

以知識探索為本之知識組織方法論及研究分析

An Analytical Review of Current Methodologies of Knowledge Organization : A Knowledge Discovery Approach

陳亞寧
Ya-ning Chen

陳淑君
Shu-jiun Chen

中央研究院計算中心
Computing Centre, Academia Sinica, Taipei, Taiwan, ROC

【摘要 Abstract】

隨著 Internet 的應用與各式電子資訊資源的出現，雖有各式網路蒐尋引擎的設計發展，但是仍然無法解決網路無序的現象。本文試圖從不同學科知識組織的觀點進行研究網路無序的可行途徑，進而達成有效的知識探索。本文係採取比較分析法，以圖書館學的分類編目、資訊科技界的網路蒐尋引擎與資料探勘，以及新近崛起的 Metadata、語義網路與主題地圖為研究對象，探討對知識組織的起源、理論與方法、現況發展與趨勢及問題。經比較分析後，本文建議知識組織除了必須以資源管理為範圍外，更應從知識的關聯性為中心，重新思索知識組織的基礎理論架構與方法，此外有關規模、經濟與變動性等問題也必須一併解決，方能達成有效的知識探索。

With the advance of the Internet and the emerging electronic information resources, chaos remains as a constant phenomenon on the Internet. The purpose of the article is to study the solutions through methodologies of knowledge organization from different domains, in order to achieve the efficiency of knowledge discovery. An analytical comparison is taken among the systems of classification & cataloguing, search engines, data mining, metadata, semantic network, and topic maps. The issues of origins, theories & methods, current status & trends are also reviewed. In the findings, authors suggest that relationship-centric should be highlighted as a key framework for knowledge organization theory. Besides, some factors related to scale, economics and dynamics should be taken into consideration to achieve knowledge discovery.

關鍵詞 Keyword

知識組織 知識探索 資源探索 知識管理 分類編目 蒐尋引擎 資料探勘 語義網路 主題地圖
Knowledge organization ; Knowledge discovery ; Resource discovery ; Knowledge management ;
Classification and cataloging ; Search engines ; Data mining ; Semantic network ; Topic maps



壹、前言

隨著 Internet 的興起與各式電子資訊資源的應用，有關知識組織 (Knowledge organization) 的議題已逐漸形成一項重要研究與應用的課題。本文試從圖書館的分類編目學、資訊科技界的蒐尋引擎 (Search engines) 與資料探勘 (Data mining)，以及人工智慧 (Artificial intelligence, AI) 的語義網路 (Semantic network) 與全球資訊網聯盟 (World-wide Web Consortium, W3C) 的主題地圖 (Topic Maps, TM) 等不同學科進行知識組織的比較、分析、探討，祈能對知識組織提出建議，達到有效的知識探索 (Knowledge discovery)，以及解決現行 Internet 無序的迫切問題。本文共分爲定義與起源、現有知識組織方法論的分析、未來趨勢、知識組織方法的比較與分析，及結論與建議等五部份。

貳、定義與起源

在探討知識組織時，本文試從名辭的定義與起源爲起始進行探討，包括資料(Data)、資訊(Information)、資源(Resource)、知識(Knowledge)、知識管理(Knowledge management)、知識組織(Knowledge organization)、資源探索(Resource discovery)、知識探索(Knowledge discovery)，如下：

一、資源

對於資源一詞，各方持有許多不同的看法，例如 L. Dempsey 認爲：「資源可能是一個資料庫、資料庫中的記錄，或是儲存的檔案，可以有眾多不同的取用方式」(Dempsey, 1994)。依澳洲分散系統科技中心(Distributed Systems Technology Centre, DSTC)的解釋：「所謂資源是一種可以辨認的實體或虛擬物件(Object)，可能是靜態或動

態、也許只短暫留存在一陣子，也可能恆久存在」(DSTC, n.d.)。總體而言，資源是一種物件，能以實體或虛擬形式存在，且隱含了可供利用的資訊。

二、知識及其類型

依據 M.H. Zack 的解釋：「資料是在某種情境下呈現的一種觀察或事實，本身並不具備任何直接的意義。資訊則是以某種訊息方式呈現，通常在有意義的情境下，自資料中產生而成。知識則是透由經驗、溝通或推理方式，以形成有意義、有組織的一種資訊累積，通常一般人已經深信，且具有某種價值」。同時 M.H. Zack 更進一步依知識的形式、類型與範圍，將知識區分爲：

- (一)形式：內隱與外顯(Tacit and explicit knowledge)。
- (二)類型：宣示性、程序性與因果關係的知識(Declarative, procedural and causal knowledge)。
- (三)範圍：一般與特定(General and specific knowledge)(Zack, 1998)。

三、知識管理

以資訊管理學科 (Management Information Systems, MIS) 而言，所謂知識管理是以某一企業組織爲範圍，就其內部有關知識的產生、傳播、儲存、應用與共享等議題爲主。依據 B. Newman 的闡釋：「所謂知識管理乃是由許多處理過程所組成的，以達成知識的創造、傳播與利用」。M. Sierhuis 則更明確、簡單說明：「知識管理就是管理知識的能力」(Newman, 1996)。其實長久以來，圖書館向來以知識的守門員(Keeper of knowledge)自居，徵集各式資料與資訊，同時予以組織，進而提供各項服務。因而圖書館從事的各項活動即是一種知識管理，只是觀點與方法不同而已。

就知識管理而言，K.E. Sveiby 指出在電腦與資訊科學領域 (Computer and/or information



science) 中，知識管理是一種資訊管理 (Management of information)，進而設計成所謂的資訊管理系統 (Information management systems)。對此一學科領域而言，知識就是一種物件 (Object)，具備可以確認 (Identified)、處理 (Handled) 的特質，以利資訊管理系統的建構。對哲學、心理學、社會學與商業管理學等而言，所謂的知識管理就是一種有關人的管理 (Management of people)。對這些學科而言，知識就是一種程序處理、複雜的技巧與方法 (Processes, a complex set of dynamic skills, know how etc.)，同時具備隨時變異的特性 (Constant changing level) (Newman, 1996)。此外，M.H. Zack 也提出：「知識可以被視為一種物件 (Thing)，或是一種程序，並指出在管理知識時，應包含物件與程序，缺一不可。同時建議知識管理應包含策略、知識、組織單位與科技等四大層面的關係」 (Zack, 1998)。由此可知，其實知識管理對不同領域而言，是有不同的實質意義。在所謂數位圖書館的網路時代下，可從英美許多相關的實驗計畫內涵中，明顯發現知識管理已經橫跨上述兩種學派，也嘗試作某種程度上的整合與融合。

