統與特定的知識領域沒有基本概念，因此著重在知識領域的概念化資訊的提供。

- (1)知識本體(知識本體與關鍵字層的骨架網路)－可協助認識領域知識本體架構。在探索平台上，使用者先進入知識本體層(圖4)，可直覺瞭解此「Data Mining 概念下的技術(Technique)」中的知識概念，其概念間透過單向箭頭連接，表示各概念間的一般化(Generalization)與特殊化(Specialization)關聯。如 Decision Tree 下有 C5.0 技術等。









▲ 圖 7 雙點擊關鍵字「fp-tree」後之引文網路層

及引文文獻作者（淡藍色單向箭頭指向）。延續上例，若使用者對「在多重支持度下有效率的挖掘與維護關聯規則」之論文節點有興趣，在瞭解節點作者資料為「國立中央大學資訊管理所」的「胡雅涵」後，透過雙點擊論文節點，而切換圖層進入人員關係網路層（圖 8），使用者會在平台上發現「陳彥良」是直接被許多人員的粉紅色箭頭所包圍，即表示「陳彥良」為指導教授。在此圖形亦可知道陳彥良教授所指導的學生「胡雅涵」所引用文獻的作者，包含如「aggarwal c. c.」、「yu s. p.」等，由於這些節點較其他大，所以可知在資料庫中，很多文獻亦會引用這些作者所寫的文章（每個節點代表該作者，其所寫得文章可能不只一篇）。再者同師門的學生，如「陳仕昇」、「余忠慶」…等人的研究也都會去引用上述「胡雅涵」所引用的這些特定國外人士所撰之文獻。由此可認知特定知識領域中作者間的專業領域關聯（蔡明月，2003；Chen and Paul, 2001）。

透過以上的使用過程，使用者瞭解「Apriori」是在「Association Analysis」中的一個特別化概念，且在實際的論文中，會以「apriori」、「apriori 演算法」或「apriori algorithm」關鍵字實際出現在論文中，而若以「apriori」關鍵字，則會使用於「房屋稅…」、「舌診…」、「高頻物…（項集合）」、「交易物…（項）」、「fp…（tree）」等應用領域。依此延伸探索，再從其共字關聯的「FP-tree」關鍵字切換到引文網路層中發現七篇論文研究。點選吸引其注意的中央大學資管所「胡雅涵」論文後，更進一步進入人員網路層，瞭解其老師「陳彥良」所指導的學生亦會去參考「aggarwal, c. c.」、「yu, p. s.」、「han, j.」等作者的文獻。這樣的過程不斷地重複，即是研究者在進行文獻分析時的一種探索流程 — 從知識領域到關鍵字、從關鍵字找論文、從論文找引用文獻、從文獻作者相關的社會網路尋找可能參考之作者。

