

國立臺灣師範大學圖書資訊學研究所
博士學位論文

指導教授：柯皓仁 博士

以關鍵字使用分析探討社會標記者與
索引專家的文獻標引心智模式

A Study on Mental Models of Taggers and Professional Indexers
for Article Indexing Based on Analysis of Keyword Usage

研究生：陳亞寧 撰

中華民國一〇二年六月

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

謝 辭

在一個重要的人生轉折中，懷著一份感恩的心情就讀臺師大圖資所。於求學期間，感謝吳美美教授、陳昭珍教授、卜小蝶教授、謝建成教授、邱銘心教授與謝吉隆教授，傾囊相授各項專業理論與知識，及柯皓仁教授在論文的指導、國際期刊論文的投稿與處事圓融的方式等，讓我學習到很多，受用無窮。

感謝黃鴻珠教授與林呈潢教授的推薦信，才得以順利考上臺師大圖資所博士班。論文計畫書與學位論文口試時，謝謝黃鴻珠教授、陳雪華教授、陳昭珍教授、陳志銘教授及唐牧群教授等，以及所上每位老師與 Dr. Diane Sonnewald 提供的各項寶貴建議，論文才得以順利完成。苑菁助教耳提面命的叮嚀與奕翔助教在設備上的協助，也是在求學過程中，不可或缺的一環，才得以順利完成各項學業課程上的修習。在論文撰寫歷程中，也感謝秋霞、國勳、美聖、慧嫻等人的協助。口試時，恆毅的海報製作、智惠的餐點安排、佳香的會議紀錄，以及修課中，阿亮、寂絹、忠諺、于第、淇龍、智惠、佳香、忠勤、宗曄，及柯柯軍團的每位學弟妹，各項問題的討論及參與，彼此間的鼓勵與協助，是驅動我持續努力完成論文的原動力之一，在此也一併致謝。

求學期間，家中父母先後住院與開刀，如果沒有家姐的付出，我是不可能完成論文，順利畢業的，特別感謝她的辛勞。最後，也感謝默默協助我，卻一點也不在意忘記提起他/她們的每一位貴人。

陳亞寧 謹識

102.6.28

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

摘 要

隨著 Web 2.0 的廣泛應用，各式社會網路平台提供了社會標記的功能，讓社會標記者以自由形式的關鍵字組織各式資訊。在圖書資訊學界方面，資訊組織的權威控制與索引典控制係以一種控制詞彙的方式標引資訊的內容屬性。無論是社會標記者或是圖書資訊學界的索引專家皆將關鍵字視為一種觀念，且依其先備經驗與知識，經由關鍵字以表徵其對資訊內容所認知的理解與心智模式。現有的研究皆著重在以個別的關鍵字為研究對象，包括關鍵字的來源與使用情形，以及社會標記與控制詞彙間關鍵字的重複情形等，並未具體提出完整的文獻標引心智模式地圖，而是零散、片斷且沒有任何關聯關係的文獻標引心智模式。如果可以更瞭解社會標記者與索引專家的心智模式及其差異，即可選用更適當的關鍵字組織各項資訊資源，進而促成資源的發掘，導引使用者找到所需的資訊。

本研究旨在以關鍵字的使用分析，探討社會標記者與索引專家的文獻標引心智模式。在樣本資料方面，本研究選取 13 種圖資期刊中 1,489 篇文獻的社會標記與控制詞彙等關鍵字為樣本，包括 CiteULike 的 3,972 個社會標記（1,672 個不重複標記）與 LISA 的 6,708 個控制詞彙（1,338 個控制詞彙）。在研究方法則是計算關鍵字的使用情形外，還包括社會網路分析與頻繁樣式成長法（含頻繁樣式樹）等方法討論隱藏在關鍵字間的關聯關係結構與樣式。從關鍵字的使用情形、冪次定律分佈、社會標記與控制詞彙間的關鍵字比對、社會網路分析（包括：中心度、階層集叢、同等角色）及頻繁樣式成長等方面而言，結果顯示社會標記者與索引專家間的文獻標引心智模相似度並不高，主要的研究結果如下：

1. 社會標記者的文獻標引心智模式比索引專家更為多元化。
2. 社會標記者直接從文獻題名中選用關鍵字的傾向高於索引專家。
3. 社會標記與控制詞彙間의 相同程度不高且彼此互補。

4. 社會標記者不經常單獨使用內容群組的關鍵字類別及其所擁有的關鍵字，卻經常與題名主題群組的關鍵字類別及其所擁有的關鍵字一起搭配使用。索引專家雖然不常單獨使用其他群組的關鍵字類別及其所擁有的關鍵字作為文獻標引之用，卻常與題名、主題與內容等群組及其所擁有的關鍵字一起搭配使用。
5. 社會標記者傾向交替使用較多組的關鍵字類別及其擁有的關鍵字，而且每組幾乎是由兩種不同的關鍵字類別所組成。然而，索引專家則是傾向交替較少組的關鍵字類別及其擁有的關鍵字，且每組幾乎是由兩種以上不同的關鍵字類別所組成（即 2、3 與 7 種）。
6. 社會標記者傾向使用較少的 FP-tree 規則且較少的關鍵字類別進行文獻標引，而索引專家則是傾向使用較多的 FP-tree 規則且較多的關鍵字類別，組織各式資訊資源。

基於前述的研究結果，本研究貢獻主要有五項：發展文獻標引的心智模式、分析文獻標引心智模式地圖及其結構與樣式、分析關鍵字的使用情形及其共同出現的關聯關係特質、從心智模式解釋社會標記與控制詞彙兩者互補現象的原因，及擴展標記類別模式（tag category model）的可行性驗證與應用解釋。研究結果亦可進一步應用在資訊系統的設計，包括關鍵字的推薦、使用者界面的設計及瀏覽分類架構的建立與運用。

關鍵字：社會標記、控制詞彙、標引、心智模式、社會網路分析、頻繁項目樣式
成長

Abstract

With the wide application of Web 2.0, various social networking platforms allow taggers to use uncontrolled, free keywords (i.e., social tags) to organize information. In library and information science, professional indexers are guided by the principles of authority control and thesaurus control to organize information with controlled vocabularies. Both social taggers and professional indexers regard keywords as concepts that represent their cognitions and mental models of information content, according to their prior experience and knowledge. Existing studies have focused on examining the sources and usage of individual keywords, and comparing the similarity between tags and controlled vocabularies. However, the results of such studies only reflect scattered debris rather than a whole picture of the mental models used by social taggers and professional indexers for article indexing. A better understanding of the mental models of taggers and professional indexers and their usage gap may inspire better selection of appropriate keywords for organizing information, facilitating resource discovery, and guiding users to find the right information.

This study explores the mental models used by taggers and professional indexers to designate keywords for article indexing. Using a dataset of 3,972 CiteULike tags and 6,708 Library and Information Science Abstracts (LISA) descriptors from 1,489 scholarly articles in 13 library and information science journals, this study attempts to analyze the keyword usage of taggers and professional indexers to capture and build up their mental models for article indexing, and generalize their structures and patterns. To achieve this end, in this study social network analysis and frequent-pattern growth methods were employed. When measured with respect to terms used, power law distribution, a comparison of terms used as tags and descriptors, social network analysis (including centrality, overall structure and role equivalence) and frequent-pattern growth analysis (including frequent-pattern tree), little similarity was found between the mental models of taggers and professional indexers in article indexing.

The results of this study are summarized as follows:

- Taggers' mental models for article indexing are more diverse than those of professional indexers.
- Social taggers have a higher preference than professional indexers to select terms for article indexing from title keywords.
- There is little similarity between social tags and controlled vocabularies and they complement each other.
- Keywords in content-related categories were not used independently by

social taggers, but they were often used with those from topic-related categories. On the other hand, keywords of other-related categories were often co-used with those of title-, topic- or content-related categories by professional indexers.

- Social taggers may prefer to assign co-occurring keywords with more sets of fewer facets' viewpoints (almost always two-facets); however, professional indexers may be inclined to offer keywords with fewer sets of more facets' viewpoints (i.e., two-, three- and seven-facets).
- Social taggers may be inclined to assign keywords with fewer path-based rules comprising fewer keyword categories. Professional indexers may tend to offer keywords with more path-based rules comprising more keyword categories.

According to the research results mentioned above, the key contributions of this study are as follows:

- Development of a generic model of mental models of social taggers and professional indexers for article indexing.
- Analysis of the structures and patterns embedded in maps of mental models of social taggers and professional indexers in article indexing.
- Analysis of the characteristics of keyword usage and co-occurring keywords' associations.
- Presentation of a theoretical basis to explain the reason why social tags complement controlled vocabularies.
- Extension of the tag category model by feasibility examination and explanation.

Furthermore, the results of this study also inform the design of information systems, including term recommendations and user interfaces for indexing, as well as frequent-pattern based classification trees for browsing and navigation.

Keywords: social tags, controlled vocabularies, indexing, mental models, social network analysis, frequent pattern growth

目次

摘要.....	i
目次.....	v
表次.....	vii
圖次.....	viii
第一章 緒論.....	1
第一節 研究背景與動機.....	1
第二節 研究目的與問題.....	4
第三節 研究的重要性.....	4
第四節 研究範圍與限制.....	5
第五節 名詞解釋.....	6
第二章 文獻探討.....	9
第一節 理論框架：心智模式、系統印象與資訊組織.....	9
第二節 關鍵字的來源及其類型.....	16
第三節 社會標記與控制詞彙間的關聯程度.....	17
第四節 關鍵字的類別及其使用情形.....	20
第五節 關鍵字類別的關聯關係結構與樣式.....	27
第六節 文獻標引心智模式的理論模式.....	28
第三章 研究方法.....	31
第一節 研究工作.....	31
第二節 研究對象.....	34
第三節 資料收集.....	35
第四節 關鍵字的類別及入類.....	37
第五節 關鍵字意義的解讀與分析.....	42
第六節 關聯關係的結構與樣式分析.....	44

第七節 心智模式相似性的評估方式.....	57
第四章 研究結果與討論.....	59
第一節 關鍵字及其類別的使用分析.....	60
第二節 社會標記與控制詞彙之間的關鍵字比對.....	68
第三節 SNA 中心度分析.....	71
第四節 SNA 階層集叢分析.....	80
第五節 SNA 同等角色分析.....	84
第六節 FP-tree 分析.....	89
第七節 FP-growth 分析.....	99
第八節 使用關鍵字的心智模式.....	103
第九節 資訊系統設計的意涵.....	105
第五章 結論與建議.....	109
第一節 結論.....	109
第二節 研究貢獻.....	114
第三節 未來研究建議.....	119
參考文獻.....	123
附錄.....	133
附錄一 FP-tree 路徑規則中的各階層社會標記關鍵字類別.....	133
附錄二 FP-tree 路徑規則中的各階層控制詞彙關鍵字類別.....	139
附錄三 以社會標記群組組成的路徑式規則階層及其關鍵字類別與規則.....	147
附錄四 以控制詞彙群組組成的路徑式規則階層及其關鍵字類別與規則.....	151

表次

表 2.1	文獻題名分別與社會標記、控制詞彙間的關鍵字比對情形.....	17
表 2.2	社會標記與控制詞彙重複比例的文獻分析.....	20
表 2.3	TCM 項下的各子類別.....	24
表 2.4	最常用的社會標記類別.....	27
表 2.5	社會標記類別的類型.....	27
表 3.1	研究樣本一覽.....	36
表 3.2	關鍵字類別及其與各文獻間的關聯關係：第一與二項研究工作.....	39
表 3.3	關鍵字類別及其與各文獻間的關聯關係：第四項研究工作.....	42
表 3.4	文獻題名與社會標記間關鍵字比對的入類情形.....	45
表 3.5	文獻題名與控制詞彙間關鍵字比對的入類情形.....	47
表 3.6	社會標記與控制詞彙間關鍵字比對的入類情形.....	50
表 3.7	FP-tree 項目集.....	56
表 4.1	RQ 和第四章研究結果與討論各節內容間的相互關係.....	60
表 4.2	社會標記關鍵字類別的使用數量與百分比.....	61
表 4.3	控制詞彙關鍵字類別的使用數量與百分比.....	62
表 4.4	社會標記與控制詞彙間關鍵字的比對情形.....	69
表 4.5	關鍵字類別三項中心度的排名名次與係數.....	77
表 4.6	共同出現的社會標記關鍵字類別及其次數.....	83
表 4.7	共同出現的控制詞彙關鍵字類別及其次數.....	84
表 4.8	同等角色的社會標記關鍵字類別及其次數.....	87
表 4.9	同等角色的控制詞彙關鍵字類別及其次數.....	87
表 4.10	社會標記者與專家間共同的路徑式規則群組及其實例.....	98
表 4.11	社會標記的 FP-growth 規則組合及其支持度與信賴度.....	99
表 4.12	控制詞彙的 FP-growth 規則組合及其支持度與信賴度.....	100

圖 次

圖 2.1	心智模式系統印象與語意落差間的關係.....	12
圖 2.2	文獻標引的心智模式及其推導的研究問題.....	30
圖 3.1	研究組成.....	32
圖 3.2	研究程序.....	32
圖 3.3	研究問題與樣本間相互關係及其分析方式.....	33
圖 3.4	星狀式社會網路.....	53
圖 3.5	環狀式社會網路.....	53
圖 3.6	FP-tree 範例圖.....	56
圖 4.1	社會標記的關鍵字使用分佈情形.....	64
圖 4.2	控制詞彙的關鍵字使用分佈情形.....	64
圖 4.3	社會標記的關鍵字類別使用分佈情形.....	65
圖 4.4	控制詞彙的關鍵字類別使用分佈情形.....	66
圖 4.5	社會標記出現在控制詞彙的類別分佈情形.....	69
圖 4.6	控制詞彙出現在社會標記的類別分佈情形.....	70
圖 4.7	社會標記關鍵字類別的三項中心度.....	75
圖 4.8	控制詞彙關鍵字類別的三項中心度.....	76
圖 4.9	社會標記關鍵字類別的 SNA 階層集叢多維尺度圖.....	81
圖 4.10	控制詞彙關鍵字類別的 SNA 階層集叢多維尺度圖.....	82
圖 4.11	社會標記關鍵字類別的 SNA 同等角色多維尺度圖.....	85
圖 4.12	控制詞彙關鍵字類別的 SNA 同等角色多維尺度圖.....	86
圖 4.13	社會標記關鍵字類別的 FP-tree 使用分析.....	91
圖 4.14	控制詞彙關鍵字類別的 FP-tree 使用分析.....	92
圖 4.15	社會標記者使用關鍵字的路徑式決策樹.....	95
圖 4.16	專家使用關鍵字的路徑式決策樹.....	96
圖 4.17	社會標記者使用關鍵字的文獻標引心智模式.....	104
圖 4.18	專家使用關鍵字的文獻標引心智模式.....	105

第一章 緒論

資訊組織 (Information Organization, IO) 的主要目的在於有效整理各式資訊資源，及組織蘊含在資訊中的各項人類知識，進而促成最佳化的資訊檢索 (Information Retrieval, IR)，引導使用者查找到所需的各項資訊及其知識。隨著資訊暨通訊技術 (Information and Communication Technologies, ICTs) 的日新月異，圖書資訊學界 (以下簡稱「圖資界」) 除了在資訊組織理論與實務等方面適時調整外，也產生了一些改變。尤其在個人資訊管理 (Personal Information Management, PIM) 的趨勢衝擊下，資訊組織除了從學術與實務等專業領域進行探討外，也必須充分瞭解使用者在個人資訊管理的行為特質與樣式 (pattern) 後，才能發展出貼近使用者需求的資訊組織理論與實務。有鑑於此，本研究旨在於探索一般使用者與索引專家 (以下簡稱專家) 以關鍵字進行文獻標引的行為，進而探究兩者間文獻標引的心智模式及其相似性。本章內容共分為研究背景與動機、研究目的與問題、研究的重要性、研究範圍與限制，以及名詞解釋等節，說明本研究的目的及其內容梗概。

第一節 研究背景與動機

資訊組織的發展與應用已累積許多的學術理論與實務經驗，以有效組織各項人類知識及其載體 (如：圖書與期刊)。從早期的館藏清單 (inventory list)，乃至於卡片目錄 (card catalog) 與書本式目錄 (book catalog)，以及近年來的線上公用目錄 (Online Public Access Catalog, OPAC)、全球資訊網型公共目錄 (Web OPAC)、社會化公用目錄 (social OPAC) 等，皆是因應不同時代潮流的需求，所產生的資訊組織系統或工具，藉以引導使用者找到所需的資訊資源。背後支撐這些資訊組織系統的理論與實務，則包括了記述編目 (descriptive cataloging) 與編目規則 (cataloging rules)、主題編目 (subject cataloging) 與分類表 (classification) 及主題標目 (subject headings)、權威控制 (authority control) 與權威規範檔

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

(authority file)，以及索引典控制 (thesaurus control) 與索引典等。經由這些資訊組織的學術理論與實務工具，各式資訊物件 (information object) 得以不同的方式與粒度 (granularity) 有效組織，進而透過呈現、瀏覽與檢索的機制，導引使用者發現、識別、選擇與獲取所需的資訊資源。在資訊組織領域中，所發展出相關的理論與規範，係一種由上而下的專家型資訊組織，為能瞭解此種資訊組織的有效性，已有相關研究從使用者與圖書館館員所使用的檢索關鍵字來檢視資訊組織的有效性。例如，Carlyle (1989) 以美國加州大學洛杉磯分校 (University of California, Los Angeles, UCLA) 171 位線上資訊系統使用者的查詢異動紀錄為研究樣本，比對美國國會主題標目 (Library of Congress Subject Heading, LCSH) 與使用者查詢內容間的異同，結果顯示近 74% 的查詢內容可以比對到單一的 LCSH 標目。另外，Voorbij (1998) 以 475 本人文與社會科學專書的書目紀錄為樣本資料，由學科館員 (subject librarian) 以專書書目紀錄的題名關鍵字與主題控制詞彙分別查詢，結果發現主題控制詞彙可以對前述專著增加約 37% 至 49% 的檢索回現率 (recall)。

另一方面，隨著 Google 等網路搜尋引擎的興起與影響，已改變使用者查找資訊資源的方式，使用者習慣以單一關鍵字 (keyword) 或兩個以上關鍵字組成的關鍵片語 (key phrase) 代表使用者的資訊需求與知識概念，查找各式資訊資源。除此之外，隨著 Web 2.0 的推波助瀾，各式社會網路平台 (如：Del.icio.us、Flickr、Last.fm、YouTube 等) 提供所謂的社會標記 (social tags) 或俗民分類 (folksonomy) 方式，進行各式資訊資源的個人資訊管理。此種社會標記應用於資訊組織的方式被稱為「民主式索引」(democratic indexing)(Rafferty & Hilderley, 2007)，係指使用者在組織這些個人化資訊資源時，並不依循資訊組織的原則與規範，而是由下而上自然形成一種集體式的資訊組織及民主現象 (Vander Wal, 2007)。儘管社會標記此種方式有別於既有的資訊組織方式，卻提供探索使用者資訊組織行為的機會，也可以從另外一種觀點重新檢視資訊組織既有學術理論與

實務規範的適用性。

就資訊組織而言，關鍵字常被視為一種基本的資訊組織方式，經由以字詞代表主題的概念或知識，藉以組織各式的資訊資源，以因應個人、機構（如：圖書館）、學科社群（如：圖資界）等不同層次的需求。在資訊組織方面，關鍵字或詞彙常經由所謂的「詞彙控制」(vocabulary control)方式，達成概念之表徵 (National Information Standard Organization [NISO], 2005)，同時關鍵字也被視為一種資訊組織的書目語言 (bibliographic language)，主要用來描述資訊資源的主題、實體與製作等屬性資訊 (subject, physical and production attributes) (Svenonius, 2000)。

在社會標記方面，Smith (2008) 指出：「在標記與組織個人資訊資源的過程中，社會標記者 (social taggers) 除了將社會標記視為一種關鍵字外，也視為一種概念。」在研究學生的資料庫查詢行為中，Holman (2011) 指出：「在資料庫查找文獻時，使用者將關鍵字視為一種概念 (concept)。」由此可知，關鍵字皆被一般使用者與專家視為對資訊資源及其內容所認知理解的一種知識表徵 (knowledge representation)。在以全球資訊網為研究對象之資訊檢索系統研究中，Zhang (2008b) 發現學生具有薄弱的資訊組織心智模式 (mental models)，並建議深入瞭解使用者的資訊組織心智模式，將有助於全球資訊網資訊資源的組織。唯有資訊組織專家與一般使用者擁有共享的經驗時 (Inskip, MacFarlane, & Rafferty, 2008)，資訊組織才能成功地達成資源發掘 (resource discovery)，導引使用者找到所需的資訊資源。反之，所謂的「語意落差」(semantic gap) (Maron, 1977; Rafferty & Hilderley, 2007) 將會存在於一般使用者與資訊組織專家之間，導致使用者無法找到所需的資訊資源。因此，如果資訊組織專家能夠清楚瞭解一般使用者的文獻標引心智模式及其行為與認知，將會消弭上述語意落差的鴻溝，且有助於資訊資源的發掘。

第二節 研究目的與問題

本研究以心智模式為理論基礎，探討一般使用者與專家間資訊組織的文獻標引方式與行為，並以關鍵字的使用分析為依據，進而建構「文獻標引心智模式」的理論模式。換言之，本研究主要目的為探討一般使用者與專家間使用關鍵字之文獻標引心智模式的相似性，及其文獻標引的行為與知識結構，期能對資訊組織系統的功能設計提出建議，茲將研究目的分述如下：

1. 以標引期刊文獻的關鍵字為基礎，建構一般使用者與專家所擁有的文獻標引心智模式。
2. 檢視一般使用者與專家之文獻標引心智模式是否相似。

為達成上述目的，除了瞭解一般使用者所運用關鍵字的文獻標引心智模式外，也必須深入探討專家使用關鍵字的文獻標引心智模式。無論是一般使用者或是專家，其文獻標引的心智模式將轉換為具體的行為，包括使用的關鍵字及關鍵字樣式。因而，本研究的主要研究問題（Research Question, RQ）如下：

RQ1：以標引文獻所使用的關鍵字而言，一般使用者與專家間的文獻標引心智模式為何？

RQ2：一般使用者與專家間的文獻標引心智模式是否相似，而且兩者是否存在不同的系統印象？

第三節 研究的重要性

以往有關心智模式的研究多數集中在資訊檢索（Zhang, 2008a, 2008b; Zhang & Chignell, 2001）、資訊尋求（Marchionini, 1989a, 1989b; Sutton, 1994）、參考服務（Michell & Dewdney, 1998）、SPSSX 軟體使用（Staggers & Norcio, 1993）、圖書館資源（Makri, Blandford, Gow, Rimmer, Warwick, & Buchanan, 2007）等方面的認知研究，而有關資訊組織方面的研究也僅限於 FRBR 方面的認知探討

(Pisanki & Žumer, 2011a, 2011b)，或是有關關鍵字的來源及兩兩關鍵字相互比對與其相同比例的研究，然而在資訊組織心智模式內容及其地圖方面的研究仍付諸闕如。截至目前為止，已有 Zhang (2008b) 完成使用者的資訊檢索心智模式地圖，但在資訊組織方面未有任何研究以共同出現的關鍵字關聯關係為基礎發展心智模式及其地圖。因而，本研究具有以下的主要重要性或貢獻：

一、建立文獻標引的心智模式地圖

本研究將導入頻繁樣式樹 (Frequent Pattern tree, FP-tree) 方法，分析關鍵字的使用，以發展出路徑式的規則地圖，藉以擷取一般使用者與專家的認知，進而建立其擁有之文獻標引的心智模式地圖及其內容。

二、分析文獻標引心智模式的結構與樣式

除了 FP-tree 方法外，本研究將同時採用社會網路分析 (Social Network Analysis, SNA) 與頻繁樣式成長 (Frequent Pattern growth, FP-growth) 等方法，分析一般使用者與專家所使用的關鍵字，進而歸納其擁有之文獻標引心智模式的結構與樣式。

三、比較一般使用者與專家之文獻標引心智模式的異同

本研究將以標引文獻所使用的關鍵字為分析基礎，建立和描繪一般使用者與專家所擁有的文獻標引心智模式內容外，同時進一步經由冪次定律 (power law) 分佈現象、SNA、FP-tree 與 FP-growth 等分析，從量化與質化兩種觀點，綜合比較兩者間文獻標引心智模式的異同。

第四節 研究範圍與限制

鑑於既有的多數研究以 Flickr、Del.icio.us 及 LibraryThing 為研究對象，其所

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

涵蓋的資料類型包括影像(image)、網路資源的書籤(bookmark)與圖書(book),藉以探討有關社會標記在文獻標引方面的相關資訊組織議題。然而,較少研究以CiteULike及其期刊文獻與社會標記為研究對象,探討文獻標引心智模式方面的研究議題。為能探索一般使用者與專家間文獻標引的方式及其異同,本研究選擇圖資界的國際期刊為研究範圍,所選擇之期刊文獻必須同時具備有社會標記與控制詞彙,其中前者基於CiteULike社會標記者提供的社會標記,後者則根據圖書資訊學文獻摘要資料庫(Library and Information Science Abstract, LISA)專家提供的控制詞彙,本研究將以這些社會標記與控制詞彙作為研究對象。除此之外,還有兩項研究限制,說明如下:

1. 特別選擇圖資學的主要原因在於圖資學是一個跨領域的學科,具備多重學科的性質,且注重資訊的使用者等前提下,因而就資訊組織的研究而言更形重要且具代表性,除了可以分析一般的文獻標引使用情形外,還可以分析具備圖資學背景之使用者與專家的文獻標引特性及其心智模式。
2. 再者,在本研究中,所謂的社會標記者係指實際針對圖資學期刊文獻提供社會標記的個人,而不是單純查詢社會標記,以發現其所需資訊物件的使用者;而專家則是指受過權威控制、索引典控制等專業訓練,以及熟悉相關主題分析方面的資訊組織理論、規範、權威檔與索引典等控制詞彙及其系統,且實際從事期刊文獻標引工作的專業人員。

第五節 名詞解釋

在本研究中,引用許多名詞,為便利名詞的統一使用,將相關名詞予以一致的定義與解釋。茲分述如下。

1. 社會標記(social tag):社會標記一詞最早出現於社會性書籤網站的應用,社會性書籤網站提供個人對其所收集的網頁或網站給予標籤或關鍵

字進行分類，便於日後搜尋、取用、分享，此種網路資源的資訊組織方式即所謂的社會標記（卜小蝶、張淇龍，2009）。社會標記是由一般使用者提供的關鍵字，藉以描述資訊資源的主題、位置、用途（intended use）、提示（reminder）或其他特性。有些社會標記屬於描述性質，有的則是意見表達（expressive）（Smith, 2008）。

2. 俗民分類（folksonomy）：係由 Thomas Vander Wal 將「俗民」（folks）和「分類」（taxonomy）兩個詞彙組合而成的新名詞，藉以描述由下而上所形成的草根性分類系統，作為建立、指派與管理社會標記，達成資訊內容的註解與分類（http://www.abc-clio.com/ODLIS/odlis_f.aspx#folksonomy）。依據 Vander Wal（2007）提出的定義：「俗民分類係由個人針對資訊或物件進行標記，作為個人檢索之用。」
3. 控制詞彙（controlled vocabulary）：係指已組配好的詞彙（陳和琴、張慧銖、江綉瑛與陳昭珍，2003），或是一組明確列舉的詞彙，且經由一個權威機構所控制與提供（NISO, 2005），而這些經過定義與規範的詞彙，可作為編目人員或索引人員標誌作品內容之用（http://www.abc-clio.com/ODLIS/odlis_c.aspx#controlled）。
4. 關鍵字（keyword）：出現在文獻內容之中，並可作為文獻標引與檢索的字（NISO, 2005）。另外，關鍵字可能出現在題名、主題標目或描述語、內容註、摘要、書目資料庫紀錄內容的重要字或片語，作為檢索紀錄之用（http://www.abc-clio.com/ODLIS/odlis_jk.aspx#keyword）。
5. 關鍵字類別（category）：所謂的關鍵字類別係指一組關鍵字具有語意或統計方面的關聯（semantically or statistically associated），例如：屬種關係（genus/species）、親子關係（parent/child）或全部與局部關係等（part/whole）（NISO, 2005）。本研究所指的關鍵字類別，係指具備某一種相同屬性的不同關鍵字，且可依此一屬性將相關關鍵字予以聚合。
6. 心智模式（mental models）：係指嵌在個人內在心理的一種工作模式，

藉以促成與不同人、技術及外在環境的互動 (Michell & Dewdney, 1998), 或是系統使用者的一種心理表徵 (Staggers & Norcio, 1993)。在本研究中, 將文獻標引的心智模式定義為: 「社會標記者與專家在面對資訊物件與文獻標引系統時, 會依其先備經驗與知識, 提供關鍵字及其內隱的知識結構和關聯關係, 分別建構社會標記者與專家所擁有的文獻標引心智模式, 以表徵其對資訊的一種認知。」

7. 心智模式一致性 (consistency of mental models): 係指不同獨立個體具有相似的知識結構或基模 (schemas), 及其意義之認知、瞭解與經驗的共享, 且以集體方式呈現, 包括事件的詮釋及其行為。
8. 系統印象 (system image): 係指設備裝置呈現給使用者的印象 (Staggers & Norcio, 1993), 或是設計者依其設計模式所完成的系統, 及使用者可以互動到的系統面向, 包括使用者界面、操作手冊、線上求助與錯誤訊息等 (Makri et al., 2007)。
9. 標引 (indexing): 由人或機器自控制詞彙中選出詞彙或主題標目, 作為表徵物件概念或屬性的一種方法 (NISO, 2005)。針對單一出版品編輯索引的過程, 通常由訓練有素的專業索引人員在閱讀或掃瞄內容後, 選擇適當的標目 (如, 名稱、地名或主題) 以促進檢索 (http://www.abc-clio.com/ODLIS/odlis_i.aspx#indexing)。

第二章 文獻探討

本章內容係以文獻分析為基礎，探討一般使用者運用社會標記及專家使用控制詞彙進行文獻標引的研究現況，同時闡述與建構本研究所需的理論模式。本章分為五節，逐一探討相關文獻與分析相關議題，進而推導研究問題。第一節，就心智模式、系統印象與文獻標引進行論述分析，以建構本研究的理論基礎；第二節，分析關鍵字的來源及其類型，包括文獻題名與社會標記間，以及文獻題名與控制詞彙間的關鍵字關聯性；第三節，探討社會標記與控制詞彙之間的關鍵字關聯關係；第四節，分析關鍵字類別及其使用情形，包括社會標記與控制詞彙的關鍵字類別等兩方面；第五節，探討社會標記與控制詞彙之關鍵字類別的關聯關係結構與樣式；第六節，依前述各節文獻探討的內容重點為基礎，提出本研究之文獻標引心智模式的理論模式。

第一節 理論框架：心智模式、系統印象與資訊組織

本節內容區分為心智模式的定義、一致性與相似性、系統印象、建構與相似性評估，及資訊組織心智模式等重點項目進行論述。

一、心智模式的定義

心智模式一詞最早可追溯至 1943 年 K. Craik 出版「大自然的解釋」(The Nature of Explanation) 一書，提出心智模式是以文字、數據或其他符號來表徵人們對外在世界物件與現象的內在認知 (Zhang, 2008b)。從現有的文獻而言，心智模式的定義存在著許多不同的界說。「心智模式」(mental models) 專書編輯者 Gentner and Stevens (1983) 指出心智模式係融合了認知心理學 (cognitive psychology) 與人工智慧 (artificial intelligence) 兩大學科。在認知心理學方面，Johnson-Laird (1989) 認為心智模式是表徵知識與形塑推理形式的一種方法。在

電腦科學方面，特別是在資訊系統設計領域，Norman (1983) 認為：「心智模式內嵌在人的腦中，藉以引導人對外在世界的行為與反應。心智模式也被視為有關實際系統中的一種信念反射。」在與現實世界相關物件的互動中，人會依據其先備經驗與知識，建立其本身擁有的心智模式 (Norman, 1983)。簡言之，所謂的心智模式係指人依其本身的內在認知與外在世界進行互動過程中所呈現的一種表徵。

心智模式存有不同的同義詞，包括：知識結構 (knowledge structure) (Brookes, 1980)、知識狀態 (knowledge state) (Beklin, Oddy, & Brooks, 1982)、認知結構 (cognitive structure) (Wang & Soergel, 1998) 與認知狀態 (cognitive state) (Ingwersen, 1996)、概念地圖 (concept maps) (Bilal & Wang, 2005; Carvalho, Hewett, & Canas, 2001; Freeman, 2004)、認知地圖 (cognitive maps) (Eden, 1992; Kearney & Kaplan, 1997; Peruch, Gaunet, Thinus-Blanc, & Loomis, 2000) 等，以描述使用者存取資訊檢索系統時的一種心理反射 (Cole, Lin, Leide, Large, & Beheshti, 2007); Winn (2001) 則是將心智模式與基模 (schemas) 視為同義。在圖資界，心智模式可被視為一種知識結構，而管理科學方面亦有類似的觀點與見解 (Mohammed, Klimoski, & Rentsch, 2000)。在圖資界的範疇中，所謂的心智模式係指：「人對資訊物件、資訊系統，及相關處理的其他資訊等之心智表徵 (mental representation)」(Zhang, 2008b)。心智模式除了應用於資訊檢索方面外，也有應用在資訊組織方面，包括：入口網站 (Ahlstron & Allendoerfer, 2004) 與書目紀錄功能需求模式 (Functional Requirements for Bibliographic Records, FRBR) (Pisanki & Žumer, 2011a, 2011b) 等。對資訊組織而言，Pisanki and Žumer (2011a) 認為：「心智模式是對外在世界的一種內在表徵。」

二、心智模式的系統印象、一致性與相似性

Norman 在界定心智模式時，定義心智模式係由設計者的心智模式概念、使

用者的心智模式、系統與系統印象等四元件所共同組成。在系統的發展階段，設計者依其心智模式概念發展系統，待系統完成設計後，連同使用者界面、操作手冊、線上求助與錯誤訊息等提供給使用者，進而促成使用者對系統產生具體的內在表徵。因而，當設計者與使用者對同一系統具有類似的內在表徵時，亦即設計者與使用者兩者間的心智模式趨於一致（consistent）時，即已具備相似的系統印象。Norman 也指出設計者與使用者間的心智模式可能是不同的（Norman, 1983），也就是當設計者與使用者間的心智模式存有差異或是不一致時，即已具備不同的系統印象，Blackwell（1996）與 Guerra, Sanz, Díaz, and Aedo（2007）則是將這種不同系統印象的情形視為一種語意落差的現象。

然而，何謂心智模式的一致性？在 Mohammed, Ferzandi and Hamilton（2010）研究團隊心智模式（Team Mental Model, TMM）時，指出：「所謂共享性（sharedness）的 TMM 係指團隊成員間的心智模式是一致的，或是聚合的（converge），而不是完全相同（identical）的。」Rentsch, Small and Hanges（2008）則是認為：「所謂的共享性是一種認知的相似性（cognitive similarity），意指團隊成員間具有相似的意義或瞭解，且用來詮釋內外事件之用，包括：行為與想法等。」在研究團隊基模時，Rentsch and Klimoski（2001）則認為：「所謂團隊成員的基模相似性（team member schema similarity）係指團隊成員間具有相似或相容的知識結構（similar or compatible knowledge structure），以組織與瞭解團隊的相關現象。」依據 Mohammed, Ferzandi and Hamilton（2010）的看法：「雖然 TMM 係由不同獨立個體的心智模式所組成，卻是一種集體式的現象。」Rouse, Cannon-Bowers and Salas（1992）則進一步說明 TMM 其中的一項功能在於團隊成員能以相似的方式進行資訊的描述與詮釋，也就是 Inskip, MacFarlane and Rafferty（2008）所提出的共享經驗。簡言之，由上述論述中，可以得知所謂心智模式的一致性係指不同獨立個體具有相似的知識結構或基模（亦即 Norman 定義的系統印象），以及其意義之認知、瞭解與經驗的共享，且以集體方式予以呈

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

現，包括事件的詮釋及其行為。反之，則意謂著不同個體間具備不同的知識結構及其意義的認知、瞭解、經驗與詮釋，進而產生所謂不同系統印象或語意落差的現象。就心智模式而言，專家一如模式中的設計者，而一般使用者一如模式中的使用者，即使針對同一文獻或資訊物件（亦即等同於圖資界學者研究的電腦系統或使用者界面），因兩者具備不同的文獻標引心智模式，也會有不同的系統印象，以及造成語意上的落差（請參見圖 2.1）。

