

關於網路時代知識組織的幾個思考

吳美美

國立臺灣師範大學圖書資訊學研究所教授兼所長

E-mail: meiwu@ntnu.edu.tw

關鍵詞：分類理論；知識組織；鏈結資料；知識本體；語意網絡

【摘要】

本文思考網路時代知識組織的相關發展，首先討論為什麼用知識組織來取代過去常用的資訊組織，並從馮道的啟發來提醒知識組織必先確定知識的範圍；其次討論網路環境中知識組織的思考，網路時代的知識組織方法，和傳統知識組織最大的不同，在於網路時代的資訊物件的界線不復存在，資訊和知識互聯的特質，促使知識組織的概念和方法產生根本變革，從傳統以單一資訊物件為知識組織單元的孤島式知識組織，演進為多面向知識組織方法，但其功能都仍以辨識和搜尋知識為主，語意網絡化的知識組織，允許知識重新建構組織，賦予新的知識詮釋，形成更全面的知識開展，這是在網路時代，知識組織方法發展之後，對人類知識理解的新的貢獻；最後列舉知識組織處理人類全體知識、領域知識、社群知識和個人知識的案例，論述和辯證知識組織新進的發展，並檢討現代知識組織所需要的新知能。

前言—為什麼是知識組織？

「知識組織」或「資訊組織」，兩者之間有說不清楚的纏繞，例如李鶴立主張採用「知識組織」一詞作為領域的正式名稱，她的理由是知識概念和知識組織核心議題都包括「觀念（concept）、詞彙（terminology）、意義（meaning）、關係（relationship）、順序（order）、結構（structure）、方式（approach）、角度（perspective）和脈絡情境（context）」，而「資訊」一詞並不一定讓我們聯想到其中幾項（譬如：意義、方式、角度或脈絡情境）。所以，我主張用知識組織一詞」（李鶴立，2015，頁 100）。這是強而有力的論述，直接指出知識的特質，以及這些特質和組織概念的強力連結關係。不過知識組織在網路時代之所以更適合取代資訊組織一詞，可能還有其他也很重要的原因。

Buckland（1991）在曾經被引用將近 1,460 次著名的 JASIS&T 文章中，提出資訊三種特質，（1）資訊是「物件」（object），例如文獻、或經過編目的文物、（2）資訊是「告知的過程」（process），例如傳播歷程、言談、（3）資訊是「知識」（knowledge），也就是李鶴立（2015）

和 Smiraglia(2014)所指出的,知識具有觀念(concept)、詞彙(terminology)、意義(meaning)、關係(relationship)、順序(order)、結構(structure)、方式/取向[1](approach)、角度/觀點[2](perspective)和情境(context)等特質。其中資訊的知識面需要組織,資訊的物件面同樣或更需要組織[3]。更具體的說法是,百餘年來的西方圖書館分類法,雖以分類人類全體知識為名,其實是圖書(文獻、物件)分類,例如杜威十進分類法,將人類的知識分成九大類,這是利用知識分類法的精神,入甲類即不入乙類,不過杜威分類法畢竟是用於人造物,例如圖書、文獻等物件,因此留了一個後路,就是放不進各類的就放到總類(0號),所以總共有十大類,而且為了更周延的表達文獻中存在知識多面向的特性,還增加了複分法,例如相同主題,增加了不同的時代或地理區域複分。因此,嚴格來說,杜威分類法並不是純粹知識分類法,而是圖書、文獻、物件面的分類法,這種圖書文獻分類法,稱為資訊組織方法,和嚴謹的知識分類相比,有更多的彈性。美國國會圖書分類法,同樣以文獻、物件為分類單元,但並不採用類似杜威十進分類法,而是用類聚(clustering)原則,則更是廣義的資訊組織方法了。

那麼為什麼在網路、數位環境中,「知識組織」一詞更適合於現代的表達呢?主要是因為數位和網路發展,承載人類知識的媒體(media, container)改變了,原來傳統出版品、文獻,作為一個分類單元的對象,不論討論的議題、範圍為何,總是有明顯的邊界(boundary),用一個蘿蔔一個坑的概念來加以分類,有其妥適性,而文獻中若有不同主題纏繞,不容易歸到適合的知識類別,也就用其他的方法加以補救,例如複分法,或是主題詞表來增加檢索點。到了網路環境中,文獻單元邊界消失了,網網之間有超連結,知識之間的觀念、詞彙、意義、關係、順序、結構、取向、觀點和情境等特質,沒有了物件邊界的束縛,知識組織也就有了許多新的可能,知識組織不但呈現原有知識的意義、結構、順序,更有機會利用「知識再現」(knowledge representation)的方法,重新詮釋和發現新的知識和意義,由此可知知識組織一詞,較之於資訊組織適合現代用法。

馮道的啓發—知識觀決定知識的寬度

馮道是誰?西雅圖華盛頓大學圖書館[4]在規劃設計和建造過程中,知識淵博、匠心獨運的選上了六位對人類知識出版和印刷卓有貢獻的人物,將姓名鑄刻在六道銅製大門之上,依時代排序,馮道為首,之後立即接著古騰堡,這位十五世紀德國活字印刷創始人就排在這位十世紀中國官員之後。不過很少人聽過馮道是誰,為什麼他排在古騰堡前面(Mccormick, 2009)?原來馮道是五代十國的官員,在戰亂中歷經多朝代和皇帝,花了21年時間上奏進行木刻經書的出版,成為世界歷史上首度大規模以官方財力印刷套書的推動者和執行者。儘管馮道在知識保存的歷史地位中佔有如此重要的一席,但在中國通史或中國文化史上,大多輕

描帶過，在圖書館學的教材中，更沒有特別提到這段歷史，可能是因為中國官場有不同二君的潛規則，史家對其評價缺缺，包括歐陽修和王安石都是負面評論，不過也有相關研究文獻指出馮道其人謙和，很受愛戴[5]。

這個案例啟發我們兩個深刻的思考，第一，究竟維護人類知識寶庫重要？還是跟隨一朝、一君重要？相較於維護一君一朝、短暫的權力傾軋，保存知識寶庫是為人類萬代服務，顯然勝負立判，馮道用盡畢生努力做到了，千年之後國際歷史學者也看到了。人類知識保存者，可以在偏狹的意識形態中寂寞，卻在人類歷史的大洪流中勝出，我們的認識和史觀有沒有受到影響？第二，我們的知識版圖包括了馮道，或是略去了馮道，是受到個人知識觀的影響，知識觀決定知識的範圍。而知識組織者首要之務在於確定知識的範圍，知識觀決定範疇，值得知識組織研究者關切。知識寬度影響我們知識分類的範圍，而對於知識寬度的視野，有賴於我們對於事物的認識。我們是否因為自己的偏見，而漏掉了重要的觀點？尤其當我們在談知識組織時，首先就要處理範疇界定的問題，是否能做比較完整的界定，當我們受到歷史觀、偏見影響而做了某種篩選時，對事物的理解有可能不盡完善，而周延的範疇界定是知識組織的第一個要務。有關圖書資訊學專業教育養成，不只要講求工具理性，更需要重視思想史與學術史的學習歷程，以及邏輯思考鍛鍊。

網路環境中知識組織的思考

網路時代的科技演進，帶來知識組織的典範變遷，以下從知識組織中的載體、知識組織的目的、分類理論、分類困難、分類應用、立體知識分類演進、知識本體應用和鏈結資料和創新知識詮釋等八個議題加以討論。

知識組織中的載體——知識載體影響知識組織的方法

網路環境改變知識載體，從單一知識物件到串聯的知識物件，知識的界線網網相連，知識組織的方法也大有不同。例如從知識組織的對象來看，有實體物件和數位物件之分，相對於實體物件和數位物件的知識組織，當資訊環境由實體物件演進到數位物件時，過去的知識組織概念就需要重新檢討，因為實體物件的時代，知識依附於承載媒體（media, container），主要的組織原則是一個蘿蔔一個坑對號入座，所以知識組織主要方法就是發展編目規則、圖書分類表、控制詞彙、索引典等為方法來加以分類、組織，主要的目的在便於盤點、排序、存放。到了網路數位時代，原生或數位化資料，因為太大量無法，也無須盤點、不須事先排序（電腦可以安排任何一種自動排序）、複製容易、無單一實體存放位置，但是會需要有獨特的辨識，並且能夠在不同系統間交換等（如表 1）。