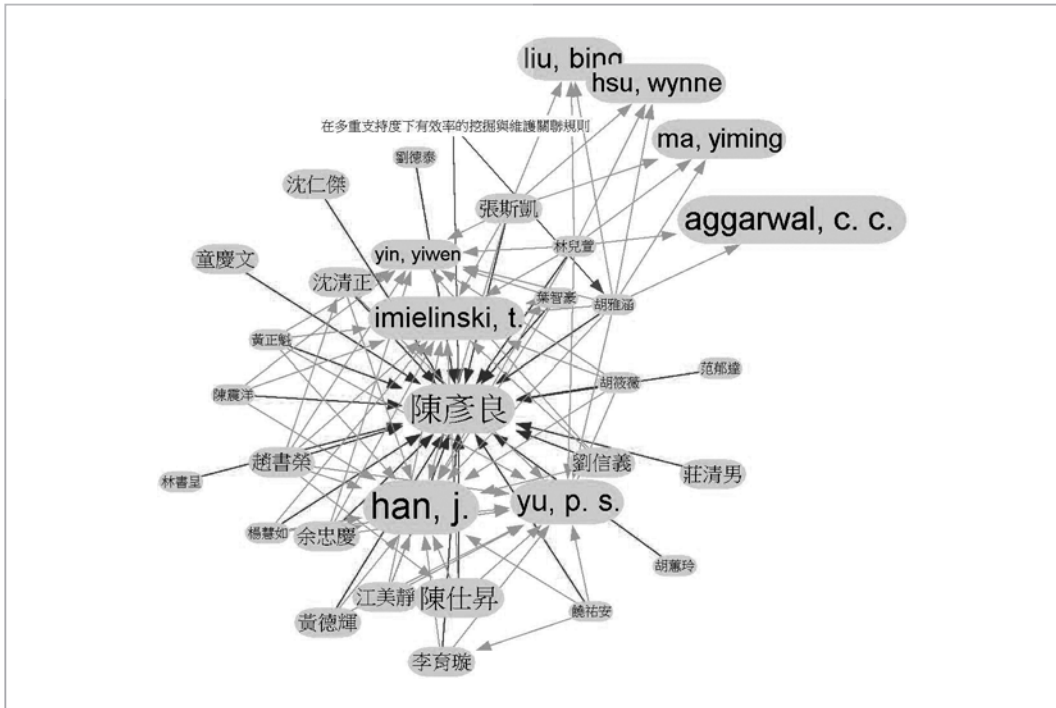
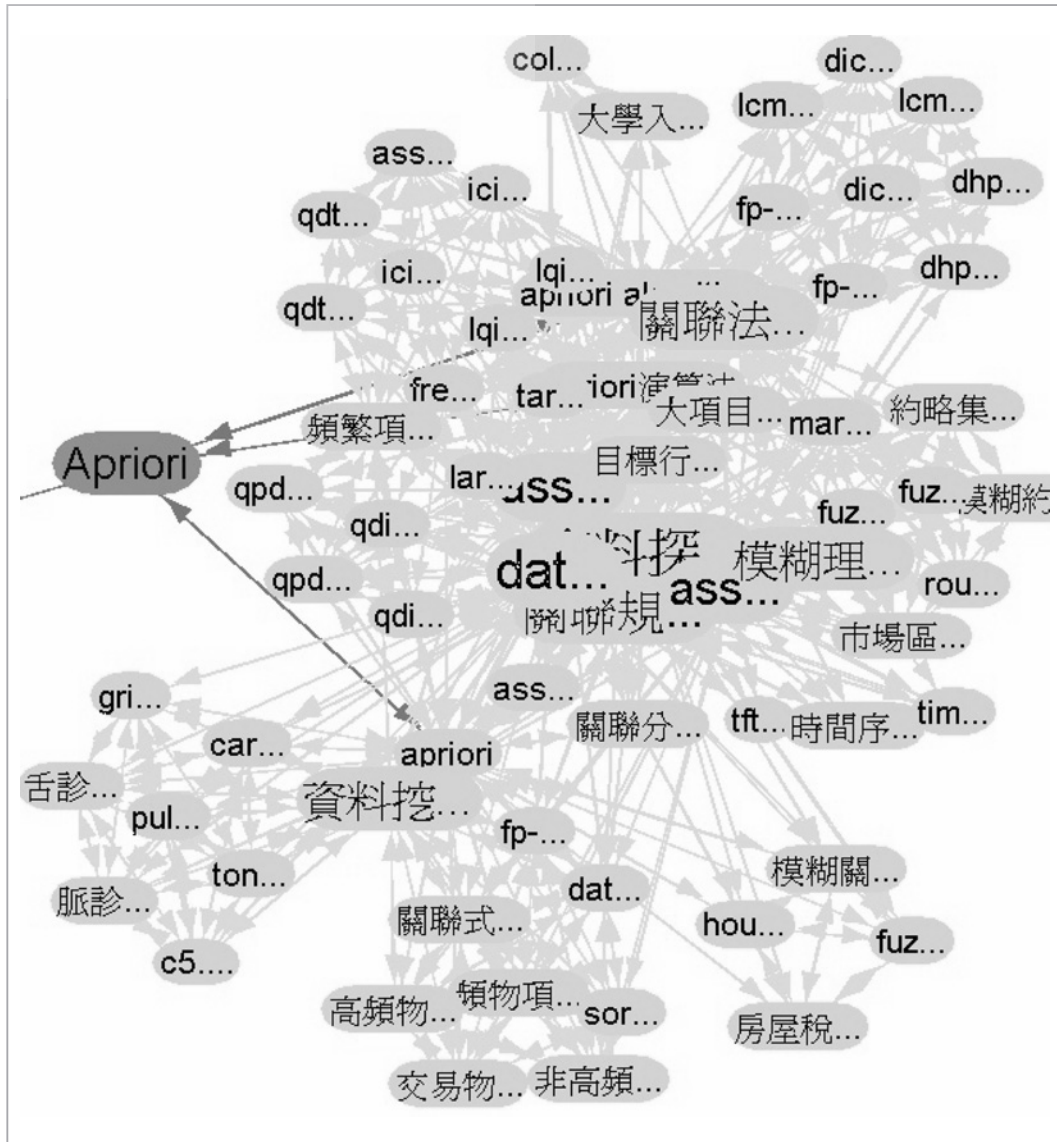


圖 8 雙點擊興趣論文節點後，進入人員網路層

#### 4.2.2 知識領域入門者

知識領域入門者對某些概念雖有瞭解，但仍不清楚在特定領域中有哪些已完成的重要研究文獻值得參考。本系統能除了提供一般查詢瀏覽者功能外，亦提供進一步互動介面功能，供入門者能透過網路中的關聯連結，在不同的屬性圖層中探索，以尋找適合的文獻參考。