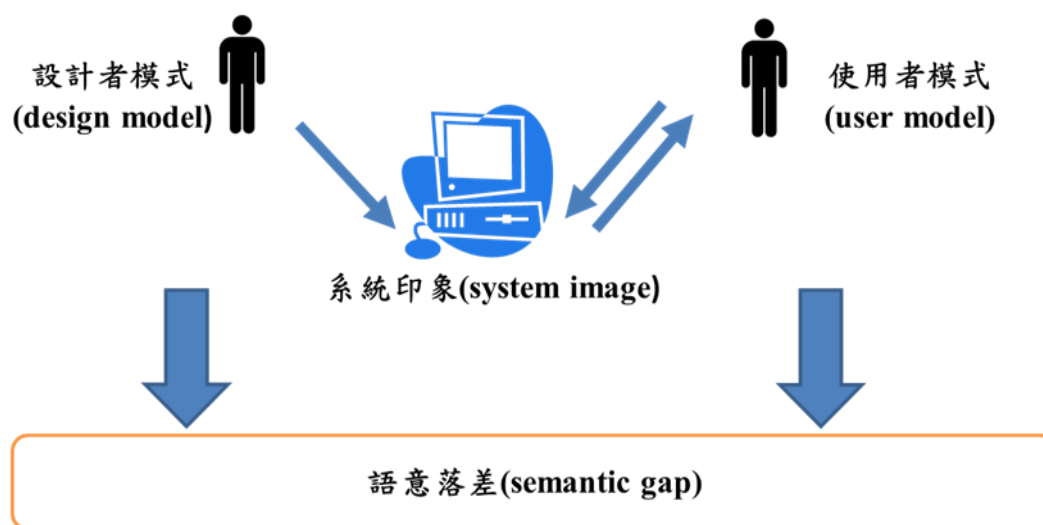


圖 2.1 心智模式系統印象與語意落差間的關係

三、圖資界有關心智模式相似性的研究

圖資界也有許多研究探討使用者的心智模式。在參考服務方面，Michell and Dewdney（1998）發現在參考服務的互動過程中，使用者與圖書館館員間的系統心智模式是有所差異的。在資料庫檢索方面，Holman（2010）指出學生並未具備與資料庫發展者相似的心智模式，因而檢索正確性並未如預期的理想。在資訊檢索系統方面，Zhang and Chignell（2001）發現不同使用者（即圖書館館員與學生）具有不同的心智模式，進而有不同的查詢效能或行為。在傳統圖書館與數位圖書館的研究方面，Makri et al.（2007）指出使用者對傳統圖書館與數位圖書館的資訊資源擁有不同的心智模式，且確切知道兩者間的差異。除此之外，在一項

資訊尋求的研究中，Cole et al. (2007) 更進一步指出：「必須將使用者資訊尋求與資訊組織系統間的心智模式落差予以消弭，如此方能將學生所使用的查詢詞彙與索引典控制詞彙予以串連在一起。」

在資訊組織方面，Quintarelli (2005) 探討使用者會依據本身的心智模式、詞彙及語言進行物件的社會標記。在一項比較社會標記與控制詞彙的使用研究中，Kipp (2006) 具體指出使用者傾向使用任務與時間 (task and time) 及一般與新興 (generalities and emergent) 等類別的關鍵字，專家則是傾向使用地理 (geographic) 與特定 (specifics, 如：使用對象、系統與研究方法等) 類別的關鍵字。另外，Inskip, MacFarlane and Rafferty (2008) 試從溝通模式 (communication model) 與符號學 (semiotics) 理論為基礎，探討音樂資料的資訊組織與檢索議題時，也指出作曲者、演奏者及編目者或聆聽者間的語意落差，包括能指 (signifier) 與所指 (signified) 間的不同。然而，在一項 FRBR 的使用者研究中，Pisanki and Žumer (2011a) 卻發現使用者具有類似的 FRBR 概念認知。

由此可知，無論是在參考服務互動過程、資訊檢索、資訊尋求與資訊資源等，乃至於資訊組織方面，多數研究認為一般使用者與圖資專家間存在有不一致的心智模式，亦即兩者間資訊組織心智模式的知識結構、語意與其詮釋，及行為等並不相似。除此之外，Zhang (2008b) 也建議深入探討使用者的資訊組織心智模式及其必要性。所以，必須深入探索與瞭解使用者的資訊組織心智模式後，方能消弭橫互在一般使用者與圖資專家間的系統印象或語意落差，以提供適當的關鍵字及其相關服務 (Rorissa, 2010; Zhang, 2008b)。依據上述文獻探討，本研究推導出下列的主要研究問題：

RQ1：以標引文獻所使用的關鍵字而言，一般使用者與專家間的文獻標引心智模式為何？

RQ2：一般使用者與專家間的文獻標引心智模式是否相似，而且兩者是否存有不同系統的印象？

四、心智模式理論的導入與應用方式

在研究設計方面，有關心智模式的導入方式亦有所不同。在一項圖書館參考服務的研究中，Michell and Dewdney (1998) 將參考服務館員視為心智模式中的設計者，而服務對象則視為使用者，藉以探討參考服務過程中，館員與讀者間對參考服務認知的異同。在一項有關 FRBR 的研究中，Pisanki and Žumer (2011a) 導入心智模式時，也是將館員視為系統設計者，而書目紀錄使用者視為使用者。簡言之，這類型的研究皆將心智模式視為一種研究的理論依據，同時將研究對象劃分為系統的設計者與使用者，藉以探討兩者間心智模式的相似性。

除此之外，有些研究在應用心智模式時，並未嚴格區分為系統的設計者與使用者。例如，在研究 SPSSX 軟體時，Staggers and Norcio (1993) 只將研究對象區分為資深與新手使用者等兩種類型。換言之，只就心智模式的使用者進行研究，探討不同經驗的使用者是否具備相似的心智模式。在一項資訊檢索系統的研究中，Zhang and Chignell (2001) 也採取類似的作法，將資訊檢索系統的使用者劃分為館員、研究生、大學生與高中生等四種類型，以探討不同類型使用者之資訊檢索系統的心智模式異同。

由上述討論可以得知，有關心智模式導入研究的方式約略可歸納為兩種：

1. 將研究對象劃分為系統的設計者與使用者，藉以探索兩種不同研究對象間的心智模式相似性。
2. 只導入心智模式中的使用者為研究對象，且進一步區分不同類型的使用者，以檢視特定系統之不同使用者間的心智模式異同。

在本研究中，則是將一般使用者與專家視為不同的使用者，藉以探索不同使用者間的心智模式異同。

五、心智模式內容的建構

由於心智模式是十分抽象化的，對許多研究而言，如何導引 (elicitation) 與與表徵 (representation) 心智模式的具體內容變成十分重要的研究工作。擷取、表徵與建置心智模式的方法很多，可以區分為質化與量化兩種類型。在質化方面，包括訪談 (interview)、畫圖 (drawing)、觀察 (observation) (Zhang, 2008b) 與卡片分類 (card sorting) (Pisanski & Žumer, 2010a) 等不同方法。在量化方面，有些研究經由問卷的評比量化資料，採取方格法 (repertory grid technique) (Zhang & Chignell, 2011) 建構心智模式。多數研究採取質化方式建構心智模式的內容，如前述的 Makri et al. (2007)、Pisanski and Žumer (2010a)，只有少數採取同時採取質量複合方式，以避免質化研究者的主觀判斷與影響，如 Zhang (2008a, 2008b)。

六、心智模式相似性的評估

在心智模式相似性的評估方面，約可區分為量化、質化與複合式等三種。在心智模式相似性的量化評估方面，係採取統計數據或是演算法方式量測心智模式的相似性。例如，Michell and Dewdney (1998) 係以五級李克特量表 (5-level Likert scale) 進行參考服務前後的問卷調查，藉以量測館員與服務對象間心智模式的差異程度。另外，Schaffernicht and Groesser (2011) 提出所謂的「距離比例法」 (distance ratio approach)，經由節點與節點間鏈結情形，計算心智模式的異同。至於心智模式相似性的質化評估則首先需要勾勒或描繪心智模式的圖形，再以心智模式圖的內容與結構為依據，評估兩兩心智模式間的相似性，包括 Pisanski and Žumer (2010a) 與 Toker (2012) 等研究個案均採質化評估進行。除了量化與質化方式外，也有些研究同時結合量化與質化的複合方式，檢驗心智模式的相似性，如 Zhang (2008a) 同時採取畫圖、問卷調查與訪談方式，藉以分析不同使用者之全球資訊網搜尋行為的心智模式相似性。本研究將採取複合方式評估文獻標引心智

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

模式的相似性，請參見本文第三章第七節之「心智模式相似性的評估方式」說明。

第二節 關鍵字的來源及其類型

有些研究已證實社會標記與控制詞彙間具有互補的現象 (Kipp, 2006, 2011a, 2011b; Lu, Park, & Hu, 2010; Rolla, 2009; Thomas, Caudle, & Schmitz, 2009; Yi & Chan, 2009)，然而有些學者更進一步探究社會標記與文獻題名間關鍵字的關聯程度。以 Flickr 為對象的研究中，Heymann, Koutrika and Garcia-Molina (2008) 發現 16% 社會標記出現在網頁題名的關鍵字。在 Bischoff, Firan, Nejdil and Paiu (2008) 的研究中發現，標註在 Delicio.us 之 2,507,688 個社會書籤的 323,294 個社會標記中，約有 44.85% 會標記與網頁內容的關鍵字相同，在 Last.fm 中，標註在 317,058 個音樂檔的 21,177 個社會標記，則約有 1.54% 社會標記與歌詞內容的關鍵字相同。此外，Hotho, Jächke, Schmitz and Stumme (2006) 同樣以 Delicio.us 為研究對象，也發現逾 50% 以上的社會標記與網頁內容的關鍵字相符。以 Connotea 為研究對象，Heckner, Mühlbacher and Wolff (2008) 發現有 26.5% 社會標記與期刊文獻題名的關鍵字相同，而 Lin, Beaudoin, Bui and Desai (2006) 以醫學期刊文獻為對象的研究中，則發現只有 19% 的社會標記取自於文獻題名的關鍵字。

在書目紀錄方面，Frost (1989) 以 2,401 筆書目紀錄的 4,264 個主題標目及所屬文獻題名為對象，分析結果顯示 73% 的文獻題名關鍵字至少可以比對到主題標目或複分標目的一部分。以醫學碩士論文為研究對象，Ansari (2005) 則發現 70% 以上的控制詞彙可以比對到學位論文題名的關鍵字。儘管如此，有關社會標記與文獻題名間的關聯程度仍存有差異，並未達成一致的共識，而文獻題名與控制詞彙間的關鍵字關聯程度也不盡相同 (請參照表 2.1)。依據上述文獻探討，本研究推導出下列研究問題：

RQ2.1：文獻題名分別與社會標記、控制詞彙間的關鍵字關聯程度為何？

表 2.1 文獻題名分別與社會標記、控制詞彙間的關鍵字比對情形

文獻	比對對象	關鍵字相同的百分比
Ansari, 2005	學位論文題名及其控制詞彙	70% 以上
Bischoff et al., 2008	Del.icio.us 社會標記與網頁內容	44.85%
	Last.fm 社會標記與歌詞內容	1.54%
Frost, 1989	書目紀錄的主題標目與其所屬文獻題名	73%
Heckner et al., 2008	Connnetta 社會標記與資訊科技期刊文獻題名	26.5%
Heymann, Koutrika, & Garcia-Molina., 2008	Flickr 社會標記與網頁題名	16%
Hotho et al., 2006	Del.icio.us 社會標記與網頁內容	50% 以上
Lin et al., 2006	Connnetta 社會標記與醫學期刊文獻題名	19%

第三節 社會標記與控制詞彙間的關聯程度

除了文獻題名與社會標記間的關鍵字關聯外，也有些學者將研究重點放在社會標記與控制詞彙間的關鍵字關聯關係。在圖書方面，Iyer and Bungo (2011) 以 40 本醫學圖書為研究對象，以分別來自 LibraryThing 與 OCLC Connexion 蒐集的 2,074 個社會標記與 162 個 LCSH 主題標目進行研究，該研究融合 Carlyle (1989)、Strader (2009) 與 Voorbij (1998) 的比對原則，將比對類別區分為完全比對 (full)、部分比對 (partial) 與無法比對 (none) 三種，進行社會標記與 LCSH 間的關鍵字比對。結果發現只有 1.21% 社會標記可以完全比對到 LCSH，5.52% 社會標記可以部分比對到 LCSH，以及 93.27% 社會標記無法比對到 LCSH。以 10 本流通最高的圖書為研究對象，Thomas et al.(2009) 以 Voorbij(1998) 與 Kipp(2006) 提出的比對類別為基礎，一共提出相同 (same)、同義 (synonym)、自然語言形式的同義 (natural language synonym)、廣義 (broader term)、狹義 (narrower term)、相關 (related term)、候選的主題標目 (LCSH not assigned)、

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

與 LCSH 不同形式的相關 (related)、不相關 (not Related) 九種比對類別，研究發現只有 6% 社會標記具有與 LCSH 相同的關鍵字，8% 社會標記出現在非 LCSH 主要款目 (used for subject headings) 的關鍵字，而同義、廣義、狹義與相關等類別的比對百分比分別為 8%、0.3%、0.3% 與 0.4%。換言之，只有約 14% 社會標記具有與 LCSH 相同的關鍵字。在以 8,652 本書的 176,105 筆 LibraryThing 社會標記與 7,628 個 LCSH 主題標目的研究中，Lu et al. (2010) 發現只有 2.2% 社會標記具有與 LCSH 相同的關鍵字，50.1% LCSH 主題標目中的關鍵字被選用為社會標記，而且約有 85% 的圖書書目紀錄中，同時具有一個以上相同關鍵字的 LCSH 與社會標記。

在期刊文獻方面，以「Journal of Molecular Biology」與「Proteins」2 種生物醫學學術期刊的 1,083 篇文獻為對象，並以 Kipp (2006) 融合 Voorbij (1998) 的分類類別為依據進行簡化與調整，Kipp (2011a) 提出了相同 (same)、同義 (synonym)、狹義或廣義 (narrower or broader term)、相關 (related term) 及未出現在索引典的相關 (related but not in thesaurus) 五大類別進行比對與分類。Kipp (2011a) 比對 3,788 個 CiteULike 網站的社會標記與 12,473 個 PubMed 網站的醫學主題標目 (Medical Subject Heading, MeSH)，發現 16% 期刊文獻所擁有的社會標記與 MeSH 具有相同關鍵字。延續上述 Kipp (2011a) 以醫學期刊文獻為研究對象的研究後，Kipp 另以 7 種圖資界學術期刊的 165 篇文獻為研究對象，比對 529 個 CiteULike 網站的社會標記與 727 個取自 Library Literature 與 INSPEC 資料庫的控制詞彙，也發現 16% 期刊文獻具有相同關鍵字的社會標記與控制詞彙 (Kipp, 2011b)。此外，上述 Kipp 兩項研究也進一步發現「未出現在索引典的相關關鍵字」(related not in thesaurus) 此一類別最常被使用。在以 2,786 篇教育文獻為研究對象，進行 3,176 個 CiteULike 社會標記與 1,083 個 ERIC 控制詞彙的研究，Bruce (2008) 發現只有 7.6% 社會標記可以比對到 ERIC 控制詞彙。以 45 篇醫學期刊文獻為研究對象，分別自 Connotea 與 PubMed 抽取 540 個社會標記與

1,034 個 MeSH 主題標目進行交叉比對，Lin et al. (2006) 則發現有 11% 相同的關鍵字出現在社會標記與 MeSH。除此之外，以 Del.icio.us 為研究對象中，Yi and Chan (2009) 以 3 組 299 個社會標記連結至 LCSH 的情形下，發現 60.9% 社會標記可以精確地比對至 LCSH。由上述討論，社會標記與控制詞彙間的關鍵字關聯程度不盡相同。從上述文獻探討（請參見表 2.2）歸納重點如下：

1. 無論是以圖書或期刊文獻等資料類型為研究對象，乃至於醫學、教育、圖資等領域，社會標記與控制詞彙間的關鍵字相同性比率偏低，且未獲一致的共識。
2. 以關鍵字相似性而言，社會標記確實能增補原來控制詞彙未能標引的關鍵字及其主題概念。
3. 就控制詞彙而言，半數以上的控制詞彙關鍵字可以比對到社會標記；如：前述的 Lu et al. (2010) 與 Yi and Chan (2009) 等研究。
4. 社會標記與控制詞彙間關鍵字的主題概念存有某種程度的相關性；如：前述的 Kipp (2011a, 2011b) 與 Thomas et al. (2009) 等研究。
5. 在比較社會標記與控制詞彙間的關鍵字重複比例方面，計算對象有所不同，進而可以歸納成兩種類型：一為關鍵字，二則期刊文獻。由於計算對象的差異性，所代表的意義亦不相同。以關鍵字為對象者，係以關鍵字的整體觀之，雖然可以明顯得知社會標記與控制詞彙間的關鍵字重複比例情形，卻容易出現 Heymann and Garcia-Molina (2009) 所發現許多現行研究忽略的一項事實：相同關鍵字可能出現在不同資訊物件上的現象。換言之，亦即相同關鍵字可能標引在不同的文獻上，即使關鍵字相同，亦不代表指涉相同的資訊物件。反之，以期刊文獻為計算與比對的對象而言，即使可以看出哪些期刊文獻具有相同關鍵字的社會標記與控制詞彙，卻無法得知關鍵字整體的重複情形。唯一例外者，只有 Lu et al. (2010) 的研究，兩者兼具。
6. 在比對社會標記與控制詞彙間的關鍵字重複比例方面，除了直接進行社

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

會標記與控制詞彙間關鍵字的兩兩比較外，也有些研究依據以往研究提出的比對原則為依據，更進一步細分比對的結果；例如，前述的 Iyer and Bungo (2011)、Kipp (2011a, 2011b) 與 Thomas et al. (2009) 等研究。

依據上述文獻探討，本研究推導出下列研究問題：

RQ2.2：社會標記與控制詞彙間的關鍵字關聯程度為何？

表 2.2 社會標記與控制詞彙重複比例的文獻分析

文獻	研究對象	社會標記與控制詞彙的相同百分比	計算單位
Bruce, 2008	CiteULike 社會標記與 ERIC 控制詞彙	7.6%	關鍵字
Iyer & Bungo, 2011	LibraryThings 的社會標記與主題標目 (含複分標目)	1.21%	關鍵字
Kipp, 2011a	CiteULike 社會標記與 MeSH	16%	期刊文獻
Kipp, 2011b	CiteULike 社會標記及 INSPEC、Library Literature 控制詞彙	16%	期刊文獻
Lin et al., 2006	Connotea 社會標記與 MeSH	11%	關鍵字
Lu et al., 2010	LibraryThing 社會標記與 LCSH	2.2% 社會標記出在 LCSH、50.1% LCSH 出現在社會標記；85% 書目紀錄具有相同的社會標記與 LCSH	關鍵字、書目紀錄
Thomas et al., 2009	LibraryThing 社會標記與 LCSH	約 14%	關鍵字
Yi & Chan, 2009	Del.icio.us 社會標記與 LCSH	60.9% 社會標記出現在 LCSH	關鍵字

第四節 關鍵字的類別及其使用情形

研究社會標記的使用情形方面，已有許多研究顯示個別社會標記的實際使用呈現冪次定律分佈現象 (Angus, Thelwall, & Stuart, 2008; Bischoff et al., 2008; Lu et al., 2010; Marlow, Naaman, Body, & Davis, 2006; Munk & Mørk, 2007; Sen, Lam, Rashid, Cosley, Frankowski, Osterhouse, Harper, & Riedl, 2006; Yi & Chan, 2009)，

而以控制詞彙為研究對象的研究中，亦有類似的發現(Chan & Vizine-Goetz, 1998; Ravari, 2012)。除此之外，也有些學者致力於社會標記與控制詞彙的類別，以及其類別的實際使用情形。

在社會標記類別的研究方面，Golder and Huberman (2006) 以 Del.icio.us 為研究對象，提出了 7 項社會標記類別，包括：指明主題 (identifying what or who it is about)、指明類型 (identifying what it is)、指明擁有者 (identifying who owns it)、修飾類別 (refining categories)、指明品質或特質 (identifying qualities or characteristics)、自我參考 (self reference) 與任務整理 (task organizing)。自此之後，引發許多社會標記類別的後續研究，探討社會標記類別及其使用情形。首先，有的研究以 Golder and Huberman (2006) 提出的 7 項社會標記類別為基礎，進一步探討社會標記類別的實際使用情形。例如，在 Thomas 等人的研究中，以 10 本流通使用率最高的圖書為研究對象，其中一項研究工作係延伸 Golder and Huberman (2006) 提出的社會標記類別進行使用情形的探討，結果發現指明主題此一類別最常被使用，修飾類別最不常被使用，而且指明主題 (37%)、指明品質或特質 (12%)、自我參考 (11%) 與任務整理 (17%) 等類別的使用率累計總和為 77% (Thomas et al., 2009)。再者，有的研究係以 Golder and Huberman (2006) 的 7 項社會標記類別為依據，濃縮成新的社會標記類別。例如，在一項電影推薦的系統設計中，Sen et al. (2006) 以 Golder and Huberman (2006) 提出的 7 項社會標記類別為基礎，濃縮成事實型 (factual tags)、主觀型 (subjective tags)、個人化 (personal tags) 等三種。第三，有些研究則是以 Golder and Huberman (2006) 的 7 項社會標記類別為比對對象，建立兩種不同社會標記類別的對照關係。例如，Bischoff et al. (2008) 以前述 Golder and Huberman (2006) 提出的社會標記類別為基礎，重新制定一套社會標記類別，並且與 Golder and Huberman (2006) 提出的社會標記類別進行比對 (Bischoff et al., 2008)。此項研究提出的類別包括主題 (topic)、時間 (time)、空間 (location)、類型 (type)、作者／擁

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

有者 (author/owner)、評論 (opinions/qualities)、使用情境 (usage context) 與自我參考, 同時以 Del.icio.us、Flickr 與 Last.fm 為測試對象, 結果顯示主題類別使用率最高, 而其他類別則會因測試對象的不同, 而有不同的使用率。

有些研究則以 Del.icio.us 與 Flickr 為研究對象, 同時導入不同的方法或理論, 發展不同的社會標記類別, 以探討各種社會標記類別的使用情形。在一項以 Del.icio.us 為對象的研究中, Munk and Mørk (2007) 以統計學的對應分析 (correspondence analysis) 為方法, 分析出內容 (content categorization)、媒體 (media categorization)、文本類型 (genre categorization)、著作權 (copyright categorization)、價值判斷 (value categorization)、後設類別 (meta categorization)、工作與過程 (process categorization)、時間 (time categorization) 與個人註記 (personal categorization) 等 9 項社會標記類別, 其中以內容類別使用率最高, 高達 79.1%。在以 Flickr 為研究對象的研究中, Angus et al. (2008) 結合前述的 Golder and Huberman (2006) 7 項社會標記類別, 及意義層次 (levels of meaning)、屬於 (of) 與關於 (about) 等理論概念, 為影像資料發展一個兩階層四大類的複合式類別, 從而進行社會標記的使用分析, 前述四大類包括: 社會標記與影像間的一般關係 (generic relationship between tag and image content)、社會標記與影像間的特定關係 (specific relationship between tag and image content)、個人或群組用途 (tag only useful to individual/university group)、其他 (miscellaneous categories)。結果發現最常用的五項子類別使用率累計高達 87%, 這些子類別及其使用率分別是: 社會標記與影像間的特定關係類別下的關於影像 (what image is about, 21%) 與指涉的地理/人/事件名稱 (what image is of place/names/events, 14%)、社會標記與影像間的一般關係類別下的屬於影像 (what image is of, 15%), 以及個人或群組用途類別下的修飾 (refining tag, 13%)、自我參考 (self-reference tag, 12%) 與複合式 (compound tag, 12%)。在 Rorissa (2010) 的研究中則是以所謂的影像屬性類別 (categories of image attributes) 為依據, 提出

10 項社會標記類別，包括：地點 (location)、內容 (content/story)、人 (people)、描述 (description)、物件 (objects)、抽象概念 (abstract concepts)、藝術史資訊 (art historical information)、與人相關的屬性 (people-related attributes)、視覺元素 (visual elements) 與顏色 (color)。該研究發現地點、內容、人與描述等 4 項類別最常被使用，同時前述這些類別使用率累計高達 76.26%。此外，該研究更進一步導入所謂的影像查詢類別 (categories of image queries)，結合上述影像屬性類別，形成一個二維矩陣的社會標記類別。Lin et al. (2006) 則是提出 16 項類別，包括：地名 (place-name)、複合式 (compound)、事物 (thing)、人 (person)、事件 (event)、無法識別 (unknown)、照片的 (photographic)、時間 (time)、形容詞 (adjective)、動詞 (verb)、一般地點 (place-general)、評比 (rating)、語言 (language)、生活事物 (living thing)、幽默 (humor)、詩文的 (poetic)、數字 (number)、情感 (emotion) 等。經歸納發現地名 (28.21%)、複合式 (14.05%)、事物 (11.37%)、人 (8.81%)、事件 (5.69%) 與無法識別 (4.79%) 等類別最常被使用，且使用率累計高達 78.61%。另外，以 12 個影像檔為測試對象，且劃分為非結構化與結構化社會標記的影像資料描述對照組實驗研究中，Bar-Ilan, Shoham, Idan, Miller, and Shachak (2008) 發現結構化社會標記提供較豐富的描述資訊。

此外，有些學者以索引典內容為依據，進行關鍵字的異同比對。以 Del.icio.us、Furl 與 Technorati 等為對象的研究中，Spiteri (2007) 以美國資訊標準組織 (National Information Standard Organization, NISO) 的「建置控制詞彙指引」(Guidelines for the construction format and management of monolingual controlled vocabularies) 文件提出的索引典類別為基準，檢視社會標記的使用情形。結果發現以名詞文法形式與事物概念等 2 類別最被常使用。在前述的 Kipp 兩項研究中，則是分別發現「未出現在索引典的相關關鍵字」(related not in thesaurus) 此一類別最常使用，且在圖資界與生醫領域中的使用率分別是 45%

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

(Kipp, 2011b) 與 35% (Kipp, 2011a)。

也有些研究嘗試發展模式，藉以分析與歸納社會標記的使用情形。如，Heckner et al. (2008) 提出所謂的「社會標記類別模式」(Tag Category Model, TCM)，且可進一步細分為「語言」(Linguistic Category Model, LTCM)、「功能」(Functional Category Model, FTCM) 與「文本」(Text to Text Category Model, T2TCM) 等三種子模式，三個子模式所涵蓋的類別如表 2.3 所示。該研究歸納出語言、功能與文本子模式最常用的類別與使用率分別是「名詞」(noun, 72%)、「主題相關的一般內容描述」(general content description of subject related tags, 79.49%) 與「關鍵字的各種變形」(variation, 54%)。

表 2.3 TCM 項下的各子類別 (資料來源：Heckner et al., 2008)

子模式	第一層級	第二層級	第三層級
LTCM	字組 (word class)	單一字的社会標記	形容詞
			副詞
			動詞
			名詞
			功能詞
			縮寫詞
			數字
		單一字以上的社会標記	複合詞
	拼字	正確	片語
		錯誤	
各式變形			
新興字 (neologisms)			
語文 (language)			
FTCM	主題相關	資源相關	創作者
			資源類型
			檔案類型
			日期
			來源
			語文
		內容相關	內容描述

子模式	第一層級	第二層級	第三層級
			探討範圍
			分類
			內容類型
			方法
			代碼
	個人相關	情感的 (affective)	正面的
			負面的
時間與任務		行動導向的	
		工作情境流程相關的	
	無須社會標記 (tag avoidance (no tag))		
T2TCM	與文本完全相同 (identical to fulltext)	題名	
		摘要	
		文本內容	
		相同的關鍵字	
	文本的不同形式 (variant from fulltext)	錯誤拼字 (spelling error)	
		字根/字尾變形 (stemming/inflection)	
	未出現在文本 (not occurring in fulltext)	同義	
		下位詞 (hyponym)	
		上位詞 (hyperonym)	

在書目資料庫方面，探討控制詞彙實際使用的相關研究並不多見。在一項以 OCLC 聯合目錄資料庫為研究對象的研究中，自 4 百萬個以上的 LCSH 中，抽取 20,473 個主題標目作為研究樣本，分析 LCSH 在前述資料庫的實際使用情形，並以機讀編目格式 (MACHINE-Readable Catalog, MARC) 紀錄的欄號 600 (subject added entry - personal name)、610 (subject added entry - corporate name)、611 (subject added entry - meeting name)、630 (subject added entry - uniform title)、650 (subject added entry - topical term) 與 651 (subject added entry - geographic name) 等主題附加款目作為使用率的分析項目。在該研究中，Chan and Vizine-Goetz (1998) 發現 LCSH 在 OCLC 聯合目錄資料庫的實際使用情形呈現著冪次定律分佈現象，亦即少數 LCSH 被大量使用，及多數 LCSH 使用率低。除

了使用的分佈現象之外，還進一步發現主題詞彙標目（即 tag 650）被應用在近三分之二以上的書目紀錄，次則為名稱標目（即 tag 600、610、611 與 651）與劃一題名（即 tag 630），使用率趨近於三分之一。

依前述探討內容，本研究歸納出下列重點，說明如下：

1. 多數研究以發展單一主軸的屬性類別為主要目標，以分析社會標記的使用情形，只有少數研究嘗試發展雙主軸的複合式屬性類別，如：前述的 Angus et al. (2008) 與 Rorissa (2010) 等研究。
2. 有些研究以索引典作為發展基礎，藉以分析社會標記的使用類別及其關鍵字間的索引典結構關係，包括：相同 (same)、廣義 (broader term, BT)、狹義 (narrower term, NT)、相關 (related terms, RT)、未出現在索引典的相關關鍵字 (related terms not in thesaurus)；如：前述的 Spiteri (2007) 與 (Kipp, 2011a, 2011b) 等研究。
3. Heckner et al. (2008) 提出 TCM，試圖從語文、功能與文本等三種不同的獨立觀點，個別分析社會標記的使用情形。截至目前為止，沒有任何研究同時混合不同子模式，以結合 2 個以上的不同屬性或觀點，進行社會標記的使用分析。
4. 以社會標記類目的使用情形而言，似乎具有 80/20 原則 (80/20 rule of distribution) 的分佈現象，且最常被使用的社會標記類別並未趨於一致 (請參照表 2.4)。
5. 依現有文獻提出的社會標記類別而言，可以進一步區分為功能型、索引典型與屬性型等三種，同時多數類別為一維類別，少數則為二維類別 (請參照表 2.5)。
6. 有關控制詞彙實際應用在資料庫書目紀錄的使用研究，截至目前為止，只有少數個案進行探討，且僅限於主題附加款目此一範圍。

表 2.4 最常用的社會標記類別

文獻	研究對象	最常用的類別	使用百分比
Angus et al, 2008	Flickr	關於影像	21%
Chan & Vizine-Goetz, 1998	OCLC 聯合目錄	主題詞彙標目	約 2/3
Heckner et al., 2008	Connotea	名詞、主題相關的一般內容描述與關鍵字的各種變形	72%、79.49%、54%
Kipp, 2011a	CiteULike	未出現在索引典的相關關鍵字	35%
Kipp, 2011b	CiteULike	未出現在索引典的相關關鍵字	45%
Lin et al., 2006	Flickr	地名	28.21%
Munk & Mørk, 2007	Del.icio.us	內容	79.1%
Rorissa, 2010	Flickr	地點	30.63%
Spiteri, 2007	Del.icio.us、Furl 與 Technorati	名詞與事物概念	18%、35% & 23% ; 76%、82% & 90%
Thomas et al., 2009	LibraryThing	指明主題	37%

表 2.5 社會標記類別的類型

社會標記類型	一維類別	二維類別
功能導向型	Golder & Huberman(2006)、Thomas et al.(2009)、Bischoff et al. (2008)、Munk & Mørk (2007)、Angus et al. (2008)、Sen et al. (2006)	
索引典導向型	Kipp (2011a, 2011b)、Spiteri (2007)	
屬性導向型	Lin et al. (2006)、Heckner et al. (2008)	Rorissa (2010)

依據上述文獻探討，本研究推導出下列研究問題：

RQ2.3：最常使用的社會標記與控制詞彙的關鍵字類別及其比例為何？其使用情形是否具備冪次定律分佈現象？

第五節 關鍵字類別的關聯關係結構與樣式

心智模式也是一種組織化的結構與關係 (Rentsch & Klimoski, 2001; Staggers & Norcio, 1993)，因而會反映在使用者文獻標引的行為與知識結構上。在以往的研究中，曾有學者以共同出現的關鍵字為基礎，導入 SNA、FP-growth 等方法探討關鍵字間的關聯關係。例如，以 2004 至 2008 年期間 36 個高出現頻率的 MeSH

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

關鍵字為研究樣本，Yang, Wu and Cui (2012) 導入 SNA 方法進行共字分析，討論醫學資訊學 (Medical informatics) 研究主題的趨勢演進與發展；李亞婷與馬費成 (2012) 以 64 篇文獻的 CiteULike 社會標記為研究對象，採用 SNA 分析社會標記的共現情形，探討關鍵字相關性，包括小世界 (small world) 與冪次定律分佈現象，以及中心度 (centrality) 分析；Huang, Li, Tan and Gao (2009) 採用 FP-growth 為方法，探勘短文中共同出現的關鍵字，以發展短文的類型類別；Yuan, Wu and Yang (2007) 也是採取 FP-growth 為方法，自詞典中探勘共同出現的關鍵字，以發展視覺化片語。

事實上，無論使用者是以個人或社群會員身份進行社會標記，就為不同的社會標記間建立隱性的結構性關聯關係。Bar-Ilan et al. (2008) 即明確指出，使用者就是經由關聯活動 (association) 進行社會標記。Rorissa (2010) 更建議「可以進一步歸納內嵌在社會標記間的關聯關係及其類別」，將有利於發掘社會標記的應用與使用。另外，Stvilia and Jørgensen (2010) 還指出「社會標記及其類別關聯關係有助於控制詞彙的存取與回現率。」由上述討論，隱約透露出社會標記類別的關聯樣式對於資訊組織的關鍵字查詢與使用是十分有用的。截至目前為止，儘管關鍵字類別的關聯關係十分重要，卻沒有任何相關研究深入探討此一議題，尤其是檢視一般使用者與專家間的異同。如果把標引同一資訊物件共同出現的社會標記與控制詞彙等關鍵字轉化為一種共同出現的屬性關係，進而建構使用者與專家的文獻標引心智模式地圖，除有助於分別探索使用者與專家文獻標引心智模式的知識結構地圖及其關聯關係外，更能深入分析與探討心智模式的結構關係及其樣式。依據上述文獻探討，本研究推導出下列研究問題，如下：

RQ2.4：一般使用者與專家間使用關鍵字類別之關聯關係的結構與樣式為何？

第六節 文獻標引心智模式的理論模式

在心智模式的理論中，主要係由設計者、使用者、系統與系統印象等四大元

件所共同組成。在實際應用時，有些研究同時包括心智模式的四大元件，有些則是將使用者加以區分成不同類型的使用者，取代原有模式中的設計者，維持四大元件（請參照第二章第一節之「四、心智模式理論的導入與應用方式」之說明）。首先，本研究將文獻標引的心智模式定義為：「社會標記者與專家在面對資訊物件與文獻標引系統時，會依其先備經驗與知識，提供關鍵字及其內隱的知識結構和關聯關係，分別建構社會標記者與專家所擁有的文獻標引心智模式，以表徵其對資訊的一種認知。」以此定義為依據，本研究係以關鍵字的使用分析為基礎，所建構之文獻標引心智模式理論模型的四大元件係由使用者（可進一步劃分為一般與專家）、資訊物件及關鍵字等共同組成，至於組成元件與其定義，以及其與原有心智模式元件間的關係，分述如下（請參見圖 2.2）：

1. 資訊物件：亦即原來心智模式中所定義的系統，可以是電腦軟體、電腦資訊系統，乃至於前述文獻探討所討論的參考服務、資訊檢索系統、資訊資源等。在本研究提出的文獻標引心智模式中，主要對象係為文獻標引所要組織的資訊物件。
2. 關鍵字：即原來心智模式中所定義的系統印象，換言之，是一般使用者與專家基於文獻標引的需求前提下，針對同一物件所產生的內在認知，進而具體化為外在表徵，予以轉換為關鍵字。
3. 專家：即從事於文獻標引的專家，負責分析資訊物件的主題分析，進而提供控制詞彙（如：描述語、權威標目等），以標引資訊物件所涵蓋主題的概念或知識。
4. 一般使用者：即社會標記者，針對所收集的資訊物件，以標籤分類，方便日後搜尋取用。

整體而言，以關鍵字的使用分析為基礎所發展的文獻標引心智模式下，一般使用者與專家面對外在的同一資訊物件時，分別依其先備經驗與知識，將其內在認知轉化為外在文字形式的關鍵字，藉以表徵其內在的一種心理反射。至於兩者

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

間所產生的系統印象是否趨於一致，則依據本研究第二章文獻探討所歸納的關鍵字的來源及其類型、社會標記與控制詞彙之間的關聯程度、關鍵字的類別及其使用情形，及關鍵字類別關聯關係的結構與樣式等四大要項進行評估，以檢驗兩者間文獻標引心智模式的異同。