這些特點改變了過去知識組織方法講求排序和單一存放位置，改為重視能在不同的系統之間交換與複製，因此在知識組織的方法上，有了更多處理的可能性，例如從結構的組織工具觀點來看，發展數位文件編碼標準和相關工具，如「都柏林核心集」(Dublin Core, DC)、「數位文件編碼先行計畫」(Text Encoding Initiative, TEI)；從內容的組織工具方面來看，有根據國際通行的「英美編目規則」(Anglo-American Cataloguing Rules, AACR)作為基礎概念發展而來的數位環境下資源描述和取用的標準，如「資源描述與檢索」(Resources Description and Access, RDA)；從語意的組織工具來看，有各種分類表、主題表、索引典、詞彙表等，如「美國國會主題標目」(Library of Congress Subject Headings, LCSH)、「藝術與建築索引典」(Art & Architecture Thesaurus, AAT)；以及最近熱門的知識本體和鏈結資料等工具，都是為了語意網而設計的知識組織工具。

表1 實體物件和數位物件特質和知識組織的方法

	實體物件	數位物件
特質	<ul style="list-style-type: none"> ● 盤點 ● 排序 ● 有實體存放位置、對號入座 	<ul style="list-style-type: none"> ● 原生或數位化 ● 無法盤點 因為太大量 ● 不須事先排序 ● 獨特的辨識號 ● 在不同系統間交換 ● 複製容易 ● 無單一實體存放位置
知識組織的方法	<ul style="list-style-type: none"> ● 編目規則 ● 分類表 ● 標題表、索引典等控制詞彙 	<ul style="list-style-type: none"> ● 結構--數位文件編碼標準和工具，例如 DC、TEI... ● 內容--RDA ● 語意/知識表徵--詞彙表 分類表，例如 LCSH, AAT... ● 管理應用和交換標準 ● 知識本體和鏈結資料

知識組織的目的—知識組織的目的引導知識組織的方法

其次從知識組織的目的來看，有管理相對於獲取的知識組織，並更進步到創新和詮釋的功能。管理的觀點，是以知識盤點為主要目的，利用事先訂定的分類表，以邏輯相關(logical relevance)為主要判定標準，例如圖書十進分類法、美國國會分類法等，分類方法是依照知識前組合，將相關知識事先集結為知識組織方法；換言之，依照知識的前組合，把分類表排列好，再將資料放進各類屬，便於管理。知識組織以集結為主，設定了一個一個的抽屜，事先做好的組合，就像中藥行一樣，一一將相屬的內容放進去。

獲取的觀點則以獲取知識為主，依個人當時的資訊需求進行檢索，而檢索是依照後組合的方式，以適切相關(pertinence relevance)為原則，相對於知識集結，是以知識分析為主要的分類方法，亦即分析事物的特質，而不是將事物直接歸入某一類；換言之，當時的檢索需

要才是知識組織的重要衡量，可以用大風吹遊戲來做比喻，每次的需求都可以重新組合。從獲取的觀點來看，知識分析是知識組織的基礎，這種以知識分析為主的知識組織方法，例如索引法、主題詞表、多面向分類號法等。而語意網的時代，透過鏈結資料和知識本體，將原本獨立的知識紀錄，可以藉由時間、空間、人物、事件，加以重新鏈結組織，呈現新的知識版圖，擴大知識理解，是一種創新的知識組織觀點（如表 2）。

表 2 管理、獲取和創新觀點的知識組織

	管理的觀點	檢索的觀點	創新的觀點
特質	<ul style="list-style-type: none"> ● 以盤點為主 ● 事先訂定的分類表 ● 邏輯相關 (logical relevance) ● 依照知識前組合呈現 ● 以知識集結為主要方法 	<ul style="list-style-type: none"> ● 以獲取為主 ● 依當時的資訊需求檢索 ● 需求相關 (pertinence relevance) ● 檢索是依照後組合 ● 以知識分析為主要方法 	<ul style="list-style-type: none"> ● 以知識創新和知識再現為主 ● 以標記和詮釋知識為主要方法
方法	<ul style="list-style-type: none"> ● 圖書十進分類法 ● 國會分類法 	<ul style="list-style-type: none"> ● 索引法 ● 主題詞表 ● 多面向分類號法 	<ul style="list-style-type: none"> ● 交換標準 ● 知識本體 ● 鏈結資料

分類理論

知識組織和知識分類的關係密切，所有的組織方法都植基於分類的概念。傳統圖書資訊學的學生在學習分類法的時候，都被要求謹記一個教條，就是做分類表要窮盡範圍、類別之間一定要互斥。圖書分類法僅能講求類別與物件的關係，至於屬性，則包含在類別的定義之中，並不特別提出來分析，處理圖書文獻物件的多元屬性之時，最多利用複分法來加以處理，例如歷史複分、地理複分等。不過從分類理論來思考分類的方法，就會發現傳統入甲類即不入乙類的知識分類概念，並非唯一的知識分類概念。Sparck Jones (1970; 2005, pp. 575-576) 提出分類的三個要素：類別 (class)、物件 (object/entity) 和屬性 (attributes)，利用這三個要素的特質，可以組成 8 種分類模式。在類別方面，思考類別和類別之間的關係是有次序排列或是沒有次序，例如傳統知識組織或知識分類非常強調順序，但是在以檢索為主的電腦資料庫的環境中，排序就並不重要了；第二個要素是物件，思考物件和類別的關係，傳統的分類法中，物件入甲類即不入乙類，特別講求分類的互斥概念，不過物件入類一定要互斥嗎？是否可以同時進入不同的類別嗎？這個分類概念在以管理觀點的環境，一定是要求互斥，但是在以檢索為目的的環境中，則檢索點越多，顯然對檢索越有幫助；第三，屬性和類別的關係，一種類別是包含單一屬性或是包含多重屬性，依此思考，得到 8 種分類的類型 (如表 3)，並指出從檢索的觀點來看，後驗 (posteriori) 分類優於先驗的 (priori) 分類。

表 3 Sparck Jones 的 8 種分類體系

	類別 (有無次序)		物件 (是否互斥)		屬性 (單一或多重)		舉例
	有	無	是	否	單一	多重	
I	X		X		X		學術分類表 (Taxonomies)
II	X		X			X	十進分類表複分
III	X			X	X		多分類號法
IV	X			X		X	多分類號法 複分法
V		X	X		X		國會分類法 四部法
VI		X	X			X	國會分類法 複分法
VII		X		X	X		多面向分類法
VIII		X		X		X	鏈結資料法 語意網

資料來源：譯自 Sparck Jones, K. (1970). Some thoughts on classification for retrieval", *Journal of Documentation*, 26(2), 89-101. (Reprinted in *Journal of Documentation*, 61(5), 2005, 571-581.)

型 I 類別有固定的次序、物件分類必須互斥、每個類別的屬性獨特單一，以學術分類表 (taxonomies) 為典型代表，例如中學時候大家都學過的生物學分類法，有界、門、綱、目、科、屬、種七個層屬，每個層屬之中的類目也有一定的邏輯次序，物件一旦編入一個類別，就不能歸入另一類別，每種類別的屬性都是獨特單一；型 II 類別有固定的次序、物件分類必須互斥、每個類別的屬性可以多重，以十進分類法中的複分功能為典型代表；型 III 類別有固定的次序、物件並不互斥，亦即可以同時入甲類或乙類、每個類別的屬性單一，以多分類號法為代表；型 IV 類別有固定的次序、物件並不互斥，亦即可以同時入甲類或乙類、每個類別的屬性多重，以多分類號法和複分法為代表；型 V 類別沒有固定的次序、物件互斥、每個類別的屬性單一，以美國國會分類法為典型代表；型 VI 類別沒有固定的次序、物件互斥、每個類別的屬性多重，以美國國會分類法和複分法為代表；型 VII 類別沒有固定的次序、物件並不互斥，亦即可以同時入甲類或乙類、每個類別的屬性單一，以多面向分類法或冒號分類法為代表；型 VIII 類別沒有固定的次序、物件入類也並不互斥，亦即可以同時入甲類或乙類、每個類別的屬性多重，過去有爭論，認為這些條件的組合不能構成分類法，不過在數位網路的環境中語意網和鏈結資料卻順理成章成為 8 個分類模型中型 VIII 的典型代表，這個分類理論的模型和意義在數位時代就比較容易顯現和理解了。