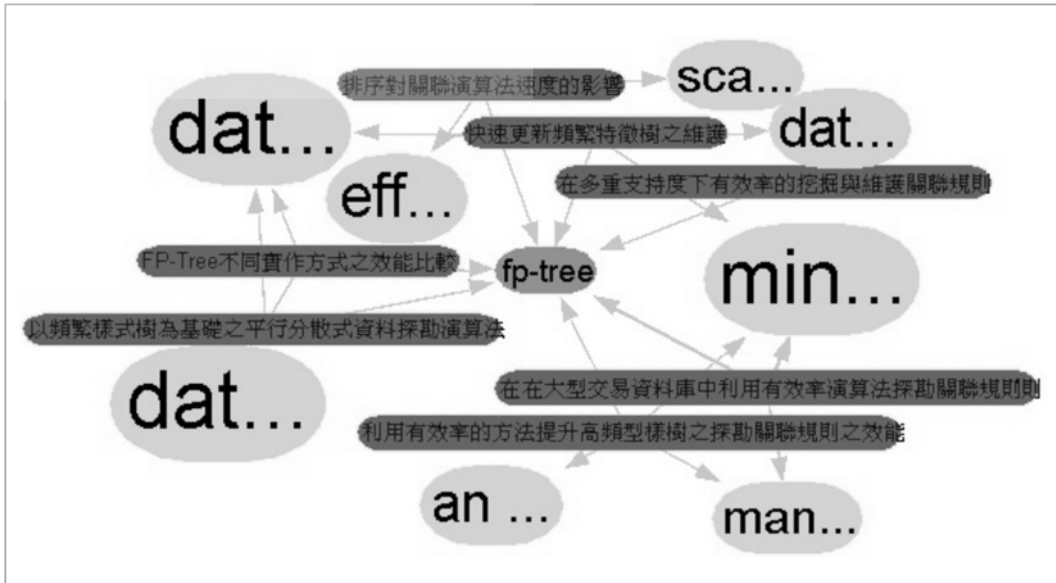
- (1)高關聯距離設定與自動群聚 — 便於視覺化挖掘。如前例中，在同樣探索路徑「Association Analysis」的特別化知識概念「Apriori」進行探索。在高關聯距離的設定下，本導覽平台會將呈現所有與知識概念「Apriori」有關聯的關鍵字網路（見圖 9），其中關鍵字與關鍵字的關聯建立是以皮膚色的共同出現關係（即共字關係）建立。在完成自動群聚後，可大體上看出幾個現況已經有人研究的應用群組，如：「col... (lege entrance exam)」與「大學入... (學)」群組、「約略集... (合理論)」與「市場區... (隔)」群組、「舌診...」與「脈診...」群組、「高頻物... (項集合)」與「交易物... (項)」群組、「模糊關... (聯性)」與「房屋稅...」群組等。這裡系統會同時去找所有圖 5 中的「apriori」、「apriori 演算法」或「apriori algorithm」關鍵字所出現的應用領域，所以入門



▲ 圖 9 知識本體「Apriori」關鍵字網路圖（入門者情境）

者看到的應用領域會比一般查詢瀏覽者還多。

- (2)引文網路層 – 可瞭解在相同關鍵字下，文獻間共引用與共被引之關聯。如前例透過雙點擊關鍵字節點「fp-... (tree)」進入引文網路層，在高關聯距離下（圖 10）可進一步看到其他引文文獻（過濾後）與論文間的關聯，所有的引文節點（淡綠色）是進行關鍵字「fp-... (tree)」研究的參考文獻。此處的過濾是以梁定澎（2003）所定之資訊管理期刊排名中優良的期刊文獻。在此網路圖