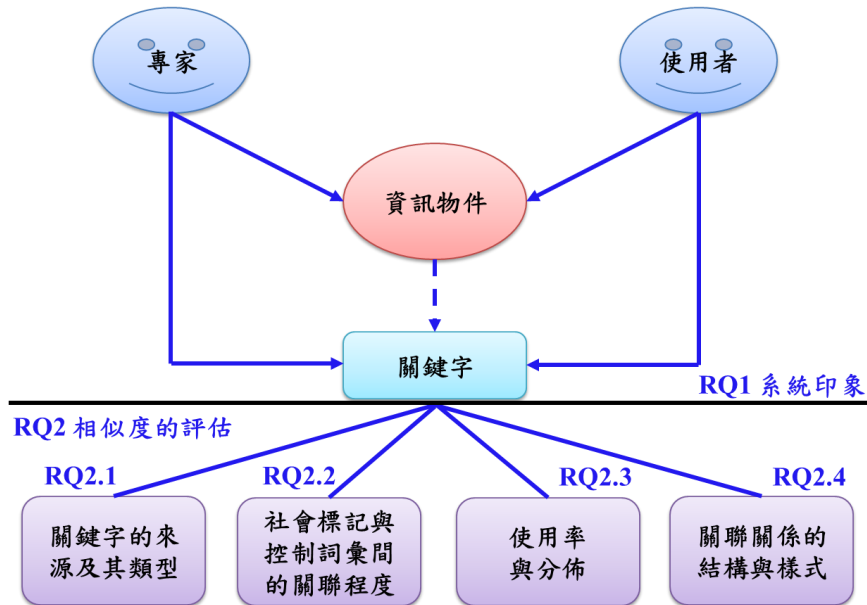


圖 2.2 文獻標引的心智模式及其推導的研究問題

第三章 研究方法

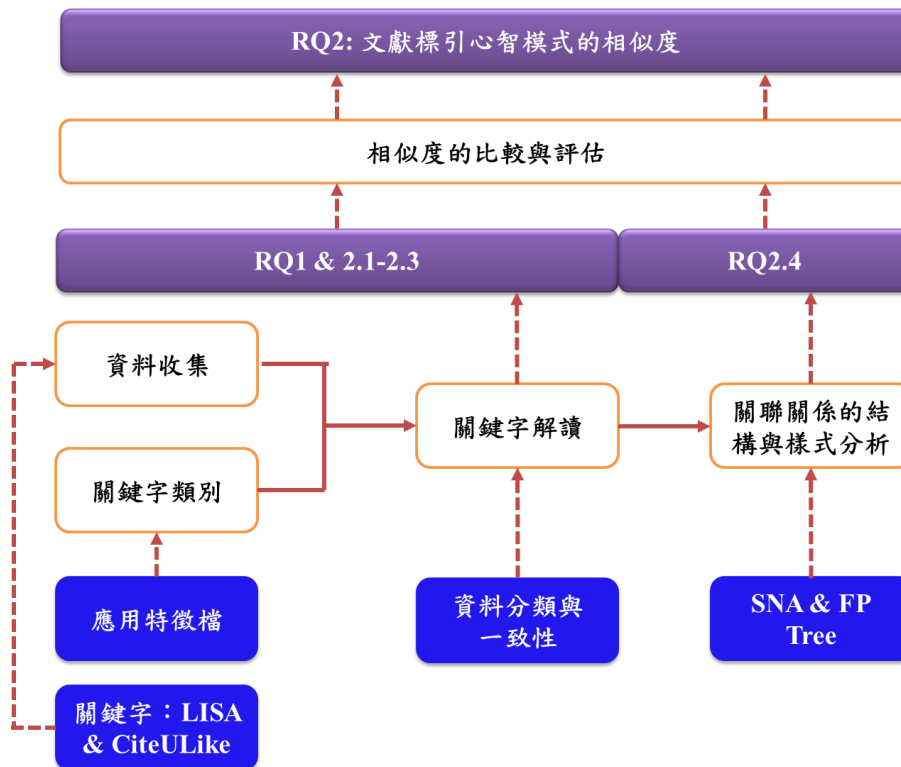
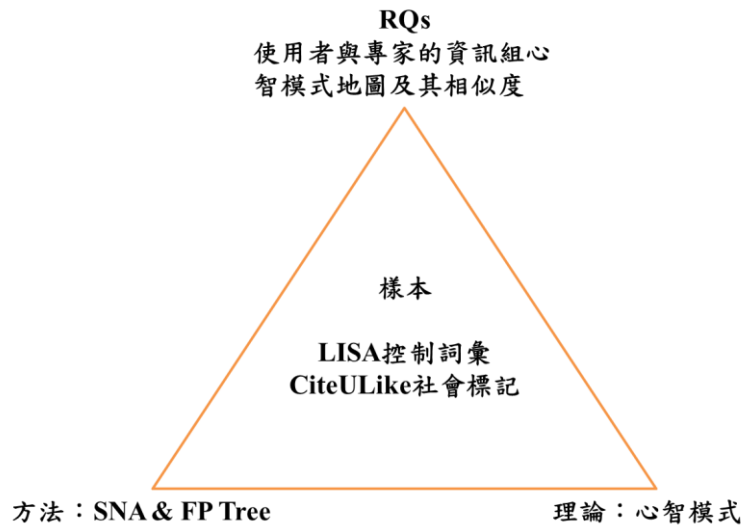
本章詳述本研究的程序與方法，進行相關資料的蒐集與分析，主要項目包括：研究工作與流程、研究對象、資料收集、關鍵字的類別與入類、關鍵字意義的解讀與分析、關聯關係的結構與樣式分析，以及心智模式相似性的評估方式。第一節係依前述第二章文獻探討所導出的 RQ 與文獻標引心智模式為基礎，說明研究工作的組成，包括理論模式、方法、RQ 與樣本間的彼此相互關係（請參見圖 3.1），以及六項研究工作與整個研究流程（請參見圖 3.2）。由於本研究係以 CiteULike 社會網路平台的社會標記與 LISA 提供的控制詞彙為研究對象，收集與分析相關資料，因此在第二節說明本研究選擇 CiteULike 與 LISA 作為研究對象的原因。第三節則詳述自 CiteULike 選取社會標記與 LISA 選取控制詞彙等關鍵字的程序，以及樣本的選取原則、正確性的查核，以及數量上的要求等。第四節說明導入應用特徵檔（Application Profile, AP）的理論與原則，採用既有文獻的研究成果，發展符合本研究需求的兩項類別表，作為關鍵字分類之用。第五節描述用來解讀關鍵字的意義時，所使用的參考工具、分類結果的入類原則，以及兩兩分類結果的一致性與內在信賴度。第六節說明發展文獻標引心智模式地圖的方法與其介紹，以及本研究方法不同於以往研究和其所採用方法的原因。第七節則說明本研究評估文獻標引心智模式相似性的方法。

第一節 研究工作

在檢視一般使用者與專家間文獻標引心智模式之相似性，依據前述文獻探討，本研究的組成如圖 3.1 所示。整體而言，本研究係由四大部分組成，包括 RQ、方法、理論與樣本兩兩相互交錯與支援。其中，前述三者以樣本為中心，經由方法分析樣本資料，運用理論解釋隱藏在樣本資料的現象，進而從樣本中找出驗證 RQ 的證據。另一方面，本研究以前述第二章圖 2.2 理論模式為基礎，進

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

而發展相關的研究程序，以探討由文獻探討推導出的相關 RQs（請參見圖 3.2）及 RQs 與樣本間的交互關係（請參見圖 3.3）。經由本研究所提出 RQs 的探討，可以清楚地呈現一般使用者與專家使用關鍵字進行文獻標引時所擁有的心智模式內容與特質，進而評估兩者間文獻標引心智模式的相似性。本研究主要涵蓋了下列幾項工作：



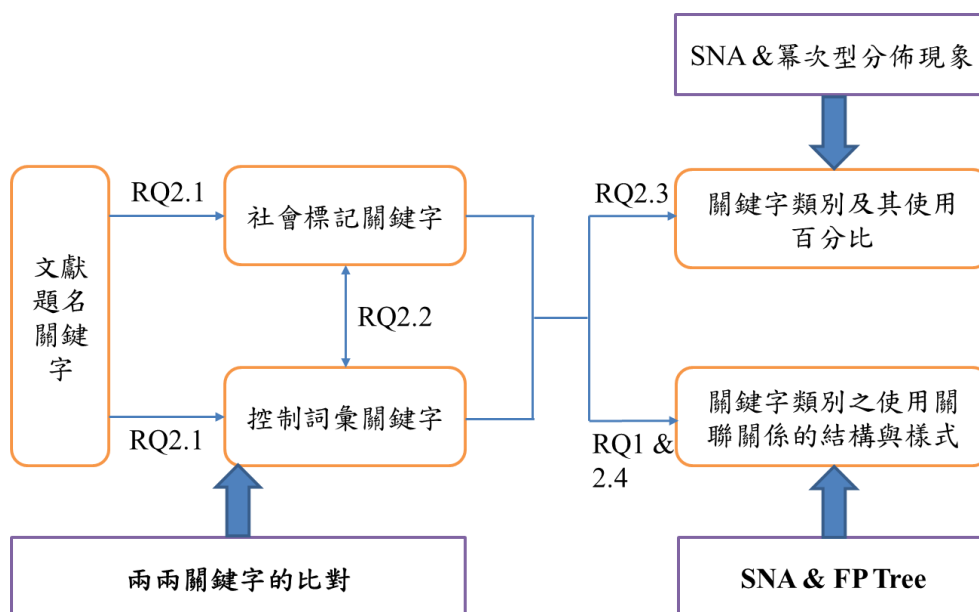


圖 3.3 研究問題與樣本間相互關係及其分析方式

1. 第一項：與文獻名稱形式相關的關鍵字比對，以文獻名稱的關鍵字為主要對象，包括文獻名稱與社會標記間，及文獻名稱與控制詞彙間的關鍵字比對與分類，採取關鍵字語法形式（syntactical form）的觀點，進行關鍵字間的比對，以檢驗 RQ2.1。
2. 第二項：與文獻名稱形式無關的關鍵字分類，以上述第一項工作無法分類的關鍵字為比對與分類對象，進行關鍵字語法形式外的關鍵字比對與分類，以檢驗 RQ2.1。
3. 第三項：依據本研究提出的類別，分別計算社會標記與控制詞彙等關鍵字在所屬各類別的使用情形及其數據，作為檢驗 RQ1 與 RQ2.3 之用。
4. 第四項：社會標記與控制詞彙間的關鍵字比對，以社會標記與控制詞彙為主要對象，進行關鍵字的兩兩比對，包括前述語法形式與非語法形式兩種，以檢驗 RQ2.2。
5. 第五項：分別建立社會標記與控制詞彙所屬的關鍵字類別關聯關係，作為檢驗 RQ1 與 RQ2.4 之用。
6. 第六項：評估社會標記者與專家間文獻標引心智模式的相似性，以檢驗 RQ2。

第二節 研究對象

本研究為能探索一般使用者與專家間運用關鍵字之文獻標引的行為及其心智模式的相似性，乃以 CiteULike 與 LISA 針對相同圖資學期刊文獻所提供的社會標記與控制詞彙為研究對象，利於探討研究之用，主要原因說明如后。首先，圖資學是一個跨學科領域的學科，除了引用其他學科領域的文獻，也經常被其他學科領域所引用 (Cronin & Meho, 2008)。因而，圖資學除具備多重學科的性質外，也特別注重資訊的使用者 (Prebor, 2010)，就資訊組織的研究而言更形重要且具代表性，以圖資學期刊文獻為研究對象，除了能探討一般使用者的文獻標引特質外，還可以分析具備圖資學背景之使用者與專家的文獻標引特性及其心智模式。再者，在本研究中，所謂的社會標記者係指具有社會標記平台帳號者，且可針對圖資學期刊論文實際提供社會標記的個人，並不包括單純查詢社會標記的使用者；而專家係指除接受過權威控制、索引典控制等專業訓練外，也具備主題分析方面的資訊組織理論知識，同時熟悉相關的規範、權威檔與索引典等控制詞彙及其系統，且實際從事期刊文獻標引的專業工作人員。第三，雖然期刊文獻型的社會網路平台有 CiteULike 與 Connotea，而且兩者皆有圖資界的相關社群，但在蒐錄範圍上，CiteULike 以期刊文獻為主，同質性較高，而 Connotea 則較多元化，還包括了網頁文獻、研究報告、技術報告等，此外 CiteULike 網站仍然繼續維持運作，持續蒐集資料期刊文獻書目資訊與社會標記，可視為一種長期型的資料。如果以研究對象的相同百分比而言，CiteULike 與 LISA 能針對同一批資訊物件提供更多的社會標記與控制詞彙作為研究分析之用，而非 Connotea 與 LISA。第四，同時選取 CiteULike 與 LISA 提供的關鍵字為研究對象，亦能以 Bruce (2008) 及 Kipp (2011b) 等研究結果為基準點，從不同的學科領域 (教育與圖資界) 與索引典 (INSPEC、Library Literature 與 LISA) 加以比較研究結果，擴展 CiteULike 的研究成果與瞭解。茲將 CiteULike 與 LISA 背景簡介說明如下。

CiteULike 是一種社會網路平台，類似 Delicio.us、Flickr 與 Last.fm 提供所謂的網路社交功能，主要目的在於提供網路使用者針對學術期刊文獻進行社會標記。在 2004 年 11 月，CiteULike 由 Richard Cameron 撰寫程式而成，之後成立 Oversity Ltd. 公司，進行 CiteULike 的營運（CiteULike, n.d.）。就資料來源而言，目前已有許多商業出版社及開放取用期刊（open access journals）與學科典藏庫（eprint archives, 如 arXiv 與 PubMed）等來源提供者，直接將其學術期刊文獻書目資訊轉入 CiteULike；反之，網路使用者也可以自行上載或輸入所欲標記的學術期刊文獻書目資訊至 CiteULike。在標記的過程中，網路使用者可以採取個人名義，或是加入某一社群會員的方式進行標記。

LISA 係以全球圖資界的學術期刊文獻為蒐錄範圍，主要包括逾 60 個國家出版的 500 種以上期刊文獻，提供專業的索引摘要服務。LISA 所涵蓋的主題除了圖書資訊學相關主題外，也包括數位學習、電子政府、電子商務、語意網、人工智慧與資訊科技等。LISA 提供摘要的資訊服務外，也依據索引典提供控制詞彙，以組織所蒐錄的期刊文獻內容（ProQuest, 2009）。

第三節 資料收集

本研究依據上述研究問題，分別自 CiteULike 與 LISA 中，以圖資學領域的國際期刊為範圍，選擇同時具有社會標記與控制詞彙的期刊文獻為研究所需的樣本資料。在期刊選擇上，期刊必須符合下列二項原則：

1. 具有一定的學術品質與聲譽：期刊同時被 LISA 與期刊引用報告（Journal Citation Report, JCR）所收錄。
2. 期刊性質：研究對象的整體期刊性質，必須涵蓋理論研究與實務應用等兩種性質。

依據上述原則，本研究在 2011 年 2 月 26 日至 3 月 2 日期間，由三位研究人

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

員分別收集本研究所需的樣本資料後，再兩兩交換資料進行交叉查核，以確認研究樣本的資料正確性。結果共選出 13 種期刊、1,489 篇文獻、3,972 個社會標記（1,672 個不重複標記）與 6,708 個控制詞彙（1,338 個不重複控制詞彙），作為研究樣本（請參照表 3.1）。本研究將收集所得之樣本資料依序列出期刊名稱、文獻題名、文獻所擁有的社會標記及控制詞彙後，依序進行編號，並以 Excel 檔案格式儲存。就社會標記的最低門檻要求而言，已達成 1,000 個以上關鍵字的關鍵數量（critical mass）（Spalding, 2007）。除此之外，本研究對象的樣本數量方面，呈現出控制詞彙數量多於社會標記的情形，雖然與期刊為研究對象的現有研究相似（Kipp, 2011a, 2011b），卻有別於現行許多研究中呈現研究樣本之社會標記多於控制詞彙的情形，此類研究諸如第二章所述的 Iyer and Bungo（2011）、Lu et al.（2010）、Thomas et al.（2009）等研究，乃是以圖書的書目紀錄為主要的研究對象。

表 3.1 研究樣本一覽

期刊名稱	文章數量	社會標記數量	控制詞彙數量
Aslib Proceedings	29	77	178
Information Processing & Management	275	778	1,184
Interlending & Document Supply	46	82	219
Journal of Documentation	57	161	227
Journal of Information Science	133	373	589
Journal of Librarianship and Information Science	48	109	240
Journal of the American Society for Information Science and Technology（包括：Journal of the American Society for Information Science）	392	1,174	1,640
Library Collections Acquisitions & Technical Services	47	107	282
Library Hi Tech	104	251	494
Online Information Review	60	143	301
Program-Electronic Library and Information Systems	17	47	89

期刊名稱	文章數量	社會標記數量	控制詞彙數量
Scientometrics	218	535	968
Serials Review	63	135	297
總計			
13	1,489	3,972	6,708

第四節 關鍵字的類別及入類

本研究以 Heckner et al. (2008) 提出的 TCM，及 Carlyle (1989) 提出的控制詞彙與使用者查詢語的關鍵字比對類別為分類框架 (framework) 外，同時導入應用特徵檔理論，分別發展符合本研究所需的 2 項分類類別，以執行前述第一節的第一項至第四項研究工作。依據 Heery and Patel (2000) 的定義，所謂的應用特徵檔係指：「由一個以上不同名稱空間之綱要所組成的資料元素，以因應特定的在地化應用」。應用特徵檔理論包括下列四大原則：

1. 直接使用一個以上的標準或規範所制定的內容：亦即以某一個或兩個以上的標準內容為依據，直接採用其內容，而不作任何變更。
2. 不產生新的資料元素：如果現有標準內容無法滿足需求，可以自行創立新的資料元素及其內容定義，並且建立名稱空間 (namespace)，宣告與維護這些資料元素，原則上仍然盡量避免創立新的資料元素。
3. 指定資料元素內的綱要或資料值：使用者可依實際需求，直接應用已存在的綱要或資料值 (即分類表或控制詞彙)。
4. 加以限定或修飾既有的標準內容與定義：亦即重新定義，且予以縮小或更專精化 (narrower or specific)。

由於本研究係探討每一類別下關鍵字的應用情形，因而前述應用特徵檔的第三項原則並不適用於本研究。在導入應用特徵檔理論時，本研究以前述第二章文獻探討為基礎，將提出關鍵字類別的每一研究文獻視為應用特徵檔定義的名稱空間，而每一研究文獻提出的類別視為應用特徵檔中的綱要。所以，本研究依循前

述的應用特徵檔原則，直接採用、重新定義現有的類別，或是創立新的類別，以符合本研究的需求。在第一個分類類別方面，本研究首先導入 Hecker et al.(2008) 提出之 TCM 的 T2TCM 模式，發展前述第一項研究工作所需的分類類別。再者，本研究同時採取 TCM 的 FTCM 與 LTCM 等子模式，發展前述第二項研究工作所需的分類類別。同時基於應用特徵檔的原則前提下，本研究除了直接採用 LTCM、FTCM 與 T2TCM 的類別外，也自其他文獻所提出的類別加以重新定義，以及提出新的類別與其定義。因而，第一個分類類別中，有 18 個類別直接沿用現有文獻所提出的類別，4 個類別係取自現有文獻提出的類別且加以重新定義，及本研究自行提出 4 個新類別，共計 26 個類別。其中，在 4 個重新定義的類別方面，則是現有文獻已提出類別，但其定義較廣，因應本研究需求必須將原有類別的定再加以縮小或專精化；如，類別 10（指明主題—首字語），在 Golder and Huberman（2006）研究中，提出指明主題此一類別，但未再進一步加以細分。在 4 個新類別方面，係無法自現有相關文獻中，找到相同或相似的類別，因而提出新類別，因應本研究需求；如類別 08（文獻題名—分離型詞彙）。另一方面，在第二個分類類別的設計，仍依循第一個分類類別的發展方式與原則，且以 Carlyle（1989）提出的分類類別為主要的依據框架，以執行前述第四項研究工作。本研究總計提出 8 個類別，其中共有 7 個類別直接採用現有文獻所提出的類別，及 1 個新類別。

為檢驗前述兩種分類類別的可行性，本研究隨機抽樣 200 筆期刊文獻及其所擁有的社會標記與控制詞彙進行檢測，結果證明可行。最後，本研究提出兩種關鍵字類別，分別由 26 類與 8 類組合而成，至於此兩種關鍵字類別所採取的應用特徵檔原則及其相關文獻來源間的關係，請參照表 3.2 與 3.3。

表 3.2 關鍵字類別及其與各文獻間的關聯關係：第一與二項研究工作

關鍵字類別	定義	文獻來源	TCM	應用特徵 檔原則	範例
01. 文獻題名－相同 (Title-exact)	直接取自文獻題名關鍵字的社會標記或控制詞彙。	Heckner et al., 2008; Kipp, 2006	T2TCM	完全採用	文：The Bibliometric Properties of Article Readership Information 關：readership
02. 文獻題名－首字語 (Title-acronym)	與文獻題名關鍵字意義相同，但以首字語等詞彙形式表達的社會標記或控制詞彙。	NISO, 2005; Spiteri, 2007	T2TCM	完全採用	文：question answering 關：qa
03. 文獻題名－縮寫 (Title-abbreviation)	與文獻題名關鍵字意義相同，但以縮寫等詞彙形式表達的社會標記或控制詞彙。	NISO, 2005; Spiteri, 2007	T2TCM	完全採用	文：information network 關：infonet
04. 文獻題名－同義 (Title-synonym)	與文獻題名關鍵字意義相同的社會標記或控制詞彙。	Heckner et al., 2008; NISO, 2005; Spiteri, 2007	T2TCM	完全採用	文：evaluation 關：assessment
05. 文獻題名－詞彙形式 (Title-variants)	與文獻題名關鍵字相似，但以名詞、過去分詞、現在分詞或形容詞等不同詞彙形式表達的社會標記或控制詞彙。	Heckner et al., 2008; NISO, 2005; Spiteri, 2007	T2TCM	完全採用	文：Search 關：Searching
06. 文獻題名－單複數 (Title-singular and plural)	與文獻題名關鍵字相似，但以單複數等詞彙形式表達的社會標記或控制詞彙。	NISO, 2005; Spiteri, 2007	T2TCM	完全採用	文：Library 關：Libraries
07. 文獻題名－非英文的同義詞彙 (Title-equivalent non-English term)	等同英文詞彙之非英文關鍵字的社會標記或控制詞彙。	NISO, 2005	T2TCM	完全採用	文：Bibliometric 關：Bibliometria (德文)
08. 文獻題名－分離型詞彙 (Title-split term)	指單一社會標記或控制詞彙必須與其他社會標記一起組合後，才具有符合原來文獻指涉的意義。		T2TCM	新類別	文：query expansion 關：query, expansion
09. 指明主題－一般 (Topic-general)	用來識別與主題相關的社會標記或控制詞彙。	Golder & Huberman, 2006	FTCM	完全採用	文：Combining full text and bibliometric information in mapping scientific disciplines 關：bibliometrics

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

關鍵字類別	定義	文獻來源	TCM	應用特徵 檔原則	範例
10. 指明主題—首字語 (Topic — acronym)	以首字語的形式識別與主題相關的社會標記或控制詞彙。	Golder & Huberman, 2006	FTCM & LTCM	重新定義	文：Centroid-based summarization of multiple documents 關：nlp (即 national language processing)
11. 指明主題—縮寫 (Topic — abbreviation)	以縮寫字的形式識別與主題相關的社會標記或控制詞彙。	Golder & Huberman, 2006	FTCM & LTCM	重新定義	文：An Information Architecture to Support the Visualization of Personal Histories 關：inforvis (即 information visualization)
12. 指明主題—分離型詞彙 (Topic — split term)	以分離詞的形式識別與主題相關的社會標記或控制詞彙。	Golder & Huberman, 2006	FTCM	重新定義	文：An implicit feedback approach for interactive information retrieval 關：adaptive, ir
13. 指明內容—品質或特性 (Content — qualities and characteristics)	針對文獻所進行個人評論或意見等評價論述的社會標記或控制詞彙。	Golder & Huberman, 2006; Heckner et al., 2008	FTCM	完全採用	關：New, Out-of-dated
14. 指明內容—空間 (Content — place)	用來識別與空間相關的社會標記或控制詞彙。	Holley, 2010; Kipp, 2006; Lin et al., 2006	FTCM	完全採用	文：The Scottish Parliament and e-democracy 關：uk
15. 指明內容—時間 (Content — temporal)	用來識別與時間相關的社會標記或控制詞彙。	Bischoff et al., 2008; Lin et al., 2006; Munk & Mørk, 2007a	FTCM	完全採用	文：2000 關：20 th century
16. 指明內容—人 (Content — person)	用來識別與人相關的社會標記或控制詞彙。	Holley, 2010; Lin et al., 2006	FTCM	完全採用	文：h-index 關：Hirsch
17. 指明內容—機構 (Content — institute)	指明內容涉及機構的社會標記或控制詞彙。	Golder & Huberman, 2006	FTCM	重新定義	文：academic libraries 關：UICU Library

關鍵字類別	定義	文獻來源	TCM	應用特徵 檔原則	範例
18. 指明內容－資源類型 (Content－resource type)	指明文獻內容涉及資源類型的社會標記或控制詞彙。	Munk, & Mørk, 2007a	FTCM	完全採用	關：survey
19. 指明內容－物件 (Content－object)	指明文獻內容涉及物件的社會標記或控制詞彙。		FTCM	新類別	關：citation-database
20. 指明內容－使用對象 (Content－audience)	指明內容所適用對象的社會標記或控制詞彙。		FTCM	新類別	關：s (即 student)
21. 指明內容－資源來源 (Content－source cue)	指明取得原始文獻相關資訊的社會標記。	Marlow et al., 2006	FTCM	完全採用	關：fromCiteULike
22. 指明功能－任務 (Function－task)	係社會標記或控制詞彙用來指明特定的任務，如閱讀 (to-read)、蒐集資料 (to-collect)。	Golder & Huberman, 2006	FTCM	完全採用	關：toread, todo
23. 指明功能－自我參考 (Function－self reference)	與社會標記者或專家有關的內容，通常以「我的」(my) 字樣出現，如：我的事物 (mystuff)。	Golder & Huberman, 2006	FTCM	完全採用	關：thesis (即作為撰寫碩士論文之用)
24. 其他－不用標記 (Other－tag avoidance)	指明無須標記的社會標記或控制詞彙。	Heckner et al., 2008	FTCM	完全採用	關：no-tag
25. 其他－複合式 (Other－hybrid)	指結合前述 01-24 等 2 種以上類目的社會標記或控制詞彙。		FTCM	新類別	關：s-collected
26. 其他－無法識別 (Other－unknown)	難以識別或辨認的社會標記或控制詞彙。	Angus et al., 2008, Lin et al., 2006		完全採用	關：1-layla

備註：範例欄位中的「文」一字代表文獻題名的原始關鍵字，而「關」一字則是代表社會標記或控制詞彙的原始關鍵字。

表 3.3 關鍵字類別及其與各文獻間的關聯關係：第四項研究工作

比對類別	定義	文獻來源	應用特徵 檔原則	範例
01. 完全相同 (Exact match)	社會標記與控制詞彙間的關鍵字完全相同。	Carlyle, 1989; Iyer & Bungo, 2011; Strader, 2009; Voorbij, 1998	完全採用	標：relevance-feedback 控：Relevance feedback
02. 部分相同：包括單一與多重 (Partial match: single & multiple)	社會標記與控制詞彙間的關鍵字部分相同，可能是單一或 2 個以上關鍵字相同。	Carlyle, 1989; Iyer & Bungo, 2011; Strader, 2009	完全採用	標：information retrieval 控：computerized information retrieval
03. 同義 (Synonym)	社會標記與控制詞彙間的關鍵字關係為同義。	Voorbij, 1998	完全採用	標：e-journal 控：electronic periodicals
04. 縮寫字 (Abbreviation)	社會標記與控制詞彙間的關鍵字關係為縮寫字形式。	Carlyle, 1989; Iyer & Bungo, 2011; Strader, 2009	完全採用	標：e-books 控：electronic books
05. 單複數 (Singular & plural)	社會標記與控制詞彙間的關鍵字關係為單複數形式。	Carlyle, 1989;	完全採用	標：digital libraries 控：digital library
06. 非英文的同義詞彙 (Equivalent non-English term)	社會標記與控制詞彙間的關鍵字關係為同義，而是英文以外的西方語文關鍵字。	Iyer & Bungo, 2011	完全採用	標：bibliometrie 控：bibliometrics (西班牙文)
07. 複合 (Hybrid)	社會標記與控制詞彙間的關鍵字關係為結合上述 01-06 等 2 種以上類型。		新類別	標：clir (即 cross-lingual information retrieval) 控：information retrieval
08. 無法比對 (None)	社會標記與控制詞彙間的關鍵字沒有上述 01-07 等任何關係。	Carlyle, 1989; Iyer & Bungo, 2011; Strader, 2009; Voorbij, 1998	完全採用	

備註：範例欄位中的「標」一字代表社會標記的原始關鍵字，而「控」一字則是代表控制詞彙的原始關鍵字。

第五節 關鍵字意義的解讀與分析

基於社會標記具有內隱性與模糊性等特質，無法單從社會標記進行內容分析，因而本研究進行社會標記分析時，同時將社會標記與學術期刊文獻名稱平行並置，以決定其所傳達的意義。在控制詞彙方面，除了參閱牛津科學摘要公司 (Cambridge Scientific Abstract, CSA) 提供的 LISA 索引典內容外，也進一步查

閱「圖資學線上詞典」(Online Dictionary for Library and Information Science, ODLIS, http://www.abc-clio.com/ODLIS/odlis_about.aspx)。若仍無法從學術期刊文獻名稱識別社會標記或控制詞彙時,本研究則更進一步參閱期刊文獻的作者關鍵字、原文摘要、LISA 摘要、原文全文,或者單一整合型詞典(OneLookDictionary, <http://www.onelook.com>)與維基百科全書(<http://www.wikipedia.org/>),乃至於使用 Google 網路搜尋引擎,尋找相關文獻的解釋與說明。

在比對與分類方面,本研究首先執行上述第一項工作,直接比對關鍵字的異同,包括期刊文獻名稱與社會標記,及期刊文獻名稱與控制詞彙,並以語法形式為主要的基準原則,只有語法形式完全無異者才視為相同的關鍵字,且歸類至表 3.2 的類別 01。其餘類似期刊文獻名稱關鍵字的社會標記與控制詞彙則分別入類至表 3.2 的 02 至 08 項類別。至於上述第二項工作,則是分別入類至表 3.2 的 09 至 26 項類別。在第四項工作方面,則是社會標記與控制詞彙間的兩兩關鍵字比對,仍然依循前述第一項與第二項工作的方式,依比對屬性分別入類至表 3.3 的各項類別中。三項比對與分類工作,共有 5 位研究人員分別執行。在期刊文獻名稱與社會標記間、期刊文獻名稱與控制詞彙間,及社會標記與控制詞彙間的關鍵字比對與分類等三項工作方面,除了由研究者執行上述三項工作外,分別由具有圖書資訊與檔案學研究所碩士背景、圖書資訊學研究所碩士背景且具有 3 年以上的分編作業經驗,及已取得直升圖書資訊學研究所碩士班資格等 3 人分別執行。在執行比對與分類工作之前,研究者分別製作三項指引說明文件(如表 3.2 與 3.3 的類別、定義與範例等欄位所示),逐一向 3 位研究人員分別說明此三項指引與其相對應之前述第一、二與四等三項研究工作的關係與注意事項,及前述可參考的相關工具與資源,達成共識後,再進行兩兩關鍵字間的比對與分類。除此之外,遇有任何問題時,採取電子郵件與電話方式進行溝通。分別完成上述工作後,再交由第五位人員兩兩比對結果與交叉核對,以確認資料的正確性。其一致性情形如下:

1. 文獻題名與社會標記間的關鍵字比對：Kappa 係數約為 93.55%（請參見表 3.4）。
2. 文獻題名與控制詞彙間的關鍵字比對：Kappa 係數約為 96.02%（請參見表 3.5）。
3. 社會標記與控制詞彙間的關鍵字比對：Kappa 係數約為 94.60%（請參見表 3.6）。

有關名目尺度 (nominal scale) 量測結果的可信賴度 (dependability) 或一致性 (consistency) 方面，Kappa 係數經常作為信度的測量指標，一般而言名目或研究對象愈多時，所得的 Kappa 係數正確性愈高 (Gwet, 2012)。本研究同時具備上述兩種條件，且名目與研究對象的數量上也遠比一般研究個案來得多，因而本研究前述各項 Kappa 係數的正確性也較高，且值得信賴。更進一步，若依據 Landis and Koch (1977) 提出的標準，由 0 至 1 的 Kappa 係數可劃分為完全不一致 (poor, <0.0)、稍微一致 (slight, 0.00 to 0.20)、有點一致 (fair, 0.21 to 0.40)、中度一致 (moderate, 0.41 to 0.60)、相當一致 (substantial, 0.61 to 0.80)、幾乎完全一致 (almost perfect, 0.81 to 1.00) 等五個範圍。若以此為依據基準，本研究兩兩比對結果的一致性屬於「幾乎完全一致」範圍，具有高度的內在可信度 (inter-reliability)。

第六節 關聯關係的結構與樣式分析

擷取、表徵與建置心智模式的方法很多，包括訪談、畫圖、觀察、方格法、卡片分類與問卷調查等不同方法。然而，前述這些方法必須具備研究者與受訪者面對面的互動環境下，方能引導與建置心智模式的內容地圖。鑑於本研究限於樣本資料的取得方式不同，且提供資料的對象趨於分散、龐雜與匿名之故，所以本研究無法直接與這些匿名資料提供者面對面地接觸與互動，所以僅能採取事後的

表 3.4 文獻題名與社會標記間關鍵字比對的入類情形

		研究者 1																										
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	總計
研 究 者 2	01	1121	1	0	0	5	7	0	0	6	0	0	12	0	0	0	0	0	0	6	0	0	0	1	0	0	0	1159
	02	7	44	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	53
	03	2	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27
	04	10	0	0	154	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	165
	05	15	0	0	2	71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	88
	06	51	0	0	0	0	57	0	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	111
	07	0	0	0	0	0	1	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
	08	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
	09	12	0	0	0	0	2	0	0	1134	0	2	2	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1154
	10	0	0	0	0	0	0	0	0	0	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95
	11	0	0	0	0	0	0	0	0	1	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27
	12	2	0	0	0	0	0	0	3	0	0	0	456	0	0	0	0	0	0	0	0	0	0	0	0	0	0	461

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

		研究者 1																										
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	總計
研 究 者 2	13	0	0	0	0	0	0	0	0	3	0	0	0	41	0	0	0	0	0	0	0	0	0	0	0	0	0	44
	14	3	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	7
	15	2	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	24
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	19
	17	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	9
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56	0	0	0	0	0	0	0	0	56
	19	2	0	0	0	0	0	0	0	0	4	0	0	0	0	0	1	14	0	46	0	0	0	0	0	0	0	67
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	9
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	28
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	77	0	0	77
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	10
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	24	

		研究者 1																											
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	總計	
	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	44	44
	總計	1232	45	25	156	76	67	11	12	1148	97	28	472	41	4	22	20	18	56	55	1	9	29	78	10	24	44	3780	

表 3.5 文獻題名與控制詞彙間關鍵字比對的入類情形

		研究者 1																										
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	總計
研究者 2	01	1207	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1210
	02	0	98	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
	03	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	04	0	8	0	489	1	3	0	0	18	0	0	3	0	0	0	0	0	0	2	0	0	0	0	0	0	0	524
	05	0	0	0	0	283	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	285
	06	0	0	0	0	1	260	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	265

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

		研究者 1																										
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	總計
研 究 者 2	07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	09	0	2	0	42	6	4	0	0	2676	0	0	13	0	0	0	0	0	2	4	0	0	0	0	0	0	0	2749
	10	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	11	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	12	0	0	0	9	4	3	0	0	9	0	0	712	0	0	0	2	0	2	0	0	0	0	0	0	0	0	741
	13	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	4
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	170	0	0	0	0	0	0	0	0	0	0	0	0	170
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
	16	2	0	0	2	4	0	0	0	0	0	0	0	0	0	0	100	0	0	1	0	0	0	0	0	0	0	109
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	192	0	2	0	0	0	0	0	0	0	197
	18	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	82	2	0	0	0	0	0	0	0	90
19	0	0	0	2	0	0	0	0	15	0	0	1	0	0	0	0	0	0	7	188	0	0	0	0	0	0	213	

		研究者 1																										
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	總計
研究者 2	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
	26	7	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	31
總計	1216	108	1	546	299	273	0	0	2731	2	3	729	4	173	1	102	194	93	200	0	0	0	0	0	2	31	6708	

表 3.6 社會標記與控制詞彙間關鍵字比對的入類情形

		研究者 1								
		01	02	03	04	05	06	07	08	總計
研 究 者 2	01	486	0	0	0	0	0	0	0	486
	02	0	819	7	1	1	0	2	65	895
	03	0	1	86	0	0	0	0	5	92
	04	0	0	1	51	0	0	0	0	52
	05	0	1	0	0	93	0	0	0	94
	06	0	0	0	0	0	6	0	3	9
	07	0	12	4	2	2	0	289	21	330
	08	0	10	2	0	0	0	2	2000	2014
總計		486	843	100	54	96	6	293	2094	3972

資料蒐集與分析。從 CiteULike 與 LISA 的樣本資料中，可以明顯得知同時出現的關鍵字，因而本研究將這些同時出現的關鍵字建立成兩兩共同出現的相互關係，以建構成一個知識結構的心智模式概念圖，表徵文獻標引的心智模式內容地圖，及呈現社會標記者與專家的知識結構。本研究除了以關鍵字的使用情形勾勒文獻標引心智模式外，也同時採取網路分析的 SNA 與關聯關係法則的 FP-growth (含 FP-tree) 等兩種方法，進行社會標記者與專家之文獻標引心智模式的擷取、表徵與建構，以及視覺化等工作。主要原因說明如下：