學術分類法以嚴謹著稱，但在應用分類方面，則仍有很多可能的爭議，例如番茄、酪梨在學術分類中固然有其位置，但是若要分類蔬菜還是水果，就有很多爭議，有可能兩者皆是，這時如果採用型 III、IV、VII、VIII 等分類模式設計，知識組織可以入多類目，就可以解決這類應用的問題。

分類困難

在大學研讀圖書資訊學時，主要學習杜威十進分類法、美國國會分類法、主題詞表等知識組織的方法；而在美國研究所學習知識組織、分類，老師一定出一個很有趣的練習題：將

波赫士著名的中國某個百科全書中的 14 類動物重新分組。這 14 類動物有：(a) 屬皇帝所有、(b) 有芬芳的香味、(c) 馴順的、(d) 乳豬、(e) 鰻鯨、(f) 傳說中的、(g) 自由走動的狗、(h) 包括在目前分類中的、(i) 發瘋似地煩躁不安的、(j) 屬不清的、(k) 渾身有十分精緻的駱駝毛刷的毛、(l) 等等、(m) 剛剛打破水罐的、(n) 遠看像蒼蠅的。看起來十分無厘頭的項目，究竟要如何歸類？更有趣的是，這道題目背後其實深有寓意。波赫士 (Jorge Luis Borges) 是阿根廷著名詩人，曾任阿根廷國家圖書館館長，他撰寫〈約翰威爾金斯的分析語言〉(The Analytical Language of John Wilkins) (Borges, 1993) 其實是用來批評 17 世紀英國數學家、語言學家約翰威爾金斯 (John Wilkins) [6] 所提出來的一套分類符號，波赫士對於這套分類符號大加抨擊，採用反諷的筆法，文章中時而偽託，時而真實，例如聲稱他聽 Franz Kuhn (1884-1961, 德國漢學家，翻譯過紅樓夢和金瓶梅) (這一部分的訊息為真)，說有一本中國的百科全書 Celestial Empire of benevolent Knowledge(帝國仁慈寶典)[7] (這部分是虛構的)，說書中將動物分為 14 類，用以表達他對於 Wilkins 提出的分類語言和思維的荒謬大不表贊同。無論如何，赫赫有名的波赫士寫的反諷偽託，又被傅柯大大的欣賞，在傅柯的成名著作《事物的次序：人類科學考古學》[8] 前言中就提到《事物的次序》一書其實是受波赫士這個好笑的分類案例所激發 (Foucault, 1994, p. xv)。波赫士的偽託，加上傅柯名著的轉述，成為學界十分有名的分類練習案例。老師出這個練習，一方面告訴我們，分類工作所面對的往往是難以理出頭緒的內容，一方面提醒我們，理想的分類法可能並不存在。

分類應用：自然物和人造物分類—知識分類、文獻分類和語意網分類

知識是將一組概念和意義用可理解的、有結構的方式呈現，以彰顯概念或事物之間的關係和意義；分類是將事物依照相同、相異、相關等屬性，分成為不同類別；而組織則是將特質相同的組合在一起，依照相關性質加以排列，分類若是一種心智的辨識、比較和區分的活動，組織便是一種根據分類結果加以組合的系統。分類既是一種辨識、比較和區分的心智活動，便需要知道分類對象的知識範疇和區分的目的，因此分類可以根據特定的目的、需要來建立分類的原則，透過比較，把分類的完整範疇確定之後，依照相同、相異、相關等屬性劃分成為不同類別的知識體系，例如可以依照效用分類，例如中國最早的藥書《本草綱目》，便是依照藥草的效用分類，也可以依照需要診治的徵狀來分類，另外也有依照知識或物質的屬性、形態、運作方式、思維方式、自然現象或社會現象、研究方法、研究對象、知識的內在聯繫，或是發展趨勢等來分類。

如果思考自西方亞里是多得以來哲學家努力以求的分類法，以及圖書文獻分類法的異同，也許可以用自然物的知識分類和人造物的文獻分類來辨別，相較於人造物，自然物或是人類的自然環境知識的屬性可能比較穩定，而人造物則屬性複雜多元，有很多變動的可

能。自然物的屬性相對容易定義，其類別依次序排列、涵蓋完整、入甲類即不入乙類，是一種規範比較嚴謹的邏輯知識組織，從分類理論來看，可以稱為知識分類法 (taxonomy)，例如生物學分類為界門綱目科屬種等七大層級，是以自然物為分類目標，用白話文來說，也就是天造之物，對天造之物設法了解和歸類。人造物和自然物在分類實踐最大的不同就在於人造物的物件屬性比較不穩定，且經常依情境和需求而有改變，在分類上的困難和複雜程度甚高。

分類法的最大宗應用應該是在人造物的分類，例如圖書和文獻分類，又可稱為文獻知識組織。在實體物件時代，現代分類號的概念是利用一組獨特號碼，用以指引該圖書文獻所在的位置，因此無法指定多重分類號。在數位物件時代，雖然還是需要一組獨特代號來聯繫物件所在位置，卻可以採用多種知識表徵的方法，提供多種檢索點來增加檢索的查全率 (recall) 和查準率 (precision)，例如同時使用多種分類法，杜威十進分類法、美國國會分類法、多面向分類法、標題表法等來協助檢索。到了語意網的環境，知識組織工具可以利用邏輯知識分類法、文獻分類法等為基礎，建立節點和關聯的知識表徵，例如知識本體 (ontologies)、主題地圖 (topic maps)、鏈結資料 (Linked Data) 等，透過標示語言來詮釋和鏈結相關的知識，進一步組織成創新知識再現 (knowledge representation)。

立體知識分類演進—「跨領域」和「超領域」

施正鋒 (2013, 頁 132-134) 曾經提出三種不同的領域研究觀點，分別是「孤島式」、「跨領域」和「超領域」(如圖 1)，其中「孤島式」是指研究者將歷史學、語言學、文學、宗教學、心理學、社會學、經濟學、政治學等單一分別研究，以分類法比喻，就如同傳統知識分類法，入甲類即不入乙類；第二種是「跨領域」研究，連結各學科主題，但是學科主題之間僅有超連結的功能，並未有交集，如同傳統知識分類法中的參照或多面向分類；第三種是「超領域」，將各學科主題連結，並能產生彼此間的交互作用和關聯，也就是各學科主題知識之間能夠有交集，使能產生新的知識發現，如同目前鏈結資料知識組織方法。相較於孤島式單一傳統知識組織，「超領域」知識觀結合、組織相關學科主題，是一種語意網的知識組織概念。



圖 1 施正鋒 (2013)「客家學」、「孤島式」、「跨領域」和「超領域」的三種知識觀

資料來源：施正鋒 (2013)。原住民族知識體系與客家知識體系。台灣原住民族研究學報，3 (2)，133-134。

以另一個知識本體結構案例來說明，汪明輝 (2009) 研究原住民族知識和保存問題，從個人主體和互為主體的群體觀來思考，提出鄒族存有及知識本體架構，將本體架構分為社會、歷史 (時間)、空間三者，其中社會性包括了經濟、文化、社會、宗教、政治司法等；歷史性包括了個體、群體歷史、歷史神話、主觀時間和生活時間等；空間性有客觀真實時間、主觀空間符號和生活空間等，可以說表面上這是依照鄒族發展的自然現象和社會現象分類，進一步看，也看出研究者試圖從知識的內在聯繫來建構鄒族的知識分類體系。利用多元的知識分類方式，在語意網的時代可以建立創新的立體的知識本體，相較於孤島式分類原理所呈現的「點」、超連結、多面向分類原理所呈現的「面」，語意網的知識本體和鏈結資料可以呈現包括時間、空間、人物和事件所共同組成的「立體」知識面向。

知識本體 (ontologies) 應用

分類法是知識組織中最重要核心方法，近年來由電腦科學家和人工智慧研究者相繼提出知識本體 (ontologies) 的概念，一般而言，知識本體 (ontology) 是哲學家用以對事物的完整描述，而電腦科學家和人工智慧研究者的知識本體 (ontologies) 則另有所指，主要用於機器可以辨識的事物描述，和分類法有很密切的關係，但是兩者究竟有何相同與相異？