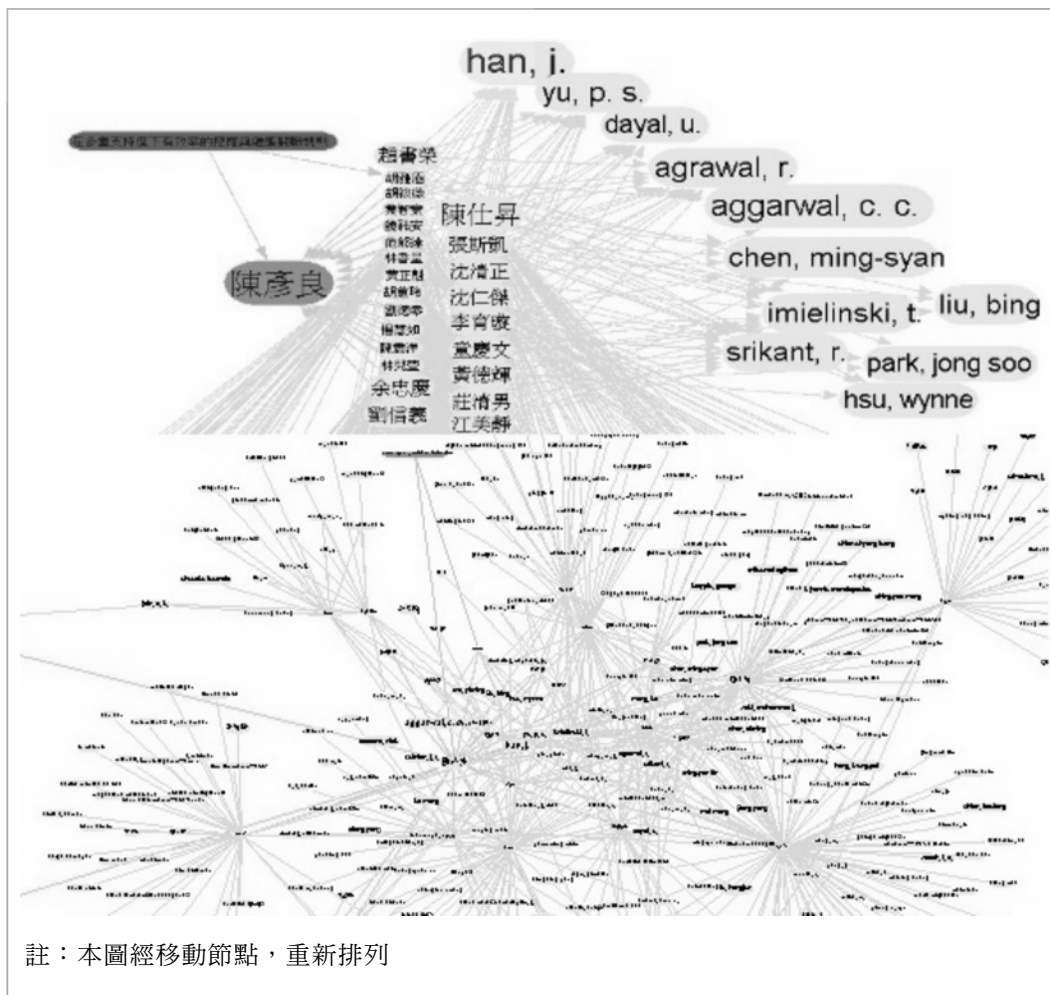


▲ 圖 10 文獻中出現「fp-tree」關鍵字節點的引文網路（入門者情境）

中，入門者可以直接參考在網路結構中被較多連入（Cited-in）的引用文獻，或是再進一步雙擊引用文獻，以取得更多論文及引文網路資料。

在圖 10 中可知，圖 7 的七篇碩博士論文引用優良期刊的中的八篇引文文獻，因此均適合入門者在研究此領域時參考。而其中「dat...」（即 data mining: an overview from a database perspective）與「min...」（即 mining sequential patterns）等引文節點有三篇碩博士論文同時引用到，即表示此兩篇引文更值得優先參考。此外在圖中這兩篇引文與另一篇「dat...（a mining- concepts and techniques）」均較其他節點大，此乃表示此三篇是在全資料庫中的文獻一併計算後權重較大，這是另一優先參考的提醒。

- (3)人員網路層 — 可進一步瞭解同師門共引用關聯。延續前例，入門者對有興趣論文節點「在多重支持度下有效率的挖掘與維護關聯規則」，切換到人員關係網路層後會呈現如圖 11 之高關聯距離的網路圖形。圖中被粉紅色箭頭指向的人員節點「陳彥良」是指導老師，另一端人員節點即為其研究生。研究生所依淡藍色線（參考表 1）向外的，是該研究生所引用全部文獻之作者。圖 11 不僅顯示出圖 8 中「胡雅涵」一人所引用文獻的作者，還包含其他同師門所有學生的各別引用。所以圖 11 不會只看到特定一位學生的引用被其他學生共同引用的狀況，還看到全部學生間相互參考的引用的狀況。由圖 11 可知，人員節點「chen, ming-syan」、「srikant, r.」、「agrawal, r.」、「dayal, u.」等的人員引用關聯十分稠密，因此可以推論在陳彥良老師所指導的研究領域中，多數會參



▲ 圖 11 興趣論文節點之人員關係網路層（入門者情境）

考到上述作者的文獻。而此關係亦是本雛型系統，在人員關係網路層中所揭露的「同師門下作者共引現象（author co-citation under same thesis advisor）」關係，即說明了老師在指導特定研究領域時，會傾向推薦學生進行研讀特定作者的作品，亦值得入門者選擇參考。

知識領域入門者在面對興趣的知識領域，往往不得其門而入。透過雛形系統高關聯距離下的網路結構配合自動群聚功能，能經由知識本體的導引，瞭解其所感興趣的目前所有應用領域的現況。再進一步針對特定的關鍵字能進行切換到引文網路層中，經過分析節點大小（節點大者）與關聯方式（關聯密度高者），尋找到整個學術領域適合的文獻進行研讀。亦或者在人員關係層中，依特定老師指導學生的模式，選擇「同師門作者共引」的較多學生參考的引文作者做為入門墊腳石。