1. 雖然個別關鍵字可以視為心智模式的一種具象表徵，然而過去的研究皆著重在以個別的關鍵字為研究對象，包括關鍵字的來源、社會標記與控制詞彙間關鍵字的重複情形，以及關鍵字的使用情形等。以往研究並未具體提出完整的心智模式地圖，而是零散、片斷且沒有任何關聯關係的心智模式內容。因而本研究採取 FP-growth 與 FP-tree 為方法，以關鍵

字共同出現的關係及其出現頻率為依據，期能建構一個框架，以描繪出社會標記者與專家的階層式文獻標引心智模式的面貌及其地圖內容。

2. 雖然 FP-growth 與 FP-tree 方法可依共同出現關鍵字的關聯關係及其頻率完成文獻標引心智模式地圖內容，乃至於樹狀的階層式結構關係，但是無法分析關鍵字所屬類別間相互影響的特定關聯關係（如：具有影響使用的關鍵字類別、經常共同使用的關鍵字類別、可以相互替代的關鍵字類別等）。所以，本研究採取 SNA 方法，將同時出現關鍵字所建立兩兩共同出現相互關係及其連結次數所形成的關聯關係圖形視為一種網路結構圖，導入 SNA 的中心度、階層集層分析與共同角色等分析方法，提出另外一種分析觀點與結果，以補充 FP-growth 與 FP-tree 未發掘的現象。同時，也交叉檢驗關鍵字間的使用是否彼此相互影響。

有關網路分析的 SNA 及關聯關係法則的 FP-growth 與 FP-tree 方法，茲分述如下：

一、網路分析：SNA

SNA 是眾多網路分析方法之一，所謂的社會網路係指一群人(a set of people)或是不同人群(a set of groups of people)間具備某種樣式的接觸或互動關係(Newman, 2003)。在社會網路中，包含了所謂的不同個體或社會行動者(social actors)，而兩兩社會行動者間具備某種關係(relations)，進而形成不同且特定的連帶關係(ties)(Haythornthwaite, 2009)，最後形成所謂的社會網路。SNA 則是一種研究方法，藉以分析社會網路中社會行動者及其社會連帶的關聯關係。在圖資界中，SNA 可用來分析與解說資訊流及其交換關係(information flow and its exchanging relationships)(Haythornthwaite, 1996)，亦可用來呈現資訊間的關聯關係及其隱含的結構與樣式。本研究將採用 SNA 為研究方法之一，從網路結構

的觀點，將原有 SNA 研究社會網路的社會行動者及其關係與連帶關係，轉換探索關鍵字類別彼此間的各式概念關係（包括：關聯關係與連帶關係），藉以瞭解文獻標引心智模式的知識結構，同時採用 NetMiner 4 軟體工具，作為分析 SNA 數據，及產生 SNA 圖形之用。在 SNA 的資料處理過程中，必須建立一個二維資料表，表中縱軸（即 Y 軸）係為每一篇文獻，而橫軸（即 X 軸）列出每一篇文獻關鍵字所屬的關鍵字類別，且每一篇文獻關鍵字所屬的關鍵字類別視為 SNA 中的節點，每一篇文獻同時出現的兩兩關鍵字類別關係則視為 SNA 的節點關係。此外，本研究將採用 SNA 進行中心度（centrality）、階層集叢（hierarchical clustering）、同等角色（role equivalent sub-structure）等分析，其中中心度方面還包括：程度中心度（degree）、接近中心度（closeness）、中介中心度（betweenness）等三種類型。分述如下：

1. 中心度：在社會網路中，衡量個體與其他行動者的接觸或關聯關係，以及其在資訊傳遞的影響情形（Haythornthwaite, 2009），也就是個體在社會網路中與其他行動者的關聯關係情形（Haythornthwaite, 1996; Wasserman & Faust, 1994）。就本研究而言，將為同一篇文獻所擁有的社會標記與控制詞彙等關鍵字進行分類，給予每一關鍵字合適的類別，而且每一關鍵字類別視為社會網路中的一個個體或節點。同一文獻不同關鍵字可分別歸入所屬的關鍵字類別，而同時出現的不同關鍵字類別間，則具備節點間的連帶或連結（link）關係。以 SNA 中心度而言，本研究主要是探討不同關鍵字類別及其關鍵字間的關聯關係，及其相互影響。
2. 程度中心度：程度中心性較高者，表示其在網路中與較多的行動者有所關聯（Haythornthwaite, 2009; Wasserman & Faust, 1994）。本研究主要用來識別出最經常使用的關鍵字類別，而且也具備相當程度的影響力，直接或間接影響其他關鍵字類別及其所擁有關鍵字的使用。換言之，連結數量愈多者，其程度中心度愈強，反之亦然。社會網路程度中心度的計算方式與範例說明如下（Wasserman & Faust, 1994）：

- (1) 計算方式：網路中所有節點與某一節點的連結總數÷（網路中的節點總數-1）
- (2) 範例 1：如圖 3.4 所示，節點 n1 的程度中心度為 1
 $([1+1+1+1+1+1] \div [7-1])$ ，n2 至 n7 分別為 0.167 $([1] \div [7-1])$ 。
- (3) 範例 2：如圖 3.5 所示，節點 n1 至 n7 的程度中心度皆為 0.333 $([1+1] \div [7-1])$ 。

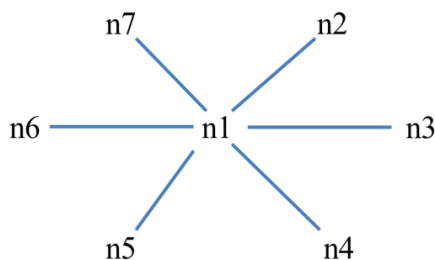


圖 3.4 星狀式社會網路（資料來源：Wasserman & Faust, 1994）

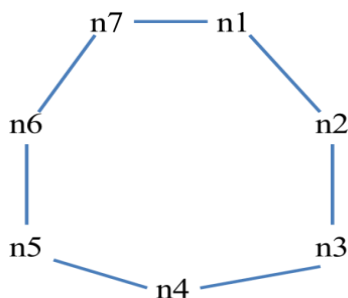


圖 3.5 環狀式社會網路（資料來源：Wasserman & Faust, 1994）

3. 接近中心度：個體與其他個體的接近程度（Wasserman & Faust, 1994）。本研究主要用來分析哪些特定關鍵字類別與其他關鍵字類別間，保持最近或較近的關聯距離。換言之，本研究延續上述關鍵字類別的連結關係，以關鍵字類別的連結距離遠近為依據，推導出不同關鍵字類別間的距離遠近。距離愈近者，其接近中心度愈強，反之亦然。社會網路接近中心度計算方式與範例說明如下（Wasserman & Faust, 1994）：
 - (1) 計算方式： $(\text{網路中的節點總數}-1) \div \text{網路中所有節點與某一節點最短距離的總和}$ 。
 - (2) 範例 1：如圖 3.4 所示，節點 n1 的接近中心度為 1 $([7-1] \div [1+1+1+1+1+1])$ ，n2 至 n7 分別為 0.545 $([7-1] \div [2 \times 5 + 1])$ 。

(3) 範例 2: 如圖 3.5 所示, 節點 n1 至 n7 的接近中心度皆為 $0.5([7-1] \div [1 \times 2 + 2 \times 2 + 3 \times 2])$ 。

4. 中介中心度: 係指個體在社會網路中居於其他成員彼此互動的橋樑, 代表了其引導資訊流通的機會, 而且也往往扮演操縱資訊流通的關鍵性位置 (Haythornthwaite, 2009; Wasserman & Faust, 1994)。換言之, 本研究延續上述關鍵字類別的連結關係, 以關鍵字類別所連結的最短路徑多寡為依據, 判別哪些鍵關字類別雖然不常被使用, 卻扮演居中的中介角色, 引導使用其他的關鍵字類別及所擁有關鍵字。社會網路中介中心度的計算方式與範例說明如下 (Wasserman & Faust, 1994):

(1) 計算方式: 網路中任一節點與其他節點的最短路徑總和 \div 網路中所有節點的路徑總和。

(2) 範例 1: 如圖 3.4 所示, 節點 n1 的中介中心度為 $1([1+1+1+1+1+1] \div [1+1+1+1+1+1])$, n2 至 n7 分別為 0。

(3) 範例 2: 如圖 3.5 所示, 節點 n1 至 n7 的中介中心度皆為 $0.2([1+1+1+0+0] \div [6 \times 5 \div 2])$ 。

5. 階層集叢: 主要是識別兩兩社會行動者彼此相互連結外, 同時也經常兩兩共同出現 (Haythornthwaite, 2009; Wasserman & Faust, 1994), 亦即兩個節點與其他節點距離的特徵 (profiles of distance) 十分相近, 進而可以形成一個集叢 (Hanneman & Riddle, 2005)。本研究主要用於找出哪些關鍵字類別經常被一起共同使用, 其計算方式與範例說明如下:

(1) 計算方式: 某些節點之間的實際連結總數 \div 網路中所有節點間可能所產生的連結總數。

(2) 範例: 假設 n1 至 n3 等三個節點中, 兩兩節點可能所產生的連結總數為 3 (即: n1 與 n2、n2 與 n3, 及 n1 與 n3 分別連結 1 次), 就 n1 與 n2 而言, 兩者間只有一個連結, 所以階層集叢指數為 $0.333(1 \div 3)$ 。

6. 同等角色 (structural equivalence)：主要是識別哪些社會行動者連結相同或相似的其他社會行動者 (Wasserman & Faust, 1994)。本研究主要用來發掘哪些關鍵字類別扮演同樣的關聯關係角色，而且共同使用的其他關鍵字類別趨於完全相同，或幾乎相似。

二、關聯關係法則：FP-growth 與 FP-tree 分析

本研究採用 FP-growth 與 FP-tree 為研究方法，進行文獻標引心智模式概念圖的建置。FP-growth 屬於關聯規則探勘的一種方法，藉以找出經常發生的關聯樣式，可應用於購物籃分析、交叉行銷、型錄設計、行銷活動分析、網路異動檔分析 (web log: click stream) 等方面。FP-growth 方法由 Han, Pei and Yin (2000) 所提出，主要包括建立 FP-tree、探勘 FP-tree，以及找出所有頻繁的項目集等階段。它是進行關聯規則探勘所需的資料結構，包含了一個根節點 (root node) 以及其他的子節點。每個節點都包含了標誌 (label, 如圖 3.6 橢圓形所示)、支持度 (support, 如圖 3.6 的橢圓形內的數字)，以及節點連結 (node-link, 如圖 3.6 的 f 與 c 等橢圓形間的連結) 等三部分。依 Han et al. (2000) 所提出的說明，以圖 3.6、表 3.7 為例，產生 FP-growth 主要的程序如下：

1. 在第一次掃描階段時，找出每個出現的項目及其次數，同時設定出現次數的門檻值 (本例中設定為 3)。例如，從表 3.7 的 5 次紀錄中，分別歸納出 (f:4)、(c:4)、(a:3)、(b:3)、(m:3)、(p:3) 等。
2. 依出現次數的多寡，作為排序的依據，將出現次數比較多的項目排在前面；如：表 3.7 第三欄所示。
3. 在第二次掃描階段時，就會將該筆交易紀錄做排序，再將排序後的資料建立在 FP-tree 中。
4. 建立 FP-tree 時，由根節點開始往下擴展與計算出現次數，如果存在同樣標誌的節點則該節點支持度的值加一。例如，TID 100 紀錄中，出現

表 3.7 FP-tree 項目集 (資料來源：Han et al., 2000)

交易紀錄 (TID)	購置項目	頻繁項目
100	f, a, c, d, g, i, m, p	f, c, a, m, p
200	a, b, c, f, l, m, o	f, c, a, b, m
300	b, f, h, j, o	f, b
400	b, c, k, s, p	c, b, p
500	a, f, c, e, l, p, m, n	f, c, a, m, p

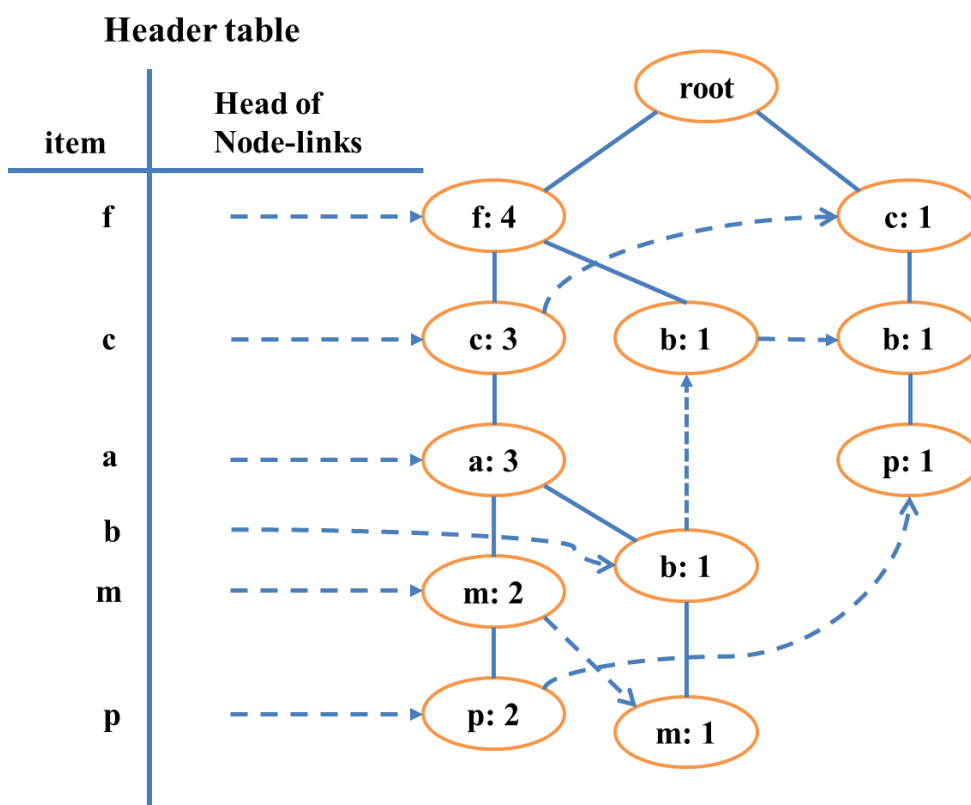


圖 3.6 FP-tree 範例圖 (資料來源：Han et al., 2000)

了 (f→c→a→m→p) 的某一 FP-tree，所以根節點為 f，以後依序為 c、a、m、p，同時每一項目計算次數為 1，亦即所謂的支持度。如果沒有與一個既存根節點及其相關項目 (即樹節點) 完全相同，則必須建立一個新節點加入 FP-tree 中，同時也在圖 3.6 中，建立 Header table 對應的項目連結。例如，TID 200 紀錄中，出現了 (f→c→a→b→m) 的某一 FP-tree，所以根節點為 f，以後依序為 c、a、b、m，同時每一項目計算次數為 1。但 TID 200 與 TID 100 之間不同處，在於項目 4 與 5，因而

必須自第 4 個項目起，另外建立新節點（即：b 與 m）。依此類推，若無存在相同標誌的節點，則必須從第一個項目起，開始建立新節點，也從圖 3.6 Header table 建立對應的項目連結。

5. 之後使用 FP-growth 來探勘 FP-tree 中的頻繁樣式，逐步選取在 Header table 中的項目。依照該項目的連結(link)建立對應的基底條件 FP-tree，如此不斷的遞迴重建 FP-tree，直到所建立出來的 FP-tree 為單一路徑(single path)或是空樹為止，接著在從 Header table 的下一個項目開始遞迴重建，直到全部完畢後，也就探勘完成所有的頻繁樣式。

本研究採取 FP-growth 為另一研究方法，分析關鍵字使用類別的各式路徑，作為表徵文獻標引的路徑式關聯規則(path-based association rules)，進而描繪與分析整體的路徑式決策樹。在 FP-growth 與 FP-tree 的資訊處理過程中，也類似 SNA 的處理方式，必須建立一個二維資料表。本研究將每一期刊文獻視為單一交易紀錄(即表格的 Y 軸)，而出現在期刊文獻中之社會標記或控制詞彙所屬的類別則被視為單一交易紀錄所出現的項目(即表格的 X 軸)。依此類推，產生社會標記與控制詞彙類別的 FP-tree。為進行 FP-tree 分析，本研究直接採用卓克索大學資訊科技學院(College of Information Science and Technology, Drexel University)所發展的 FP-tree 軟體工具，作為分析關鍵字類別使用的關聯關係，及產生 FP-tree 圖之用。在 FP-growth 方面，則是採用加拿大麥克敦大學(University of Moncton, Canada)提供的「次序樣式探勘框架」(Sequential Pattern Mining Framework, SPMF, <http://www.philippe-fournier-viger.com/spmf/index.php?link=download.php>) 開放原始碼軟體分析關鍵字類別使用的關聯關係，以產生本研究所需的 FP-growth 規則。

第七節 心智模式相似性的評估方式

基於相同的資訊物件或研究對象前提下，本研究將社會標記者提供的社會標記與專家提供的控制詞彙視為兩種不同類型的心智模式。雖然每一使用者與專家

擁有獨立不同的心智模式，無形中分別對所要組織的資訊物件呈現出一種集體式的現象。換言之，雖然不同的社會標記者具有不同獨立的心智模式，但整體的社會標記結果可以視為一種集體式的心智模式現象。依此類推，專家提供的控制詞彙可視為另外一種集體式的心智模式現象。基於此故，在面對同一資訊物件的情形下，社會標記者與專家依其所擁有的知識與經驗產生關鍵字進行文獻標引，本研究以此為基礎，藉以探討兩者間文獻標引的心智模式相似性及其異同。

本研究採取複合式的評估方式，以檢視一般使用者與專家間不同心智模式的相似性及其異同，說明如下：

1. 量化：以使用率與冪次定律分佈現象，及 SNA 的中心度、階層集叢與同等角色等分析的數據與結果等依據，評估兩兩心智模式間的相似性。
2. 質化：本研究將 FP-tree 分析所產生之路徑式規則及其概念化樣式視為一種樹狀的階層式圖形與結構，進而彙整成心智模式地圖，並以此為依據，分析與比較不同心智模式間的異同。

在量化的心智模式相似性評估方面，本研究以 Fleiss (1981) 提出 Kappa 係數的三種範圍一致性作為評估基準，以檢視心智模式一致性之用。前述三種範圍如下：輕度一致性 (poor, less than 40%)、中度一致性 (fair to good, between 40% and 75%) 與高度一致性 (excellent, more than 75%) 等。

第四章 研究結果與討論

本章詳述依據第三章說明的研究程序及方法，以 CiteULike 社會網路平台的社會標記與 LISA 提供的控制詞彙為研究對象，分析相關資料，提出各項的研究結果，以及從資訊組織之主題分析的權威控制與索引典控制，乃至於層面分析等觀點，闡述各項分析結果及其原因。第一節，主要以關鍵字及其類別的使用描述統計數據為依據，探討社會標記與控制詞彙在個別關鍵字與關鍵字類別兩個層次上的使用情形，包括：冪次定律分佈、80/20 原則等不同現象的分析，以及評估社會標記者與專家之文獻標引心智模式的相似性，主要項目包括個別關鍵字的使用情形、關鍵字類別及其相關群組的使用情形與排名等。第二節，則是社會標記與控制詞彙的關鍵字比對分析，範圍包括了個別關鍵字比對結果，及其在關鍵字類別與其所屬期刊文獻等方面的分佈情形，進而評估社會標記者與專家之文獻標引心智模式的相似性。第三節，則是有關 SNA 中心度分析，除了包括程度中心度、接近中心度與中介中心度等三項數據係數與排名分析外，同時就關鍵字類別的使用情形進行交叉分析，並以關鍵字類別的三項 SNA 中心度的係數與排名位置情形為依據，評估社會標記者與專家間之文獻標引心智模式的異同。第四節，有關 SNA 階層集叢分析，探討哪些關鍵字類別經常一起共同出現，亦即一起被社會標記者或專家所使用，形成所謂的集叢現象，以及從關鍵字類別的共同使用現象為觀點，比較與分析社會標記者與專家間使用關鍵字進行文獻標引的差異之處。第五節，為 SNA 同等角色分析，主要目的是探索哪些關鍵字類別具有相互交換或取代的現象，並以 SNA 同等角色功能的分析結果為基礎，探討社會標記者與專家兩者文獻標引心智模式的取向及其異同。第六節，是經由 FP-tree 分析，藉以建立與描繪社會標記者與專家的路徑決策樹型文獻標引心智模式地圖，並討論社會標記者與專家之文獻標引心智模式的異同，包括規則數量、規則樣式，以及組成規則與樣式的階層數量等。第七節，採取 FP-growth 分析方式，進一步確

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

認可支持與可信賴的 FP-growth 路徑規則，及分析與比較社會標記者與專家之文獻標引心智模式的異同，主要項目包括 FP-growth 規則與規則涵蓋範圍、規則樣式，以及關鍵字類別及其關鍵字間的連動使用關係等。第八節，以關鍵字類別的使用情形為基礎，勾勒出另外一種一般使用者與專家的文獻標引心智模式地圖。第九節，以前述第一至第八節的內容為基礎，探討研究結果對資訊系統設計的影響，包括關鍵字的推薦、使用者界面的設計及瀏覽分類架構的建立與運用。至於前述 RQ 與本章節內容的相互關係，如表 4.1 所示。

表 4.1 RQ 和第四章研究結果與討論各節內容間的相互關係

研究問題	研究結果與討論
RQ1	第六節與第八節
RQ2.1	第一節
RQ2.2	第二節
RQ2.3	第一節
RQ2.4	第三節至第八節

第一節 關鍵字及其類別的使用分析

以關鍵字的來源而言，從表 4.2 與 4.3 可以得知，約有 31.97% 的社會標記及 18.38% 的控制詞彙分別與期刊文獻題名關鍵字完全相同，亦即直接取用期刊文獻題名中的關鍵字作為社會標記與控制詞彙。以直接取用期刊文獻題名關鍵字作為社會標記方面而言，本研究結果高於 Heckner et al. (2008) 的 26.5% 與 Hyemann, Koutrika and Garcia-Molina (2008) 的 16%，究其原因可能在於標記物件類型的不同，Heckner et al. (2008) 以資訊科技期刊文獻為標記物件，Hyemann, Koutrika and Garcia-Molina (2008) 以相片影像為為標記物件，而本研究則為圖資界的期刊文獻；再者，由於同以期刊文獻為標記物件，故本研究結果與 Heckner et al. (2008) 提出的研究結果較為相近。在直接取用期刊期刊文獻題名關鍵字為控制詞彙方面而言，本研究結果皆小於 Ansari (2005) 的 70% 以上，以及 Frost (1989)

的 73%，即使將表 4.3 類別 01 至 08 等題名群組的使用比率予以累計，也只達 36.34%，仍與前述兩項研究結果間存有相當的差異。造成此種差異的潛在原因可能是研究樣本的資料類型及其控制詞彙樣本的數量，因為 Frost (1989) 研究對象為書目紀錄，而 Ansari (2005) 的研究對象為學位論文，原則上一般在建立此種資料類型的書目紀錄主題標目數量以 3 個為限。然而，本研究樣本中，控制詞彙數量多於社會標記的原因主要在於研究對象為期刊文獻，常提供較多的控制詞彙擴大文獻的回現率及涵蓋範圍 (cover categories, 如: 研究方法) (Kipp, 2011a)，而且並未受到圖書主題分類規範「將控制詞彙的最高數量限制為 3 個以下」的影響，所以平均每筆期刊文獻的控制詞彙數量皆高於書目紀錄的控制詞彙數量。除此之外，本研究結果顯示「社會標記者具有直接選用期刊文獻關鍵字作為標引資訊資源的主題概念或知識之用」的現象，而且傾向高於專家。如果進一步與 Ansari (2005) 和 Frost (1989)，以及 Heckner et al. (2008) 和 Hyemann, Koutrika and Garcia-Molina (2008) 的研究結果兩兩對比，則有與本研究結果不同的相反現象，亦即前述這些研究指出：「專具有直接選用期刊文獻關鍵字作為標引資訊資源的主題概念或知識之現象，而且此一傾向高於社會標記者。」至於是何種原因造成不同的結果，則有待進一步的研究。

表 4.2 社會標記關鍵字類別的使用數量與百分比

關鍵字類別名稱	社會標記數量	百分比	排名
01. 文獻題名—相同	1270	31.97%	01
02. 文獻題名—首字語	54	1.36%	11
03. 文獻題名—縮寫	27	0.68%	16
04. 文獻題名—同義	167	4.20%	04
05. 文獻題名—詞彙形式	93	2.34%	07
06. 文獻題名—單複數	121	3.05%	05
07. 文獻題名—非英文的同義詞彙	12	0.30%	21
08. 文獻題名—分離型詞彙	12	0.30%	21
09. 指明主題—一般	1168	29.41%	02
10. 指明主題—首字語	98	2.47%	06
11. 指明主題—縮寫	28	0.71%	15

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

關鍵字類別名稱	社會標記數量	百分比	排名
12. 指明主題—分離型詞彙	477	12.01%	03
13. 指明內容—品質或特性	44	1.11%	12
14. 指明內容—空間	7	0.18%	25
15. 指明內容—時間	24	0.60%	17
16. 指明內容—人	20	0.50%	20
17. 指明內容—機構	23	0.58%	19
18. 指明內容—資源類型	56	1.41%	10
19. 指明內容—物件	76	1.91%	09
20. 指明內容—使用對象	1	0.03%	26
21. 指明內容—資源來源	9	0.23%	24
22. 指明功能—任務	29	0.73%	14
23. 指明功能—自我參考	78	1.96%	08
24. 不用標記	10	0.25%	23
25. 複合式	24	0.60%	17
26. 無法識別	44	1.11%	12

表 4.3 控制詞彙關鍵字類別的使用數量與百分比

關鍵字類別名稱	控制詞彙數量	百分比	排名
01. 文獻題名—相同	1233	18.38%	02
02. 文獻題名—首字語	106	1.58%	10
03. 文獻題名—縮寫	1	0.02%	18
04. 文獻題名—同義	532	7.93%	04
05. 文獻題名—詞彙形式	295	4.40%	05
06. 文獻題名—單複數	270	4.03%	06
07. 文獻題名—非英文的同義詞彙	0	0.00%	20
08. 文獻題名—分離型詞彙	0	0.00%	20
09. 指明主題—一般	2760	41.14%	01
10. 指明主題—首字語	2	0.03%	16
11. 指明主題—縮寫	3	0.04%	15
12. 指明主題—分離型詞彙	724	10.79%	03
13. 指明內容—品質或特性	4	0.06%	14
14. 指明內容—空間	171	2.55%	09
15. 指明內容—時間	1	0.02%	18
16. 指明內容—人	100	1.49%	11
17. 指明內容—機構	192	2.86%	07
18. 指明內容—資源類型	89	1.33%	12
19. 指明內容—物件	192	2.86%	07

關鍵字類別名稱	控制詞彙數量	百分比	排名
20. 指明內容－使用對象	0	0.00%	20
21. 指明內容－資源來源	0	0.00%	20
22. 指明功能－任務	0	0.00%	20
23. 指明功能－自我參考	0	0.00%	20
24. 不用標記	0	0.00%	20
25. 複合式	2	0.03%	16
26. 無法識別	31	0.46%	13

以個別關鍵字的使用情形而言，社會標記與控制詞彙皆呈現著所謂的冪次定律分佈現象（請參見圖 4.1 與 4.2），換言之，亦即少數的社會標記與控制詞彙經常被使用，多數的社會標記與控制詞彙並不常被使用。在關鍵字類別的使用而言，無論是社會標記或是控制詞彙之關鍵字類別的使用情形，也類似呈現著冪次定律分佈的現象，亦即少數關鍵字類別經常被使用，多數的關鍵字類別並不常被使用（請參見圖 4.3 與 4.4）。除此之外，8 個常被使用的關鍵字類別，其累計的使用率總和也約略符合 80/20 原則，進一步比較 8 個常被使用的社會標記與控制詞彙關鍵字類別，可發現兩者之常用關鍵字類別多數是相似的，只有兩個類別是不同的，包括：社會標記的關鍵字類別 10（指明主題－首字語）與 23（指明功能－自我參考），及控制詞彙的關鍵字類別 17（指明內容－機構）與 19（指明內容－物件）。至於 8 個常被使用的關鍵字類別分別如下：

1. 8 個常被使用的社會標記關鍵字類別：01（文獻題名－相同）、09（指明主題－一般）、12（指明主題－分離型詞彙）、04（文獻題名－同義）、06（文獻題名－單複數）、10（指明主題－首字語）、05（文獻題名－詞彙形式）與 23（指明功能－自我參考）（使用率總和：85.45%）。
2. 8 個常被使用的控制詞彙關鍵字類別：09（指明主題－一般）、01（文獻題名－相同）、12（指明主題－分離型詞彙）、04（文獻題名－同義）、05（文獻題名－詞彙形式）、06（文獻題名－單複數）、17（指明內容－機構）與 19（指明內容－物件）（使用率總和：92.39%）。

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

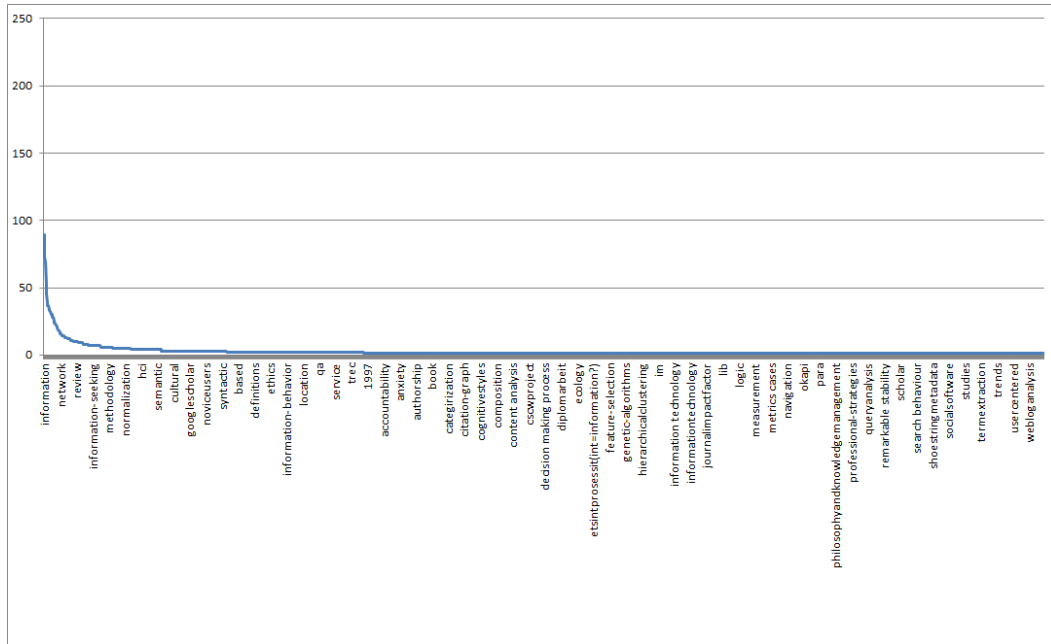


圖 4.1 社會標記的關鍵字使用分佈情形

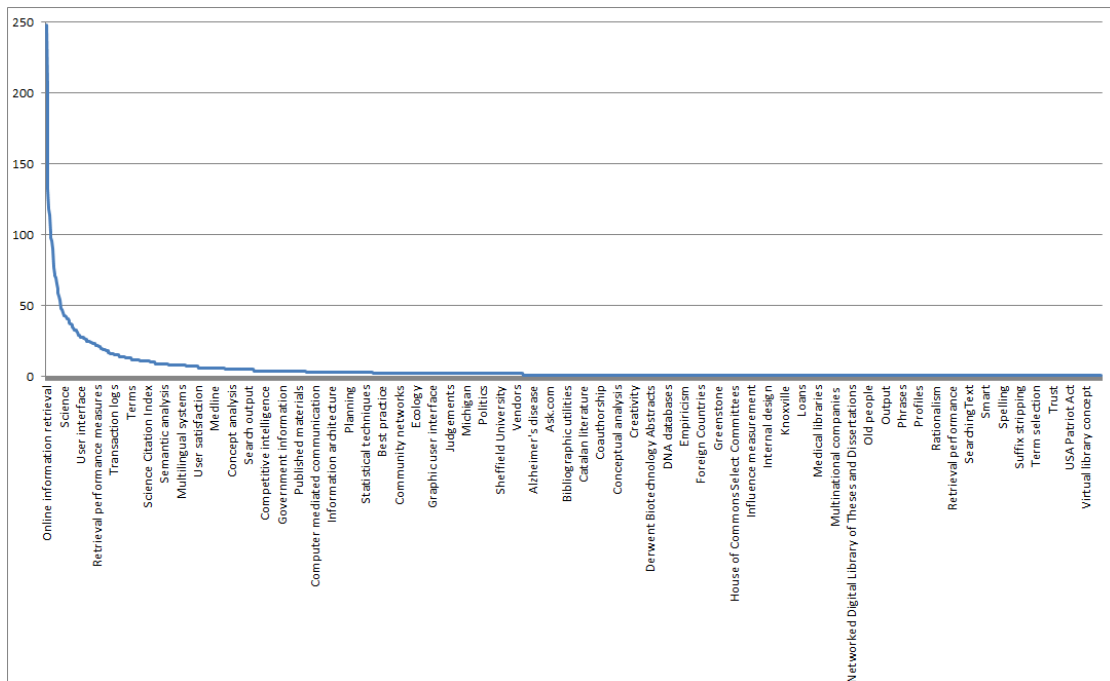


圖 4.2 控制詞彙的關鍵字使用分佈情形

比較圖 4.3 與 4.4，可發現圖 4.4 比圖 4.3 的曲線更為陡峭，這種現象顯示控制詞彙關鍵字類別遠比社會標記關鍵字類別更集中使用少數類別，而表 4.2 與 4.3 數據也可以交叉解釋此一現象。在關鍵字類別的使用排序方面，社會標記的關鍵字類別 01（文獻題名—相同）百分比（31.97%）最高；反之，控制詞彙關鍵字

類別則是類別 09（指明主題—一般）的百分比（41.14%）最高。如果進一步將相關類別予以組合且加以排序，可以發現在關鍵字類別的使用上，社會標記以主題群組最經常被使用，次之為題名群組，且前述兩者的使用率相差無幾。在控制詞彙關鍵字類別方面，主題群組仍是最經常被使用，次之為題名群組，但兩者間之使用率存在較大的差距，且功能群組使用率為 0。由此可知，社會標記與控制詞彙間的關鍵字類別使用上存有差異，且呈現出不同的現象，其關鍵字類別群組及使用率如下：

1. 社會標記：主題群組（類別 09-12，使用率：44.60%）、題名群組（類別 01-08，使用率：44.20%）、內容群組（類別 13-21，使用率：6.55%）、功能群組（類別 22-23，使用率：2.69%）、其他群組（類別 24-26，1.96%）。
2. 控制詞彙：主題群組（類別 09-12，使用率：52.00%）、題名群組（類別 01-08，使用率：36.34%）、內容群組（類別 13-21，使用率：11.17%）、其他群組（類別 24-26，使用率：0.49%）、功能群組（類別 22-23，使用率：0%）。