四、知識組織

所謂知識組織，依 A. Sigel 的解釋：「知識組織是將含有知識的集合物加入資訊價值的一種跨學科領域的文化活動」，而 J.D. Anderson 則認為：「知識組織是有關文獻的描述、內涵、特色、目的及將前述這些活動予以組織，以利於使用者的尋找。知識組織包含了索引、摘要、編目、分類、記錄管理、書目，以及相關文獻資訊的產生或檢索用的書目資料庫」 (Sigel, 2000a)。由上述定義加以分析，不難發現這些觀點認為知識必須具體化身為各項文獻或文本後，再經由特定方式將文

獻內涵的知識層面予以抽離、標引，進而建立一套系統化作業模式與原則理論，藉以有效整理現有的文獻；而這些活動正是圖書館學中分類編目、索引典等理論與實務企圖涵蓋的層面。以目前現況而言，知識必須具體落實在某種資訊載體上是無庸置疑的，但是載體除了傳統紙本與視聽媒體外，而各項數位資訊載體也正蓬勃發展，分編與索引典等方式也只是眾多知識組織方式之一而已。

五、資源探索 (Resource discovery)

廣義而言，資源探索係指在電腦網路上的資訊資源蒐尋。若再深入仔細分析資源探索，所謂探索是涉及資源的找尋 (Finding) 與擷取 (Retrieving)，而資源則是泛指一個實際或概念性的物件 (Object)，而這個物件可能包含資訊 (如文獻、資料庫或只是一種指標 (Pointer)，如 URL 串聯)，或是一種服務 (如查詢引擎)，而資源探索涵蓋層面除了資源與探索外，也包括了使用者與服務提供者 (Service providers) (DSTC, n.d.)。若從現實環境中，其實資源探索的工具一直推陳出新，如 Archie、Gopher 與 World-Wide Web 蒐尋引擎等皆是。

六、知識探索 (Knowledge discovery)

如果依上述資源探索的定義為基礎，進而推演所謂的知識探索，意義應為「在數位網路的大環境下，涉及知識的尋找與擷取」。若依 M.M. Owrang 與 F.H. Grupe 兩人的闡釋，其實知識探索應具備兩種層面的涵意：一為從既有資料庫中推演、找出有用的資訊，二則包含關聯性關係的辨識 (Identification of relationships)，而達成知識探索的關鍵在於學科領域的知識 (Owrang, and Grupe, 1996, pp. 173-174)。



上述各種辭彙其實是環環相扣的，必須將彼此關係加以釐清。就整理而言，知識管理是一個全面性的概念與管理，內涵包含了以知識為中心相關的管理與服務，而其手段為知識組織，最終目的是要達成資源探索，更進一步者則為知識探索。如果圖書館仍以知識守門人自居為前提，這也似乎意謂著圖書館必須達成的是知識探索，而非單純的資源探索。

參、現有知識組織方法論的分析

隨著時代的演進與資訊科技不斷的推陳出新，知識組織面對的環境是不同於以往的，若從對象而言，除了傳統的資訊資源外，尚包括了純然的數位資訊資源與從傳統資源轉化為數位化的資訊資源。若依 G. Hodge 的分析與歸納，目前知識組織的系統(Systems of knowledge organization)約略可以分為辭彙(Term lists)、分類與主題(Classifications and categories)及關係(Relationship lists)等三大體系(Hodge, 2000, pp. 3-7)，但是不同學科對於知識組織的方式皆根源於不同的理念與觀點。本文試從圖書館、資訊科技與學科社群等為範圍，探討三種截然不同領域對知識組織的起源、實際應用、特質、現況發展與未來趨勢等各方面進行分析、探討。

一、圖書館界

有關圖書館界論及知識組織方面，首以分類編目學為首的書目控制最為相關，本節試從分類編目學的觀點進行知識組織的探討。

(一)歷史回顧

在圖書館學界，分類編目學與知識組織相關度最高。圖書館學尚未正式形成前，並沒有所謂的分類編目學。此一時期內，圖書館是由學科專

家管理的。其中與分類編目最為相關的就是登錄清單或目錄，係由學科專家依據其理解的知識體系與架構進行分類與登記，但與後來發展的目錄有著截然不同的面貌。接下來，隨著杜威創立圖書館學校，分類編目也隨著圖書館學的形成而逐漸發展、建立。傳統上分類編目是由記述編目(Descriptive cataloging)與主題編目(Subject cataloging)共同組成，前者著重在記載、描述，後者則是內容的主題分析與標引。因而在圖書館界就發展、制訂出相關的理論與工具，這些包括了編目規則、主題表、分類表等，同時為了能對不同辭彙予以標準化與相關性的建立，也發展出了權威控制與索引典等機制。

隨著圖書館自動化的發展與應用，原本分類編目的理論、實務並未受到極大的影響，在此時期內的機讀編目(Machine Readable Cataloging, MARC)即是其中極為重要的一項成果。近年來，面對網路與電子資訊資源的衝擊，圖書館界也嘗試以分類編目學為基礎，發展出兩種不同的因應方式。第一是以機讀編目為發展途徑，在既有的欄位加入電子資源(Electronic location and access)，同時結合原有圖書館自動化系統功能與界面，將網路電子資源納入館藏的一部份。第二種方式即是利用蒐尋軟體(Agents, crawlers, robots, spiders)與全球資訊網等功能，建立出所謂的「主題式轉接站」(Subject-Based Gateway, SBG)，負盛名者如英國的 BUBL (Bulletin Board for Libraries, <<http://bubl.ac.uk>>)與美國的 IPL (Internet Public Library, <<http://www.ipl.org>>)。

(二)現況發展與趨勢

依據 L.M. Chan 的分析、報導，近年來圖書館界面對網路與電子資源的出現，圖書館界已著手探討相關研究的可行性，以找出合宜的解決途徑，這些研究範疇包括自動索引(Automatic



indexing)、不同主題辭彙、分類表與主題表間的轉換及串聯使用(Mapping terms and data from different sources)、不同主題使用工具的整合(Integrating different subject access tools)(Chan, 2000)。