### 4.2.3 釐清研究方向者

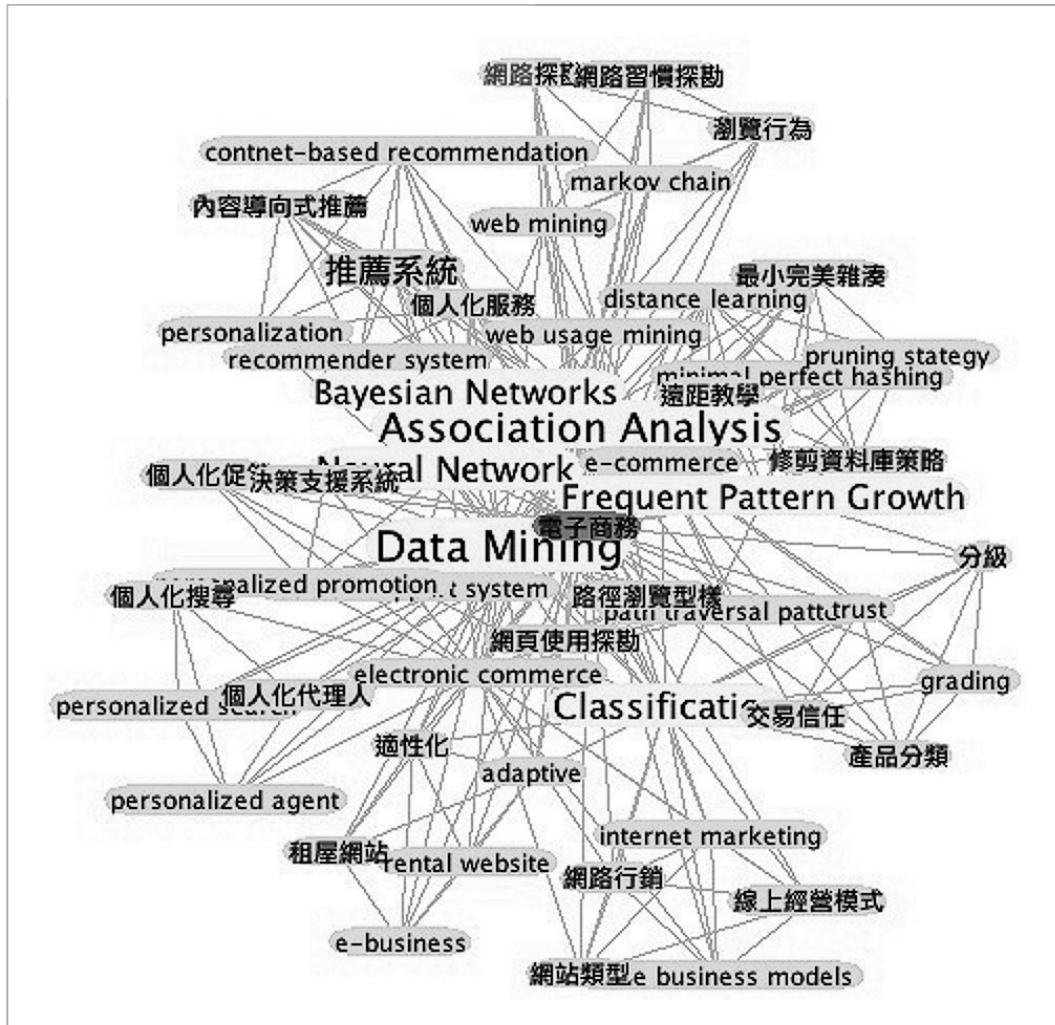
釐清研究方向者對某知識領域已有一定的瞭解與心得，使用本系統時較著重在進一步釐清目前此知識領域發展方向，及該領域現在已發表的脈絡，期能確認可能的研究方向及目標。因此較一般查詢瀏覽者與研究入門者更重視節點的關係，會很快地離開系統一開始所預設的知識本體層，深入去釐清研究方向。

- (1)全局檢索 — 以分析目前研究領域現況。透過雛形視覺化平台左上角的觀點切換，釐清研究方向者可以快速直接進入「知識本體層與關鍵字網路層（全圖）」和「關鍵字網路層（全圖）」，配合自動群聚功能、不同的關聯距離設定及探索互動功能，讓釐清研究方向者能分析網路圖中的各節點及關聯方式，並深入瞭解所產生網路結構的各項參數的視覺化義涵，並且能在不同的網路圖層中切換，進行來回地探索及分析。

「知識本體層與關鍵字網路層（全圖）」（如圖 12）是不經過濾的知識本體與同義關鍵字圖層，能釐清研究方向者在已知所欲研究問題的領域時，減少其不必要的互動功能。在此圖層中，配合「查文獻」功能可以直接提供釐清研究方向者所有屬於同知識領域的文獻。而「關鍵字網路層（全圖）」則是表現所有關鍵字間的「共字關係」（如「電子商務」會與「Neural Network」中的「Bayesian Networks」與「Association Analysis」中的「Frequent Pattern Growth」共同出現）。由於此網路層的節點多且關係複雜，因此在此層進行探索時，需對本領域的研究問題有充分的認識。再配合系統所提供之自動群聚與互動縮放機制，能讓使用者取得最好的觀測分析視角。如圖 12，在自動群聚後，可大體上區分為「推薦系統」、「個人化服務」群組、「瀏覽行為」與「網路習慣探勘」群組等。

- (2)進階互動功能。系統亦在右方工具區塊中提供查文獻功能，當使用點選某一特定的知識概念之後，點選「查文獻」按鈕即可馬上切換到引文網路層；「停止群聚」功能則是停止目前拉、推引力間的作用，讓所有群聚效果停止；關閉「過濾模式」功能，會顯示許多非資管期刊排名中的文獻，包括書、研討會論文等。比較圖 13 與圖 14，研究人員經由關鍵字節點「電子商務」切換進入引文網路層後，由於「過濾模式」會依梁定澎（2003）之排序而過濾可參考之文獻，因此在圖 13 中的文獻會包含「fas…」（即 fast algorithms for mining association rules in large databases）、「new…」（即 new sweeder: learning to filter netnews）等參考資訊。

在此情境下，雛形系統將資料庫中的資料依關聯與脈絡完全展現，輔以即有的互動機制，供研究人員能快速地在不同網路層級中比較節點性質、關聯方式及網路結構，瞭解已發表的文獻和引文資訊，以確定可能研究的方向或可進一步協助研究的參考文獻。