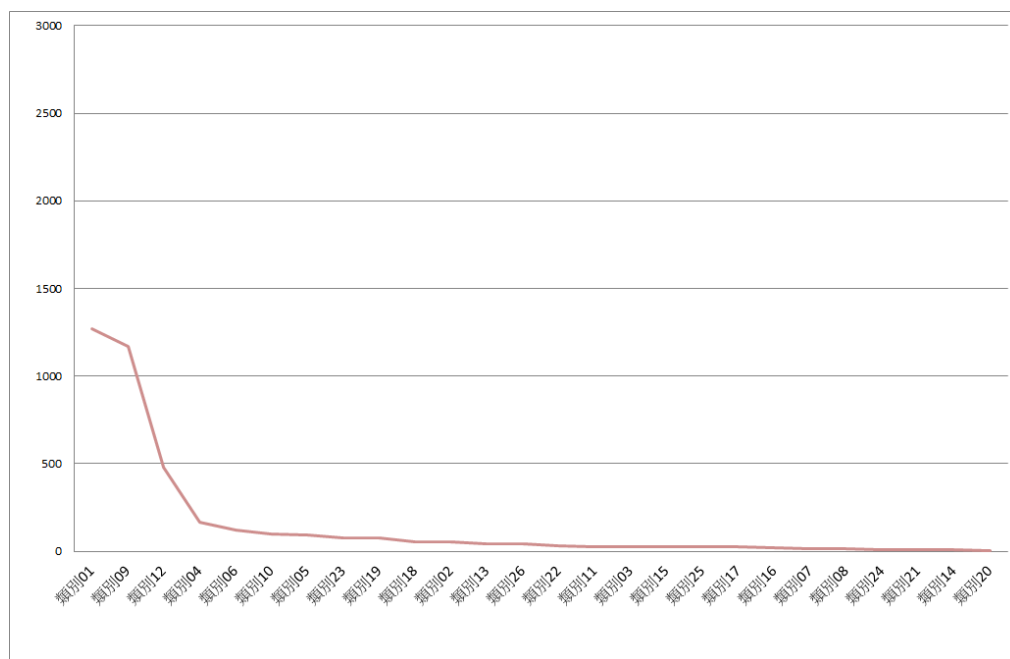


圖 4.3 社會標記的關鍵字類別使用分佈情形

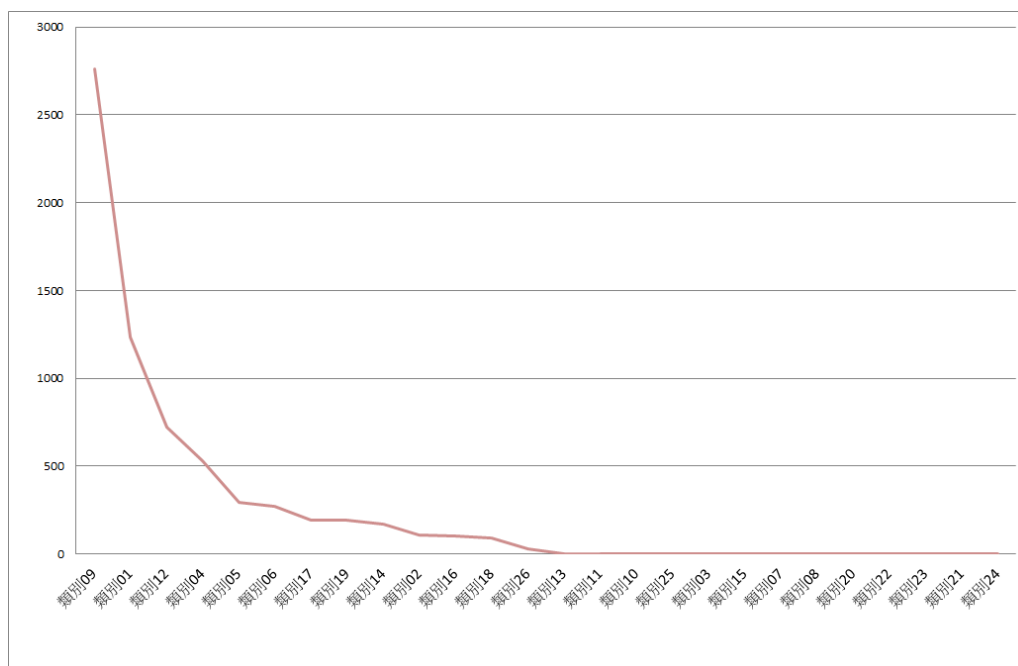


圖 4.4 控制詞彙的關鍵字類別使用分佈情形

在 Angus et al. (2008)、Munk and Mørk (2007) 與 Thomas et al. (2008) 等研究中，皆是歸納指明主題類別為最常使用的類別，而 Rorissa (2010) 與 Lin et al. (2006) 之研究則是提出以地點或地名作為最常使用的類別。就本研究結果而言，社會標記者與專家兩者最常使用主題群組此一類別進行資訊資源的標引，此種現象與 Angus et al. (2008)、Munk and Mørk (2007) 與 Thomas et al. (2008) 等研究結果趨於一致，其主要原因可能在於標引資訊物件的主題仍是社會標記者與專家首選的文獻標引重點，並不因文本資料類型或身份別（即：社會標記者與專家）而異。另外，就 80/20 原則而言，本研究結果與 Angus et al. (2008)、Lin et al. (2006)、Munk and Mørk (2007)、Rorissa (2010) 及 Thomas et al. (2008) 等研究發現趨於一致，這也顯示 80/20 原則的現象並不因資料類型或身份別的差異而有所不同。

就關鍵類別的使用率而言，社會標記者與專家間並未達成任何一致的現象。以關鍵字類別的使用排名名次而言，只有三個類別同時被社會標記者與專家使用且排名位置相同者，這些關鍵字類別包括：04（即：文獻題名—同義）、11（指

明主題—縮寫)與 12 (指明主題—分離型詞彙),其排名位置分別為 4、15 與 03 (請參見表 4.2 與 4.3)。所以,就關鍵字類別的使用排名名次而言,社會標記者與專家間心智模式的相似性達到 11.54% (3/26),屬於「輕度一致性」的範疇。更進一步,如果以關鍵字類別的屬性為依據,將相關屬性的關鍵字類別予以彙整在一起,可以發現社會標記者傾向自主題群組或題名群組的關鍵字類別中,選取關鍵字作為社會標記,進行文獻標引之用。同時,主題群組與題名群組兩者的使用百分比十分接近,而且兩者使用率總和超過 80%,約略符合 80/20 原則。就控制詞彙關鍵字類別的使用情形而言,專家也是傾向自主題群組或題名群組的關鍵字類別中,選取關鍵字作為控制詞彙,進行文獻標引之用。但是主題群組與題名群組兩者的使用比率並不相近,兩者間存有差異,而且兩者的使用率總和超過 80%,也約略符合 80/20 原則。從文獻標引觀點而言,其潛在的原因在於專家可能知道字根機制 (stemming mechanism) 能將不同形式的關鍵字予以聚合進行檢索,或是受到編目規則、權威控制或索引典控制等類似規範指引的影響,除了以主題分析為主要的文獻標引工作外,也直接採用官方來源的既有控制詞彙為首要原則後,再依一定的原則與程序選取文獻題名相關形式的關鍵字,因而降低專家自資訊資源題名中,選取關鍵字作為文獻標引之用。例如,英美編目規則的檢索款目導論 (Gorman & Winkler, 1998) 與中國編目規則乙編標目第二十一章檢索款目之擇定的編目規範條文 (即 21.0.3 編目規則) 中,即明確指引專家自作品的標題與題名之外建立檢索款目 (圖書館自動化作業規劃委員會、中國編目規則研訂小組、中華民國圖書館學會分類編目委員會, 2005)。即使如此,專家仍無法避免自資訊資源題名中,摘取重要關鍵字,標引資訊資源的主題或內容。再者,有關功能群組的關鍵字類別的使用上,此類關鍵字有被社會標記者使用,然而使用百分比並不高,但專家並不使用這類的關鍵字,作為文獻標引之用。其潛在原因可能在於專家認為這類關鍵字與一般資訊系統之個人化功能的訴求相似,可由資訊系統的功能達成,因而並未採用功能群組關鍵字作為文獻標引之用。基於上述原因,才造成社會標記者傾向使用主題或題名群組的關鍵字進行文獻標引,而

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

專家則傾向以標引資訊物件的主題為主。

就關鍵字類別群組的使用情形而言，只有主題、題名與內容等群組關鍵字類別的使用排名名次相同，因而社會標記者與專家間心智模式的相似性為 60% (3/5)，屬於「中度一致性」範疇之內。依據上述討論，就關鍵字類別的使用百分比而言，社會標記者與專家間並未達成一致性的共識。然而就關鍵字類別的使用排名名次，以及關鍵字類別群組的使用排名名次等方面而言，社會標記者與專家間之文獻標引心智模式的相似性並不高，而且皆未超過 60%，介於在「輕度一致性」與「中度一致性」之間的範疇。

第二節 社會標記與控制詞彙之間的關鍵字比對

在社會標記與控制詞彙間的關鍵字比對方面，從表 4.4 可以得知類別 08 (無法比對) 的數據最高，且超過半數以上。排名第二的是類別 02 (部分相同)，第三名為類別 01 (完全相同)。就比對類別 01 而言，計算單位為關鍵字的前提下，比對結果分別為 12.23% 與 7.25%。如果比對的計算單位更改為期刊文獻篇數時，則有 417 篇期刊文獻同時擁有相同關鍵字的社會標記與控制詞彙，百分比約為 28.01% ($417 \div 1,489$)。若從關鍵字比對類別的百分比分佈情形而言，除了呈現著冪次定律分佈的現象外，也具備 80/20 原則的現象；換言之，社會標記與控制詞彙的關鍵字比對結果集中在少數的比對類別。更進一步，若從 80/20 原則的觀點而言，可以發現控制詞彙的比對結果可由類別 08 (無法比對) 與 02 (部分相同) (百分比總和：84.5%) 累計達成 80/20 原則的現象，而社會標記的關鍵字比對結果則是類別 08 (無法比對)、02 (部分相同) 與 01 (完全相同) (百分比總和：86.2%)，也就是控制詞彙的關鍵字比對結果遠比社會標記關鍵字的比對結果更集中在少數類別 (請參照圖 4.5 與 4.6)。就社會標記與控制詞彙間的關鍵字比對而言，無法比對者 (unmatch) 的比率高達一半以上。約有 28.01% 的期刊文獻同時具備相同關鍵字的社會標記與控制詞彙，即使將表 4.4 的類別 01 與 03-06

(即：完全相同、同義、縮寫字、單複數與非英文的同義詞彙)使用百分比予以累計，使用率總和分別為 18.65% (tag-2-descriptor) 與 11.06% (descriptor-2-tag)，前述三項數據皆未超過 40%，因而也屬於「輕度一致性」範疇之列。

表 4.4 社會標記與控制詞彙間關鍵字的比對情形

比對類別：關鍵字字數	比對百分比：社會標記 (Tag-to-Descriptor)	比對百分比：控制詞彙 (Descriptor-to-Tag)
01. 完全相同：486	12.23%	7.25%
02. 部分相同：包括單一與多重：843	21.22%	12.57%
03. 同義：100	2.52%	1.49%
04. 縮寫字：54	1.36%	0.80%
05. 單複數：96	2.42%	1.43%
06. 非英文的同義詞彙：6	0.15%	0.09%
07. 複合：293	7.38%	4.37%
08. 無法比對：2094 vs. 4830	52.72%	72.00%

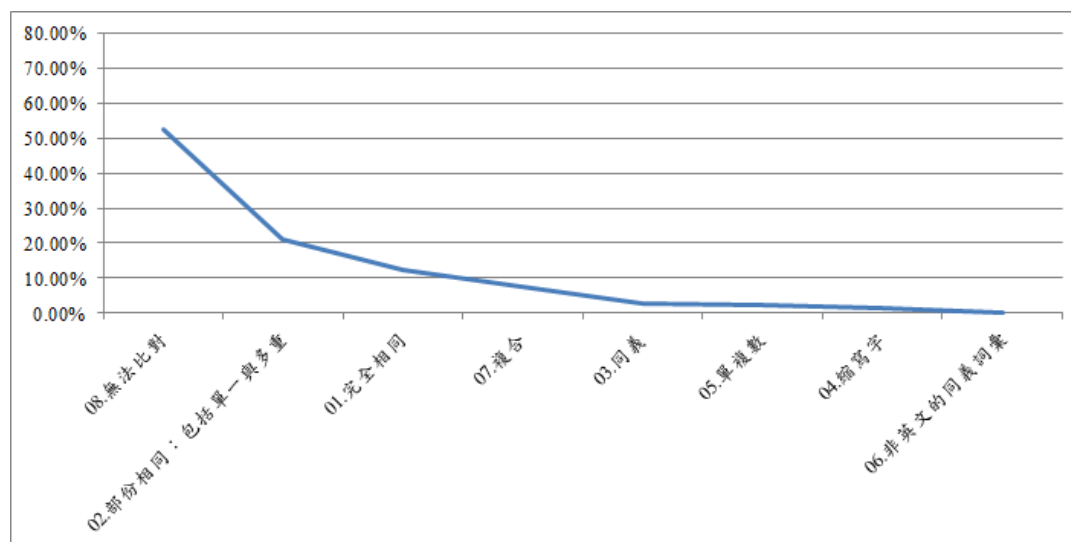


圖 4.5 社會標記出現在控制詞彙的類別分佈情形

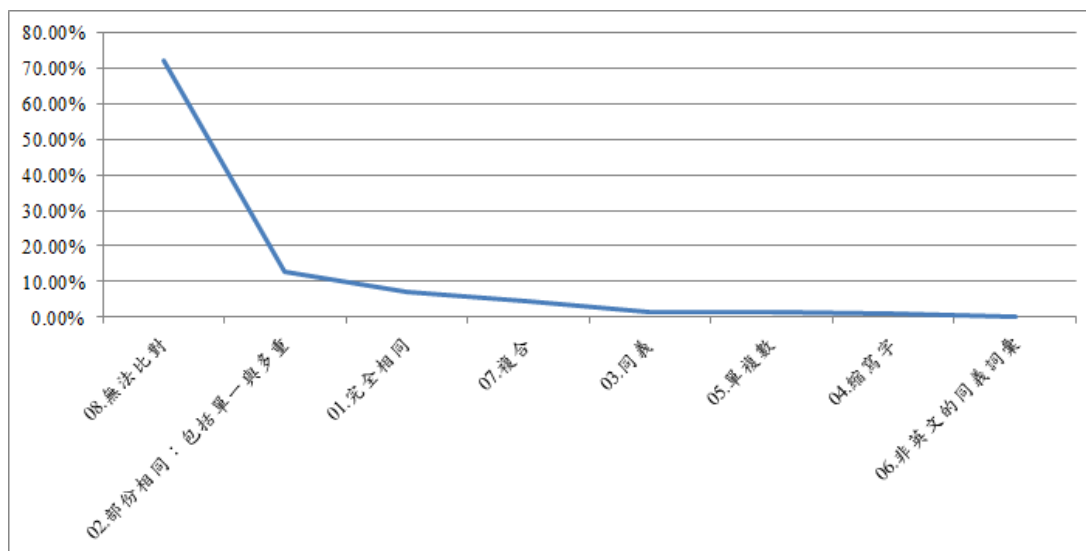


圖 4.6 控制詞彙出現在社會標記的類別分佈情形

造成上述社會標記者與專家兩者間使用相同關鍵字的比例如此低或不同的潛在原因，可能是專家特別注重資訊組織的主題分析，因而以標引資訊物件的主題為主要的優先考量，造成使用主題群組關鍵字的使用率過半，而且超過題名群組關鍵字的使用比例甚多。同時專家也會盡量直接採用官方維護的既有控制詞彙，如 LCSH 等。即使自文獻題名中萃取關鍵字，專家也受到編目規則、權威控制與索引典控制等規範指引的影響，必須依遵循一定的原則，才能自題名中直接選取關鍵字，作為標引資訊物件的控制詞彙；如，NISO (2005) 的「建置控制詞彙指引」與中國編目規則乙編標目第二十一章檢索款目之擇定的編目規範條文（即 21.1.1.1 至 21.1.1.2 編目規則）中，即有相關的規範原則明確指引專家自作品的標題與題名相關名稱建立檢索款目（圖書館自動化作業規劃委員會、中國編目規則研訂小組、中華民國圖書館學會分類編目委員會，2005）。除此之外，社會標記並不如控制詞彙必須依循一定的原則與處理程序後，才能增加新的關鍵字，反而具備即時性的新字，以滿足新的需求。因而，大幅降低社會標記與控制詞彙間的關鍵字相同程度。

在社會標記可以完全對應到控制詞彙，且以關鍵字為計算單位的前提下，本

研究結果 (7.25%、12.23%) 相近於 Bruce (2008) 的 7.6%、Lin et al. (2006) 的 11% 及 Thomas et al. (2009) 的 14%，但與 Iyer and Bungo (2001) 的 1.25% 相差較大。本研究結果相近於 Bruce (2008)、Lin et al. (2006) 的主因，可能在於前述兩項研究對象皆是期刊文獻，而且其提供的社會標記平台分別是 CiteULike 與 Connotea。至於對象同為 LibraryThing 的 Thomas et al. (2009) 與 Iyer and Bungo (2001) 研究中，是何種原因造成 Thomas et al. (2009) 的研究發現與本研究結果較為相近，而 Iyer and Bungo (2001) 卻與本研究結果存有較大的差距，則需要進一步深入的研究。若以期刊文獻為計算單位，則本研究發現約有 28% 的期刊文獻同時擁有相同關鍵字的社會標記與控制詞彙，這與 Kipp (2011a, 2011b) 16% 的研究結果仍有一段差距，可能原因在於本研究樣本以圖資界為範圍，除了提供 LISA 控制詞彙的專家具備圖資的資訊組織專業背景外，CiteULike 圖資社群的社會標記者也可能具備與專家類似的專業背景，才造成本項研究的百分比高於 Kipp (2011a, 2011b) 的研究結果。除此之外，Kipp (2011a) 的研究對象為生物醫學的期刊文獻，以及文獻的 MeSH 控制詞彙與 CiteULike 社會標記，而 Kipp (2011b) 則為圖資學的期刊文獻，以及文獻的 INSPEC 與 Library Literature 等索引典控制詞彙，也是可能造成不同的研究結果。另一方面，在控制詞彙數量遠大於社會標記數量的相似前提下，社會標記與控制詞彙間關鍵字完全相同的比率方面，本研究與 Yi and Chan (2009) 兩者間的研究結果卻存有極大的差異 (約 5 倍，12.23% 與 60.9%)，其潛在原因可能是 Yi and Chan (2009) 是以整個 LCSH 為比對的對象，但本研究則不是。

第三節 SNA 中心度分析

在本研究中，SNA 主要是用來評估不同關鍵字類別間的相互關係，以檢視三項文獻標引的過程與行為特質，包括：

1. 哪些類別的關鍵字除了經常被使用外，也影響其他類別關鍵字的使用。

2. 哪些類別的關鍵字雖不被經常使用，一旦被使用後，卻影響其他類別關鍵字的使用。
3. 哪些類別的關鍵字經常被單獨使用，卻不影響其他類別關鍵字的使用。

SNA 中心度共有程度中心度、接近中心度與中介中心度等三種，其中程度中心度用來衡量某一關鍵字類別與其他類別的鏈結情形，接近中心度則是量測某一關鍵字類別與其他類別的距離遠近，而中介中心度主要是檢視某一關鍵字類別引導使用其他類別的情形。如圖 4.7 與表 4.5 所示，程度中心度排名前 8 名的社會標記關鍵字類別分別是「01 與 12」（文獻題名—相同、指明主題—分離型詞彙）、09（指明主題—一般）、05（文獻題名—詞彙形式）、「04、06、19」（文獻題名—同義、文獻題名—單複數、指明內容—物件），及「02 與 10」（文獻題名—首字語、指明主題—首字語）等類別，而接近中心度排名前 8 名的社會標記關鍵字類別與上述程度中心度完全相同。在中介中心度排名前 8 名的社會標記類別，則不同於上述程度中心度與接近中心度，包括的類別有「01 與 12」（文獻題名—相同、指明主題—分離型詞彙）、09（指明主題—一般）、02、17、05、19 及 10（文獻題名—首字語、指明內容—機構、文獻題名—詞彙形式、指明內容—物件、指明主題—首字語）等。依本章第一節內容所述，類別 01（文獻題名—相同）、04-06（文獻題名—同義、文獻題名—詞彙形式、文獻題名—單複數）、09-10（指明主題—一般、指明主題—首字語）、12（指明主題—分離型詞彙）與 23（指明功能—自我參考）是 8 個最經常被使用的社會標記關鍵字類別。若從 SNA 中心度觀點而言，雖然類別 23（指明功能—自我參考）是經常被使用的社會標記關鍵字類別之一，在 SNA 的程度中心度、接近中心度與中介中心度等三方面的重要性，不若類別 01（文獻題名—相同）、05（文獻題名—詞彙形式）、09-10（指明主題—一般、指明主題—首字語）與 12（指明主題—分離型詞彙）；亦即類別 23 只是經常被使用的社會標記關鍵字類別而已，不若 01（文獻題名—相同）、05（文獻題名—詞彙形式）、09-10（指明主題—一般、指明主題—首字

語) 與 12 (指明主題—分離型詞彙) 等類別, 較不具備影響其他社會標記關鍵字類別及其所擁有社會標記的使用, 具有偏向獨自使用的現象。再者, 類別 04 (文獻題名—同義) 與 06 (文獻題名—單複數) 則僅在程度與接近中心度具備較大的重要性, 並未在中介中心度具有同等的影響力, 較不具備引導使用其他社會標記關鍵字類別及其所擁有社會標記的中介功能。相較之下, 01 (文獻題名—相同)、05 (文獻題名—詞彙形式)、09-10 (指明主題—一般、指明主題—首字語) 與 12 (指明主題—分離型詞彙) 等類別不僅是經常被使用的社會標記關鍵字類別, 也在 SNA 中心度方面極具重要性, 除了類別本身及其所擁有社會標記常被使用外, 也影響其他社會標記關鍵字類別及其所擁有之社會標記的使用。儘管類別 02 (文獻題名—首字語)、17 (指明內容—機構) 與 19 (指明內容—物件) 並未列入 8 個最經常被使用的社會標記關鍵字類別之內, 然而從 SNA 中心度觀點而言, 類別 02 (文獻題名—首字語) 與 19 (指明內容—物件) 在三項中心度極具重要性, 並且影響其他社會標記關鍵字類別及其所擁有社會標記的使用; 類別 17 (指明內容—機構) 只在中介中心度具備較大的重要性, 在社會標記關鍵字類別及其所擁有社會標記的使用中, 扮演極為重要的居中功能, 具有導引使用其他社會標記關鍵字類別及其所擁有社會標記的現象。

在控制詞彙方面(請參照圖 4.8 與表 4.5), 以排名前 8 名的程度中心度而言, 包括的關鍵字類別分別是 09 (指明主題—一般)、01 (文獻題名—相同)、「04、06、12、14」(文獻題名—同義、文獻題名—單複數、指明主題—分離型詞彙、指明內容—空間)、02 (文獻題名—首字語), 及「16-18」(指明內容—人、指明內容—機構、指明內容—資源類型) 等類別。排名前 8 名接近中心度與程度中心度的控制詞彙關鍵字類別十分相似, 包括的關鍵字類別及其排名分別是 09 (指明主題—一般)、01 (文獻題名—相同)、「04、06、14」(文獻題名—同義、文獻題名—單複數、指明主題—分離型詞彙、指明內容—空間)、02 (文獻題名—首字語), 及「16-18」(指明內容—人、指明內容—機構、指明內容—資源類型)

等類別。以排名前 8 名的中介中心度而言，類別與前述的程度與接近中心度類別並不太相同，包括 09（指明主題—一般）、06（文獻題名—單複數）、01（文獻題名—相同）、12（指明主題—分離型詞彙）、04（文獻題名—同義）、02（文獻題名—首字語）、14（指明內容—空間）與 16（指明內容—資源類型）等類別。依本章第一節內容所述，類別 01（文獻題名—相同）、04-06（文獻題名—同義、文獻題名—詞彙形式、文獻題名—單複數）、09（指明主題—一般）、12（指明主題—分離型詞彙）、17（指明內容—機構）與 19（指明內容—物件）是 8 個最經常被使用的控制詞彙關鍵字類別。若從 SNA 中心度觀點而言，雖然 05、12、17 與 19（文獻題名—詞彙形式、指明主題—分離型詞彙、指明內容—機構、指明內容—物件）等類別屬於 8 個經常被使用的控制詞彙關鍵字類別，但是在 SNA 的程度中心度、接近中心度與中介中心度等三方面的重要性，不若類別 01、04、06 與 09（文獻題名—相同、文獻題名—同義、文獻題名—單複數、指明主題—一般），較不具備影響其他控制詞彙關鍵字類別及其所擁有控制詞彙的使用。相較之下，01（文獻題名—相同）、04（文獻題名—同義）、06（文獻題名—單複數）與 09（指明主題—一般）等類別不僅是經常被使用的控制詞彙關鍵字類別，也在 SNA 中心度極具重要性，除了類別本身及其所擁有控制詞彙常被使用外，也影響其他控制詞彙關鍵字類別及其所擁有控制詞彙的使用。主要差異的情形如下：

1. 雖然類別 12（指明主題—分離型詞彙）與 17（指明內容—機構）是 8 個最經常被使用的控制詞彙關鍵字類別之一，但類別 12（指明主題—分離型詞彙）在接近中心度的影響力，以及類別 17（指明內容—機構）在中介中心度的影響力，不如 01、04、06 與 09（文獻題名—相同、文獻題名—同義、文獻題名—單複數、指明主題—一般）等類別。

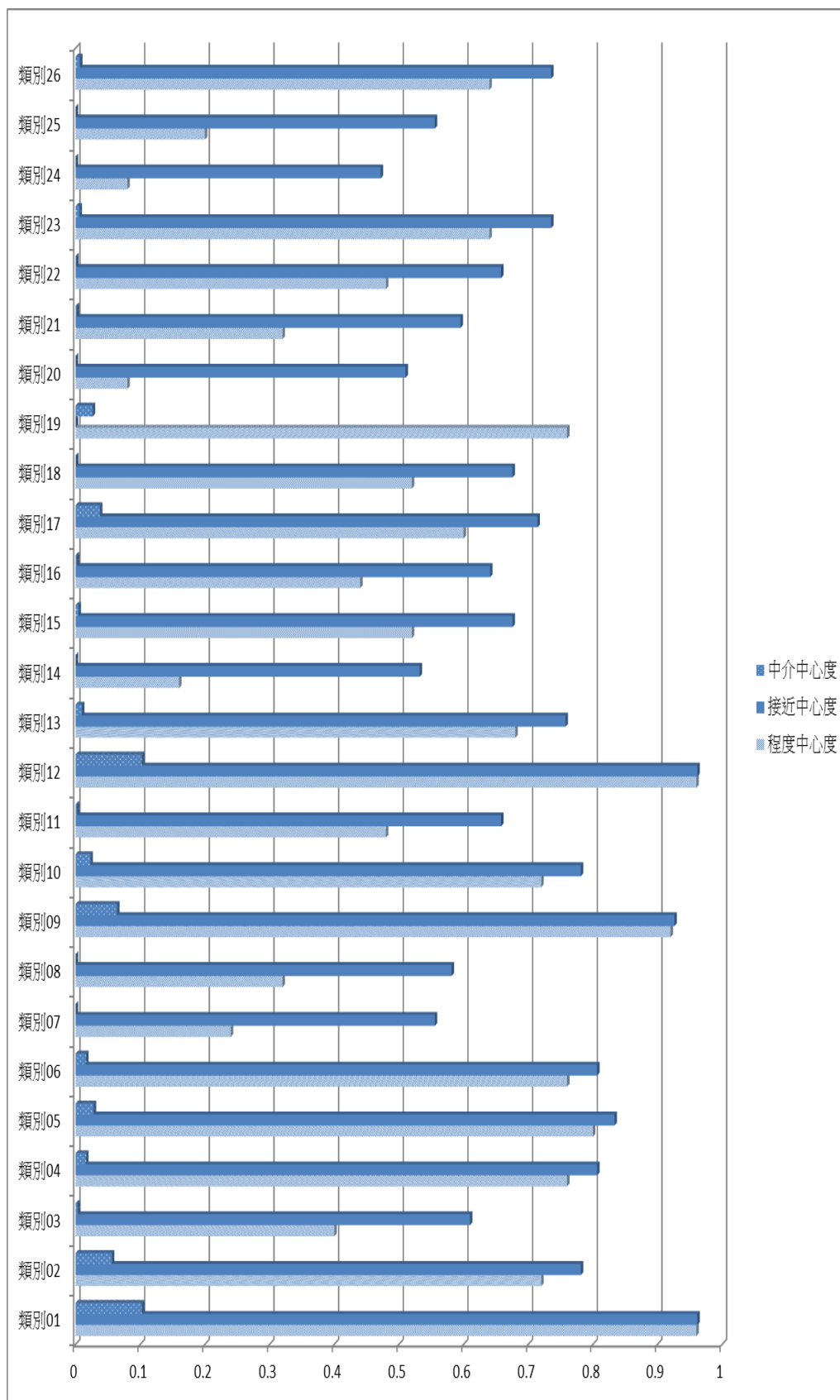


圖 4.7 社會標記關鍵字類別的三項中心度

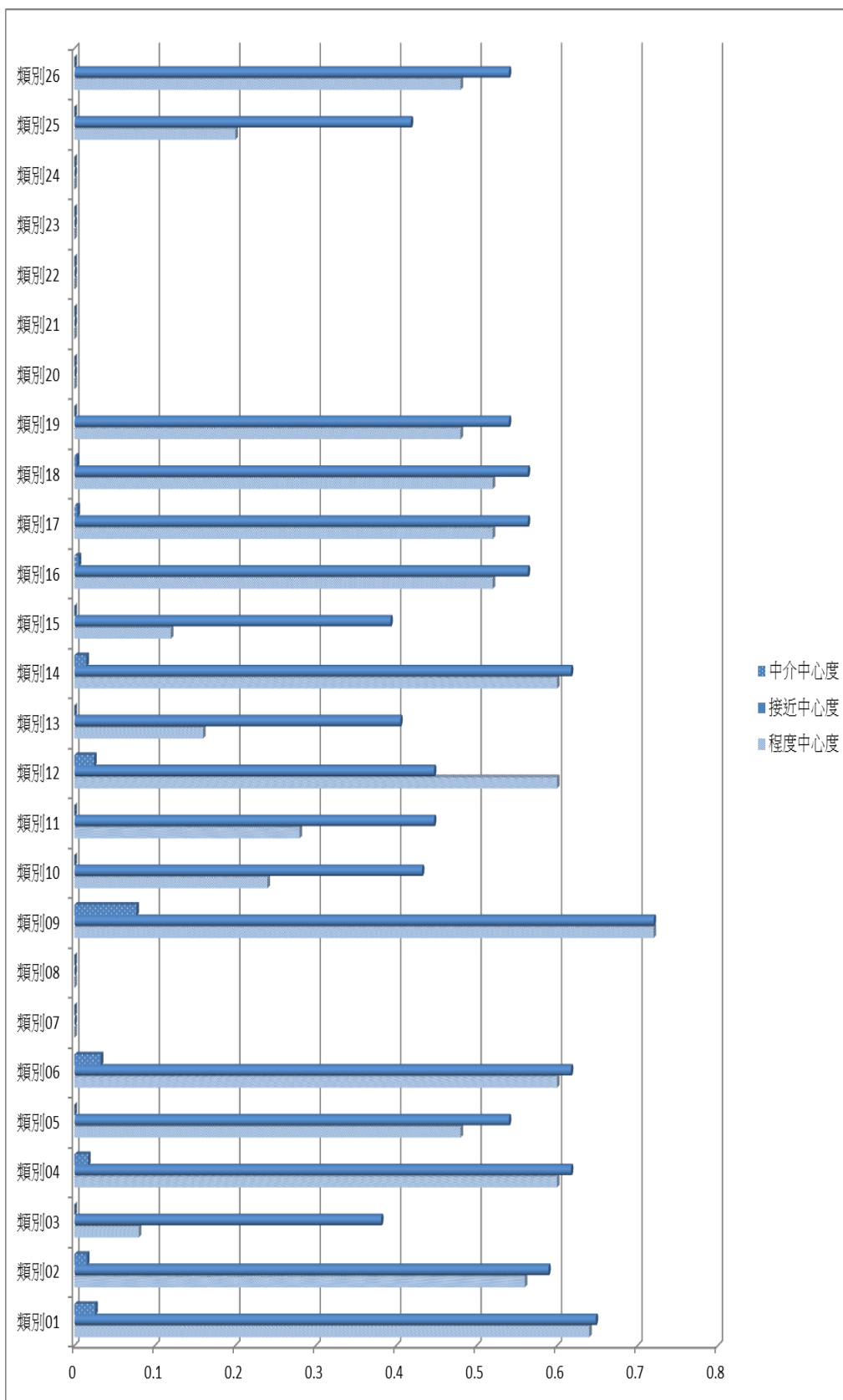


圖 4.8 控制詞彙關鍵字類別的三項中心度

表 4.5 關鍵字類別三項中心度的排名名次與係數

關鍵字類別	社會標記			控制詞彙		
	程度中心度	接近中心度	中介中心度	程度中心度	接近中心度	中介中心度
01. 文獻題名－ 相同	01 (0.960000)	01 (0.961538)	01 (0.102623)	02 (0.640000)	02 (0.648000)	03 (0.025968)
02. 文獻題名－ 首字語	08 (0.720000)	08 (0.781250)	04 (0.056055)	07 (0.560000)	06 (0.589091)	06 (0.015635)
03. 文獻題名－ 縮寫	19 (0.400000)	19 (0.609756)	15 (0.003147)	19 (0.080000)	19 (0.381176)	11 (0.000000)
04. 文獻題名－ 同義	05 (0.760000)	05 (0.806452)	09 (0.016079)	03 (0.600000)	03 (0.617143)	05 (0.017024)
05. 文獻題名－ 詞彙形式	04 (0.800000)	04 (0.833333)	06 (0.027858)	11 (0.480000)	10 (0.540000)	11 (0.000000)
06. 文獻題名－ 單複數	05 (0.760000)	05 (0.806452)	09 (0.016079)	03 (0.600000)	03 (0.617143)	02 (0.032857)
07. 文獻題名－ 非英文的同 義詞彙	22 (0.240000)	22 (0.555556)	21 (0.000000)	20 (0.000000)	20 (0.000000)	11 (0.000000)
08. 文獻題名－ 分離型詞彙	20 (0.320000)	21 (0.581395)	21 (0.000000)	20 (0.000000)	20 (0.000000)	11 (0.000000)
09. 指明主題－ 一般	03 (0.920000)	03 (0.925926)	03 (0.064290)	01 (0.720000)	01 (0.720000)	01 (0.077079)
10. 指明主題－ 首字語	08 (0.720000)	08 (0.781250)	08 (0.022920)	15 (0.240000)	15 (0.432000)	11 (0.000000)
11. 指明主題－ 縮寫	16 (0.480000)	16 (0.657895)	16 (0.002242)	14 (0.280000)	13 (0.446897)	11 (0.000000)
12. 指明主題－ 分離型詞彙	01 (0.960000)	01 (0.961538)	01 (0.102623)	03 (0.600000)	13 (0.446897)	04 (0.024500)
13. 指明內容－ 品質或特性	10 (0.680000)	10 (0.757576)	11 (0.009797)	17 (0.160000)	17 (0.405000)	11 (0.000000)
14. 指明內容－ 空間	24 (0.160000)	24 (0.531915)	21 (0.000000)	03 (0.600000)	03 (0.617143)	07 (0.014857)
15. 指明內容－ 時間	14 (0.520000)	14 (0.675676)	14 (0.004444)	18 (0.120000)	18 (0.392727)	11 (0.000000)
16. 指明內容－ 人	18 (0.440000)	18 (0.641026)	17 (0.002004)	08 (0.520000)	07 (0.563478)	08 (0.005333)
17. 指明內容－ 機構	13 (0.600000)	13 (0.714286)	05 (0.037649)	08 (0.520000)	07 (0.563478)	09 (0.003889)

關鍵字類別	社會標記			控制詞彙		
	程度中心度	接近中心度	中介中心度	程度中心度	接近中心度	中介中心度
18. 指明內容－資源類型	14 (0.520000)	14 (0.675676)	20 (0.000556)	08 (0.520000)	07 (0.563478)	10 (0.002857)
19. 指明內容－物件	05 (0.760000)	05 (0.806452)	07 (0.026352)	11 (0.480000)	10 (0.540000)	11 (0.000000)
20. 指明內容－使用對象	25 (0.080000)	25 (0.510204)	21 (0.000000)	20 (0.000000)	20 (0.000000)	11 (0.000000)
21. 指明內容－資源來源	20 (0.320000)	20 (0.595238)	18 (0.001733)	20 (0.000000)	20 (0.000000)	11 (0.000000)
22. 指明功能－任務	16 (0.480000)	16 (0.657895)	19 (0.000884)	20 (0.000000)	20 (0.000000)	11 (0.000000)
23. 指明功能－自我參考	11 (0.640000)	11 (0.735294)	13 (0.005958)	20 (0.000000)	20 (0.000000)	11 (0.000000)
24. 其他－不用標記	25 (0.080000)	26 (0.471698)	21 (0.000000)	20 (0.000000)	20 (0.000000)	11 (0.000000)
25. 其他－複合式	23 (0.200000)	22 (0.555556)	21 (0.000000)	16 (0.200000)	16 (0.418065)	11 (0.000000)
26. 其他－無法識別	11 (0.640000)	11 (0.735294)	12 (0.006708)	11 (0.480000)	10 (0.540000)	11 (0.000000)

2. 類別 05 與 19 (文獻題名－詞彙形式、指明內容－物件) 也是 8 個最經常被使用的控制詞彙關鍵字類別之一，在程度、接近與中介等三項中心度影響力皆不如 01、04、06 與 09 (文獻題名－相同、文獻題名－同義、文獻題名－單複數、指明主題－一般) 等類別。

另一方面，類別 02、14、16 與 18 (文獻題名－首字語、指明內容－空間、指明內容－人、指明內容－資源類型) 雖然未列入 8 個經常被使用的控制詞彙關鍵字類別之內，從 SNA 中心度觀點而言，類別 02、14 與 16 (文獻題名－首字語、指明內容－空間、指明內容－人) 則在三項中心度極具重要性，影響其他控制詞彙關鍵字類別及其所擁有控制詞彙的使用。至於類別 18 (指明內容－資源類型)，只在程度與接近中心度具備較大的重要性。