Noy 與 McGuinness (2001) 指出建立知識本體的目的是研究者群為共享某個領域知識而建立的正規語言，透過正規語言的建立，研究者或軟體之間可以了解共同的語言、可以重複使用領域知識、領域中的假設可以更明確、有助於區分領域知識和操作知識，以及有助於分析領域知識。

Lassila 與 McGuinness (2001) 提出知識本體詞彙表，認為在知識本體的詞彙光譜上面，從最左邊的目錄、詞彙表到索引典，以及最單純的屬於關係，是屬於人可以辨識的知識本體，從中間右邊開始的計算機正規屬於關係、實例關係、結構和屬性關係、值限、部分、一般邏

輯侷限等，就是正規語言（formal language）的知識本體（如圖 2）。因此可知知識本體是用以表達知識再現，是知識表徵（knowledge representation）的方法，有一部分是人可以閱讀，有一部分是給機器閱讀。而分類法雖然也是知識表徵的一種方法，實際上比較偏向知識引道（surraget），就是透過分類，指引到知識原件的方法。

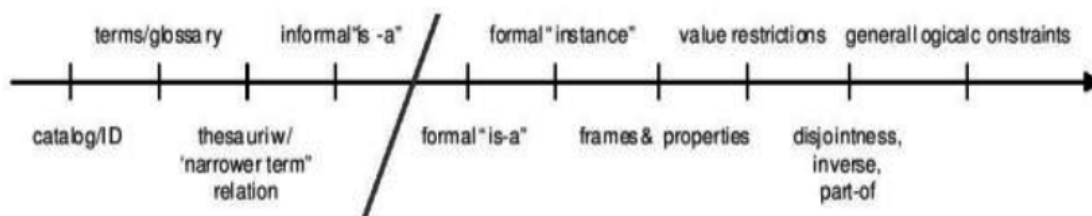


圖 2 Lassila 與 McGuinness (2001) 知識本體詞彙表

資料來源：Lassila, O., & McGuinness, D. L. (2001). The Role of Frame-Based Representation on the Semantic Web (Knowledge Systems Laboratory Report, KSL-01-02). Stanford University. Retrieved from http://www-ksl.stanford.edu/pub/KSL_Reports/KSL-01-02.html

Legg (2007) 進一步將語意網中使用的知識本體（ontologies）工具依照三種層次說明，第一種是索引典類的知識本體(thesaurus ontologies)，包括都柏林核心集、詮釋資料(Metadata)和要素集(element set)，相當於分類架構；第二種是知識本體表達(Ontologies with (Largely) DL Expressivity)，利用「主題地圖」(topic maps)、「資源描述架構」(RDF Schema)、DAML+OIUO WL 和 SHOE 等語意網語言；第三種是本體最高層的邏輯知識表徵或表達(Ontologies with First-Order Logic Expressivity (or Higher)，最流行的語言有 Cyc、SUMO (Suggested Upper Merged Ontology) 和 SUO (Standard Upper Ontology) 等三種，Legg 的上述說明，也同樣是將知識本體依照人可以閱讀和理解的層次，一直發展推演到機器可以辨識和演算的程次。夏翠娟(2016)也主張知識本體需要先解決三個問題，包括提出領域知識本體模型、抽象資料模型，以及資料編碼格式，其中領域知識本體模型是提供應用領域中涉及到的概念定義，以及概念間關係，是建立資料間關聯關係的基本依據，這是以一般人可以閱讀的語言來處理的；第二個問題是建立抽象資料模型，定義資料與資料間的邏輯結構，是知識組織方式的實踐，抽象資料模型的表現，使用「資源描述架構」(RDF)來處理；第三是資料編碼格式，利用系統邏輯語言表達，由機器讀取、處理和理解資料的語義，可以呈現知識本體的表達形式。

Giunchiglia、Zaihrayeu 與 Farazi (2008) 是義大利多倫多大學的研究團隊，則從使用者、目的、語言、節點、邊緣、實例、案例等方面，比較分類法和知識本體的不同，認為分類表

是給人讀的，而知識本體是給機器讀的，前者是自然語言，後者是電腦正規語言，前者作為文獻館藏的知識組織，後者是在建立領域模型等（如表 4）。

表 4 Giunchiglia, et. al. (2008) 分類表和知識本體之比較

範疇	分類表 (classification schemes)	知識本體 (ontologies)
使用者	人	機器
目的	文獻館藏組織	建立領域模型 (modeling of a domain)
語言	自然語言，例如英語	計算機正規語言，例如 OWL
節點	通常是複雜的概念或個體	通常是最小的概念
邊緣	語意定義不清	語意定義明確
實例	不一定填滿類目	填滿類目
案例	DDC, LCC, Colon classification	Gene ontology, OpenCyc ontology, MeSH ontology

資料來源：譯自 Giunchiglia, et. al. (2008). Converting classifications into OWL ontologies, Technical Report. Department of Information Engineering and Computer Science, University of Trento, Italy. p.5)

中央研究院資訊科學研究所利用 WordNet 和自然語言的技術，發展領域詞彙分類表和雙語知識本體詞網及兩者的對照（陳永祥、黃居仁，2006），便是典型的知識本體系統，該系統的主要應用領域包括雙語查詢、拓展查詢檢索點、提供知識地圖等，例如利用中英雙語對應詞彙，提供聯合目錄展示系統發展中英雙語查詢介面，透過連結「領域辭典詞彙表」（Domain Lexico-Taxonomy, DLT）、「中央研究院中英雙語知識本體詞網」（Bilingual Ontological Wordnet, Sinica BOW）與「詞網」（WordNet），有利於拓展多種知識組織方法，包括提供瀏覽或檢索的拓展查詢模式，而在知識地圖的應用方面，則可以透過主題分類與「建議上層共用知識本體」（Suggested Upper Merged Ontology, SUMO）節點的對應，描繪數位典藏計畫中典藏品在整體人類知識架構中所在之位置，有助於了解相關領域知識之定位。

鏈結資料和創新知識詮釋

檢視知識分類、文獻分類，再到語意網絡的知識組織，是隨著時間、科技和載體變遷而有不同的發展。嚴謹的邏輯知識分類非常不容易實踐，一本書不可能只論及一種概念，不論是原住民族的知識觀、客家民族的知識觀，都無法只套用一種分類法。每一個物件都可以有很多的分類法或知識組織的方法，不能拘泥於一種分類方法。圖書館界一向以收藏人類全體知識為職志，對於知識組織方法的最新發展最為關注，美國國會圖書館可以做為代表，為提供鏈結資料組織服務，釋出了 54 種各式知識本體系統、鏈結資料服務項目，包括（1）詞彙表（例如 LC subject headings, LC Name Authority File, LC Children's Subject Headings）、（2）分類表（例如 LC Classification, Classification Schemes）、（3）機讀編目格式（例如 MARC Relators, MARC Countries）、（4）語言國際標準（例如 ISO639-1 Languages, ISO639-2 Languages,

ISO639-5 Languages)、(5) 詮釋資料 (例如 Extended Data/Time Format, Identifiers)、(6) 知識本體等 (例如 BIBFRAME Ontology, Event Type Ontology for LC Classification) 六大類型 [9], 主要目的在協助人和機器都可以利用國會圖書館所提供的權威檔資料集。Ford (2013) 指出國會圖書館的鏈結資料服務的第一個工作, 應該就是將國會分類號 (LCC) 和杜威分類號 (DCC) 做互相連結, 而該項資料互相鏈結的重要性, 並非只在協助指引到書架找書, 而是在於書目資料的指引, 也就是相關知識的鏈結和檢索。美國國會的鏈結資料服務系統背後的企圖應該是服務人類全體的知識組織、知識組織的對象文獻, 該鏈結服務提供了語意網的基礎, 可以達到超領域知識組織的要求。