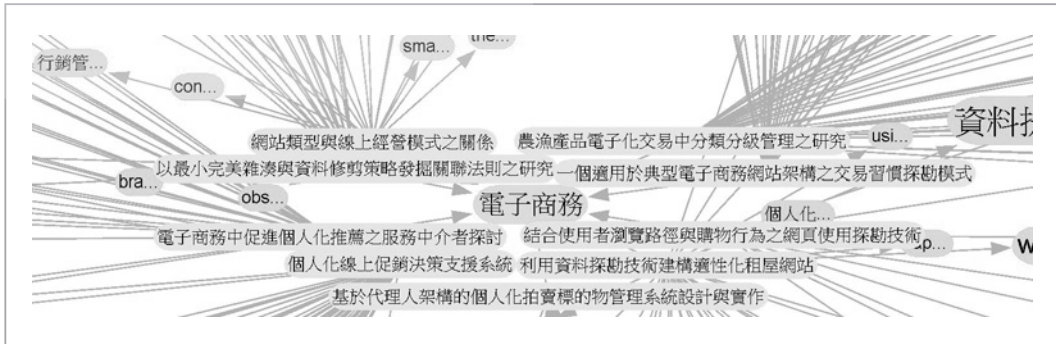
▲ 圖 12 關鍵字「電子商務」的知識本體與關鍵字網路（釐清研究方向者情境）

#### 4.3 對系統雛形之意見回饋

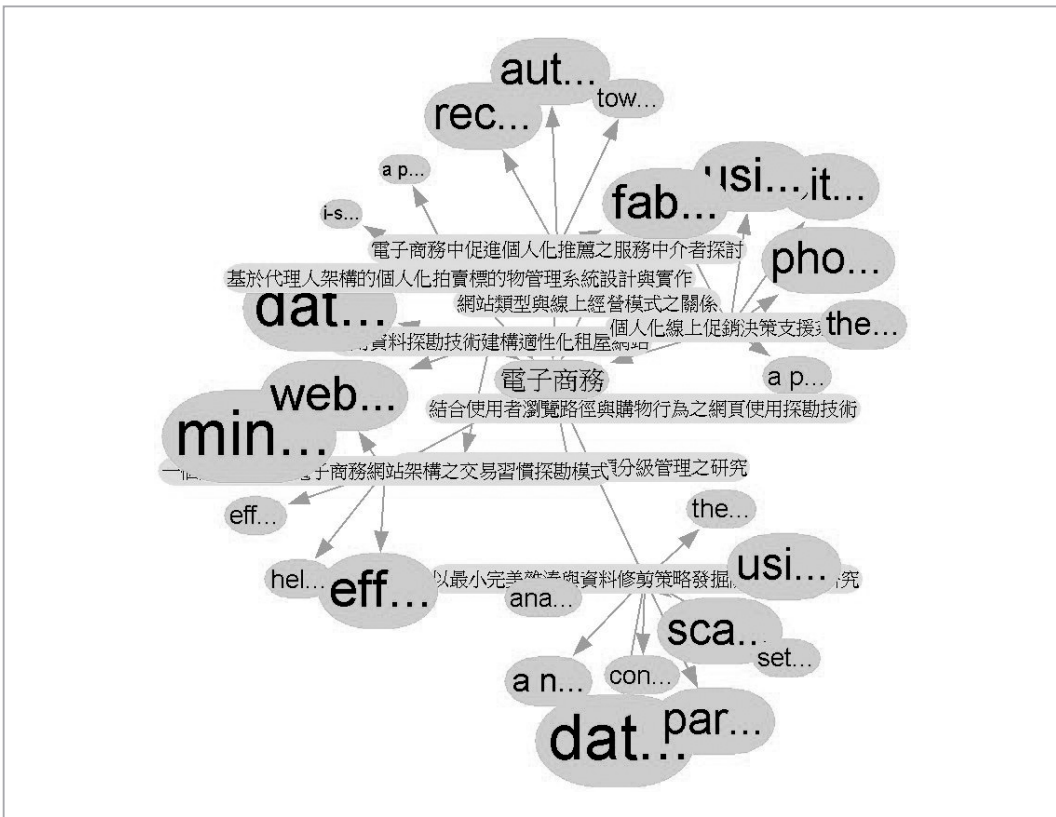
本研究所提出之觀念與架構，進行雛形系統實作後，請了某大學資訊管理碩士班五位同學進行小規模試用，以提供對該系統意見回饋，整理如下。

- (1) 對系統功能可進一步加強之建議：(a) 由於雛形系統互動功能操作（即單點、雙擊等）與一般介面不進相同，所以需要更多時間熟悉才能在此導覽平台中自由探索。如何協助快速上手使用，可再思考；(b) 目前雛形系統，尚無法提供博碩士論文、或其他資管期刊資料庫論文全文下載，未來可與相關單位合作。
- (2) 對系統效益評估意見回饋：(a) 可透過自動群聚後的知識概念與關鍵字網路的分佈情況，瞭解現在已經完成的研究為何，以進一步思考在尚未產生關聯的節點