就排名前 8 名的中心度而言，社會標記者三項中心度在前 8 名者涵蓋了 3 至 5 項題名群組、3 項主題群組，以及 1 至 2 項內容群組的關鍵字類別，而專家三項中心度皆在前 8 名者涵蓋了 4 項題名群組、1 至 2 項主題群組與 2 至 4 項內容群組的關鍵字類別。這意謂著社會標記者除了傾向經常選取題名與主題群組關鍵字類別及其所擁有的關鍵字外，也影響了其他關鍵字類別及其所擁有關鍵字的使用。然而，專家則是傾向經常使用題名群組關鍵字類別及其所擁有的關鍵字外，也影響了其他關鍵字類別及其所擁有關鍵字的選取。在選取文獻題名關鍵字時，社會標記者僅就文獻題名的相同、同義與詞彙形式等關鍵字進行組織，而專家則是選取文獻題名的相同、首字語、同義與單複數等不同關鍵字標引資訊物件之用。換言之，專家仍遵守權威控制與索引典規範原則的指引，及以直接採用官方來源的既有控制詞彙為優先原則，謹慎選取與文獻題名相關關鍵字的控制詞彙，而社會標記者也可能具備類似資訊組織背景，但不若專家對權威控制與索引典規範來熟悉與瞭解，以及採用官方維護的既有控制詞彙。例如 NISO (2005) 的「建置控制詞彙指引」規範中，對於首字語、同義、單複數等關鍵字的選取與維護皆有關相關的原則說明與規定，以指引專家選用適當的題名關鍵字。另一方面，雖然專家與社會標記者皆經常使用主題群組關鍵字類別及其所擁有的關鍵字外，專家不若社會標記者影響其他關鍵字類別及其所擁有關鍵字的使用，造成社會標記者與專家在選取關鍵字上的此種差異，其潛在原因可能在於社會標記者與專家皆受到權威控制與索引典控制等規範的原則影響，只是影響不同，專家更專注且單獨標引資訊物件的主題，達成主題分析的目的。除此之外，在權威控制與索引典控制等規範的影響下，專家也較著重在資訊物件的內容標引，包括空間、人與物件等名稱。如，中國編目規則乙編標目第二十一章檢索款目之擇定的編目規範條文（即 21.1.1.1 至 21.1.1.2 編目規則）中，即列出人名、團體與地名等標目的選取原則，在 NISO (2005) 的「建置控制詞彙指引」也有類似的指引原則，同時一般的權威控制除了將主題標目列為重點項之一外，也將人名、團體與地名等標目列為權威款目的管理範圍之內。

就 26 種關鍵字類別的 SNA 中心度而言，社會標記與控制詞彙間的三項 SNA 中心度全然不同。如果改從 SNA 中心度的排名位置而言，只有少數的關鍵字類別的 SNA 中心度排名位置是相同。從上述討論，在關鍵字類別的 SNA 中心度排名位置方面，可以推論社會標記者與專家間的心智模式相似性介於 3.8% ($1 \div 26$) 至 19.23% ($5 \div 26$) 的範圍之間，屬於「輕度一致性」的範疇。至於三項 SNA 中心度相同排名位置的關鍵字類別如下：

1. 程度中心度：計有類別 03、08、21 與 26（文獻題名—縮寫、文獻題名—分離型詞彙、指明內容—資源來源、無法識別）。
2. 接近中心度：只有類別 03 與 21（文獻題名—縮寫、指明內容—資源來源）。
3. 中介中心度：只有類別 13（指明內容—品質或特性）。

如果從 8 個最高的 SNA 中心度為基礎進行分析，其結果可以代表另外一種關鍵字類別的實際使用現象。即使在 8 個最高 SNA 中心度的範圍中，社會標記與控制詞彙的關鍵字類別排名位置皆不相同，但是有幾個關鍵字類別是共通的。以 SNA 的程度與接近中心度而言，有 6 個最高 SNA 中心度的關鍵字類別是相同的，包括類別 01-02、04、06、09 與 12（文獻題名—相同、文獻題名—首字語、文獻題名—同義、文獻題名—單複數、指明主題—一般、指明主題—分離型詞彙），百分比高達 75%。以 SNA 的中介中心度而言，有 4 個最高 SNA 中心度的關鍵字類別是相同的，包括類別 01-02、09 與 12（文獻題名—相同、文獻題名—首字語、指明主題—一般、指明主題—分離型詞彙），相似性高達 50%。因而，以 8 個最高 SNA 中心度的範圍而言，社會標記者與專家間之文獻標引心智模式的相似性介於 50% 與 75% 之間，屬於「中度一致性」的範疇。

第四節 SNA 階層集叢分析

SNA 除了可使用前述三項中心度分析各關鍵字類別的影響力外，也可導入

階層集叢分析，進一步檢視哪些關鍵字類別的使用上，呈現集叢的屬性現象。換言之，在文獻標引過程中，有哪些關鍵字類別及其所擁有的關鍵字共同出現，或是經常一起被選取使用。這也意謂著集叢內的某一關鍵字類別一旦被使用時，屬於同一集叢內的其他關鍵鍵字類別及其所擁有關鍵字將被選取作為社會標記或控制詞彙之用。如圖 4.9 所示，在社會標記的關鍵字類別方面，共產生 8 組集叢，每一集叢分別由 2 至 7 個成員所組成，其中有 6 組集叢皆由 2 個成員所組合而成。在控制詞彙方面，將關鍵字類別分為 13 組的情形下，產生 3 組集叢，每一集叢分別由 2、4 與 10 等不同成員數所構成（請參見圖 4.10）。

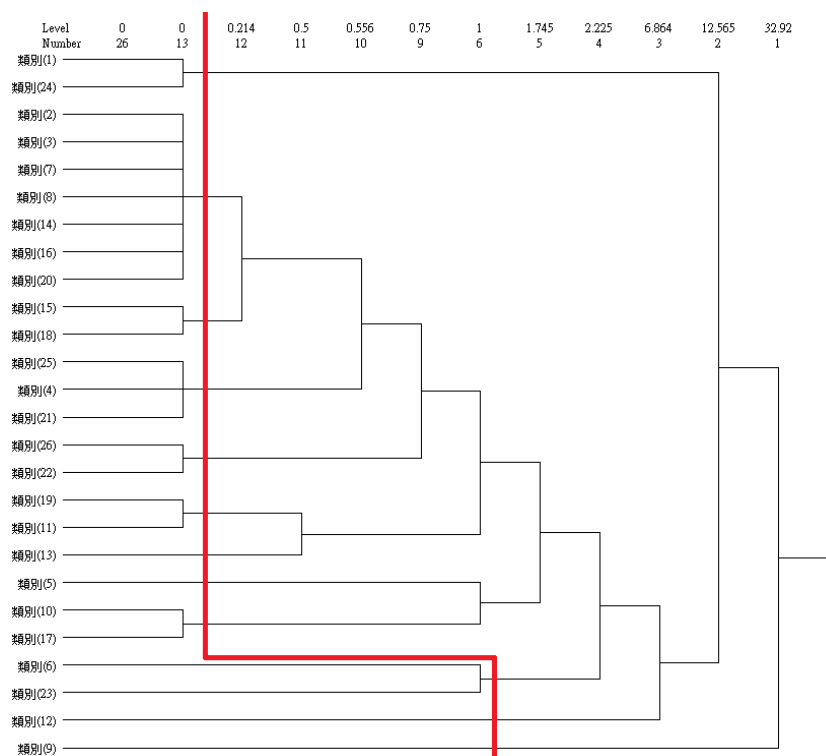


圖 4.9 社會標記關鍵字類別的 SNA 階層集叢多維尺度圖

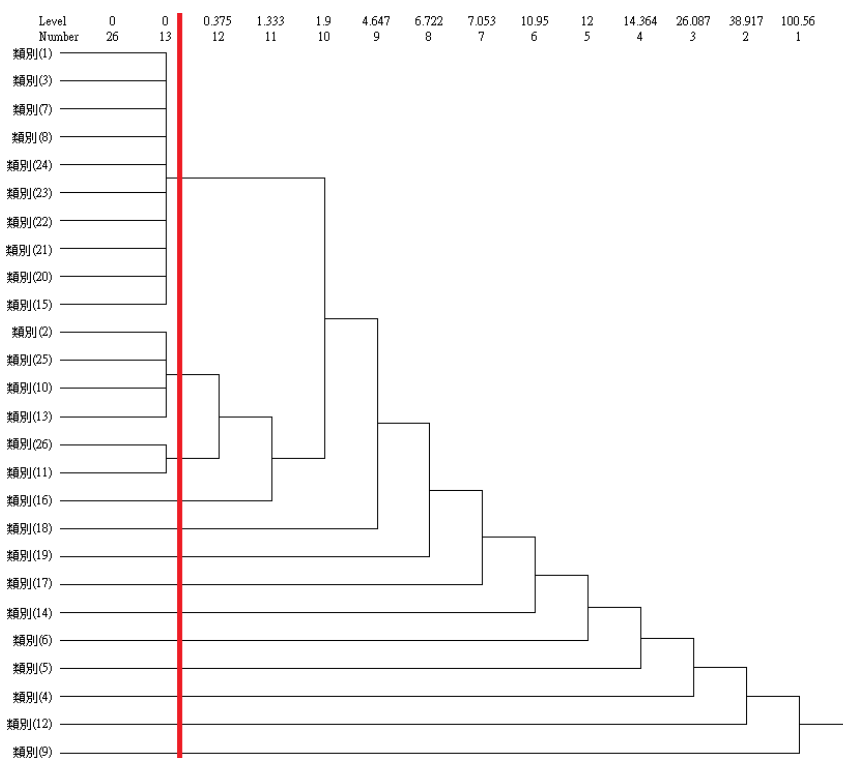


圖 4.10 控制詞彙關鍵字類別的 SNA 階層集叢多維尺度圖

SNA 階層集叢分析的結果對於社會標記者與專家如何一起使用哪些關鍵字類別的現象，對資訊組織的關鍵字標引而言，具有極大的功用與意涵。在社會標記方面，共產生 8 組的 SNA 階層集叢，而且多數由不同屬性的群組所共同組合而成（請參見表 4.6），其中內容群組的關鍵字類別最常被組合成 SNA 階層集叢，其頻率高達 5 次。第二個最常被組合成 SNA 階層集叢的社會標記關鍵字類別，則是題名群組與其他群組的關鍵字類別，頻率為 4 次。上述分析結果意謂著社會標記者經常選取內容、題名或其他群組的關鍵字類別及其所擁有的關鍵字與其他類別的關鍵字搭配使用外，其中題名群組與內容群組常一起共同使用，而內容群組也常和主題群組的關鍵字類別及其所擁有的關鍵字搭配使用。除此之外，社會標記者也常一起使用題名群組的關鍵字類別及其所擁有的關鍵字，內容群組亦有相同的現象。在控制詞彙方面，產生的 SNA 階層集叢數量少於社會標記，共有 3 組，且組成 SNA 階層集叢的成員數高於社會標記的組成成員數（即 2 個），成員數分別為 2、4 與 10。再者，多數控制詞彙的階層集叢成員係由 2 種以上的不同屬性群組組合而成（請參見表 4.7），其中其他群組的關鍵字類別最常被組合成

SNA 階層集叢，其頻率高達 3 次。這也透露著其他群組的關鍵字類別及其所擁有的關鍵字雖不常被專家單獨使用作為文獻標引之用，卻常與題名或內容等群組之一及其所擁有的關鍵字一起搭配使用。另外，專家常一起使用題名與內容群組的關鍵字類別及其所擁有的關鍵字外，也常一起搭配使用不同類型的題名或不同類型的內容關鍵字類別及其所擁有的關鍵字。依上述結果，顯示出專家較社會標記者更著重在主題的分析，所以才會經常一起使用題名、主題與內容群組的關鍵字類別及其所擁有的關鍵字，而社會標記者則是一起搭配使用題名、內容與其他群組的關鍵字類別及其所擁有的關鍵字。

以 SNA 階層集叢分析所顯示的共同出現或使用關鍵字類別而言，整體上社會標記者傾向以較多組且較少的層面分析觀點提供關鍵字，進行文獻標引，同時多數層面分析幾乎為兩種不同的觀點。然而，專家則是傾向使用較少組且較多的層面分析觀點提供關鍵字，且多數層面分析超過兩種以上的觀點，以組織各項資訊資源。另外，社會標記者也具備使用同質關鍵字類別，以形成 SNA 階層集叢，亦即社會標記者會同時使用單一類型群組的關鍵字類別及其所擁有的關鍵字，進行資訊資源的主題標引，專家也具備此一傾向。以資訊組織的觀點而言，專家可能受到層面分析理論的影響，採取較多且不同的面向進行文獻標引，以標引資訊物件的不同屬性特徵，並專心且連續標引資訊物件的屬性。然而，社會標記者則是採取分散、跳躍方式組織各項資訊物件。就關鍵字類別的共同使用現象而言，社會標記者與專家間之文獻標引心智模式的相似性幾乎完全不同，屬於「輕度一致性」的範疇。

表 4.6 共同出現的社會標記關鍵字類別及其次數

社會標記關鍵字類別	題名群組	主題群組	內容群組	功能群組	其他群組
01+24	● (1)				● (1)
02+03+07+08+14+16+20	● (4)		● (3)		
15+18			● (2)		

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

社會標記關鍵字類別	題名群組	主題群組	內容群組	功能群組	其他群組
25+04+21	● (1)		● (1)		● (1)
26+22				● (1)	● (1)
19+11		● (1)	● (1)		
10+17		● (1)	● (1)		
06+23	● (1)				● (1)
總計：不重複出現的次數	4	2	5	1	4

備註：圓括弧內數字代表出現次數

表 4.7 共同出現的控制詞彙關鍵字類別及其次數

控制詞彙關鍵字類別	題名群組	主題群組	內容群組	功能群組	其他群組
01+03+07+08+24+23+22+21+ 20+15	● (4)		● (3)	● (2)	● (1)
02+25+10+13	● (1)	● (1)	● (1)		● (1)
26+11		● (1)			● (1)
總計：不重複出現的次數	2	2	2	1	3

備註：圓括弧內數字代表出現次數。此外，實際上功能群組的關鍵字並未被專家使用，即使 SNA 階層集叢分析出功能群組，亦不能納入集叢成員之一。

第五節 SNA 同等角色分析

以 SNA 所產生的網路結構圖為基礎，除了可以進行中心度與階層集叢分析外，亦可從既有之 SNA 網路圖形的結構中，找出同等地位的功能角色，藉以分析與歸納出有哪些特定的關鍵字類別共同鏈結相同或類似的其他關鍵字類別。換言之，以 SNA 網路圖形的結構而言，這些特定的關鍵字類別扮演相同或類似的功能定位，彼此間是可以相互交換或替代。就文獻標引而言，意謂著社會標記者或專家在選取關鍵字標引資訊物件時，這些關鍵字類別及其所擁有的關鍵字經常

被社會標記者或專家交替運用。在社會標記方面，共有 10 組，而且皆由 2 種關鍵字類別所組成（請參見圖 4.11）。另外在控制詞彙方面，則有 8 組，其中 6 組係由 2 種關鍵字類別所組成，其餘則分別由 3 與 7 種關鍵字類別所構成（請參見圖 4.12）。

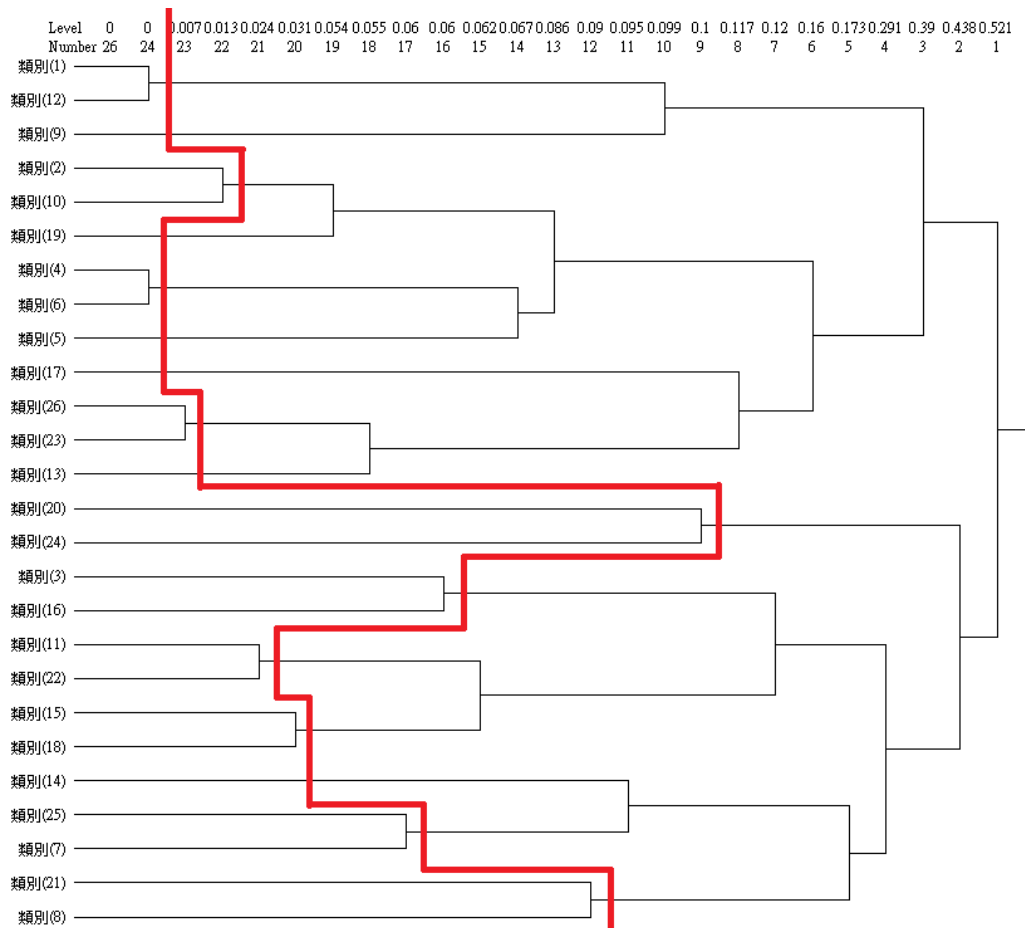


圖 4.11 社會標記關鍵字類別的 SNA 同等角色多維尺度圖

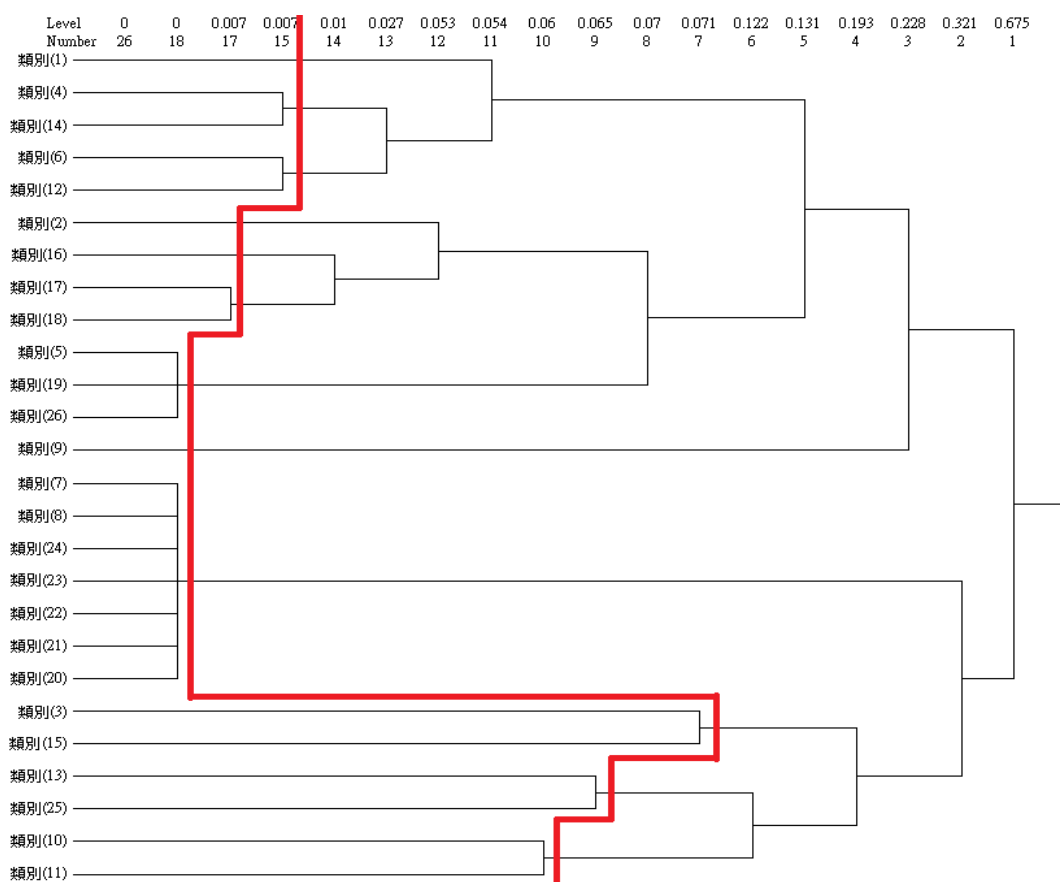


圖 4.12 控制詞彙關鍵字類別的 SNA 同等角色多維尺度圖

以 SNA 圖的網路結構而言，有些節點共同鏈結至相同或幾乎相同的節點，因而這些節點是可以相互交換或替代。以 SNA 同等角色的分析結果而言，在社會標記方面，共產生 10 對的組合（請參見表 4.8），而且這 10 對組合中，幾乎由不同屬性的群組所共同組合而成，只有 2 對是同質屬性的群組所構成，分別是題名與內容群組。在 10 對的同等角色組合中，以題名群組出現頻率最高，其頻率為 6（60%）。排名第二與第三的 SNA 同等角色的組合成員，分別是內容，以及主題與其他群組，其頻率分別為 4 與 3（40% 與 30%）。在控制詞彙方面，共產生 8 組的同等角色組合（請參見表 4.9），其中 6 組係由 2 種不同屬性的群組所共同組合成對，只有 2 組成員數超過 2，成員數分別是 3 與 7。在 6 對組合中，只有 3 對是同質屬性的群組所構成，分別是題名、內容與主題群組。另外，在 8 組的同等角色組合中，以題名、內容及其他群組出現頻率最高，其頻率分別為 5

(83.33%)、5 (83.33%) 與 3 (33.33%)，代表著題名、內容及其他等群組是 3 個最常出現在 SNA 同等角色的組合成員。由以上討論，可以歸納出下列要點：

表 4.8 同等角色的社會標記關鍵字類別及其次數

社會標記關鍵字類別	題名群組	主題群組	內容群組	功能群組	其他群組
01+12	● (1)	● (1)			
02+10	● (1)	● (1)			
04+06	● (2)				
26+23				● (1)	● (1)
20+24			● (1)		● (1)
03+16	● (1)		● (1)		
11+22		● (1)		● (1)	
15+18			● (2)		
25+07	● (1)				● (1)
21+08	● (1)		● (1)		
總計：刪除重複出現次數	6	3	4	2	3

備註：圓括弧內數字代表出現次數。

表 4.9 同等角色的控制詞彙關鍵字類別及其次數

控制詞彙關鍵字類別	題名群組	主題群組	內容群組	功能群組	其他群組
01+04	● (2)				
06+12	● (1)	● (1)			
17+18			● (2)		
05+19+26	● (1)		● (1)		● (1)
07+08+24+23+22+21+20	● (2)		● (2)	● (2)	● (1)
03+15	● (1)		● (1)		
13+25			● (1)		● (1)
10+11		● (2)			
總計：刪除重複出現次數	5	2	5	1	3

備註：圓括弧內數字代表出現次數。此外，實際上功能群組的關鍵字並未被專家使用，即使 SNA 同等角色分析出功能群組，亦不能納入集叢成員之一。

1. 以 SNA 同等角色的分析結果而言，社會標記的組合成員數皆為 2，而控制詞彙則趨於多元化，成員數分別是 2、3 與 7，其中以 2 個成員佔多數。這項分析結果意謂著社會標記者經常交替使用兩種關鍵字類別及其所擁有的關鍵字，並搭配其他類別及其所擁有的關鍵字進行文獻標引，而專家則是較為多元化，經常交替使用 2 至 7 種關鍵字類別及其所擁有的關鍵字。
2. 以 SNA 同等角色的組合而言，社會標記以題名群組的關鍵字類別最常出現或被組合成 SNA 同等角色，而控制詞彙則以題名與內容群組的關鍵字類別最常出現或被組合成 SNA 同等角色。上述結果透露出社會標記者除了自文獻題名中直接取用關鍵字外，也交替使用不同形式的文獻題名關鍵字進行文獻標引，而專家則是交替使用題名與內容群組的關鍵字類別及其所擁有的關鍵字標引資訊。
3. 在同質的 SNA 同等角色且組合成員為 2 的前提下，社會標記集中在題名與內容群組的關鍵字類別，而控制詞彙則是集中在題名、內容與主題群組的關鍵字類別。
4. 以異質的 SNA 同等角色組合數量或頻率而言，社會標記者較專家傾向交替使用不同的異質群組關鍵字組合成 SNA 同等角色的組合成員，以鏈結至相同或幾乎雷同的其他關鍵字類別。

以資訊組織的觀點而言，可能造成上述前 3 點的差異在於專家遵循層面分析的原理，試從不同面向標引資訊物件的屬性，同時也受到權威控制與索引典控制之類規範（如，NISO（Anderson, 1997）與 Getty Research Institute 的各項索引典編輯指引）的影響，指引專家依一定原則謹慎選取文獻題名相關形式的關鍵字外，也引導專家標引機構、地名空間、時間，及註明資源類型、物件、使用對象與資源來源等。至於造成上述第 4 點的差異，其潛在原因可能源自於資訊組織強調主題分析的重要性，而影響專家會傾向單獨選用主題群組的關鍵字類別及其所

擁有的關鍵字標引資訊物件的不同主題屬性，較少與其他群組的關鍵字類別及其所擁有的關鍵字一起交替使用。

就 SNA 同等角色之集叢組數及其組成成員而言，社會標記與控制詞彙間的結果完全不相同。再者，就最常出現的 SNA 同等角色集叢之群組而言，只有題名群組完全相同。三則，就組合的 SNA 角色集叢的群組而言，只有 2 種類型的組合方式相同（即：題名與主題群組、題名與內容群組）。第四，在 SNA 同等角色集叢的同質群組中，只有 2 個群組相同（即：題名與內容群組）。整體而言，在 SNA 同等角色方面而言，社會標記者與專家間之文獻標引心智模式的相似性並不高，屬於「輕度一致性」的範疇。

第六節 FP-tree 分析

除了 SNA 方法外，本研究也採取 FP-tree 方法，建立路徑式規則的心智模式地圖，以探索與分析社會標記與控制詞彙之關鍵字使用情形。在文獻標引心智模式地圖的擷取與建構方面，首先本研究將 FP-tree 的支持度與信賴度皆設定為 0，期能完整描繪出文獻標引心智模式的地圖內容。在社會標記關鍵字類別方面，共產生 206 條規則，且規則分散在 21 個集叢（請參見圖 4.13 與附錄一）。如果依使用頻率而言，以類別 01（文獻題名—相同）與 09（指明主題—一般）為起始點所構成的 FP-tree 集叢是兩個最經常使用的集叢。若從路徑式規則的階層而言，最短路徑為單一階層，且分屬在 8 個集叢。換言之，共有 8 個集叢係為單一階層的路徑式規則，也代表著這些集叢只由單一社會標記的關鍵字類別所組成，這些類別包括：03、07、11、16、18、21、22 與 24（文獻題名—縮寫、文獻題名—非英文的同義詞彙、指明主題—縮寫、指明內容—人、指明內容—資源類型、指明內容—資源來源、指明功能—任務、不用標記）。在最長路徑方面，只有 1 條路徑式規則，係由 7 種不同的社會標記關鍵字類別所組成，其組成類別與階層如后：01→09→12→13→17→16→21（文獻題名—相同→指明主題—一般→指

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

明主題—分離型詞彙→指明內容—品質或特性→指明內容—機構→指明內容—人→指明內容—資源來源)。就路徑式規則的組成階層而言，平均每一規則係由 3.00 個階層所組成 (618÷206)，亦即至少由 3 種不同的關鍵字類別及其擁有的關鍵字所組合而成。如果本研究以關鍵字類別屬性為依據，予以重新歸納群組，可以發現以每一種群組為路徑規則起始點所形成集叢的規則涵蓋情形如下：

1. 題名群組：由類別 01 至 08 所組成，共有 126 條路徑式規則，規則涵蓋率為 61.17%。
2. 主題群組：由類別 09 至 12 所組成，共有 65 條路徑式規則，規則涵蓋率為 31.55%。
3. 內容群組：由類別 13 至 21 所組成，共有 6 條路徑式規則，規則涵蓋率為 2.91%。
4. 功能群組：由類別 22 至 23 所組成，共有 6 條路徑式規則，規則涵蓋率為 2.91%。
5. 其他群組：由類別 24 至 26 所組成，共有 3 條路徑式規則，規則涵蓋率為 1.46%。

在控制詞彙方面，共產生 222 條規則，且規則分佈在 7 個集叢（請參見圖 4.14 與附錄二）。如果依使用頻率而言，以類別 09（指明主題—一般）與 01（文獻題名—相同）為起始點所構成的 FP-tree 集叢是兩個最經常使用的集叢。若從路徑式規則的階層而言，最短路徑為兩個階層，而且只有 1 個集叢。換言之，只有 1 個集叢係為兩個階層的路徑式規則，也代表著這些集叢只由兩個控制詞彙的關鍵字類別所組成，此一兩階層的路徑式規則組合情形為：類別 17→02（指明內容—機構→文獻題名—首字語）。在最長路徑方面，只有 1 條 8 個階層的路徑式規則，係由 8 種不同的控制詞彙關鍵字類別所組成，其組成類別與階層如后：09→12→04→06→19→14→02→18（指明主題—一般→指明主題—分離型詞彙→文獻題名—同義→文獻題名—單複數→指明內容—物件→指明內容—空間→文

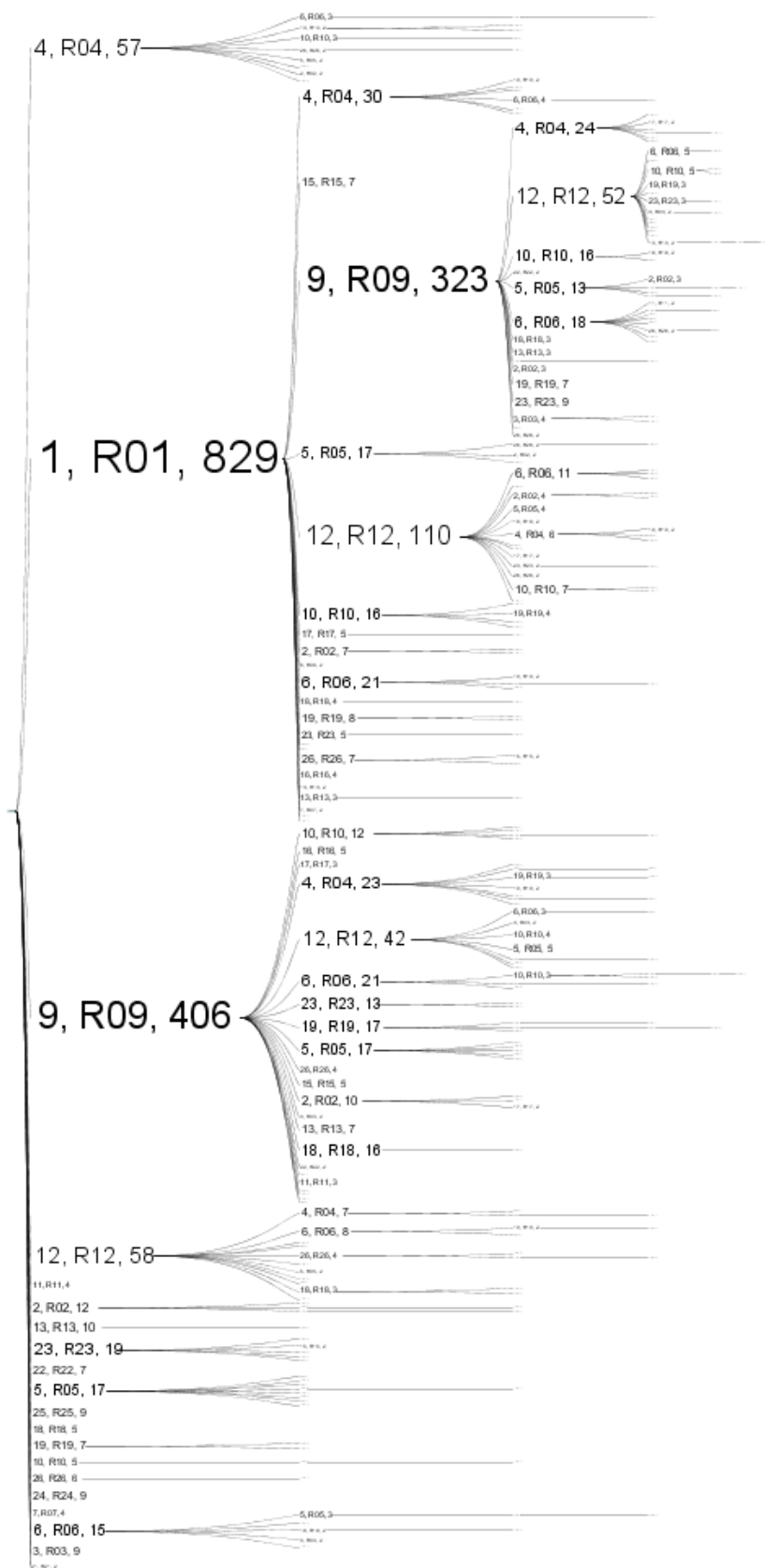


圖 4.13 社會標記關鍵字類別的 FP-tree 使用分析

備註：FP-tree 圖的三變項組合 (x, y, z) 分別代表規則序號、關鍵字類別與使用頻率。

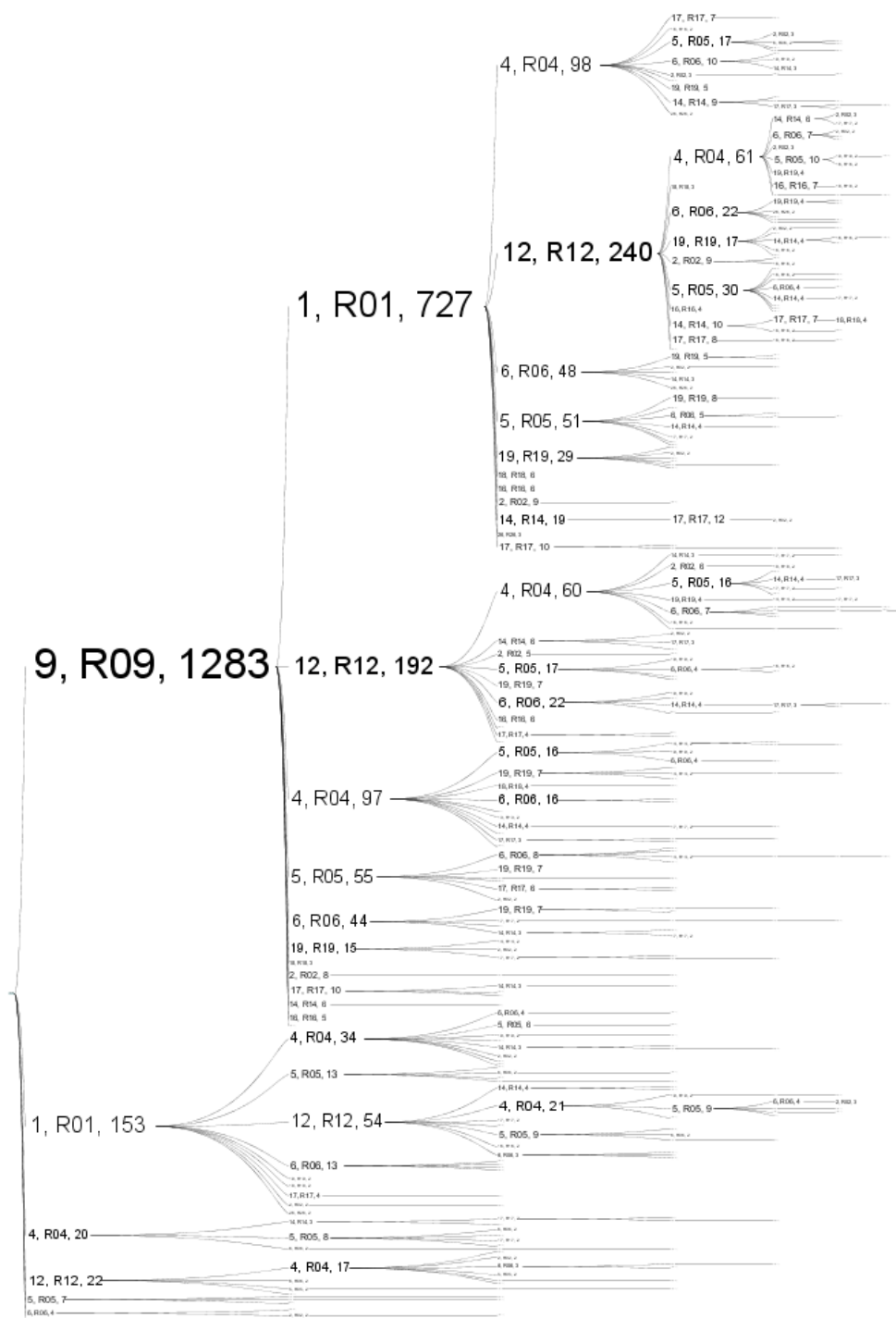


圖 4.14 控制詞彙關鍵字類別的 FP-tree 使用分析

備註：FP-tree 圖的三變項組合 (x, y, z) 分別代表規則序號、關鍵字類別與使用頻率。

獻題名—首字語→指明內容—資源類型)。就路徑式規則的組成階層而言，平均每一規則係由 4.33 個階層所組成 (962 ÷ 222)，亦即至少由 4 種不同的關鍵字類別及其擁有的關鍵字共同組成。如果本研究以關鍵字類別屬性為依據，予以重