除此之外,分類編目學因應時勢需求,也作了不同修正;例如杜威分類法(Dewey Decimal Classification, DDC)就逐漸朝向學科知識為主體,同時輔以剖面分析(Facet analytical)方向發展。此外,依據 B. Hjørland 與 H. Albrechtsen 兩人分析,未來分類法將朝下列方向發展:第一是以學科主體為主要架構的原則;第二則是分類需要更深入的單位(Unit),如書的章節、期刊的文章;第三則為由 C. Beghtol 提出的文化保證(Culture warrant)原則(Hjørland, and Albrechtsen, 1999)。此外, I.C. McIlwaine 與 N.J. Williamson 兩人則建議主題法分析應正視知識結構(Knowledge structure)的問題,才能達成有效、正確的主題分析(McIlwaine, and Williamson, 1999)。

(三)特質

在既有的分類編目體系下,相關理論與實務也基於不同的觀點而發展。一為作品保證(Literature warrant)原則,一定要有相關文獻的事實存在,方有相關的原理與應用機制,如美國國會分類法與主題法。另一種是以既有知識體系為來源,再依實際應用的需求而進行調整,如杜威分類法、國際十進分類法(Universal Decimal Classification, UDC)等。就總體而言,仍以管理為主要中心的訴求重點。此外,品質佳、精確度高則是此一學科的特點。

(四)問題

首先就分類法而言,在設計分類法體系時,其實在分類表中已依知識體系的架構進行展現與

標示,並標引相關知識的關聯性。從實際的卡片目錄與電腦目錄(Online Access Public Catalog, OPAC)實例個案中,並無法以立體方式實質展現分類法的上述特色與知識結構體系,因而即使圖書館進入自動化階段,分類法的查詢使用並未達到預期的效果,且多數僅限於圖書館人員使用而已。但是此種情形卻隨著 SBG 的應用與出現,分類法已達到某種程度上的效應;如前述的 BUBL。此外,分類號常是圖書館作為書架排架依據的一部份(即索書號),因而無論是卡片目錄或 OPAC 記錄中,分類號必須保持唯一(Unique)的現況,反而無法展現單一物件包涵不同主題的標引、分析,而許多 SBG 亦是如此,也限制了分類法應有的效能。

其次,就作業處理的結果而言,即使主題法與分類法已被圖書館界長期應用,但是彼此間的異質性卻是長久以來無法解決的眾多問題之一。換言之,處理結果並不具備一致性(Consistency),尤其是圖書館邁入自動化、網路化及書目中心的出現、應用,更是突顯此項問題的困難度。第三,隨著 Internet 的廣泛應用,近年來的學科發展十分迅速,包括了新興學科的創立、出現,以及學科間的整合,無論是分類法或主題法往往在修訂上十分緩慢,無法即時反應此種需求。第四,雖然分類編目十分重視精確度,也達成某種程度的品質要求,但是成本高、速度慢卻是一直無法解決的根本問題與挑戰。第五則是有關不同分類與主題體系間的互通性(Interoperability),雖然圖書館十分重視不同分類表與主題表間的轉換、使用,但是從以往的計畫成果中,不難發現往往無法提出具體的可行方案,未來此一問題能否解決,尚待研究。尤其是網路與電子資源具備大規模(Scalability)、易變性(Dynamic)、多元化(Diversity)等特質,此一問題是否應從另一觀點切入,頗值得思索與研究。



二、資訊科技界

廿世紀末，Internet 革命所激發的大量網路資源，讓傳統圖書館的資源組織方式面臨捉襟見肘的困窘，以資訊科技界為主導的網路蒐尋引擎及資料探勘等資源探索及知識探索技術卻以多樣化的方式持續發展，並不斷推陳出新。

(一)網路蒐尋引擎

根據“Search Engine Showdown” 2000 年對蒐尋引擎的定義，凡是用來蒐尋資料庫的任何軟體皆可稱為蒐尋引擎。就 Internet 而言，係指由機器建立的龐大網站資料庫，藉由軟體自動於 Internet 的網頁中尋覓並予以索引。Internet 蒐尋引擎的誕生約可溯於一九九四年春天，如：WebCrawler 是源自華盛頓大學（University of Washington）電腦與工程系的師生的討論、Lycos 則是由 M. Mauldin 在卡耐基大學（University of Carnegie-Mellon）實驗室發展而成、Yahoo! 係史丹佛大學（Stanford University）電子工程系兩位博士生所創造。蒐尋引擎通常由下列三部分所組成（Westera, 1999）：

1. 由機器人（即 Robots 軟體）漫遊整個網路並收集資訊。
2. 再由資料庫對收集的資訊進行編輯。
3. 最後由查詢模組協助使用者找到所需的資訊。

發展至今，蒐尋引擎可概分為三大類型：第一類是「一般性蒐尋引擎」，也是最早出現的搜尋引擎，通常嘗試涵蓋知識的所有領域，最負盛名者如 Yahoo!、AltaVista、Excite 等。第二類是「專門性蒐尋引擎」，係為符合特別需求而發展，像是某一主題（如：商業資訊）、某一地區（如：台灣）、某一類型資訊（如：討論群）、或某一特定族群（如：小學生）等。第三類為「聯合性蒐尋引擎」，容許 Internet 的使用者傳送一個查詢，即可同時在數個搜尋引擎進行檢索；例如：MetaFind、

MetaCrawler、Gogpile 等。就知識組織而言，蒐尋引擎嘗試以主題指引（Directory）的方式，根據資源的主題與屬性，給予分類與組織。例如：Yahoo! <www.yahoo.com> 即為經典之例，它以十四大類的主題為基礎架構，由編輯決定每一網站應列在什麼大類之下，以供使用者瀏覽、查詢。

根據“Search Engine Watch” 在 2001 年 8 月的最新統計，Google 為目前索引量最大的搜尋引擎，已達 10 億頁的網頁，其次為 FAST、AltaVista、Inktomi、Excite、Northern Light 等（Sullivan, 2001）。然而，在檢索結果的正確率方面，根據“Search Engine Shootout” 於 1999 年 6 月的評估報告，發現 HotBot 名列前茅，平均每一次搜尋可獲 55% 的有用資訊，其次為 Excite 可獲 30% 的有用資訊；而 Excite 又被測試為發現最少無效連結（Invalid URL link）的搜尋引擎（Clyde, 2000）。為解決低精確率的蒐尋結果，許多檢索技巧不斷出籠，例如：布林邏輯、加權比重（如出現頻率、出現位置先後、被串聯次數等）、依地理區域、網路領域名稱類型等。