開放鏈結資料 (Linked Open Data, LOD) 是以有意義的方式, 連接不同而相關資料的方法, 可以經由電腦讀取, 並傳播到網路上。鏈結資料需要幾個基本條件, 包括定義物件的獨特性和標籤, 以及採用「資源描述架構」RDF 三元組 (triple) 的方式提出事物的關聯性, 例如一個字串「蔣復璁認識梁啟超», 我們當然看得懂, 但是對電腦而言, 只是一串文字, 如果要讓電腦能夠辨識特定的蔣復璁先生和梁啟超先生這兩組符號, 就必須為這兩個人個設定一個「獨特資源識別碼」(Unique Resource Identifier, 以下簡稱 URI), 並且定義「認識」(know of), 這樣電腦透過辨識兩個獨特的人名標籤和兩者關係的定義, 就能用電腦的關聯圖方式表達出來。URI 識別碼用於個人識別, 需要長久使用, 為保持穩定, 通常由美國國會圖書館鏈結資料服務和 DBpedia, 以及線上電腦圖書館中心 (Online Computer Library Center, OCLC) 所發展的「國際虛擬權威檔」(Virtual International Authority File, VIAF) 等機構所規範和控制。

在鏈結資料的技術發展中最重要的工具應該就是「資源描述架構」(RDF), 「資源描述架構」從 2004 年提出概念以來, 已經發展了好幾個應用工具, 稱為 RDF 模式 (schema) 或 RDF 詞彙 (RDF vocabularies), 「資源描述架構」(RDF) 的規範十分結構化, 包括完整的類別和屬性的說明, 比較著名的有「朋友關係」(FOAF)、都柏林核心集、Schema.org、Simple Knowledge Organization System (SKOS) 等。

以「朋友關係」(FOAF) 為例, 在類別方面共有代理人 (agent)、文件 (document)、集團 (group)、圖像 (image)、標籤 (labelproperty)、線上帳號 (onlineaccount)、線上聊天帳號 (onlinechataccount)、線上金融帳號 (onlineecommerceaccount)、線上遊戲帳號 (onlinegamingaccount)、機構 (organization) 人 (person)、個人簡歷 (personalprofiledocument)、計畫 (project) 等 13 種。每個類別有完整的定義和內容舉例、屬性、會一起出現的類目, 以及子項目等詳細使用說明, 例如「代理人」是指個人 (person)、集團 (group)、軟體 (software) 或實體物件 (physical artifact), 代理人範圍內的屬性有: 性別 (gender)、雅虎聊天帳號 (yahooChatID)、帳號 (account)、生日 (birthday)、icq 聊天帳號 (icqChatID)、目標 (aim)、

聊天帳號 (aimChatID)、網路電話帳號 (jabberID)、製作 (made)、信箱 (mbox)、興趣 (interest)、付款或獎勵 (tipjar)、skype 帳號 (skypeID)、主題 (topic)、興趣 (interest)、年齡 (age)、個人主要信箱 (mbox_sha1sum)、狀況 (status)、msn 聊天帳號 (msnChatID)、公開帳號 (openid)、個人帳號 (holdsAccount)、網誌 (weblog) 等；「代理人」這個類別通常和製作者 (maker)、會員 (member) 一起使用；而「代理人」的子項目有個人、集團、機構等。

FOAF 目前大約定義了 60 個屬性，包括帳號 (account)、帳號名稱 (accountName)、帳號服務首頁 (accountServiceHomepage)、年齡 (age)、aim 聊天帳號 (aimChatID)、接近 (based_near)、生日 (birthday)、正在從事的計畫 (currentProject)、描述 (depiction)、提供描述 (depicts)、dna (dnaChecksum)、姓 (family_name)、名 (firstName)、聚焦 (focus)、贊助 (fundedBy)、(geekcode)、性別 (gender)、名 (givenName)、個人帳號 (holdsAccount)、首頁 homepage)、聊天帳號 (icqChatID)、圖像 (img)、興趣 (interest)、主要議題 (isPrimaryTopicOf)、網路電話號碼 (jabberID)、認識 (knows)、姓 (lastName)、標誌 (logo)、製作 (made)、製造者 (maker)、信箱 (mbox)、個人主要信箱 (mbox_sha1sum)、會員 (member)、會員等級 (membershipClass)、msn 聊天帳號 (msnChatID)、性向 (myersBriggs)、名稱 (name)、暱稱 (nick)、公開帳號 (openid)、頁面 (page)、執行過的計畫 (pastProject)、電話 (phone)、計劃 (plan)、主題 (primaryTopic)、出版品 (publications)、學校主頁 (schoolHomepage)、skype 帳號 (skypeID)、狀態 (status)、姓 (surname)、主題 (theme)、縮略圖 (thumbnail)、付款或獎勵 (tipjar)、(sha1)、頭銜 (title)、主題 (topic)、有興趣的主題 (topic_interest)、網誌 (weblog)、職場資訊首頁 (workInfoHomepage)、職場首頁 (workplaceHomepage)、雅虎聊天帳號 (yahooChatID) 等，每個屬性有定義，以及使用的類別規範。

如果說「朋友關係」(FOAF)的主要分析和組織的對象是人，那麼都柏林核心集主要就是分析和組織文獻作品，共包含 15 基本要素，題名 (title)、創作者 (creator)、主題 (subject)、描述 (description)、出版者 (publisher)、貢獻者 (contributor)、日期 (date)、類型 (type)、形式 (format)、辨識號 (identifier)、來源 (source)、語言 (language)、關係 (relation)、範圍 (coverage)、版權 (rights)。每個要素都會包含以下的屬性：標籤名稱 (name)、識別號 (identifier)、版本 (version)、註冊權威號 (registration authority)、語言 (language)、定義 (definition)、義務 (obligation)、資料類型 (datatype)、次數限制 (maximum occurrence) 和評論 (comment)。

相較於 FOAF 和 DC，由 Google，Microsoft，Yahoo 和 Yandex 所共同合作開發創建的 Schema.org 詞彙則用於組織網頁和電子郵件等資源，其分析和組織的範圍很廣，包括以下類

別：(A) 創作品 (creative works)：書籍、電影、音樂錄音、食譜、電視系列等；(B) 非文本物件 (embedded non-text objects)：聲音 (audioobject)、圖像 (imageobject)、視訊 (videoobject)；(C) 事件 (event)；(D) 健康和醫療類型 (health and medical types)：關於醫療保健和醫療類型的說明；(E) 組織 (organization)；(F) 人 (person)；(G) 地方 (place)、本地企業、餐廳等；(H) 產品 (product)；(I) 評鑑 (review)；(J) 行動 (action) 等。每個類別都有定義、屬性、應用的類型和描述，以事件為例，事件是指「在某個時間和地點發生的事件，如音樂會，講座或節日等」，而其中若有購票訊息可以透過添加提供 (offer) 這個屬性來顯示，目前事件已經有多達 25 萬個應用領域[10]。

從案例中學習——知識組織處理全體知識、領域知識、社群知識、個人知識

知識分類很困難，但是仍然可以試圖找出系統化的解決途徑，譬如首先要瞭解要分類的知識，其知識範圍是甚麼？分類人類全體的知識和分類某個領域的知識，或是個人層次的知識，需要應用不同的知識本體。過去圖書館以收藏人類全體知識為職志，所用的分類法和後來領略知識爆炸，必須各館分工，不再以收藏全人類的知識為目標，而是強調為社群服務，以社群為需求的知識組織，就會和收藏全人類知識的知識組織架構有所不同。

對於某個主題、知識領域或文類的研究者而言，語意網透過鏈結資料的知識組織方法，可以帶來對傳統事件和人物關係的全新了解，以美術史為例，陳淑君（2016）和中央研究院的團隊研究並建構以鏈結資料為基礎的藝術史數位研究環境，以近代畫家陳澄波作為研究案例，從需求面，例如使用案例分析 (case study)，以及資料面，例如藏品、人物、組織和事件等詳加收集、分析，透過需求和問題，以及對資料特質和內涵的深入理解，掌握了彼此的深層關係之後提出資料模型，然後遵從各相關可用的標準詞彙，例如 Schema.org、DBpedia、SKOS、Getty Vocabularies、DCMI Metadata Terms、RDF Schema 等，並利用資源描述架構 (RDF) 三元組格式加以編寫，使機器可以辨識處理資料。該研究成果除了提出藝術史知識本體的 30 個類別和 57 個屬性，並且由於研究材料的脈絡化，包括人、事、時、地、物，以及對於畫作內容的詮釋資料，例如一幅畫作，除了提出創作者、畫作地點、畫作內容類型、材料、畫法、對於畫作的內容物也加以描繪，對於知識物件屬性的描繪，加上鏈結資料的內部和外部資料的重新集結，可以促成新知識發現的可能，在該研究中稱為「增加發現力」(discoverability)。這是鏈結資料和知識本體做為新的知識組織方法，和過去以管理和檢索為目的的知識管理工具最為不同的地方。