▲ 圖 13 未使用「過濾模式」之引文網路



▲ 圖 14 使用「過濾模式」之引文網路



間，是否存在可行的研究的方向以建立該節點間的關聯；(b)能提供給使用者瞭解該知識領域的概念，與其概念間的關聯，並能讓使用者能夠清楚其所感興趣、「雙擊」點選後的文獻是位於那個知識概念下，較不易迷失在知識領域中；(c)資管知名期刊「過濾模式」所提供的引用文獻，不僅能協助使用者能瞭解對某關鍵字在研究時應參考的文獻，也能透過該再「雙擊」引文節點瞭解該引文可以被應用到的研究；(d)可以在人員網路層中，瞭解同一老師的不同研究生在撰寫論文的時候，會引用那些人員的文獻，並透過「雙擊」人員節點瞭解該人員所寫的引文或文獻被那些研究領域引用；(e)在平台中所呈現的節點資料，不僅可依節點大小判定值得參考的文獻，透過節點的連聯密度與叢集間的連繫節點，亦可知在該領域中較值得參考的文獻。

## 5. 結論

本研究提出一個文獻導覽平台系統架構，以博碩士論文網為例子建立導覽平台雛形，所展現的除了文獻計量學中的共引用、共字、共作者引用等在博碩士論文網中的顯性知識關係，亦透過知識本體、梁定澎（2003）的論文評比進行文論過濾原則及人員網路層中的師徒關係，展現了文獻中其他較顯性的知識分類、知識過濾對知識傳遞的關係。最後透過特別設計的互動機制，供三種不同觀點的使用者能在三個不同性質的網路層級間切換，並依雙擊後所查詢節點網路結構導覽使用者參考特定文獻。在本研究中的知識本體與關鍵字層級，表現已完成的研究中所使用的關鍵字與依本體論關聯，與論文所使用的共字網路為何；引文網路層，呈現依不同的查詢節點現存的博碩士論文間的文獻與關鍵字引用關聯為何；最後人員網路層，則是顯示師徒關聯、同師門的共被引關係及人員間的引用關係。

本研究所提出之架構，與現有的一般學術搜尋或資料庫搜尋引擎方式不同點在於下列：(1)導覽與查詢的方式不同：在本導覽平台中，使用者的查詢是依循著知識本體「雙擊」後進行層級間的切換網路層級，而一般搜尋引擎是依特定的領域、期刊或是主題（Topic）查詢後，依列表方式呈現。(2)全域與區域權重：雖然本導覽平台的「雙擊」動作與一般搜尋引擎的動作都具觸發搜尋的義涵，但是一般搜尋引擎列表的呈現方式是依計算過後所有文獻的特定指標權重為準，讓較重要的文獻列表在最上層。反觀本導覽平台不僅能提供全域計算後的權重，亦在資料呈現端使用推力與拉力的引力作用，以表現區域權重。(3)複合的資訊：在一般搜尋引擎上多是使用網頁為基礎超連結（Hyper-link）瀏覽方式，缺點是單一次的查詢行為必需依照一組特定的查詢條件。而本導覽平台在三層級中，雖然查詢原則相同，但提供自該查詢節點延關聯網路向外

延伸的資訊（如查詢特定作者，同時提供其同門師兄弟的引用作者關係）。(4)網路導覽原則：本研究的所有層級都是依網路圖形進行導覽，除配合自動群聚讓使用者能一目了然，更依不同的使用者需求設計過濾機制，讓使用者能隨需求調整資訊量的大小。(5)本研究查詢文獻的呈現需要透過 Java Applet 才能表現出引文網路圖型，一般搜尋引擎則是「Click-Search-Listing」。而相較陳銘翔（2006）與高孔明（2008）只是提供一層引文文獻網路的視覺化分析，本研究建議的文獻間多維度關聯展現能讓不同使用角色之研究人員，找到適合的研究方向與參考文獻協助更大。

現階段礙於時間經費限制，本研究雛形只能針對碩博論文網的一小部分論文（「90~97 學年」、「資訊管理」、「Data Mining 概念下的技術 (Technique)」）來建立 Meta-Data，並且對非博、碩士論文網的外部文章。雖然未進一步納入這些外部文章的次級引用文獻，但雛形系統架構亦能包含這些文獻，未來本研究將進一步擴大 Meta-Data 範圍，並透過更多的網路服務以提供研究者能探索更多的文獻資料，如期刊、研討會論文、維基百科參照等。並且希望透過對取得的文獻資料進行資料探勘達到知識本體自動修正，以更精確的呈現知識本體層探索平台。更進一步在此雛型系統架構的協助之下，針對一般者、入門者及釐清研究方向者不同的使用習性深入探討，以提供對其研究特性最幫助的系統功能，並建構與其它文獻搜尋平台之客觀比較準則。最後更希望能對文獻內容進行全文掃描後，對引用文獻之權重做字義上的分析，配合已修正的知識本體，將該文獻分類、分群，並且正確地解釋其引用文獻之權重，以協助研究者瞭解其引用關係。

## 參考文獻

- 丁一賢、陳牧言（2005），《資料探勘 Data Mining》，台中：滄海書局。
- 王奕涵（2006），《正規化概念分析的資訊管理領域理論之知識本體建構》，碩士論文，國立雲林科技大學資訊管理所。
- 曾信誠（2004），《以本體論為基礎之使用者喜好萃取、隱私權控管與側解建構》，碩士論文，國立東華大學資訊工程學系。
- 蔡明月（2003），《資訊計量學與文獻特性》，台北：國立編譯館。
- 蔡旺典（2006），《以知識本體論來輔助個人化排序》，碩士論文，朝陽科技大學資訊管理所。
- 陳俊彰（2001），《從網頁中發掘教師知識分佈》，碩士論文，國立中山大學資訊管理學系研究所。
- 高孔明（2008），《引文網路視覺化系統之設計與實作》，碩士論文，國立台北大學