由此可以發現，以蒐尋引擎作為網路資源的導覽輔助工具，具有下列特點：

1. 省力：由機器自動執行。
2. 效率高：可索引全世界七成的網頁內容，約 10 億頁網頁。
3. 效果低：常包括許多不相關的資訊，搜尋後的精確度/有效度最高只有 55%。
4. 索引層次深度化，且單位更為細緻（如全文式檢索）。

(二)資料探勘

根據“Free On-Line Dictionary of Computing”（FOLDOC <<http://foldoc.doc.ic.ac.uk/foldoc/index.html>>）記載，資料探勘最早由 IBM 所發明，終極目標是



爲了知識探索。主要方式係從龐大的資料庫中，粹取出隱藏的資訊，以找出可能的趨勢、不規則性或關聯性。因此，它可以說是從不同角度分析資料，並粹取出有用資訊的處理過程，讓資訊能夠用來增加收入或節流成本 (Palace, 1996)。資料探勘軟體可以根據使用者的查詢，分析出被儲存的交易資料 (Transaction data) 之關聯性與模式。分析的軟體類型包括：統計型 (Statistical)、機器學習型 (Machine learning)、以及神經網路型 (Neural networks) 等三種。目的在於尋找出下列任何一種關係 (Palace, 1996)：

1. 群組 (Classes)：將資料歸屬到預估的群組。以餐飲業爲例，從顧客消費的內容資料庫，可以找出什麼時候消費者會點購什麼樣的食物。
2. 叢集 (Clusters)：資料可以根據邏輯的關係或消費者的喜好爲屬性予以歸類，例如：上班族的消費者叢集、嗜好海鮮類的消費者叢集。
3. 關聯 (Associations)：資料能夠被探勘出某種關聯性。最爲經典的例子是「尿布與啤酒」個案，在探勘某大超級市場的顧客消費資料庫後，發現在傍晚時分購買尿布的男性消費者，也經常會連帶購買啤酒。於是該超市調整賣場的位置，在尿布區的附近增加啤酒的擺設，以方便提醒此類消費者的持續購買。
4. 序列模式 (Sequential patterns)：預測出行爲模式與趨勢，例如：某用品店根據消費者購買的睡袋及登山鞋，可以預測出他會喜歡的登山背包。

目前資料探勘主要應用於以消費者爲導向的企業或組織，如零售業、金融、大眾傳播、行銷業等，因爲它能夠讓這些企業發現到價格、產品定位、員工技能等「內在因素」的關聯性，以及經濟指標、競爭力、消費者特質等的「外在因素」。

舉例而言，錄影帶出租店根據顧客租借歷史檔的資料探勘，可以爲消費者推薦他們會有興趣而尚未看過的片子；美國運通卡公司，分析持卡人每月消費帳單，可以爲個別持卡人建議適合購買的產品。

由上述可以分析出，資料探勘在整個知識探索的過程具備相當核心的功能，因爲它本身包含下述四項特質：

1. 分析出使用者的行爲模式 (Behavior patterns)。
2. 找出使用者的特質及使用關聯性 (Associations / correlations)。
3. 發現最符合使用者需求的服務內容。
4. 預測未來的使用行爲與需求。

然而值得注意的是，傳統的資料探勘多限於單一資料庫的有限範圍，近年來 WWW 技術發展所伴隨的大量網路文獻及開放性資料庫環境，爲資料探勘的研究帶來另一嶄新的挑戰與機會。

三、學科社群：以 Metadata 爲例說明

除了圖書館與資訊科技界從事於知識組織外，尚有一種方式是以社群組織 (Community) 爲單位進行，而此種方式從 Metadata 的現況發展可以窺知一二。

(一) 歷史回顧

截至目前爲止，有關 Metadata 的起源發展，仍是莫衷一是，並未有統一的說法。如果依 FOLDOC <<http://foldoc.doc.ic.ac.uk/foldoc/index.html>> 的解釋，所謂“Meta-data”與“Metadata”是有所差異的。所謂 Meta-data 一詞的定義是：「描述資料的資料」(Data about data)。就資料處理方面而言，是一種在管理應用的環境前提下，提供資訊或記載的定義式資訊 (is definitional data that provides information about or documentation of other data



managed within an application or environment)。一般而言，Meta-data 是用來描述資料元素(Data elements)或屬性(Attributes，如名稱、大小、資料類型等)、有關記錄的資料、資料結構(如長度、欄位等)。Meta-data 可以是描述資訊，以描述資料的關係(Context)、品質、情境(Condition)與特性(Characteristics)等。美國電機電子工程師學會(IEEE)出版的標準辭典中，更是明確指出：「是一種描述資料的資料，如資料庫中的資料字典(Data dictionary)」(Booth, 1992, p. 799)，不過使用辭彙是 Metadata，而非 Meta-data。至於 Metadata 一詞則是在一九六九年由 J.E. Myers 所創出的辭彙，此一名詞最主要是來自 J.E. Myers 提出 Meta Model 觀念及相關的發展產品，後來 J.E. Myers 創設了“Metadata Information Partners”(後改名為 The Metadata Company)，且在一九七三年以 Metadata 一詞登記為商品，且擁有 Metadata 一詞相關的著作權(FOLDOC, <<http://foldoc.doc.ic.ac.uk/foldoc/index.html>>)。

如果追溯現行 Metadata 一詞的現況發展，可自一九九五年 OCLC 與 NCSA 聯合舉辦第一屆的 Metadata Workshop 談起，Metadata 才逐漸引起重視，同時此次會議也將 Metadata 定義為「描述資料的資料」。因而，可以明顯發現 Metadata 與 Meta-data 已被混淆一起使用。此外，前述第一屆 Metadata Workshop 中，將 Metadata(Dublin Core) 定位在網路的電子資源，但後來的實際應用發展中，現行 Metadata 範圍已擴大包含了傳統媒體資訊，而原本 Metadata 預定處理的網路電子資源也逐漸變成配角，而非主角。