新歸納群組，可以發現以每一種群組為路徑規則起始點所形成集叢的規則涵蓋情形如下：

1. 主題群組：由類別 09 至 12 所組成，共有 169 條路徑式規則，規則涵蓋率為 76.13%。
2. 題名群組：由類別 01 至 08 所組成，共有 52 條路徑式規則，規則涵蓋率為 23.42%。
3. 內容群組：由類別 13 至 21 所組成，共有 1 條路徑式規則，規則涵蓋率為 0.45%。

本研究更進一步以附錄一與附錄二為依據，以關鍵字類別屬性作為路徑式規則的聚合，重新歸納路徑式規則的樣式，產生以關鍵字使用分析為基礎所建構而成的文獻標引心智模式地圖。例如，在附錄一的 002 與 003 等 2 條規則，即可合併成單一樹狀結構。換言之，在內容群組（content）之下，可以再進一步劃分成「內容群組→內容群組」（Content→Content）與「內容群組→功能群組」（Content→Fucntion），而此兩種樣式的路徑式規則全部隸屬在內容群組此一樹狀結構。依此類推，社會標記一共可以產生 5 種群組 89 種樣式的文獻標引心智模式地圖（請參見附錄三與圖 4.15）。就路徑式規則樣式階層的平均數量而言，平均每一樣式係由 3.29（ $293 \div 89$ ）個階層的路徑式規則所組成，亦即至少由 3 種不同的關鍵字類別共同組合而成。至於社會標記者之文獻標引心智模式地圖的樣式分佈情形如下：

1. 題名群組：由附錄三的序號 10 至 57 所組成，共有 48 種樣式，樣式涵蓋率為 53.93%。
2. 主題群組：由附錄三的序號 58 至 89 所組成，共有 32 種樣式，樣式涵蓋率為 35.96%。
3. 內容群組：由附錄三的序號 01 至 03 所組成，共有 3 種樣式，樣式涵蓋率為 3.37%。

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

4. 功能群組：由附錄三的序號 04 至 07 所組成，共有 4 種樣式，樣式涵蓋率為 4.49%。
5. 其他群組：由附錄三的序號 08 至 09 所組成，共有 2 種樣式，樣式涵蓋率為 2.25%。

同上，依此類推，以附錄二進行歸納與聚合，控制詞彙總共可以產生 3 種群組與 88 種樣式的文獻標引心智模式地圖（請參見附錄四與圖 4.16）。就路徑式規則樣式階層的平均數量而言，平均每一樣式係由 4.67（ $411 \div 88$ ）個階層的路徑式規則所組成，亦即至少由 4 種不同的關鍵字類別群組所組合而成。至於專家之文獻標引心智模式地圖的樣式分佈情形如下：

1. 主題群組：由附錄四的序號 26 至 88 所組成，共有 63 種樣式，樣式涵蓋率為 71.59%。
2. 題名群組：由附錄四的序號 02 至 25 所組成，共有 24 種樣式，樣式涵蓋率為 27.27%。
3. 內容群組：由附錄四的序號 01 所組成，共有 1 種樣式，樣式涵蓋率為 1.14%。

以 FT Tree 路徑式規則的集叢數量而言，代表著社會標記者傾向使用較多種不同的關鍵字類別，而專家則傾向集中使用較少的關鍵字類別及其所擁有的關鍵字。如果以屬性的群組加以聚合後，可以發現社會標記者使用的關鍵字類別依序集中在題名、主題、內容、功能與其他等 5 個相關群組，而專家則依序集中在主題、題名與內容等 3 個相關群組，這也交叉呼應社會標記者傾向使用較多的關鍵字類別與群組，而專家則是傾向使用較少數的關鍵字類別與群組，並且社會標記者集中選用題名群組的關鍵字類別及其所擁有的關鍵字，專家則是集中選用主題群組的關鍵字類別及其所擁有的關鍵字。造成上述差異的原因可能源自於資訊組織的主題分析過程中，權威控制與索引典控制等規範的原則指引專家專注選題

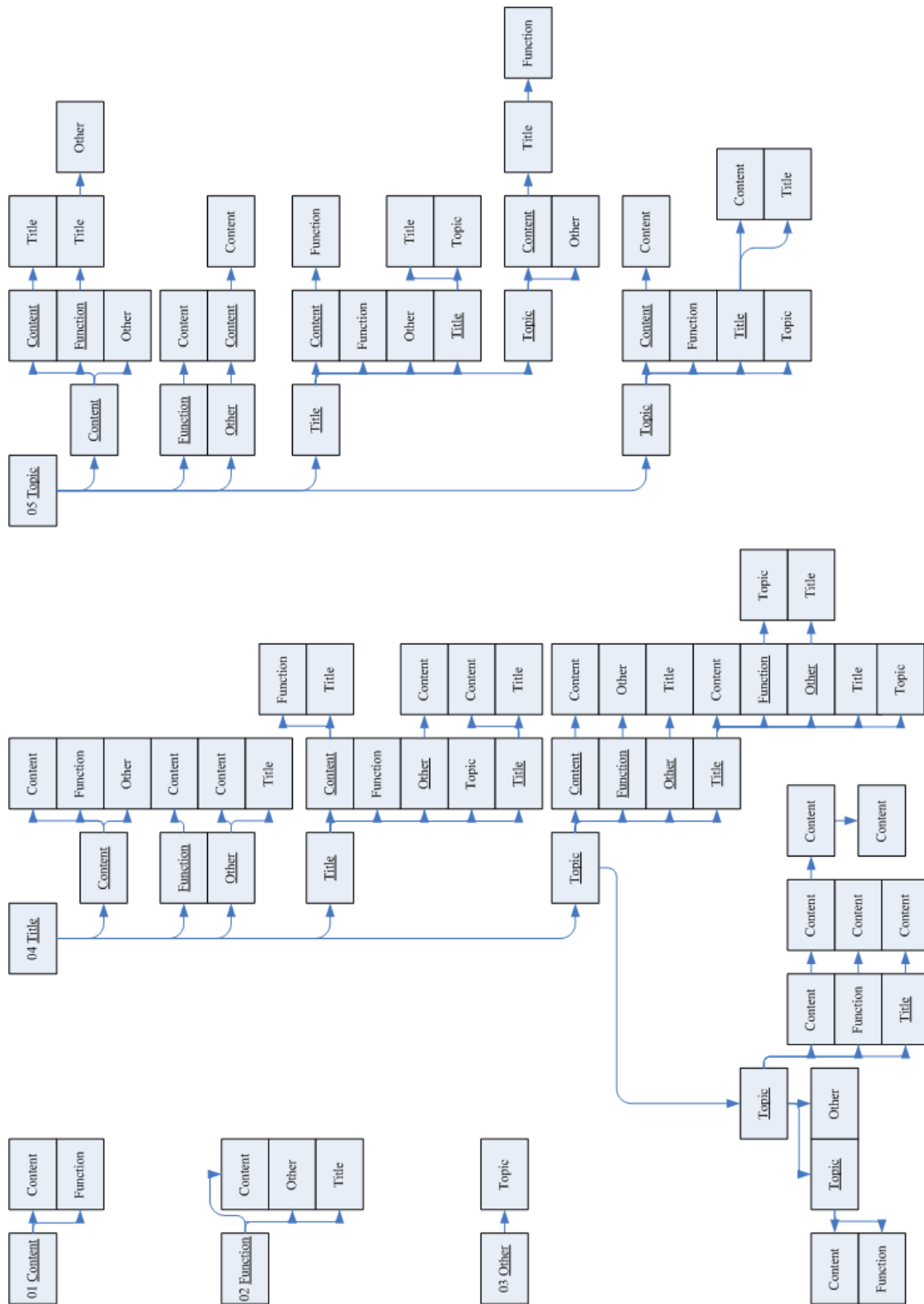


圖 4.15 社會標記者使用關鍵字的路徑式決策樹

名外的主題與內容群組等類型關鍵字作為檢索款目，並且引導專家謹慎使用不同形式的題名關鍵字外，以及直接採用官方既有的控制詞彙與減少使用複合式關鍵字（如，前述 NISO 的建置控制詞彙指引）標引資訊物件。同時，專家也熟悉資訊系統的個人化功能，避免使用指明功能類別的關鍵字。

再者，以最短路徑式規則方面而言，社會標記共產生 8 條的最短路徑式規則，而且階層為一階；然而控制詞彙則有 1 條 2 個階層的最短路徑式規則。由此可知，上述此種現象顯示著社會標記者傾向至少使用 1 種以上不同的單一關鍵字類別及其所擁有關鍵字進行文獻標引，而專家則是傾向至少使用 2 種以上不同的關鍵字類別及關鍵字組織各項資訊資源。此一現象也交叉驗證前述「關鍵字類別的共同出現／使用現象」一節內容，指出：「社會標記者傾向以較多組且較少的層面分析觀點，而專家則是傾向使用較少組且較多的層面分析觀點，且多數層面分析觀點超過兩種以上的觀點提供關鍵字，進行文獻標引。」更進一步，以最長路徑式規則方面而言，社會標記共產生 1 條的最長路徑式規則，而且階層為 7 階；然而控制詞彙則有 1 條 8 個階層的最長路徑式規則。若從規則的組成階層或關鍵字類別而言，無論是單一路徑式規則或是樣式型規則，組成社會標記 FP-tree 規則的平均階層數量皆低於組成控制詞彙 FP-tree 規則的平均階層數量。社會標記的路徑式規則數量也遠少於控制詞彙的路徑式規則數量。綜合上述的最長路徑式規則、路徑式規則數量，以及組成單一路徑式規則與樣式型規則的平均組成階層而言，可以顯示出社會標記者傾向使用較少的路徑式規則且較少的關鍵字類別進行文獻標引，而專家則是傾向使用較多的路徑式規則且較多的關鍵字類別，組織各式資訊資源。若從資訊組織的觀點而言，專家可能受到層面分析理論及專業訓練等共識基礎的影響，以比較多元化面向與集中又連續的方式標引資訊物件的屬性。最後，從 FP-tree 規則的樣式及實際內容而言，兩者之間只有 25 種 FP-tree 規則樣式與 12 條 FP-tree 規則內容是完全相同的（請參照表 4.10），相同比率也在 40% 以下，屬於「輕度一致性」的範疇。

表 4.10 社會標記者與專家間共同的路徑式規則群組及其實例

共同的路徑式規則群組	實際相同路徑式規則的類別
01. Title→Content	None
02. Title→Content→Content	05→19→17
03. Title→Other	None
04. Title→Title	None
05. Title→Title→Content	01→06→19
06. Title→Title→Other	01→02→26
07. Title→Title→Title	None
08. Title→Title→Title→Content	None
09. Title→Title→Topic	None
10. Title→Topic→Content	01→12→16 及 01→12→17
11. Title→Topic→Content→Content	None
12. Title→Topic→Title→Content	None
13. Title→Topic→Title→Title	None
14. Topic→Content	09→13 09→16
15. Topic→Content→Content	None
16. Topic→Title	None
17. Topic→Title→Content	09→04→16 及 09→05→19
18. Topic→Title→Other	None
19. Topic→Title→Title	09→05→02 及 12→04→05
20. Topic→Title→Title→Title	None
21. Topic→Title→Topic→Content	None
22. Topic→Topic→Content	09→12→18
23. Topic→Topic→Content→Content	None
24. Topic→Topic→Title→Content	None
25. Topic→Topic→Title→Title	None

第七節 FP-growth 分析

無論是社會標記或是控制詞彙之個別關鍵字及其關鍵字類別的使用情形，兩者皆是呈現 80/20 原則，集中在少數的關鍵字與關鍵字類別。因而，在分析 FP-growth 規則時，本研究以 26 類關鍵字類別的中位數為基準，作為最小支持度 (minimal support) 的值，而不是採取平均數，避免受到少數幾個極端關鍵字類別使用的影響。另外，由於關鍵字類別為偶數，因而本研究係將使用排名第 13 與第 14 位置的關鍵字類別使用比率予以加總後求其平均數，作為 FP-growth 最小支持度的設定值，以產生 FP-growth 的相關規則。在社會標記方面，中位數平均數為 2.149% ($[1.813\%+2.485\%]\div 2$)，最小支持度設定為 2%，以及最小信賴度設定為 40% 等前提下，共產生 12 組 FP-growth 規則，其中有 5 組 FP-growth 規則係以關鍵字類別 09 (即：指明主題—一般，涵蓋率為 41.67%) 為起始，7 條 FP-growth 規則以關鍵字類別 01 (即：文獻題名—相同，涵蓋率為 58.33%) 為起始，其所支持的其他關鍵字類別，請詳表 4.11。

表 4.11 社會標記的 FP-growth 規則組合及其支持度與信賴度

FP-growth 規則	支持度	信賴度
23 ==> 9	0.021490933512424447	0.5079365079365079
19 ==> 9	0.024848891873740765	0.578125
10 ==> 9	0.03022162525184688	0.5232558139534884
10 ==> 1	0.03626595030221625	0.627906976744186
5 ==> 1	0.026192075218267292	0.43333333333333335
5 ==> 9	0.026863666890530557	0.4444444444444444
6 ==> 9	0.03290799194089993	0.44144144144144143
6 ==> 1	0.040295500335795834	0.5405405405405406
4 ==> 1	0.041638683680322364	0.4105960264900662
12 ==> 1	0.1081262592343855	0.6168582375478927
9 ==> 1	0.21692411014103424	0.44368131868131866
9,12 ==> 1	0.03492276695768973	0.5531914893617021

有鑑於專家在進行文獻標引時，必須熟悉相關的作業規範，而且基於一定的共識前提下，才能提供控制詞彙作為標引資訊資源的主題概念或知識。因而在控制詞彙方面，中位數平均數約為 1.01% ($[0.269\%+1.746\%]\div 2=1.0075\%$)，本研究將最小支持度設定為 1%，最小信賴度設定為 70% 等前提下，共產生 73 組 FP-growth 規則，其中有 71 組 FP-growth 規則係以關鍵字類別 09 (指明主題—一般，涵蓋率為 97.26%) 為起始，2 組 FP-growth 規則以關鍵字類別 01 (文獻題名—相同，涵蓋率為 2.74%) 為起始，其所支持的其他關鍵字類別，請詳表 4.12。

以社會標記與控制詞彙所產生的 FP-growth 規則類型而言，社會標記以關鍵字類別 01 (文獻題名—完全相同) 與 09 (指明主題—一般) 為主要的涵蓋範圍，而且類別 01 ($7\div 12=58.33\%$) 大於類別 09 ($5\div 12=41.67\%$)，兩者百分比差距約為 18%。在控制詞彙方面，同樣以以關鍵字類別 01 與 09 為主要的涵蓋範圍，但是類別 09 (指明主題—一般，涵蓋率為 97.26%) 大於類別 01 (文獻題名—完全相同，涵蓋率為 2.74%)，兩者百分比差距高達約為 95%。就 FP-growth 規則所涵蓋的關鍵字類別而言，社會標記者與專家皆集中在類別 01 (文獻題名—完全相同) 與 09 (指明主題—一般) 兩大類別，但社會標記者傾向集中在類別 01 (文獻題名—完全相同)，而專家則是集中在類別 09，兩者間仍存有差異。

表 4.12 控制詞彙的 FP-growth 規則組合及其支持度與信賴度

FP-growth 規則	支持度	信賴度
6 ==> 9	0.13633310946944258	0.8285714285714286
5 ==> 9	0.14237743451981194	0.7910447761194029
4 ==> 9	0.2122229684351914	0.7745098039215687
12 ==> 9	0.28945601074546673	0.8500986193293886
1 ==> 9	0.48757555406312963	0.82593856652901
9,26 ==> 1	0.010745466756212223	0.8
1,26 ==> 9	0.010745466756212223	0.7619047619047619
17,18 ==> 9	0.012760241773002015	0.95
5,18 ==> 9	0.011417058428475487	0.8095238095238095
14,18 ==> 9	0.014103425117528543	0.8076923076923077

FP-growth 規則	支持度	信賴度
4,18 ==> 9	0.018132975151108125	0.84375
12,18 ==> 9	0.021490933512424447	0.8888888888888888
1,18 ==> 9	0.027535258562793822	0.8367346938775511
4,16 ==> 9	0.016118200134318333	0.8571428571428571
12,16 ==> 9	0.03089321692411014	0.9387755102040817
1,16 ==> 9	0.02887844190732035	0.8775510204081632
2,4 ==> 9	0.018132975151108125	0.7105263157894737
2,12 ==> 9	0.026192075218267292	0.8125
1,2 ==> 9	0.029550033579583614	0.7719298245614035
6,17 ==> 9	0.010073875083948958	0.8333333333333334
5,17 ==> 9	0.013431833445265278	0.7407407407407407
4,17 ==> 9	0.02216252518468771	0.7333333333333333
12,17 ==> 9	0.03089321692411014	0.8070175438596491
14,17 ==> 9	0.036937541974479515	0.8870967741935484
1,17 ==> 9	0.0409670920080591	0.8133333333333334
14,19 ==> 9	0.010073875083948958	0.9375
5,14 ==> 9	0.014103425117528543	0.84
6,14 ==> 9	0.014775016789791807	0.8461538461538461
4,14 ==> 9	0.026192075218267292	0.7647058823529411
12,14 ==> 9	0.03223640026863667	0.8421052631578947
1,14 ==> 9	0.04633982538616521	0.8518518518518519
6,19 ==> 9	0.01544660846205507	0.8846153846153846
5,19 ==> 9	0.01544660846205507	0.8846153846153846
4,19 ==> 9	0.01880456682337139	0.8235294117647058
12,19 ==> 9	0.029550033579583614	0.88
1,19 ==> 9	0.051712558764271326	0.8850574712643678
5,6 ==> 9	0.018132975151108125	0.7105263157894737
4,6 ==> 9	0.03089321692411014	0.7540983606557377
6,12 ==> 9	0.04432505036937542	0.825
1,6 ==> 9	0.06581598388179986	0.7777777777777778
4,5 ==> 9	0.039623908663532575	0.7023809523809523
5,12 ==> 9	0.049026192075218265	0.7684210526315789
1,5 ==> 9	0.0725319006044325	0.7448275862068966
4,12 ==> 9	0.08126259234385494	0.7610062893081762
1,4 ==> 9	0.10678307588985897	0.7429906542056075
1,12 ==> 9	0.16118200134318333	0.8163265306122449

FP-growth 規則	支持度	信賴度
1,12,18 ==> 9	0.013431833445265278	0.8695652173913043
4,12,16 ==> 9	0.010745466756212223	0.9411764705882353
1,12,16 ==> 9	0.018132975151108125	0.9310344827586207
2,4,12 ==> 9	0.011417058428475487	0.7083333333333334
1,2,12 ==> 9	0.014103425117528543	0.75
4,12,17 ==> 9	0.010073875083948958	0.7894736842105263
1,4,17 ==> 9	0.010073875083948958	0.75
1,12,17 ==> 9	0.0167897918065816	0.78125
4,14,17 ==> 9	0.011417058428475487	0.7727272727272727
12,14,17 ==> 9	0.018132975151108125	0.9
1,14,17 ==> 9	0.020147750167897917	0.9090909090909091
4,12,14 ==> 9	0.011417058428475487	0.8095238095238095
1,4,14 ==> 9	0.013431833445265278	0.7692307692307693
1,12,14 ==> 9	0.018132975151108125	0.8181818181818182
1,12,19 ==> 9	0.01880456682337139	0.8484848484848485
1,4,6 ==> 9	0.012760241773002015	0.7037037037037037
1,6,12 ==> 9	0.02216252518468771	0.7857142857142857
4,5,12 ==> 9	0.017461383478844864	0.7027027027027027
1,4,12 ==> 9	0.0409670920080591	0.7439024390243902

再者，就以 FP-growth 規則內的關鍵字類別範圍而言，社會標記多數對應至 1 個關鍵字類別，只有 1 組規則對應至 2 個關鍵字類別，亦即 1 對 1 為主要的現象。除此之外，在 12 組 FP-growth 規則中，其中 7 組規則是異質群組，只有 5 組規則是同質群組，而同質群組中，包括 4 組題名群組與組主題群組。在控制詞彙方面，則是具有 1 對 1、1 對 2（如：6,12 ==> 9）與 1 對 3（如：1,4,12 ==> 9）等三種現象，而且涵蓋百分比分別為 6.85%（5 ÷ 73）、67.12%（49 ÷ 73）與 26.03%（19 ÷ 73）。另外，在 73 組規則中，只有 1 組為同質群組（1.37%）、24 組為部分異質群組（32.88%），及 48 組異質群組（65.75%）。以 FP-growth 規則的群組性質而言，社會標記者使用社會標記時，傾向選取異質群組且不同關關鍵字類別及其所屬關鍵字，但同質且不同關關鍵字類別及其所擁有關字類的傾向亦不低，而且幾乎是選用某一種關鍵字類別的關鍵字後，接連引出使用另外一種關鍵字類

別的關鍵字。就專家使用控制詞彙進行文獻標引時，則是集中選取異質群組且不同關關鍵字類別及其所擁有關鍵字，而且九成以上（ $68 \div 73 = 93.15\%$ ）是使用某一種關鍵字類別的關鍵字後，後續連動使用另外兩種以上關鍵字類別的關鍵字。此一現象也呼應了前述 FP-tree 分析結果的意涵：「社會標記者傾向使用較少的 FP-tree 規則且較少的關鍵字類別進行文獻標引，而專家則是傾向使用較多的 FP-tree 規則且較多的關鍵字類別，組織各式資訊資源。」

第三，就相同的 FP-growth 規則數量而言，社會標記與控制詞彙兩者間只有 2 組完全相同的 FP-growth 規則（即： $5 \Rightarrow 9$ 與 $6 \Rightarrow 9$ ），而且全數皆分佈在類別 09（指明主題—一般）為範圍。換言之，就社會標記與控制詞彙間的 FP-growth 相同規則而言，以主題群組為主要的使用範圍外，社會標記者與專家的共同取向是使用「指明主題—一般」類別的關鍵字後，會選取題名群組的關鍵字作為文獻標引之用。從資訊組織觀點而言，FP-growth 結果顯示社會標記者傾向使用題名群組的關鍵字類別及其所擁有的關鍵字，而專家則是傾向使用主題群組的關鍵字類別及其所擁有的關鍵字。造成此種差異的潛在原因可能是專家受到主題分析的權威控制與索引典控制等規範影響，指引專家專注標引資訊物件的主題外，專家也可能熟知資訊系統可以依據字詞典或特定索引典，提供字根機制或同義字環（synonym ring）聚合不同形式的關鍵字，而減少使用文獻題名的不同形式關鍵字。另外，從 FP-growth 規則內容而言，社會標記者傾向選用某一關鍵字類別與所擁有的關鍵字後，會再使用另一種關鍵字類別與所擁有的關鍵字（即 1 對 1），而專家則是後續連同使用 2 種（含）以上的關鍵字類別與所擁有的關鍵字（即 1 對 2 與 1 對 3）。此種結果也顯示專家受到層面分析的影響，傾向選用較多元化面向或類別的關鍵字標引資訊物件的屬性。

第八節 使用關鍵字的心智模式

除了以 FP-tree 路徑式規則方式呈現社會標記者與專家使用關鍵字類別的決

策樹外，關鍵字的實際使用情形也可以呈現出社會標記者與專家的另一種文獻標引心智模式地圖。依據關鍵字類別的使用情形而言，社會標記者關鍵字的使用情形共涵蓋了五種群組，使用排序分別是：主題、題名、內容、功能與其他。如果將使用率門檻值設定為 1% 時，仍為主題、題名、內容、功能與其他等群組（請參見圖 4.17）。在專家方面，一共使用了四種群組，其中功能群組並未被使用，使用排序分別是：主題、題名、內容與其他。如果將使用率門檻值設定為 1% 時，只剩下主題、題名與內容等群組（請參見圖 4.18）。另一方面，以使用的類別數量而言，除了內容群組的類別外，整體上，社會標記者較專家使用較多的關鍵字類別進行文獻標引。更進一步，計有 2 個類別同時被社會標記者與專家使用，而且使用排序也相同，這個類別是：04（文獻題名—同義）與 12（指明主題—分離型詞彙）。

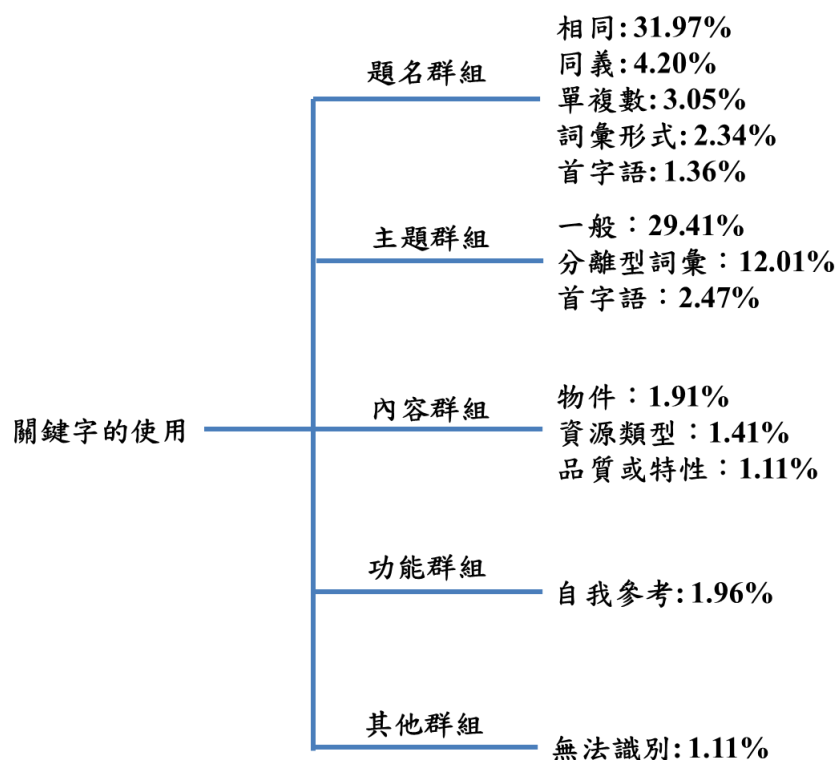


圖 4.17 社會標記者使用關鍵字的文獻標引心智模式

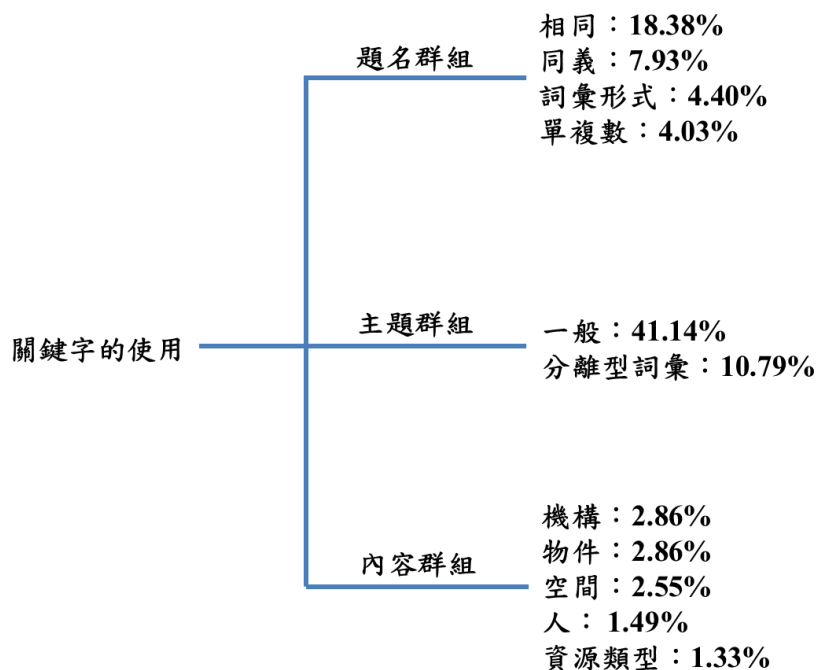


圖 4.18 專家使用關鍵字的文獻標引心智模式

第九節 資訊系統設計的意涵

前述的 SNA、FP-tree 與 FP-growth 等討論中，各節所探討的內容對資訊組織系統的設計方面具有一定的意義。本節重點以前述第一節至第八節的研究探討內容為出發點，延伸探討其對資訊系統設計的考量與影響，而且可依社會標記者與專家的分析結果為基礎，發展適合一般使用者與專家等兩種類型的資訊系統功能與使用者界面。

一、關鍵字的推薦

無論對一般使用者或是專家而言，選用適合的關鍵字以組織或查詢資訊物件一直都是熱門的研究議題。如果能適時提供適當的關鍵字建議，將有助於資訊的組織與查詢建構 (query formulation)。本研究依據前述的關鍵字使用情形與 SNA 分析結果，針對關鍵字的推薦方面提供下列延伸性的應用說明：

1. 萃取文獻題名中的關鍵字作為標引或查詢之用

依據個別的社會標記與控制詞彙，以及其關鍵字類別的使用情形而言，結果皆呈現冪次定律分佈與 80/20 原則的現象外，社會標記者與專家皆會從文獻題名中，摘取重要的關鍵字作為文獻標引之用，而且比率分別為 44.20% 與 36.34%。換言之，近三分之一以上的社會標記、控制詞彙分別與文獻題名關鍵字間具有來源關係，包括直接採用，或是同義、縮寫或是其他形式等。就資訊系統的設計方面而言，此項分析結果顯示出：「如果能適當導入相關的自動化機制，將有助於社會標記者與專家進行資訊組織的關鍵字標引作業，乃至於查詢之用。」例如，在提供關鍵字的標引或查詢作業程序中，資訊系統可以提供自動化的關鍵字斷字機制，直接萃取文獻題名中的關鍵字，供社會標記者或專家直接選用，作為標引或查詢各項資訊資源之用。

2. 提供字根相關關鍵字的推薦機制

除了自文獻題名中萃取關鍵字之外，適時導入字根機制，自動提供同義、首字語、縮寫或是其他不同形式的文獻題名關鍵字，也能擴大關鍵字的提供範圍。至於字根的依據基礎除了一般辭典（如：WordNet）外，可視學科領域的獨特性，加入適當的專門型詞彙表（如：ALA Glossary of Library and Information Science 與 ODLIS），以推薦適當的關鍵字，提升關鍵字標引或查詢的作業效能。

3. 提供相關關鍵字的關聯導引

從本研究的 SNA 中心度、階層集叢與同等角色集叢等分析結果中，可以明顯發現關鍵字彼此間存在著某種關聯關係。若從 SNA 中心度的觀點而言，關鍵字間存在著三種類型的關聯關係：經常被單獨使用、經常被使用且引發影響其他關鍵字的使用，以及雖然不常被使用，一旦被使用時，即影響其他關鍵字的使用。對關鍵字的推薦機制而言，前述第一種類型的關鍵字被使用時，系統無須自動推薦任何類別的關鍵字，但屬

於前述第二與第三種類型者，系統可依 SNA 中心度的分析結果，推薦相關類別的關鍵字，作為文獻標引或查詢之用。依此類推，若從 SNA 的階層集叢分析觀點而言，圖 4.9 與 4.10 所列的這些集叢可視為共同使用的關鍵字類別集叢，一旦集叢內某一類別關鍵字被使用時，其他類別關鍵字極有可能被一起選用。若從 SNA 的同等角色集叢等分析觀點而言，圖 4.11 與 4.12 所列的這些集叢可視為相互交替使用的關鍵字類別集叢，一旦其中一個同等角色的關鍵字類別及其所擁有關鍵字被使用時，即可引發同一集叢內的其他同等角色的關鍵字類別及其所擁有關鍵字將被使用；或是某一同等角色集叢內所有共同使用的關鍵字類別及其所擁有關鍵字被使用時，同一集叢內的某一同等角色的關鍵字類別及其所擁有關鍵字將被使用。對關鍵字推薦機制而言，前述這些 SNA 中心度、階層集叢與同等角色集叢等分析結果皆可被進一步納入關鍵字推薦機制之中，經由關鍵字類別的關聯關係，自動推薦關鍵字作為文獻標引或查詢之用。

二、使用者界面的設計

以 FP-tree 分析而言，社會標記者與專家使用關鍵字的各組路徑規則皆可顯示標引關鍵字使用頻率之連動或銜接的關係，進而作為設計資訊組織系統提供輸入關鍵字的功能參數或是提示建議。然而，將全數路徑規則予以納入資訊系統的設計並不是一項具有效益的作法或發展方式。因而，可以依據 FP-growth 的分析結果作為設計資訊系統的主要參考基礎。換言之，以表 4.11 與 4.12 的分析結果為依據，將社會標記者的關鍵字標引功能設定為以 1 對 1 的方式，亦即選用某一關鍵字類別及其所擁有關鍵字後，再後續連動帶出另外一種關鍵字類別及其所擁有關鍵字。對專家而言，則是以 1 對 2 以上的方式為主，亦即選用某一關鍵字類別及其所擁有關鍵字後，後續至少連動帶出另外兩種以上的關鍵字類別及其所擁

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

有關鍵字。以「9,12==>1」這組 FP-growth 規則為例，當社會標記者直接選用文獻題名關鍵字（即：01 文獻題名—相同）時，資訊系統即可後續連動提供類別 09（即：指明主題—一般）與類別 12（即：指明主題—分離型詞彙）等兩種類型關鍵字的功能選項，供使用者參考選用。FP-growth 所產生的規則除了可以單獨作為使用者界面的設計依據外，也可以進一步與前述的 SNA 階層集叢與同等角色集叢分析結果相互搭配，運用在文獻標引或查詢。例如，在社會標記的 SNA 同等角色集叢分析結果中，得知類別 01（文獻題名—相同）與類別 12（指明主題—分離型詞彙）相互交替使用後，再與類別 09（指明主題—一般）交替使用。如果進一步搭配前述「9,12==>1」這組 FP-growth 規則，當社會標記者選用類別 01（文獻題名—相同）關鍵字時，資訊系統即可進一步優先提供類別 09（指明主題—一般）的提示畫面，供社會標記者參考使用。

三、瀏覽分類架構的建立與運用

除了關鍵字的推薦與使用者界面的設計方面外，本研究成果對於瀏覽分類架構的建立與運用也具有某種的應用效益。以 FP-tree 分析結果而言，本研究已依頻繁項目的共同出現的次數與關聯關係提出一個樹狀分類架構。此一路徑式規則的樹狀分類架構可以進一步轉化為層面型的瀏覽分類樹，藉以指引使用者經由不同類別關鍵字的相互關係，適時提供相關的層面型關鍵字類別及所擁有的關鍵字與資訊物件，以找到所需的資訊。如，以附錄一的 001 路徑式規則而言（即：04 文獻題名—同義→06 文獻題名—單複數→19 指明內容—物件→03 文獻題名—縮寫），當使用者或專家由關鍵字類別 04（文獻題名—同義）為起始點開始瀏覽時，資訊系統即可引導使用者或專家至其他層級的關鍵字類別之一，以瀏覽相關的資訊物件，包括關鍵字類別 06（文獻題名—單複數）、19（指明內容—物件）或 03（文獻題名—縮寫）。

第五章 結論與建議

本研究以 1,489 篇圖資界的期刊論文，及其 3,972 個社會標記與 6,708 個控制詞彙為研究樣本，除了檢視社會標記與控制詞彙等關鍵字的來源與類型、使用情形，以及社會標記與控制詞彙彼此間關鍵字的相同比例外，並以共同出現的關鍵字關聯關係為依據，採取 SNA 中心度（包括：程度、接近與中介等三種）、SNA 階層集叢、SNA 同等角色、FP-tree 與 FP-growth 等方法建構文獻標引心智模式地圖及分析其結構與樣式後，提出結論、研究貢獻與未來的研究建議。第一節，就研究問題提出總結說明。第二節，提出本研究的貢獻，主要項目包括：文獻標引心智模式的發展、文獻標引心智模式地圖的分析、關鍵字的使用分析、一般使用者與專家之間文獻標引心智模式的相似性，以及 TCM 可行性的擴展驗證與應用解釋等項目。第三節，本研究試從理論模式與實務驗證提出建議，包括：發展一體兩面的文獻標引系統、擴增文獻標引心智模式的研究範圍、展延文獻標引心智模式的驗證對象，以及融合資訊檢索與文獻標引的心智模式及其資訊系統的應用設計等方面進行說明。

第一節 結論

經由使用率、冪次定律分佈、SNA 中心度、SNA 階層集叢、SNA 同等角色、FP-tree 與 FP-growth 等方法分析相關資料後，本節內容針對文獻標引心智模式、關鍵字的來源、社會標記與控制詞彙間的關聯程度，關鍵字類別的使用情形，及關鍵字類別間關聯關係的結構與樣式等方面提出說明，以總結社會標記者與專家兩者間文獻標引心智模式的異同，並從資訊組織之主題分析、權威控制、索引典控制，及層面分析等理論與規範的觀點說明造成兩者文獻標引心智模式之間差異的潛在原因。

一、社會標記者的文獻標引心智模式比專家更為多元化

在標引文獻時（請參見圖 4.17 與 4.18），社會標記者使用的關鍵字類別涵蓋了主題、題名、內容、功能與其他等五種群組，其中主題與題名兩種群組的使用率十分相近。社會標記者在標引文獻時，除具備多元化的特質外，也具有個人化的需求傾向，並以功能群組的關鍵字標引僅供個人使用的資訊物件。另一方面，專家只使用了四種群組的關鍵字標引文獻，並未使用功能群組的關鍵字，而且主題群組與題名群組間的使用率差距幾近 16%。造成社會標記者與專家間文獻標引心智模式的差異，潛在原因除了專家可能熟悉資訊系統可以提供個人化功能，而避免使用功能群組的關鍵字外，也受到主題分析與相關資訊組織規範及採用官方來源的控制詞彙等影響，以選取主題群組關鍵字為主，並謹慎選取與文獻題名相關的關鍵字，以標引資訊物件的屬性。