- 資訊管理研究所。
- 陳銘翔 (2006), 《複雜網路有效視覺化-以引文網路為例》, 碩士論文, 國立台北大學資訊管理研究所。
- 許明章 (2006), 《以本體論為基礎建構具情境感知之博物館導覽系統》, 碩士論文, 國立臺灣科技大學自動化及控制研究所。
- 梁定澎 (2003), “管理一及管理二學門國際學術期刊分級及排序專案計畫”, 《行政院國家科學委員會專題研究計畫》。
- Berry, M.J.A. and Linoff, G.S. (1999), *Mastering Data Mining: The Art and Science of Customer Relationship Management*, New York: John Wiley & Sons.
- Braam, R., Moed, H., and van Raan, A. (1991), "Mapping of science by combined co-citation and word analysis. I. structural aspects," *Journal of the American Society for Information Science*, 42(4), 233-251.
- Bunge, M. (1977), *Treatise on Basic Philosophy: Vol.3: Ontology I: The Furniure of the World*, Boston: Springer.
- Börner, K., Chen, C., and Boyack, K.W. (2003), "Visualizing knowledge domains," *Annual Review of Information Science and Technology*, 37(1), 179-255.
- Chen, C. and Paul, R.J. (2001), "Visualizing a knowledge domain's intellectual structure," *Computer*, 34(1), 65-71.
- Chen, C. (2006), "CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature," *Journal of the American Society for Information Science and Technology*, 57(3), 359-377.
- Cronin, B., Mckenzie, G., and Stiffler, M. (1992), "Patterns of acknowledgement," *Journal of Documentation*, 48(2), 107-122.
- Garfield, E., Pudovkin, A.I., and Istomin, V.S., (2003), "Why do we need algorithmic historiography?" *Journal of the American Society for Information Science and Technology*, 54(5), 400-412.
- Gori, M, and Pucci, A. (2006), "Research paper recommender systems: a random-walk based approach," in *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, Hong Kong: Convention and Exhibition Centre, 778-781.
- Giles, C., Bollacker, K., and Lawrence, S. (1998), "CiteSeer: an automatic citation indexing system," *Digital Libraries*, 98, 89-98.
- Hand, D.J., Mannila H., and Smyth, P. (2001), *Principles of Data Mining*, The MIT Press.
- Heer, J., Card, S. K., and Landay, J. A. (2005), "Prefuse: A toolkit for interactive information visualization," in *Proceedings of the SIGCHI conference on Human factors in*

- computing systems*, Oregon: Oregon Convention Center, 421-430.
- Keim, A. (2002), "Information visualization and visual data mining," *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1-8.
- Ke, W., Börner, K., and Viswanath, L. (2004), "Major information visualization authors, papers and topics in the ACM library," in *Proceedings of the IEEE Symposium on Information Visualization*, Texas.
- Lawrence, S. and Bollacker, K. (1999), "Digital libraries and autonomous citation indexing," *IEEE Computer*, 32, 67-71.
- Li, H., Councill, I., Lee, W., and Giles, C. (2006), "CiteSeerx: an architecture and web service design for an academic document search engine," in *Proceedings of the 15th international conference on World Wide Web*, Southampton: University of Southampton, 883-884.
- Noy, N.F. and McGuinness, D. (2001), "Ontology development 101: a Guide to creating your first ontology," *Stanford Medical Informatics*, Technical Report SMI-2001-0880.
- Peters, J.F. and Pedrycz, W. (1999), *Software Engineering: an Engineering Approach*, New York: John Wiley & Sons.
- Petinot, Y., Giles, C., Bhatnagar, V., Teregowda, P., Han, H., and Councill, I. (2004), "A service-oriented architecture for digital libraries," in *Proceedings of the 2nd International Conference on Service Oriented Computing*, New York: University of Trento, 263-268.
- Roiger, R.J. and Geatz, M.W. (2002), *Data Mining: A Tutorial-based Primer*, New York: Addison-Wesley.
- Scott, J. (2000), *Social Network Analysis: A Handbook*, SAGE Publications.
- Shneiderman, B. and Plaisant, C. (2005), *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, New York: Addison-Wesley.
- Sure, Y., Stabb, S., and Studer, R. (2004), "On-To-Knowledge Methodology (OTKM)," in *Handbook on Ontologies*, edited by Staab, S. and Studer, R., Boston: Springer, 117-132.
- Tan, P.N., Steinbach, M., and Kumar, V. (2005), *Introduction to Data Mining*, New York: Addison-Wesley.
- Uschold, M. and Gruninger, M. (1996), "Ontologies: principles, methods and applications," *The Knowledge Engineering Review*, 11(2), 93-136.
- Williams, G.J. and Simoff, S.J. (2006), *Data mining: Theory, methodology, techniques, and applications*, New York: Springer.