(二)現況發展與趨勢

Metadata 的發展方式約略可以歸納成下列四種類型：第一是由電腦資訊界為起點進行發展，如 IAF/WHOIS++ Templates、LDIF、SOIF 與

URC 等皆是。第二種則是由資料擁有或管理單位所主導，如博物館(如 CIMI、CDWA)、檔案館(如 ISAD)、政府機構(如 GILS)與圖書館(如 EAD、Dublin Core)等皆是。第三種則是以學科領域(Discipline domain)為主，例如 TEI 是由人文學與語言學界所共同制訂的，而 CSDGM 則是有關地理空間資訊方面的領域。第四則是著重於架構的建立，以利不同 Metadata 間的互通，如 Dublin Core 提出的 Warwick Framework 與 W3C 的 RDF。

(三)特質與問題

上述這些 Metadata 的發展，可以剖析出四種特質：第一種是強調高度的概念化，如 Dublin Core、RDF。前者 Dublin Core 是一種高度抽象化的結果，十分適用於跨學科間的整合、應用，但是精確度則是必須詳加考量的。後者如 RDF，此種類型強調建立共通性架構，以達成不同 Metadata 格式間的互通性，變成一種共通性的機制。此種理念與前述圖書館界試圖將不同分類法與主題法體系予以結合、互通，頗有異曲同工之趣。第二種則將焦點集中在以自動方式達成有效的資訊擷取，如上述 LDIF 等。此種方式延襲了上述資訊科技界進行知識組織的特質，是以一種通訊協定(Protocol)觀點切入，自動化、簡易化、迅速是其特點。第三種則是著重在品質、詳盡等層面，一如圖書館界的分類編目方式，反之速度慢、成本高是其缺點。第四種則是以某一學科領域為主體，以標示文本資料的內涵結構作為知識組織的基礎，如 TEI。但是此法能否達到高度效率與知識組織，頗值得進一步觀察與驗證。

肆、未來趨勢

除了上述圖書館學的分類編目、資訊科技界的網路蒐尋引擎與資料探勘，以及有關 Metadata



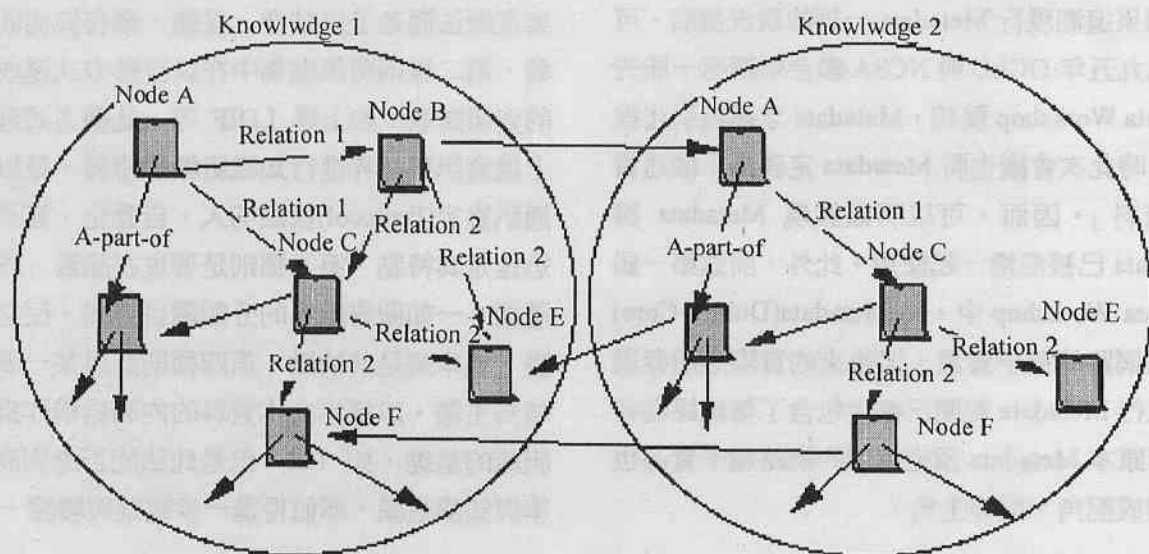
方式進行知識組織的探討與實作外，另外尚有以既有理論基礎與相關研究理論予以重新闡釋，試圖以另一種新觀點進行知識組織的建構，其中以語義網路、主題地圖與本體論 (Ontology) 等三個流派最受矚目；本文茲就語義網路與主題地圖進行探討。

一、語義網路

在人工智慧的研究領域中，語義網路的觀念便被使用作為展現結構化知識的方式 (Knowledge representation)。它以圖形呈現資源及其關聯性，目的是為了有效率地展現資料，避免重複 (FOLDOC, <<http://foldoc.doc.ic.ac.uk/foldoc/index.html>>)。近來語義網路的發展，主要目的係嘗試解決目前網路資源以階層式組織資訊的缺憾。鑑於網路資源的巨量成長之際，許多商業性的蒐尋服務及索引服務紛紛出籠，蒐尋服務以自然語言處理技術來找尋資訊，索引服務則由人工維護方式協助浩瀚資訊的

找尋。在資訊被收集後，直到有意義被使用的階段，則需要對資訊進行有效的組織。因此，諸如書籤 (Bookmark) 或熱門清單 (Hotlist) 常被應用在商業性 WWW 瀏覽器，讓使用者以樹狀的階層架構組織自己所找到的資訊。然而，階層式的資訊組織架構只能呈現某一面向，卻無法充分展現出網路資源與知識的結構。

語義網路的結構，係藉由一組「節點」 (Nodes)、「標記」 (Labels) 與「弧線」 (Arcs) 的結構展現出一組的知識。節點代表在學科領域下的概念、物件或情境，而弧線與標記則代表著連結 (Links)，是負責指示不同概念、物件或情境之間的關係 (Relationship)。語義網路常被運用在人工智慧的應用中，以展現學科專家的知識，或稱為專家系統 (Expert systems) (Freese, 2000)，請詳圖一以語義網路呈現兩組知識結構的關係 (Kuwata, and Yatsu, 1997)。



圖一：語義網路的知識結構展現圖

(資料來源：Kuwata, and Yatsu, 1997)