另以口述歷史為例，進行人物的標記、人際之間的關係來對資料進行更細緻的分析，是十分可行的方向。以美國羅格斯大學爵士口述歷史計畫[11]為例，傳統資料庫收集的演奏家

的各項資料都是平面的，但是 Linked Jazz 計畫[12]透過「資源描述架構」三元組工具，描述了 19,197 人及其相關資料，包括演奏的樂器、出生／死亡日期、位置，以及職業，並利用 RDF 三元組描述 154 個爵士樂事件，包括表演者和樂團名稱、日期、地點（如主廳），以及該演奏會節目的主標題等。

Linked Jazz 計畫最早是由 OCLC 和美國圖書資訊學教育學會（Association for Library and Information Science Education, ALISE）所贊助的計畫，利用鏈結資料要素中的三元組工具來加以組織知識，採用資料所產生新的三元組，而不是轉換現有的詮釋資料（metadata），而用來產生鏈結資料的原始材料包括 50 多個採自羅格斯大學圖書館爵士研究檔案中的口述歷史紀錄、史密森博物館爵士口述歷史、漢密爾頓大學爵士檔案館、加州大學洛杉磯分校，以及密歇根大學檔案館的口述歷史等。

Linked Jazz 計畫首先發展一系列的工具有，以便於創建鏈結資料（LOD），包括語料分析器（transcript analyzer）、人名對照和策展工具（name mapping and curator），以及大眾分析（crowdsourcing）工具等。利用這些工具，例如語料分析器用以確認口述歷史的問答結構，一旦偵測到受訪者提到人名，受訪者和提到的人名之間的簡單三元組就自動集中到「認識」（KnowOf）相關區域。而利用人名對照和策展工具，一旦在口述訪談中發現人名，便使用 DBpedia 或「國際虛擬權威檔」（Virtual International Authority File, VIAF）、美國國會圖書館（LC）「獨特資源識別碼」（Unique Resource Identifier, URI）等線上權威人名資源先加以核對，來確認該身份證明的唯一表徵。這是利用工具來自動產生鏈結資料的第一個步驟。接著將這些訪談文本片段送到大眾分析工具，由 52 街的志願者提供詳細的詞彙來描述文本中受訪者和他們所提到的人物之間的關係，這些關係經由定義，共包括：遇到（has met）、是熟識者（is an acquaintance of）、是朋友（is a friend of）、是好朋友（is a close friend of）、受到影響（is influenced by）、是導師（is a mentor of）、是合作者（collaborated with）、屬同一樂團（was in a band together with）、一起演奏（played with）、是樂團的團員（was a member of the band of）、一起旅行（toured with）、是帶領者（was the bandleader for）等[13]。

在資料處理的最後階段，將三元組傳遞到視覺化網絡，視覺化資料是採用經過標準化「認識」（know of）的三元組定義，而不是由志願者所定義的關係，從視覺化網絡中可以呈現爵士音樂家的圖像、影音和簡短的傳記。該計畫說明，雖然爵士音樂家的名單很長，不過視覺化計劃僅處理口述歷史中有提及的爵士音樂家，而且音樂家節點呈現的大小，是由口述歷史記錄中提及次數所反映的，這是很有趣的設計。

除了口述歷史的文本作為鏈結資料的分析對象，另外一群研究者發現了一批爵士樂照片館藏，共有 1,787 圖像，有 681 個獨特的個人，和超過 2,700 照片「描述」（depiction），其中人物十分適合繼續利用「認識」（know of）和「朋友關係」（FOAF）[14]的三元體來定義人際

關係，特別在同一張照片中「提供描述」(depict)的人，可以說是以某種方式彼此認識的。

哈佛大學、中央研究院和北京大學合作的「中國歷代人物傳記資料庫」(以下簡稱 CBDB)是另一個鏈結資料產生的新知識系統案例，以人物傳記為主，大約有收錄超過 36 萬筆傳記資料。CBDB 主要的功能並不在提供史料傳記全文，而是透過建立知識本體和鏈結資料，提供有別於以往資料庫檢索僅有點和面的資料查詢功能，可以藉由重構傳主的社交圈，探索其社會關係、旅行路線，進而重新認識當時的歷史和文化，亦即將一個歷史人物其多面向的生命歷程，透過親屬關係(kinship)、非親屬社會關係(non-kinship associations)、社會地位(status)、入仕途徑(modes of entry into government)、職官/宦歷(offices/postings)、所在地(places)、著述(writings)等資訊重新立體化的建構起來。這個案例的知識本體鏈結資料的結構組成和爵士樂口述歷史音樂家的資料結構很不相同，共有人名(names)、時間(time)、所在地(places)、職官/宦歷(offices)、入仕途徑(modes of entry into office)、著述(writing)、社會區分(social distinctiveness)、親屬關係(kinship)、社會關係(social associations)、財產(possessions)、事件(events)等 11 大類。

新的知識組織方法不但用於組織全人類的知識、社群或是領域知識，也可以用於個人知識組織和知識創新，其中有一些基礎規範和建設需要先完成。上海圖書館就利用開放鏈結資料技術，發展鏈結資料化的家譜、古籍、檔案、期刊報紙等文獻知識庫，改善、規範了傳統數位館藏的格式，並以人、事、時、地四個構面建立「歷史人物/名人規範庫」、「歷史地理知識庫」、「歷史紀年知識庫」、「歷史事件知識庫」等四個基礎知識庫(夏翠娟，2016)。以「歷史事件知識庫」為例，主要是提供事件的定義，以及事件與人、時、地、文獻之間的關聯。在歷史事件知識庫中，是利用約定俗成的歷史名稱來標識事件，例如「戊戌變法」、「洋務運動」、「中日甲午戰爭」等，其中，人物、地點、時間作為事件的三個必要屬性。在事件本體概念模型中，相關人物、事件發生時間、地點作為事件的屬性值與之相關聯，事件本身則作為文獻的主題與之相關聯。建立這樣的關聯關係，從事件出發，就能到達與之相關聯的人、地、時和文獻資源，而結合視覺化技術，使傳統文獻資料庫的檢索有了創新知識呈現的可能。上海圖書館利用已建立的基礎知識庫提供開放鏈結資料服務，已經應用在「盛宣懷檔案館知識庫」和「家譜知識庫」兩種文獻知識庫，其鏈結資料可以提供民眾和研究者從事相關的歷史或譜系研究。

結語

加州洛杉磯大學(UCLA)數位人文學教授 Johanna Drucker 曾經說「沒有任何分類系統是價值中立的，客觀的，或自證的，所有的分類系統都是承自自己內部生產的思想印記」(Drucker, 2013)。從以上案例，可以看到知識觀和知識組織的關係如此密切，而又如此幽微。

知識分類或知識本體受到知識觀的影響，好在現代研究者多能先清楚界定研究的領域，以領域來框架知識的範圍，這樣在窮盡類別方面，比較容易實現。現代研究者將分類法稱為知識本體，知識本體有類別和屬性兩個部分，需要由分類者來加以提出和定義，其中的類別雖然講求窮盡，但是對於出現的順序已經不予計較；對於屬性，可以在類目中單獨存在，也可以在不同的類別中共用，這些在傳統的分類法中無法被理解和接受的特質，在電子計算機時代卻通行無阻，可見 Karen Sparck Jones 在 1970 年代提出來以類目、物件和屬性之分類要素來解釋八種可能的分類法模式確實可行，特別能夠說明知識本體和鏈結資料在分類法中的位置。上述各項知識組織的發展對圖書資訊學教育和教學的啟發，除了強調知識組織的方法和工具之外，以提醒分類理論思考訓練的重要。