其實目前在 WWW 上使用頻繁的超文字 (Hypertext) 即為語義網路的一種型式，只是缺乏明確的標記，例如：標示出連結資源間的關係屬性，因此常易造成使用者迷失在浩瀚的網路資源中，而不知身在何處。另一方面，在語義網路架構下，由於每人皆可依其認定方式給予語義的表達，所以形成百家爭鳴、莫衷一是的無序現象。由此可以推知，以語義網路作為知識展現的限制在於它並沒有正式的語義 (Formal semantic)，對於展現的結構方式也沒有一致性概念。雖然語義網路的理論已經突破傳統超文字缺乏明確標記的弱點，以資源間的關聯性為主軸，嘗試建立一套更具結構性的知識展現架構，但似乎需要借助其他的機制與方式，以改善目前實際應用上的困境，主題地圖便是其中的一種機制。

二、主題地圖

主題地圖是 2000 年出版的國際標準 (ISO/IEC 13250)，由國際標準組織 SGML 委員會第三工作小組發展而成，用 SGML 或 XML 描述知識結構與聯繫至相關的資訊資源，並藉由「主題」及「連結相關資訊集」達成知識探索的目標。(Biezunski, Bryan, and Newcomb, 1999)。主題地圖原是基於合併不同索引之需求而發展的，後來逐漸延伸為傳統紙本式索引、目次、辭典、索引典、參照等的電子式導覽工具。它的應用範疇相當廣泛，除了作為更有效的導覽工具外，也是資訊管理領域中建立與維護資訊的基本組織原則。其主要觀念，包括「主題/類型」(Topic and topic type)、「關係實例」(Occurrences)、以及「關聯性」(Association)。茲分別簡略闡述如下：

(一)主題

可以是任何東西，包括人、實體、觀念等，以「歌劇字典」為例內容所涵蓋的主題可以包含：

1. 托斯卡 (Tosca)
2. 蝴蝶夫人 (Madame Butterfly)
3. 羅馬 (Rome)
4. 義大利 (Italy)
5. 普契尼 (Giacomo Puccini)
6. 盧卡 (Lucca)

再由「主題類型」(Topic types) 給予主題屬性化，就如同每本書的索引 (如：人名索引、作品索引、地名索引等)，每一主題都可能歸屬於某一主題類型。延續上述例子，如：

1. 普契尼是屬於「劇作家」類型
2. 托斯卡及蝴蝶夫人是屬於「歌劇」類型
3. 而羅馬及盧卡是屬於「城市」類型、義大利則屬於「國家」類型

(二)關係實例

主題被連結至一種或以上相關聯的資訊資源時，這些資源即稱為此主題的「關係實例」。一個關係實例可能是百科全書中關於某主題的一篇專論或文章、也可能是描繪此主題的圖案或影像、或為一段簡述以表達此主題等資源。關係實例通常是與主題地圖文件分開來，而「關係實例的角色」則是用來標引出某一主題出現在某些資源，或是某一資源具備了哪些主題的功能指標。再續上述例子，關係實例是指「歌劇字典」的，

1. 專文
2. 圖案
3. 一段文字
4. 評論

(三)關聯性

以連結元素表達出兩個或以上主題間的關係。例如，

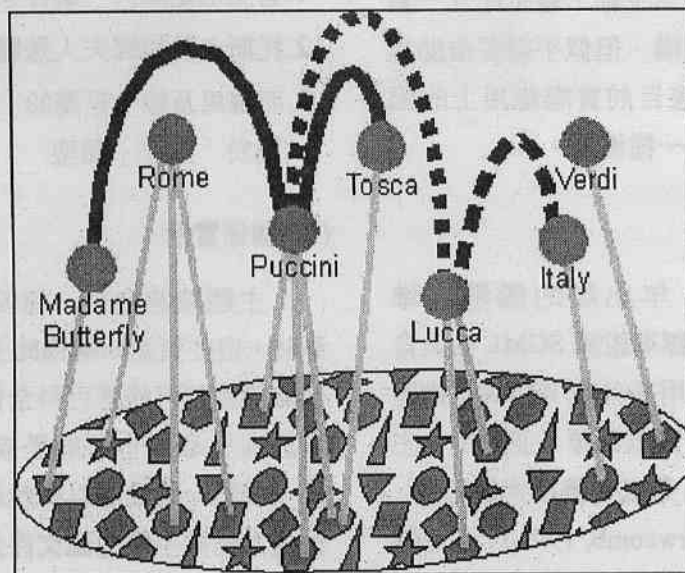


1. 「托斯卡的作者是普契尼」
2. 「托斯卡發生在羅馬」
3. 「普契尼出生於盧卡」
4. 「盧卡位於義大利」
5. 「普契尼受韋耳第 (Verdi) 所影響」

「關聯性類型」(Association types) 於此例包括：「作者是」、「發生在」、「出生於」、「位於」、「受影響於」等類型。而「關聯角色」(Association roles) 則是：「人」與「地點」表達出「普契尼出

生於盧卡」的角色關聯性。以圖二為例，

1. 本圖上方的七個圓形(●)，代表著不同的「主題」。
2. 橢圓形內所含括的眾多形狀符號(如，★●▼■)，係為「關係實例」。
3. 主題與關係實例間的淺色連接線，是指「關係實例的角色」。
4. 主題與主題間的連接弧線(包括實線及虛線)，則表示主題間的各种「關聯性」。



圖二：主題地圖知識結構圖

(資料來源：Pepper, n.d.)

不過，主題地圖與知識結構也共同面臨著下列的挑戰：

1. 如何以原則性的方式組織知識？
2. 如何讓不同的知識系統能夠互通？
3. 如何處理異質的主題資料？

知識組織的目的在於讓使用者更容易檢索、建立以及共享知識，而主題地圖則是達成此目標的支援工具之一。由於它非常吻合 Internet 環境下知識組織的需求與趨勢，因此被視為最具潛力提供創新的知識組織服務。就 Internet 環境下，德國資訊科學研究員 A. Sigel 認為知識組織



的未來需求應包括下列六項特色 (Sigel, 2000b) :

1. 能夠詮釋資訊資源之可攜性及形而上的知識結構。
2. 能夠以分散式、聯合式來建立與維護知識的組織結構，並以動態及多元的方式容許學科專家參與。
3. 高度發展的知識組織將需建立模式化 (Model-based) 的索引語言之複雜知識結構。
4. 不斷增長中的全球網路資訊資源，非常迫切需要「規模性」的考量。
5. 知識組織的理論逐漸傾向允許著錄多重選擇性的觀點及表徵，以便不同領域的使用者可以在相同的資料集建置知識。因此，創造以彈性為導向的使用環境與機制益顯重要。
6. 倘若知識工作者要以緊密交織的合作方式來產生、分享與維護概念性知識結構，則需先有語義式的知識結構，以供檢索及導覽。

伍、知識組織方法的比較與分析

經本文以上知識組織方法論的探討後，有下列的研究發現：(請詳表一)

- 一、資源的組織源自十九世紀末圖書館界的管理需求而發展，特色在於對資源有著非常嚴謹而系統化的管理及控制，例如：編目規則、分類系統表、權威控制等。相對的，這樣的管理機制成本相當高、更新速度慢，且限於單一組織單位的資訊資源為組織對象。就程度上而言，很清楚的是「資源」，而非「知識」的組織與管理控制。
- 二、廿世紀末，Internet 革命所激增的大量網路資源，讓傳統圖書館的資源組織方式面臨捉襟見肘的困窘。以資訊科技界為主導的網路蒐

尋引擎、資料探勘、語義網路、以及主題地圖等多樣化的方式，卻持續發展，並一再推陳出新。

- 三、網路蒐尋引擎特長以低成本、高效率的方式應用在大規模的網路資訊搜尋，相較於一個大型圖書館約數百萬的館藏量，目前網路蒐尋引擎已可索引高達 10 億的網路電子資源，其規模性自不可言喻，而且層次更為深度化、細緻化。
- 四、資料探勘的貢獻則是突破僅止於「資源探索」的廣度，邁向「知識探索」的深度，藉由使用者行為模式的分析與技術，智慧地找出資源間的關聯性，進一步剖析、預測使用者的需求。
- 五、Metadata 則是由圖書館、資訊科技以及學科領域共同研擬，因應 Internet 資訊資源的有效組織與探索的新方法，對於 Metadata 到底是屬於「資源探索」，抑或「知識探索」的議題，目前專家仍持有不同看法。倘使 Metadata 是定位在資源探索，在描述的層次上只須著墨在物件的表徵方面，例如：物件名稱、作者、出版或提供者、發行日期、格式大小、網址、內涵大意等。若 Metadata 是定位在知識探索，則在描述層次上需更加深入與內涵化，例如：資源內涵所涉及的人、事、時、地、物、主題等不同主軸。因此，Metadata 的規模性、經濟性、新穎性會依循賦予 Metadata 的結構繁簡、表達層次的深淺而有不同需求、考量。
- 六、語義網路及主題地圖的最大特徵是將知識結構化，並發展出語義的描述機制，以及著重知識間的關聯性，讓知識如滾雪球般不斷自動地累積。



表一：知識組織方法比較表

| 知識方法類型 | 起源時間 | 提出的學科或社群 | 目的與功能 | 方法 | 應用範圍 |
|----------|-------------------------|------------------|----------|--------------------------------|------------------------------|
| 分類編目 | 19 世紀末 | 圖書館界 | 資源的組織與管理 | 編目規則、分類法、主題法、索引典與權威控制等 | 圖書館館藏目錄的建立與管理，近年來擴展至網路上的電子資源 |
| 網路蒐尋引擎 | 1990 年代 | 資訊科技界 | 資源探索 | 電腦程式軟體(如 Robots 等) | 網路上的電子資源 |
| 資料探勘 | 約 1980-1990 年代 | 資訊科技 (IBM)與資訊管理界 | 知識探索 | 電腦軟體在資料庫中粹取有用的資訊及發掘資源間的相關性 | 以使用者為導向的企業或組織 |
| Metadata | 1995 | 圖書館、資訊科技與各學科領域 | 資源的組織與探索 | 綜合上述分類編目與蒐尋引擎的方法，但狀態仍在變化與發展中 | 數位圖書博物館 |
| 語義網路 | 不詳 | 資訊科技 (人工智慧) | 知識的組織 | 以節點、標記、弧線展現出知識的結構與關聯性 | 數位資訊資源 |
| 主題地圖 | 1990 開始發展；2000 提 ISO 標準 | 資訊科技 (W3C) | 知識的組織與探索 | 以主題、關係實例、關聯性形成核心概念以達成知識的結構與關聯性 | 數位資訊資源 |

陸、結論與建議

長久以來，人類的知識必須依附在各式的資訊載體上，而早期各類百科全書的制訂即是一種知識的組織方式。隨著資訊載體出版數量的激增，只有圖書館機構有能力將各式出版品予以徵集與整理後，方可提供給人們使用，因而在圖書館界所進行的知識組織方法也以單一圖書館為主要範圍。事實上，圖書館以分類編目學為中心的

書目控制等各項作業活動，重點除了資源組織(即記述編目)外，也進行了知識組織(即主題編目)。從 F.W. Lancaster 論文中指出，圖書館目錄無論在資料查詢與主題尋找等方面，並無法達成預期的功效(Lancaster, 1999, p. 49)，以滿足、吸引使用者的使用。換言之，圖書館界在資源探索與知識探索等雙重層面上已面臨到瓶頸，有待改善。事實上，美國國會圖書館(Library of Congress, LC)嘗試以 Metadata 進行的「國家數位圖書館計畫」(The



National Digital Library Program)中，也面臨了同質的挑戰與質疑(O'Donnell, 2000)。從計畫的評鑑報告中，美國國會圖書館資訊科技委員會(Committee on an Information Technology Strategy for the Library of Congress)對 LC 提出了一項建議：「LC 以 Metadata 方法從事的知識組織，本質上仍以 MARC 為基礎理念發展出來的。所以是將各項記錄予以獨立處理與分離，然而此種作法卻無法將資訊資源間的豐富關聯性予以標示出來」。美國國會圖書館資訊科技委員會也指出國際圖書館聯盟(International Federation of Library Associations, IFLA)提出的「書目記錄功能需求」(Functional Requirements for Bibliographic Records, FRBR)，就是正視此種需求的一種具體事實(Committee on an Information Technology Strategy for the Library of Congress, 2000)。此外，從「尿布與啤酒」的實例中，卻可以發現資料探勘已達成某種程度上的知識探索功能，但範圍也僅限於某一個企業組織或資料庫的框架內。因而，要達成有效的知識探索，除了必須對知識組織方法重新審視外，而知識關聯性也是必須留意的重點之一。因而，從資訊科技界提出的語義網路與主題地圖等兩項概念架構與應用機制，正是反應出此種趨勢的走向，先從知識組織、結構及關聯性著手切入，再擴展至相關的各式資源，而非從資源組織開始。換言之，知識關聯性也是知識組織方法論中不可或缺的元素，知識組織方法論不能只著重在知識的組織結構與元素而已。