分類和知識組織的方法與時俱進，從早期知識組織用於辨識、安置、管理，到後來多面向分類檢索，以至於近十餘年來知識組織領域藉由電腦科技的輔助發展迅速，透過鏈結資料方法，連結超領域創新知識成為可能。這些發展從 Lassila 與 McGuinness (2001) 所提出來的知識本體詞彙表光譜看出來，最左邊的是早期使用的目錄和詞彙表，是孤島式的知識組織方法，逐漸發展到多面向的知識組織，一直到後期最右邊電腦語言表達，提供知識鏈結實踐。由知識組織最近發展的案例來看，未來知識組織工作者應該會需要具備以下的知識技能：

1. 領域知識方面

需要能對領域知識中各項相關資訊資源有廣泛理解。由於處理的知識物件形態不限於文獻、文類或文本，在收集資訊資源、典藏品方面，各種人類的知識產品，不限類型，不限載體都會觸及。

2. 深度分析能力方面

需有洞察力和深度詮釋的能力。建構知識本體需能提煉該知識領域的詳盡類別和屬性，並能發展其間創新的關係，對事物有深刻的觀察和理解是必備的能力。

3. 計算機能力方面

要能撰寫程式語言，運用電腦能讀的語言編譯、處理資料，熟悉關聯式平台運作和查詢語言 SPARQL 等操作，使能呈現新的資料關聯。

未來在規劃相關課程宜加以注意，將上述三方面之知能，設計到系列課程之中，希望能對培養知識組織新的人才有所啟發。

致謝

本文從 2016 年 6 月 2 日中央研究院主辦「語意網視野下的知識組織與文化資產」研討會發表論文改寫而來。感謝與會者提供意見並鼓勵，並感謝匿名審查者的修改卓見。一併致謝。

附註

[1]按 approach 有方式或取向的意思，取向比方式易於表達知識的抽象特質，因此兩者並陳。

[2]按 perspective 有角度和觀點的意思，觀點比角度易於表達知識的抽象特質，因此兩者並陳。

[3]有一些人將知識面以內容（content）稱之，將物件面以承載物（container）稱之。本文則以資訊的知識面、物件面和過程面表示知識的內容、承載和傳播三個特質。

[4]華盛頓大學圖書館建築最早是在 1926 年完工，該館建築頗負盛名，最著名的大約就是外牆支柱有由教授推薦的 24 位全球著名的哲學家 藝術家 思想家等，以及高處另有三座代表思想、啟發和專精等三座雕像 但是有趣的是文獻上很少看到對這六道門，以及對門上銅版鑄刻的描述，在舊金山的 Tim McCormick 注意到了！“Strangely, I can t find mention of the inscriptions which are easily the most visible to any person actually entering the library: the brasswork directly over the six doors, representing six famous figures in printing and publishing: Tao Feng, Gutenberg,..., and Elsevier.” <<http://tjm.org/2009/10/03/secrets-in-the-facade-of-univ-washingtons-library/>>

[5]例如張偉保（年代不詳）。從〈馮道傳〉看五代政權之文官通朝現象。<<http://www.wangngai.org.hk/42-cheung.html>>

[6]John Wilkins, 1614--1672 曾經擔任過牛津和劍橋大學的校長，也是倫敦皇家學院的創辦人之一。<https://en.wikipedia.org/wiki/John_Wilkins>

[7]另有幽默者將之翻譯成《天朝仁學廣覽》，並依波赫士捏造之 14 種動物分為 14 章。

[8]原著法語出版於 1966 年，英譯本出版於 1994 年，將書名從《詞與物》改為《事物的次序》。

[9]總共 54 種項目，研究者將之區分為六大類型。

[10]<http://schema.org/Event>

[11]<http://newarkwww.rutgers.edu/IJS/OralHistory.html>

[12]<https://linkedjazz.org/about-the-project/>

[13]<https://linkedjazz.org/about-the-project/>

[14]<https://linkedjazz.org/tag/foaf/>

參考文獻

Borges, J. L. (1993). *The Analytical Language of John Wilkins*. In *Other inquisitions 1937-1952*. University of Texas Press. Retrieved from <https://crma.stanford.edu/courses/155/assignment/ex1/Borges.pdf>

Buckland, M. (1991). Information as thing. *Journal of the American Society for Information Science (JASIS)*, 42(5), 351-360.

- Drucker, J. (2013). *Classification systems. Introduction to Digital Humanities DH101*. Retrieved from http://dh101.humanities.ucla.edu/?page_id=33
- Ford, K. (2013). Library of Congress Classification as linked data. *JLIS.it.*, 4(1), 162-175. doi: 10.4403/jlis.it-5465
- Foucault, M. (1994). *The Order of Things: An Archaeology of the Human Sciences*. New York: Vintage Books.
- Giunchiglia, F., Zaihrayeu, I., & Farazi, F. (2008). *Converting Classifications into OWL Ontologies*. Technical Report. Department of Information Engineering and Computer Science, University of Trento, Italy. Retrieved from <http://eprints.biblio.unitn.it/1439/1/027.pdf>
- Lassila, O., & McGuinness, D. L. (2001). *The Role of Frame-Based Representation on the Semantic Web* (Knowledge Systems Laboratory Report, KSL-01-02). Stanford University. Retrieved from http://www-ksl.stanford.edu/pub/KSL_Reports/KSL-01-02.html
- Legg, C. (2007). Ontologies on the Semantic Web. *Annual Review of Information Science and Technology (ARIST)*, 41(1), 407-451.
- Mccormick, T. (2009). *Secrets in the facade of Univ. Washington's library*. Retrieved from <http://tjm.org/2009/10/03/secrets-in-the-facade-of-univ-washingtons-library>
- Noy, N. F. & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first Ontology*. Retrieved from http://protege.stanford.edu/publications/ontology_development/ontology101.pdf
- Smiraglia, R. P. (2014). *The elements of knowledge organization*. New York, NY: Springer.
- Sparck Jones, K. (1970). Some thoughts on classification for retrieval. *Journal of Documentation*, 26(2), 89-101. (Reprinted in *Journal of Documentation*, 61(5), 2005, 571-581.)
- 李鶴立 (2015)。原住民族、孔子及知識組織--專題編者序。《圖書資訊學刊》，13 (2)，99-105。
- 【Lee, Hur-Li (2015). Indigenous peoples, cofucius, and knowledge organization: Guest editorial. *Journal of Library and Information Studies*, 13(2), 99-105.】
- 注明輝 (2009, 5月)。臺灣原住民族知識論之建構—以鄒族傳統領域資源管理知識為例。在臺灣原住民教授學會、東華大學原住民族學院主辦，第一屆原住民族知識體系研討會，花蓮縣。
- 【Wang, Ming-Huey (2009, May). *Building Taiwan Indigenous Epistemologies: A Case of Cou's Traditional Knowledge of Territorial Resource Management*. In Taiwan Indigenous Professor Society & College of Indigenous Studies, National Dong Hwa University (Eds.), *The First Symposium on the Aboriginal System of Knowledge*, Hualien County.】
- 施正鋒 (2013)。原住民族知識體系與客家知識體系。《臺灣原住民族研究學報》，3 (2)，115-142。
- 【Shih, Cheng-Feng (2013). Indigenous knowledge system and hakka knowledge system. *Journal of the Taiwan Indigenous Studies Association*, 3(2), 115-142.】
- 夏翠娟 (2016, 6月)。基于關聯數據的數字人文實踐：上海圖書館的探索。在中央研究院數位文化中心、中央研究院歷史語言研究所主辦，語意網視野下的知識組織與文化資產研討會，台北市。
- 【Xia, Cuijuan (2016). *Jiyu guanlian shuju de shuzi renwen shijian: Shanghai tushuguan de tansuo*. In Academia Sinica Center for Digital Cultures & Institute of History and Philology, Academia Sinica (Eds.), *Knowledge Organization and Cultural Heritage: Perspectives of Semantic web*, Taipei.】
- 陳永祥、黃居仁 (2006)。利用領域詞彙分類與雙語知識本體詞網輔助主題資訊搜尋。第五屆數位典藏技術研討會。檢自：<http://www.ling.sinica.edu.tw/eip/FILES/publish/2007.7.18.56698244.2431219.pdf>

【Chen, Yun-Hsiang, & Huang, Chu-Ren (2006). *Liyong lingyucihui fenlei yu shuangyu zhishibenticiwang fuzhu zhuti zixun souxun*. Diwujie shuwei diancang jishu yantaohui. Retrieved from <http://www.ling.sinica.edu.tw/eip/FILES/publish/2007.7.18.56698244.2431219.pdf>】

陳淑君（2016，6月）。**建構以鏈結資料為基礎的藝術史數位研究環境**。在中央研究院數位文化中心、中央研究院歷史語言研究所主辦，語意網視野下的知識組織與文化資產研討會，台北市。

【Chen, Shu-Jiun (2016, June). *Jian gou yi lian jie zi liao wei ji chu de yi shu shi shu wei yan jiu huan jing*. In Academia Sinica Center for Digital Cultures & Institute of History and Philology, Academia Sinica (Eds.), *Knowledge Organization and Cultural Heritage: Perspectives of Semantic web*, Taipei.】

Some Thoughts on Knowledge Organization in the Web Era

Mei Mei Wu

Professor and Chairperson, Graduate Institute of Library and Information Studies,
National Taiwan Normal University, Taiwan (R.O.C.)