此外，隨著 Internet 與 WWW 的應用發展，

散置在網路上的各式數位資訊正極速增長中，而數量上也逐漸超越傳統出版品。從本文前述探討的網路蒐尋引擎內容中，不難發現各式網路蒐尋引擎以並駕齊驅的速度即時地將 Internet 上的數位資訊資源予以徵集、索引與提供查詢外，更能建立主題指引達成原本圖書館學要求的主題檢索與瀏覽等功能。除此之外，成本低、數量多、規模大也是網路蒐尋引擎已達成的功效、事實。就範圍而論，知識組織不能再以單一機構為苑囿，必須以全球(Global)為考量範圍。除此之外，數位資訊變異性(Dynamic)的特質，也是有別於傳統資訊載體的各項資訊資源。

本質上，有效的知識組織是達成知識探索的不二法門。經本文綜合上述比較、分析、研究後，作者建議無論使用何種知識組織方法，知識組織除了注重知識的組織結構與各項元素外，更應注重下列重點：

- 一、知識的關聯性 (Association) 是知識組織的必備要素，而知識組織方法論必須將之納入，也是近年來知識組織領域上，極為重要的一項事實需求與發展趨勢。
- 二、知識組織的範圍應以全球為考量，因而知識組織方法論一定要能有效解決知識組織的規模性(Scale)、經濟性(Economic)等根本性的特質與問題。
- 三、知識載體數位化後，所具備變異性的特殊問題，知識組織方法也必須一併提出有效的解決方案。

備註：本文係依據下列文章修改而成

陳亞寧、陳淑君。(民 90 年 1 月 9-10 日)。以知識探索為本之知識組織方法論及研究分析。在第三屆海峽兩岸圖書資訊研討會論文集 (頁 213-234)。臺南市：成功大學。



參考資料：

- Biezunski, M., Bryan, M., & Newcomb, S. (1999). ISO/IEC FCD 13250:1999 - Topic Maps. <<http://www.ornl.gov/sgml/sc34/document/0058.htm>> (1 Oct. 2000).
- Booth, C.J. (Ed.). (1993). The New IEEE standard dictionary of electrical and electronics terms: Including abstracts of all current standards. New York: Institute of Electrical and Electronics Engineers.
- Chan, L.M. (2000). Exploiting LCSH, LCC, and DDC to retrieve networked resources: Issues and challenges. <http://lcweb.loc.gov/catdir/bibcontrol/chan_paper.html> (21 Nov. 2000).
- Clyde, A. (2000). Search engines: An overview. Teacher Librarian, 27(4). <http://www.teacherlibrarian.com/pages/27_4feature.html> (26 Oct. 2000).
- Committee on an Information Technology Strategy for the Library of Congress. (2000). LC21: A digital strategy for the Library of Congress. Washington, DC: National Academy of Science. <<http://www.nap.edu/openbook/0309071445/html/85.html>> (1 Oct. 2000).
- Dempsey, L. (1994). Network resource discovery: A Europe library perspective. IN N. Smith, (Ed.), Libraries, networks and Europe: A European networking study. London: British Library Research & Development Department. <http://www.lub.lu.se/UB2proj/LIS_collection/lorcan.html> (23 Nov. 2000).
- Distributed Systems Technology Centre. (n.d.). Resource discovery: A definition. <http://www.dstc.edu.au/RDU/rd_define.html> (23 Nov. 2000).
- Freese, E. (2000). Using topic maps. <<http://www.infoloom.com/geaconfs/WEB/paris2000/S22-01.HTM>> (14 Nov. 2000).
- Hjorland, B., & Albrechtsen, H. (1999). An analysis of some trends in classification research. Knowledge Organization, 26(3), 131-139.
- Hodge, G. (2000). Systems of knowledge organization for digital libraries: Beyond traditional authority files. Washington, DC: The Digital Library Federation and Council on Library and Information Resources. <<http://www.clir.org/pubs/reports/pub91/pub91.pdf>> (16 May 2000).
- Kuwata, Y., & Yatsu, M. (1997). Managing knowledge using a semantic-network. <<http://spuds.cpsc.ucalgary.ca/AIKM97/kuwata/kuwata97a.html>> (20 Oct. 2000).
- Lancaster, F.W. (1999). Second thoughts on the paperless society. Library Journal, 124(15), 48-50.
- McIlwaine, I.C., & Williamson, N.J. (1999). International trends in subject analysis research. Knowledge Organization, 26(1), 23-29.
- Newman, B.D. (1996). What is knowledge management. IN The Knowledge Management Forum. <http://www.km-forum.org/what_is.htm> (17 Nov. 2000).
- O'Donnell, J.J. (2000). LC21: Hopes and cautions for the Library of Congress. D-Lib Magazine, 6(10). <<http://dlib.ejournal.ascc.net/dlib/october00/odonnell/10odonnell.html>> (19 Nov. 2000).



Owring, M.M., & Grupe, F.H. (1996). Using domain knowledge to guide database knowledge discovery. Expert Systems with Application, 10(2), 173-180.

Palace, B. (1996). Data mining.

< <http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/index.htm>> (23 Oct. 2000).

Pepper, S. (n.d.). The TAO of topic maps: Finding the way in the age of infoglut.

< <http://www.gca.org/papers/xmleurope2000/papers/s11-01.html>> (1 Oct. 2000).

Sigel, A. (2000a). The knowledge organization on Internet mini-FAQ(Ver. 0.33).

< <http://index.bonn.iz-soz.de/~sigel/ISKO/wiss-org.faq.html>> (19 Oct. 2000).

Sigel, A. (2000b). Towards knowledge organization with topic maps.

< <http://www.gca.org/papers/xmleurope2000/papers/s22-02.html>> (12 Oct. 2000).

Sullivan, D. (2001, Aug.,15). Search engine sizes. Search Engine Report.

< <http://searchenginewatch.com/reports/sizes.html>> (12 Nov. 2001).

Westera, G. (1999). Effective web searching tutorial.

< <http://lisweb.curtin.edu.au/staff/gwpersonal/searchtut/se.html>> (18 Nov. 2000).

Zack, M.H. (1998). Managing codified knowledge.

< <http://www.cba.neu.edu/~mzack/articles/fourprob/fourprob.htm>> (19 Oct. 2000).