E-mail: meiwu@ntnu.edu.tw

Keywords: Classification Theory; Knowledge Organization; Linked Data; Ontology;
Semantic Web

【Abstract】

This article argues that the biggest difference between the knowledge organization method and the traditional information organization in the Internet age is that the boundaries of information objects in the Internet age no longer exists, the interconnected characteristics of information and knowledge, and the concept and method of knowledge organization are fundamentally changed. From the island-like knowledge organization for the single information object as the knowledge organization unit, evolving to the multi-faceted knowledge organization method, of which the major functions are to identify and search for information. The Semantic Web approach of knowledge organization allows remapping and reconstructing new knowledge in terms of interpretation and thus offering new insights of knowledge. This article addresses the related theories and cases of such knowledge organization to highlight the new development of dialectical knowledge organization.

【Long Abstract】

Introduction

This paper proposes that there has been a paradigm shift in the way knowledge is organized (knowledge organization, KO); the traditional reasons for organizing information and the methods used have changed with the development of the Internet. It first argues that in the Internet era, the term “knowledge organization” is more appropriate than the term “information organization” as in our digital world, information can no longer be confined to a specific physical location as in the past. Lee (2015, p.100) supports the use of the term “knowledge organization” rather than

“information organization” because embedded within the term are concepts, terminology, meaning, relationships, order, structure, approaches, perspectives, context, etc. These are not included in an understanding of “information”. By applying the new methods of knowledge organization, such as ontologies, Linked Data, and triple RDF, knowledge will have enhanced meaning, structure, and order. Furthermore, in "knowledge representation" there will be the opportunity for reinterpretation and finding new knowledge and meaning. This paper also discusses the work of Dao Feng, the 10th century Chinese officer who made the printed Chinese classic sutra available to highlight that the bias of knowledge affects the scope of knowledge and knowledge organization.

Observation and thoughts

The evolution of information technology in the Internet age has encouraged the shift in the KO model. The Semantic Web approach of KO allows the remapping and reconstructing of knowledge in terms of interpretation, thereby offering new insights. The following eight issues, which highlight the movement of knowledge organization in the digital environment, are considered:

1. Characteristics of media affect KO methods: The methods used to organize knowledge in the Internet era differ from traditional methods in that there are no longer boundaries to the single knowledge object. This shift is the result of the interconnection between information and the characteristics of knowledge.
2. New KO methods expand KO functions: The major function of traditional information organization was to manage the information in order to be able to rapidly retrieve it. KO in the Semantic Web environment allows new interpretations, new representation. The reconstruction of knowledge is thus a new function.
3. New KO methods can be explained by Classification theory: Sparck Jones (1970, 2005) proposed three basic elements of classification: class, object/entity, and attributes. By using these three basic elements, eight classification models are developed. It is usually assumed that academic classification, the first category, is rigorous, but there are always still some grey areas that lead to differences of opinion as well as some difficulties. For example, the problem of tomatoes and avocados: there is always some controversy about whether they should be classified as vegetables or fruit as they could be either. These sorts of problems can be solved in Sparck Jones' classification model as it allows objects to be represented in multiple categories with compound attributes.
4. Classification is difficult: Jorge Luis Borges wrote “John Wilkins's Analytical Language” (Borges,

1993) to criticize a set of classification symbols developed by John Wilkins, the 17th century British mathematician and linguist. Borges made up 14 categories of animals from a fictional Chinese encyclopedia called the “Celestial Empire of benevolent Knowledge.” The story inspired Foucault and was documented in the foreword of *The Order of Things* (1994). The 14 categories have become a famous classification exercise to show LIS students that classification is difficult and that a perfect classification system may not exist.

5. Current classification in practice: There are different levels of complexity in classifying natural objects and artifacts in regard to scientific knowledge and the classification of documents and digital objects in the Internet. KO tools such as ontologies, topic maps, Linked Data, etc., are now available for innovative knowledge representation.
6. Evolution of KO: The traditional way of organizing knowledge is basically an island-like type for a single object of information. The development of the multi-oriented KO method allows multi-faceted classification; however, its function is still identifying and searching for knowledge. The triple RDF Linked Data concepts establish the innovative knowledge representation functions to weave events, people, time, space, characters, and facts together.
7. Developing all types of ontologies: To reconstruct knowledge and obtain in-depth understanding, case studies and modeling are required for developing more usable ontologies so that knowledge can be identified and translated from human to machine.
8. Transforming to Linked Data: Those who work in libraries are most concerned about the latest development of KO methods. The Library of Congress released 54 ontologies and Linked Data, such as subject headings, classification schemes, machine-readable catalogs, international language standards, metadata, ontologies, etc., to support and improve KO.

Learning from the individual cases

Chen (2016) and researchers at Academia Sinica Center for Digital Cultures’ Linked Data project employing a case study, used the works of modern painter, Chen Chengbo, to link collections, people, organizations, and events. They did this by using the relevant standard vocabulary, such as Schema.org, DBpedia, SKOS, GettyVocabularies, DCMII Metadata Terms, RDF Schema, etc. In the study, 30 categories and 57 attributes of the knowledge ontology of art history were identified. The significant contribution of this project is that a new KO can be represented by the “discoverability” through the link between the information and the knowledge ontology.

The Linked Jazz Oral History project Led by Pattuelli at Pratt Institute in New York is another

example of a successful project. Based on the Rutgers University Jazz Oral History Project, the Linked Jazz project transferred the traditional flat data to RDF triples by applying the Resource Description Framework. The triad describes 19,197 people and their relevant information, including musical instruments, dates of birth/death, locations, and occupations. It uses the RDF triples to describe 154 jazz events, including the names of performers and orchestra, dates, venues (such as the main hall), and the main titles of the concert programs.

The China Biographical Database Project (CBDB) project, which includes 36 million items of biographical information, is another case generated by the new KO system. The main function of CBDB is not to provide the full text of historical biography; rather, through the establishment of knowledge ontology and link information, its function is to provide indications of social circles, explore social relationships, and present travel routes. Then it can offer a re-understanding of the history and culture, that is, a historical portrayal of its multi-oriented life course, through kinship and non-kinship associations, social status (status, entry into Government offices / postings, places, writings), and so on. This differs from traditional database searches which have only limited access points and superficial information query functions.

The fourth case is Shanghai Library's Genealogy project which uses the Linked Open Data (LOD) technology to develop a database of genealogies, including genealogy records, ancient books, archives and periodicals, to improve and standardize the format of traditional digital collections. The Shanghai Library has provided the Open Link Information Service with four established basic knowledge bases, namely, historical figures / celebrity knowledge base, historical geographic knowledge base, historical chronology knowledge base, and historical event knowledge base. Among many, "Sheng Xuan-huai Archives Knowledge Base" and "Genealogical Knowledge Base" are two of the applications. The links provide information on the people and researchers related to the history or genealogy study. Combined with visualization technology, the retrieval of traditional literature databases with the possibility of innovation and knowledge can be presented.

Conclusions

Based on the observation on the development of information technology and the cases learned, three professional areas of knowledge and skills in KO are suggested for people who will be working in the field of KO in the future:

1. Developing domain knowledge with a broad understanding of the related information resources
2. Developing insights and in-depth analytical skills as well as good interpretative ability

3. Becoming familiar with machine languages, as well as platform and query languages to be able to operate Semantic Web KO

【Romanization of Chinese references is offered in the paper.】