二、社會標記者直接從文獻題名中選用關鍵字的傾向高於專家

從表 4.2 可以得知，社會標記者會從期刊文獻題名中，直接萃取關鍵字作為社會標記之用，而且 31.97% 是直接取自期刊文獻題名的關鍵字，44.20% 取自題名相關的關鍵字（包括：相同、首字語、縮寫、同義、詞彙形式、單複數、非英文的同義詞彙與分離型詞彙等）。另一方面，從表 4.3 可以得知，專家也會從期刊文獻題名中，直接萃取關鍵字作為控制詞彙之用，而且 18.38% 是直接取自期刊文獻題名的關鍵字，36.34% 取自題名相關的關鍵字。就關鍵字的來源而言，社會標記者較專家傾向直接自期刊文獻題名中，萃取關鍵字作為文獻標引之用，而專家則依權威控制及索引典控制的作業原則從題名關鍵字中，萃取關鍵字作為文獻標引之用；如，首字語、縮寫字與同義字等。換言之，社會標記者與專家從期刊文獻題名中，萃取關鍵字作為文獻標引的傾向並不相同，其原因除了專家可能熟知資訊系統可以結合字根機制，藉以聚合不同形式的關鍵字外，也可能與圖資界資訊組織的權威控制及索引典控制的學術理論與實務規範極有關係，因為這些

規範指引專家直接採用官方來源的既有控制詞彙(如 LCSH 與 LISA 索引典)外，也必須依循一定的原則使用不同形式的題名關鍵字標引資訊物件的屬性。因而，才降低專家從題名中直接萃取關鍵字標引資訊物件的機率。

三、社會標記與控制詞彙間的相同程度不高且彼此互補

從表 4.4 中，可以得知只有 12.23% 的社會標記與控制詞彙完全相同，卻有 52.72% 的社會標記無法比對至控制詞彙。換言之，至少有 50% 以上的社會標記與控制詞彙完全不相同。如果比對的計算單位為期刊文獻時，也只有 417 篇同時具備 1 個以上相同關鍵字的社會標記與控制詞彙，亦即約有 28% 的期刊文獻擁有相同關鍵字的社會標記與控制詞彙。因而，無論以單一關鍵字或是單一期刊文獻為比對的計算單位而言，社會標記與控制詞彙兩者間的相同性並不高，而且呈現著互補的現象，所以社會標記者與專家在進行文獻標引時，兩者選用相同關鍵字的比率並不高。造成如此差異的潛在原因可能是圖資界資訊組織特別強調主題分析的重要性，因而專家係以標引資訊物件的主題為主要的優先考量，造成使用主題群組關鍵字的使用率超過 50%。即使專家自文獻題名中萃取關鍵字，也受到權威控制與索引典控制等規範指引，必須依遵循一定的原則，才能直接選取題名關鍵字，作為標引資訊物件的控制詞彙，以及盡量直接採用官方來源等為原則。除此之外，官方維護的控制詞彙必須遵循一定的原則與處理程序後，才會增加新的關鍵字。因而，就關鍵字的新穎性而言，社會標記也較控制詞彙新穎，以反應新興的需求。基於前述原因，所以大幅降低社會標記與控制詞彙之間的相同程度。

四、社會標記者與專家使用關鍵字的傾向不同：使用比率及其影響

以 26 個關鍵字類別的使用率而言(請參照表 4.2 與 4.3)，社會標記者與專家間沒有任何一項類別的使用率完全相同。就關鍵字類別的使用率排名而言，只有 3 個類別使用排名的位置相同，即 04 (文獻題名一同義)、11 (指明主題一縮

寫)與 12 (指明主題—分離型詞彙),其排名位置分別為 4、15 與 03。在最常用關鍵字類別方面而言,社會標記者最常使用的類別為 01 (文獻題名—相同),意指社會標記者最常自題名中,直接選取關鍵字作為社會標記,而專家則是以標引主題為最常用使用的類別(09 指明主題—一般),意謂著專家傾向以標引資訊物件的主題為主,而非直接自題名中選取關鍵字作為文獻標引之用。除此之外,專家也不使用類別 20 至 24 (指明內容—使用對象、指明內容—資源來源、指明功能—任務、指明功能—自我參考及不用標記)。

若從前 8 名的三項 SNA 中心度而言,可以更清楚瞭解社會標記者與專家在選用關鍵字的傾向及其使用影響性。在前 8 名的三項 SNA 中心度中,社會標記者涵蓋了 3 至 5 項題名、3 項主題群組,以及 1 至 2 項內容群組的關鍵字類別,而專家則涵蓋了 4 項題名群組、1 至 2 項主題群組與 2 至 4 項內容群組的關鍵字類別。前述這些 SNA 分析結果顯示社會標記者經常使用題名與主題群組關鍵字類別及其所擁有的關鍵字外,也影響了其他關鍵字類別及其所擁有的關鍵字的使用,而專家則是經常選取題名群組關鍵字類別及其所擁有的關鍵字外,也影響了其他關鍵字類別及其所擁有的關鍵字的選用。在選取文獻題名關鍵字時,專家以遵守權威控制與索引典規範,及直接採用官方來源的既有控制詞彙為優先原則,慎選與文獻題名相關關鍵字的控制詞彙。另一方面,雖然專家與社會標記者皆經常使用主題群組關鍵字類別及其所擁有的關鍵字外,但社會標記者傾向也會連同影響其他關鍵字類別及其所擁有的關鍵字的使用,而專家則是傾向專注主題分析,而單獨標引資訊物件的主題。此種差異的潛在原因可能在於社會標記者與專家皆受到權威控制與索引典控制等規範的原則影響,只是影響不同。除此之外,在權威控制與索引典控制等規範的指引影響下,專家也較著重在資訊物件的地名/空間、人與物件等內容標引。

五、社會標記者與專家使用關鍵字的傾向不同：共同與交替使用

以 SNA 階層集叢分析的觀點而言，在社會標記方面共產生 8 組共同使用關鍵字類別的集叢，其中以內容群組的類別最常被組合，而題名與其他群組次之。這意謂著社會標記者不經常單獨使用內容群組的關鍵字類別及其所擁有的關鍵字，卻經常與題名或主題群組的關鍵字類別及其所擁有的關鍵字一起搭配使用。在控制詞彙方面只有產生 3 組共同使用關鍵字類別的集叢，其中以其他群組的類別最常被組合。這也透露出專家雖然不常單獨使用其他群組的關鍵字類別及其所擁有的關鍵字作為文獻標引之用，卻常與題名與內容等群組及其所擁有的關鍵字搭配一起使用。若從 SNA 同等角色分析而言，在社會標記方面共產生 10 組集叢，以題名群組出現頻率最高，內容、主題與其他群組次之。在控制詞彙方面則產生 8 組集叢，以題名與內容群組出現頻率最高，其他群組次之。這項分析結果意謂著社會標記者經常交替使用兩種關鍵字類別及其所擁有的關鍵字，並搭配其他類別及其所擁有的關鍵字進行文獻標引，而專家則是可能受到層面分析的影響，採取比較多元化面向標引資訊物件的屬性，且經常交替使用 2 至 7 種關鍵字類別及其所擁有的關鍵字。

六、社會標記者與專家使用關鍵字的傾向不同：關聯的結構與樣式

從 FP-tree 產生的集叢數量觀點而言，分析結果顯示社會標記者傾向使用較多的關鍵字類別，而專家則是傾向集中較少的關鍵字類別，前者包括題名、主題、內容、功能與其他等五種群組，而後者只有題名、主題與內容群組而已。若從 FP-tree 路徑式規則的組成數而言，社會標記者會使用 1 至 7 種關鍵字類別組織資訊物件，而專家則是會使用 2 至 8 種關鍵字類別進行文獻標引。另外，從 FP-growth 產生的規則而言，社會標記共產生 12 組規則，其中 11 組係為 1 對 1 的規則，涵蓋範圍以類別 01（文獻題名—相同）與類別 09（指明主題—一般）為主，兩者涵蓋率相差約 17%。控制詞彙共產生 71 組規則，涵蓋範圍也是以類

別 01（文獻題名—相同）與類別 09（指明主題—一般）為主，涵蓋率以類別 09 為最高，比率高達 97.26%。但是規則類型包括 1 對 1、1 對 2 與 1 對 3，其中 1 對 2 規則的涵蓋率接近一半。造成上述差異的潛在因素，可能是專家受到資訊組織的主題分析與層面分析等影響，以標引資訊物件的主題為主，同時採取不同面向標引資訊物的主題屬性有關。整體而言，社會標記與控制詞彙之間只有 25 種樣式與 12 組 FP-tree 規則，以及 2 組 FP-growth 規則完全相同。

整體而言，從關鍵字的相同比率、關鍵字類別的使用情形、SNA，以及 FP-tree 與 FP-growth 等分析觀點檢驗上述五項問題，可從其結果可以得知：一般使用者與專家間的文獻標引心智模式相似性多數落在「輕度的一致性」範圍內，只有少數在「中度的一致性」範圍，因而兩者間的文獻標引心智模式並不相似，而且兩者間存有不同的系統印象。

第二節 研究貢獻

本節內容以前述研究結果為基礎，並不從關鍵字的來源與類型、關鍵字的使用比例，以及社會標記與控制詞彙間的相同比例等觀點探討使用者與專家間的心智模式異同，而是從關鍵字的共同出現關係為基礎所發展的心智模式理論觀點出發，就文獻標引心智模式的發展、地圖內容及其樣式、關鍵字的使用分析及其特質，以及一般使用者與專家間文獻標引心智模式的相似性等五方面提出本研究之研究貢獻。

一、發展文獻標引心智模式

截至目前為止，在資訊組織心智模式方面的研究僅限於 FRBR 認知異同的探討而已。Pisanki and Žumer (2011a, 2011b) 導入心智模式，藉以探討一般使用者與專家對 FRBR 的認知或系統印象是否趨於一致。因而，在 Pisanki and Žumer (2011a, 2011b) 的 FRBR 心智模式研究中，實際上兩人並未發展資訊組織方面

的心智模式，只是以心智模式為理論基礎，進一步解釋、說明與比較 FRBR 在一般使用者與專家的心理認知。然而，本研究則以關鍵字共同出現關聯關係為基礎，進而發展出文獻標引的心智模式，藉以探索、解釋、說明與比較社會標記者與專家間之文獻標引的心理認知，並非以關鍵字的來源、使用比例與相同百分比等方式零散呈現心智模式的片斷內容。就本研究提出之文獻標引心智模式而言（請參照圖 2.2），本研究已就資訊物件，以及社會標記者與專家具體化的關鍵字與系統印象等模式元件，逐一建立彼此間的邏輯關係，同時以圖資界期刊文獻及其社會標記與控制詞彙等關鍵字為研究樣本，驗證其可行性及意涵。因而，就 Pisanki and Žumer (2011a, 2011b) 的 FRBR 心智模式研究而言，亦可融合至本研究提出的文獻標引心智模式之中，只是資訊物件可以替換為 FRBR，而具體化的關鍵字可改換為一般使用者與專家對相對應 FRBR 書目紀錄的實際認知行為，只是有關評估心智模式相似性的項目與方法，端視各研究目的與樣式而調整。換言之，本研究提出的文獻標引心智模式可視為一種理論框架，進一步擴展成為其他不同面向或其他資訊組織的心智模式等相關研究的基礎，藉以導入不同資訊組織的研究目的與樣本，逐一豐富與累積有關資訊組織心智模式方面的理論基礎，以及實務驗證之用。

二、分析文獻標引心智模式地圖及其樣式

有關心智模式在圖資界的研究，多數集中在資訊檢索與查詢方面，同時也提供相關的心智模式地圖與樣式。例如，在 Zhang (2008b) 的研究中，就歸納出在全球資訊網之使用者搜尋行為的心智模式地圖。另外，Cole et al. (2007) 則是以史學與心理學的課程為研究範圍，分析與歸納出資訊尋求行為的心智模式及其 12 種樣式。反觀在文獻標引心智模式方面的研究，僅局限在心智模式的導入與解釋，或是僅以關鍵字的異同、使用情形與來源等單一觀點，片斷式呈現心智模式內容，並未發展出任何的概念或理論模式、心智模式地圖，乃至於心智模式的

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

樣式。如果依前述第二章文獻探討與圖 2.2 文獻標引的心智模式，以及第四章研究結果等內容，可以發現本研究除了完成文獻標引心智模式之概念模式的發展與實證外，更進一步以關鍵字共同出現關聯關係為基礎，提出文獻標引心智模式的地圖內容與樣式（請參照第四章第六節與第八節等內容），進而比較與評估社會標記者與專家間之文獻標引心智模式的一致性或異同。

三、分析關鍵字的使用情形及其共同出現的關聯關係特質

除了社會標記與控制詞彙的使用統計外，本研究採用 SNA 與 FP-growth 等分析方法探討關鍵字的使用特性及關鍵字共同出現的關聯關係等，因而本研究具備多項的研究特殊性，主要項目說明如下：

1. 擴展研究樣本的粒度 (granularity)：將原來的研究樣本從個別關鍵字，進一步擴展至關鍵字類別。除了使用率外，也包括冪次定律分佈、80/20 原則，更進一步驗證關鍵字類別具有類似的使用現象，同時也將研究範圍擴大至專家的關鍵字類別及其所擁有關鍵字使用等分析。
2. 闡述關鍵字類別及其所擁有關鍵字的使用影響關係：經由 SNA 的程度、接近與中介等三項中心度，加以深入說明與解釋不同關鍵字類別及其所擁有關鍵字之間的彼此流動或影響關係。
3. 發掘關鍵字類別及其所擁有關鍵字的共同使用現象：整體而言，社會標記者傾向以較多組且較少的層面分析觀點提供關鍵字，同時多數層面分析幾乎為兩種不同的觀點；但是專家則是傾向使用較少組且較多的層面分析觀點提供關鍵字，且多數層面分析超過兩種以上的觀點，以組織各項資訊資源。另外，社會標記者也具備使用同質的關鍵字類別，亦即社會標記者會同時使用同一群組的關鍵字類別及其所擁有關鍵字，進行資訊資源的主題標引，而專家也具備此一傾向。
4. 歸納關鍵字類別及其所擁有關鍵字之同等角色的對等關係：以 SNA 同

等角色的組合而言，社會標記者最常使用題名群組的關鍵字類別組合成 SNA 同等角色，而專家則最常以題名與內容群組的關鍵字類別組合成 SNA 同等角色。

5. 分析關鍵字類別及其所擁有關鍵字的使用涵蓋範圍：以 FT-tree 路徑式規則的集叢數量而言，社會標記者傾向使用較多元化的關鍵字類別，而專家則傾向集中使用較少的關鍵字類別。如果以屬性的群組加以聚合後，可以發現社會標記者使用的關鍵字類別集中在題名、主題、內容、功能與其他等 5 個相關群組，而專家則集中在題名、主題與內容等 3 個相關群組。
6. 探討關鍵字類別及其所擁有關鍵字的最低使用基準：以最短路徑式規則而言，社會標記者傾向至少使用 1 種以上不同的單一關鍵字類別及其所擁有關鍵字進行文獻標引，而專家則是傾向至少使用 2 種以上不同的關鍵字類別及其所擁有關鍵字組織各項資訊資源。
7. 比較關鍵字類別及其所擁有關鍵字的使用決策樹類型：以最長路徑式規則、路徑式規則數量，以及組成單一路徑式規則與樣式型規則的平均組成階層而言，社會標記者傾向使用較少的路徑式規則且較少的關鍵字類別進行文獻標引，而專家則是傾向使用較多的路徑式規則且較多的關鍵字類別，組織各式資訊資源。
8. 發現關鍵字類別及其所擁有關鍵字的使用連動關係：經由 FP-tree 與 FP-growth 的分析，可以由 FP-tree 路徑式規則及其 FP-growth 規則組合中，發掘社會標記者與專家之使用關鍵字的決策樹及其路徑，乃至於樣式等。以 FP-growth 規則的群組性質而言，社會標記者使用社會標記時，傾向選用異質群組且不同關鍵字類別及其所擁有關鍵字，但同質且不同關鍵字類別及其所擁有關鍵字的傾向亦不低，而且幾乎是選用某一種關鍵字類別的關鍵字後，接連引出使用另外一種關鍵字類別的關鍵字。就專家使用控制詞彙進行文獻標引時，則是集中選用異質群組且不同的關

鍵字類別及其所擁有關鍵字，而且八成以上是使用某一種關鍵字類別的關鍵字後，後續連動使用另外兩種以上關鍵字類別的關鍵字。另外，就 FP-growth 分析結果而言，社會標記者與專家的共同取向是使用「指明主題—一般」類別的關鍵字後，會選取題名群組的關鍵字作為文獻標引之用。

四、從心智模式解釋社會標記與控制詞彙兩者互補現象的原因

雖然有些研究已證實社會標記與控制詞彙間具有互補的現象（Kipp, 2006, 2011a, 2011b; Lu, Park, & Hu, 2010; Rolla, 2009; Thomas, Caudle, & Schmitz, 2009; Yi & Chan, 2009），但是這些研究並未進一步說明社會標記與控制詞彙間具備互補的原因。然而，本研究提出的文獻標引心智模式除了是一種理論框架外，更能從原始的心智模式論理中為社會標記與控制詞彙間的互補現象提供一種理論基礎，進而解釋其異同的根本原因。依據本研究提出的文獻標引心智模式而言，雖然社會標記者與專家面對同一資訊物件時，會進而具體轉化為關鍵字，藉以表徵其對資訊物件的一種心理認知。然而，由於先備經驗與知識的不同，使用完全相同關鍵字的比率約為 13% 以下，同一文獻具備相同的關鍵字比率約為 28%（請參見前述第四章第二節社會標記與控制詞彙之間的關鍵字比對）。因而，就關鍵字的使用而言，社會標記者與專家兩者間的文獻標引心智模式的相似性並不高。另外，從冪次定律分佈、SNA 中心度、階層集叢與同等角色，以及 FP-tree 與 FP-growth 等分析結果而言，社會標記者與專家兩者間的文獻標引心智模式的相似性亦不高。

五、擴展 TCM 的可行性驗證與應用解釋

在原始的 TCM 模式中，係由 LTCM、FTCM 與 T2TCM 等三個子模式共同組成，同一樣本分別獨立應用在三個子模式且重複入類，藉以分析社會標記的使

用情形 (Heckner et al., 2008)。在本研究中，則是同時採用上述三個子模式，但是同一研究樣本只歸類至三個子模式之一，並不分別重複歸類至三個不同的子模式之中。除此之外，本研究也進一步應用至控制詞彙，以探索專家使用關鍵字的情形。因而，本研究已擴展 TCM 的原始應用方式與範圍，同時驗證其可行性，進而比較社會標記者與專家間的異同之處。更進一步，本研究結果也證實 TCM 的 LTCM、FTCM 與 T2TCM 等三個子模式可以相互支援，以提供更多元化的分析觀點，而非單一面向的分析。此外，經由應用特徵檔的導入，可以更彈性調整 TCM 及其三個子模式的類別，擴大其適用的範圍與情境。

第三節 未來研究建議

如前所述，雖然本研究已發展文獻標引心智模式，以及分析出社會標記者與專家的文獻標引心智模式地圖內容與其樣式，但也有其限制，可提供後續學術研究與應用之參考。

一、發展一體兩面的文獻標引系統：一般使用者與專家

以往的文獻標引系統皆為專家設計而成，藉以管理某一領域或機構的資訊資源及其知識。隨著 Web 2.0 與的廣泛應用，資訊組織也朝向所謂的個人資訊管理的方向發展，而社會標記正是因應此種潮流趨勢的其中應用之一。正如許多研究結果指出個人化的社會標記與學科領域型的控制詞彙兩者間呈現著互補的關係，如何發展一體兩面的文獻標引系統，藉以聯合 (align) 兩種不同取向的文獻標引方式與系統，提供不同面向的文獻標引服務及其相互支援，達成更好的資訊組織，則有待未來研究加以深入探討。例如，以資訊組織既有的文獻保證原則 (literary warrant) 為基礎，進一步從使用者保證原則 (user warrant) (NISO, 2005) 衍化擴展成所謂的關鍵字保證原則 (keyword warrant)，以現有的社會標記與控制詞彙為依據，為兩種不同類型的關鍵字間建立鏈結，進一步發展成共同出現的

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

關鍵字網路。前述此種關鍵字網路除了有助於圖資界專家瞭解個人化關鍵字的應用，以及列入控制詞彙的候選清單之外，而且有助於瞭解一般使用者與專家間使用關鍵字及其表徵主題的概念與知識等異同，更可以進一步作為社會標記者的關鍵字查詢與標引之參考基礎，也同時可以將吳美美（1998）提出之個人與領域等兩種層次的資訊組織予以聯合在一起。

二、擴增文獻標引心智模式的研究範圍：加入作者提供的關鍵字

雖然本研究以社會標記者的社會標記與專家的控制詞彙為研究樣本，同時經由冪次定律分佈、SNA 的中心度、階層集叢與共同角色，以及 FP-tree 與 FP-growth 等不同觀點與方法分析外，以歸納與探討社會標記者與專家使用關鍵字的特質，及所擁有的心智模式地圖。整體而言，本研究只包括了一般使用者與專家所使用的關鍵字及其文獻標引心智模式，並未將文獻作者提供的關鍵字列入研究，以探討其文獻標引心智模式的特質與地圖。如第二章文獻探討所述，一個完整的心智模式包括了設計者與使用者，如果未來的研究能夠同時涵蓋一般使用者、作者與專家等三種不同的角色，及其使用的關鍵字，屆時的研究成果與文獻標引心智模式地圖除了更趨於完整外，也更能瞭解三者間的異同，而且介於彼此兩兩之間的文獻標引落差才能予以消弭，以提供更佳的文獻標引服務。

三、展延文獻標引心智模式的驗證對象：記述編目的著錄項目

就傳統分類編目的範圍而言，可以區分成記述編目與主題編目等兩種不同類型的資訊組織，兩者是一體兩面的，同時也有利於資訊檢索之用。以本研究的範圍而言，係屬於主題編目的範疇之內。如果未來的研究能以本研究提出的文獻標引心智模式為基礎，將研究範圍擴展至記述編目，並以記述編目的著錄資料項目為研究標的對象，除了可以擴大證實文獻標引心智模式的可行性，以及展延 Bar-Ilan et al.（2008）的既有研究對象與範圍外，亦可進一步充實更多的文獻標

引心智模式理論基礎與案例。另外，未來的研究可就一般使用者、作者與專家的資料著錄需求加以檢視，以及比較其異同，除了有助於記述編目的研究外，也能消弭三者間記載資料的文獻標引落差，更有助於資訊檢索與網路資源搜尋，乃至於深入探討電子化研究（e-Research）的文獻記載與註解等需求。主要的研究重點項目可以包括：著錄的資料項目，以及其著錄的頻率、順序與其連動關係，乃至於著錄的目的或功能等。

四、融合文獻標引心智模式及其特性的資訊系統應用設計

在 Pisanki and Žumer (2011a, 2011b) 的 FRBR 研究中，提出一般使用者與專家間具有類似的 FRBR 心智模式，此項結果似乎與本研究結果相互衝突，其實不然。在 Pisanki and Žumer (2011a, 2011b) 的 FRBR 研究設計過程中，其中有專家的介入，提供受測者有關 FRBR 專業知識的解說導引。如果一個良好的資訊組織系統能被一般使用者充分瞭解與應用，專業知識的轉化與指引是不可或缺且十分重要的。基於此故，可能才造成本研究結果與 Pisanki and Žumer (2011a, 2011b) 的研結果有所不同。然而，資訊組織的文獻標引與資訊檢索是一體兩面的，未來研究亦可就既有的文獻標引心智模式地圖為基礎進行展延（如前述第四章第九節資訊系統設計的意涵一節內容），與現有的資訊檢索心智模式地圖予以串連與整合（如前述的 Cole et al. (2007) 與 Zhang (2008b)），發展出相關的功能作業與流程，以設計一個完全以使用者需求為導向的資訊系統，作為個人化資訊組織與檢索之用。

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

參考文獻

- 卜小蝶、張淇龍 (2009)。社會性書籤網站之使用者與標籤特性初探。圖書資訊學研究。4 卷 1 期，頁 1-26。
- 吳美美 (1998)。網際網路資源組織的三個層次。大學圖書館。2 卷 1 期，頁 27-35。
- 李亞婷、馬費成 (2012)。基於標籤共現的社會網路分析研究。情報雜誌。31 卷 7 期，頁 103-109。
- 陳和琴、張慧銖、江綉瑛與陳昭珍 (2003)。資訊組織 (初版)。臺北縣蘆洲市：國立空中大學。
- 圖書館自動化作業規劃委員會、中國編目規則研訂小組 (研訂)、中華民國圖書館學會分類編目委員會 (增修) (2005)。中國編目規則 (第三版)。臺北市：中華民國圖書館學會。
- Ahlstrom, V., & Allendoerfer, K. (2004). *Information organization for a portal using a card-sorting technique*. Retrieved 20 April, 2011 from <http://hf.tc.faa.gov/technotes/dot-faa-ct-tn04-31.pdf>
- Anderson, J.D. (1997). *NISO technical report 2 : Guidelines for indexes and related information retrieval devices*. Bethesda, MD: NISO Press.
- Angus, E., Thelwall, M., & Stuart, D. (2008). General patterns of tag usage among university groups in Flickr. *Online Information Review*, 32(1), 89-101.
- Ansari, M. (2005). Matching between assigned descriptors and title keywords in medical theses. *Library Review*, 54(7), 410-414.
- Bar-Ilan, J., Shoham, S., Idan, A., Miller, Y., & Shachak, A. (2008). Structured versus unstructured tagging: A case study. *Online Information Review*, 32(5), 635-647.
- Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, 38(2), 61-71.
- Bilal, D., & Wang, P. (2005). Children's conceptual structures of science categories and the design of Web directories. *Journal of the American Society for Information Science and Technology*, 56(12), 1303-1313.
- Bischoff, K., Firan, C. S., Nejdil, W., & Paiu, R. (2008). Can all tags be used for search? In *Proceeding of the 17th ACM conference on Information and knowledge management* (pp. 203-212). New York: ACM Press.
- Blackwell, A.F. (1996). Metacognitive theories of visual programming : What do we

- think we are doing. In *Proceedings of IEEE Symposium on Visual Languages*(pp. 240-246).
- Brookes, B.C. (1980). The foundations of information science: Part I. Philosophical aspects. *Journal of Information Science*, 2(3-4), 125-133.
- Bruce, R. (2008). Descriptor and folksonomy concurrence in education related scholarly research. *Webology*, 5(3). Retrieved 26 July, 2011 from <http://www.webology.org/2008/v5n3/a59.html>
- Carlyle, A. (1989). Matching LCSH and user vocabulary in the library catalog. *Cataloging & Classification Quarterly*, 10(1/2), 37-63.
- Carvalho, M., Hewett, R., & Canas, A.J. (2001). Enhancing Web searches from concept map-based knowledge models. Proceedings of the SCI2001 5th Multi-Conference on Systems, Cybernetics and Informatics(pp. 1–5). Orlando, FL: International Institute of Informatics and Systematics. Retrieved 16 November, 2012 from <http://www.ihmc.us/users/acanas/Publications/EnhancedWebSearches/Enhancing%20Web%20Searches%20from%20Concept%20Map-based%20Knowledge%20Models.pdf>
- Celma, Ò., Herrera, P., & Serra, X. (2005). Bridging the music semantic gap. Retrieved 22 December, 2012 from <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=B6C832AEFF8A42AE74BB4932EDC6D05F?doi=10.1.1.89.4319&rep=rep1&type=pdf>
- Chan, L.M., & Vizine-Goetz, D. (1998). Toward a computer-generated subject validation file: Feasibility and usefulness. *Library Resources & Technical Services*, 42(1), 45-60.
- CiteULike. (n.d.). *Frequently asked questions*. Retrieved 8 May, 2012 from <http://www.citeulike.org/faq/faq.adp>
- Cole, C., Lin, Y., Leide, J., Large, A., & Beheshti, J. (2007). A classification of mental models of undergraduates seeking information for a course essay in history and psychology: Preliminary investigations into aligning their mental models with online thesauri. *Journal of the American Society for Information Science and Technology*, 58(13), 2902-2104.
- Cronin, B., & Meho, L.I. (2008). The shifting balance of intellectual trade. *Journal of the American Society for Information Science and Technology*, 59(4), 551–564.
- de Abreu Moreira, D. (1998). Object oriented computing. Retrieved 22 December, 2012 from http://java.icmc.usp.br/books/ooc/html/object_oriented_programming_semantic_gap.html
- Eden, C. (1992). On the nature of cognitive maps. *Journal of Management Studies* 29(3), 261-265.

- Fleiss J.L. (1981). *Statistical methods for rates and proportions*(2nd ed.). New York: Wiley.
- Freeman, L.A. (2004). The effects of concepts maps on requirements elicitation and system models during information systems development. In A.J. Cañas, J.D. Novak, & F.M. González, (Eds.), *Concept maps: Theory, methodology, technology*. Proceedings of the 1st International Conference on Concept Mapping, Pamplona, Spain.
- Frost, C.O. (1989). Title words as entry vocabulary to LCSH: Correlation between assigned LCSH terms and derived terms from titles in bibliographic records with implications for subject access in online catalogs. *Cataloging & Classification Quarterly*, 10(1-2), 165-179.
- Gentner, D., & Stevens, A. (Eds.). (1983). *Mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gilliland, A.J. (2008). *Setting the stage*(ver. 3.0). Retrieved 26 July, 2011 from http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.pdf
- Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198-208.
- Gorman, M., & Winkler, P. (Eds.).(1998). *Anglo-American cataloging rules*(2nd rev. ed.). Ottawa: Canadian Library Association.
- Guerra, E., Sanz, D., Díaz, P., & Aedo, I. (2007). A transformation-driven approach to the verification of security policies in web designs. In *ICWE'07 Proceedings of the 7th International Conference on Web Engineering*(pp. 269-284). Retrieved 22 May, 2012 from http://astreo.ii.uam.es/~eguerra/papers/Guerra_2007_ICWE.pdf
- Gwet, K.L. (2012). *Handbook of inter-rater reliability*(3rd ed.). Gaithersburg: Advanced Analytics, LLC.
- Hanneman, R.A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside. Retrieved 25 December, 2011 from http://faculty.ucr.edu/~hanneman/nettext/Introduction_to_Social_Network_Methods.pdf
- Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*(pp. 1-12). New York, NY: ACM.
- Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Library and Information Science Research*, 18(4), 323-342.

- Haythornthwaite, C. (2009). Social networks and information transfer. In *Encyclopedia of Library and Information Sciences*(3rd ed., pp. 4837-4847). New York: Taylor and Francis.
- Heckner, M., Mühlbacher, S., & Wolff, C. (2008). Tagging tagging: Analysing user keywords in scientific bibliography management systems. *Journal of Digital Information*, 9(2). Retrieved 26 July, 2011 from <http://epub.uni-regensburg.de/6839/1/HeckneretalTaggingTaggingJoDI2008.pdf>
- Heery, R., & Patel, M. (2000). Application profiles: mixing and matching metadata schemas. *Ariadne*, 25. Retrieved 8 August, 2011 from <http://www.ariadne.ac.uk/issue25/app-profiles>
- Heymann, P., & Garcia-Molina, H. (2009). Contrasting controlled vocabulary and tagging: Do experts choose the right names to label the wrong things? In *Proceedings of Second ACM International Conference on Web Search and Data Mining (WSDM '09)*. Retrieved 10 June, 2011 from http://www.wsdm2009.org/heyman_2009_tagging.pdf
- Heymann, P., Koutrika, G., & Garcia-Molina, H. (2008). Can social bookmarking improve web search? In *Proceedings of the Second International Conference on Web Search and Web Data Mining* (pp. 195-206). New York: ACM Press.
- Holman, L. (2011). Millennial students' mental models of search: Implications for academic librarians and database developers. *Journal of Academic Librarianship*, 37(1), 19-27.
- Hotho, A., Jächke, R., Schmitz, C., & Stumme, G. (2006). Information retrieval in folksonomies: Search and ranking. In *Proceedings of ESWC 2006*. Retrieved 8 August, 2011 from <http://www.kde.cs.uni-kassel.de/stumme/papers/2006/hotho2006information.pdf>
- Huang, W., Li, S.F., Tan, Y.J., & Gao, B. (2009). Association rules based short text feature extension. *International Journal of Computer Science and Network Security*, 9(10), 227-230.
- Ingwersen, P. (1996). Cognitive perspective of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 52(11), 3-50.
- Inskip, C., MacFarlane, A., & Rafferty, F. (2008). Meaning, communication, music: Towards a revised communication model. *Journal of Documentation*, 64(5), 687-706.
- Iyer, H., & Bungo, L. (2011). An examination of semantic relationships between professionally assigned metadata and user-generated tags for popular literature in complementary and alternative medicine. *Information Research*, 16(3). Retrieved

- 16 September, 2011 from <http://informationr.net/16-3/paper482.html>
- Johnson-Laird, P.N. (1989). Mental models. In M.I. Posner (Ed.). *Foundations of Cognitive Science*(pp. 469-499). Cambridge, MA: MIT Press.
- Kearney, A.R., & Kaplan, S. (1997). Toward a methodology for the measurement of knowledge structures of ordinary people. *Environment and Behavior*, 29(5), 579-617. Retrieved 16 November, from http://deepblue.lib.umich.edu/bitstream/2027.42/66794/2/10.1177_0013916597295001.pdf
- Kipp, M.E. (2006). *Complementary or discrete contexts in online indexing: A comparison of user, creator and intermediary keywords*. Retrieved 31 December, 2010 from <http://eprints.rclis.org/bitstream/10760/8771/1/mkipp-caispaper.pdf>
- Kipp, M.E.I. (2011a). Tagging of biomedical articles on CiteULike: A comparison of user, author and professional indexing. *Knowledge Organization*, 38(3), 245-261.
- Kipp, M.E.I. (2011b). User, author and professional indexing in context: An exploration of tagging practices on CiteULike. *The Canadian Journal of Information and Library Science*, 35(1), 17-48.
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Lin, X., Beaudoin, J.E., Bui, Y., & Desai, K. (2006). Exploring characteristics of social classification. In J. Furner and J.T. Tennis. (Eds.), *Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop*. Retrieved 22 February, 2011 from <http://faculty.cis.drexel.edu/~xlin/papers/ASIS200>
- Lu, C., Park, J., & Hu, X. (2010). User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress subject headings. *Journal of Information Science*, 36(6), 763-779.
- Marchionini, G. (1989a). Information-seeking strategies of novices: Using a full-text electronic encyclopedia. *Journal of the American Society for Information Science*, 40(1), 54-66.
- Marchionini, G. (1989b). Making the transition from print to electronic encyclopaedias: Adaptation of mental models. *International Journal of Man-Machine Studies*, 30(6), 591-618.
- Makri, S., Blandford, A., Gow, J., Rimmer, J., Warwick, C., & Buchanan, G. (2007). A library or just another information resource: A case study of users' mental models of traditional and digital libraries. *Journal of the American Society for Information Science and Technology*, 58(3), 433-445.

- Marlow, C., Naaman, M., Body, D., & Davis, M. (2006). Position paper: Tagging, taxonomy, Flickr, article, toread. In *Proceedings of the Collaborative Web Tagging Workshop at the WWW 2006*. New York: ACM Press. Retrieved 30 December, 2010 from, <http://www.danah.org/papers/WWW2006.pdf>
- Maron, M.E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, 28(1), 261-266.
- Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Cannon-Bowers, J.A., & Salas, E. (2005). Scaling the quality of teammates' mental models: Equifinality and normative comparisons. *Journal of Organizational Behavior*, 26(1), 37-56.
- Michell, G, & Dewdney, P. (1998). Mental models theory: Applications for library and information science. *Journal of Education for Library and Information Science*, 39(4), 275-281.
- Mitchell, E. (n.d.). *The role of information organization and metadata in digital documents*. Retrieved 21 May, 2012 from http://erikmitchell.info/uploaded_files/dissertation/2_inforg_metadata_mitchell.pdf
- Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management*, 36(4), 876-910.
- Mohammed, S., Klimoski, R., & Rentsch, J.R. (2000). The measurement of team mental models: We have no shared schema. *Organizational Research Methods*, 3(2), 123-165.
- Munk, T. B., & Mørk, K. (2007). Folksonomies, tagging communities, and tagging strategies: An empirical study. *Knowledge Organization*, 34(3), 115-127.
- Newman, M.E.J. (2003). *The structure and function of complex networks*. Retrieved 03 November, 2010 from <http://www-personal.umich.edu/~mejn/courses/2004/cscs535/review.pdf>
- NISO. (2005). *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Bethesda, MD: NISO Press.
- Norman, D.A. (1983). Some observations on mental models. In D. Gentner, & A. Stevens (Eds.). *Mental models*(pp. 7-14). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pisanski, J., & Žumer, M. (2010a). Mental models of the bibliographic universe. Part 1: Mental models of descriptions. *Journal of Documentation*, 66(5), 643-667.
- Pisanski, J., & Žumer, M. (2010b). Mental models of the bibliographic universe. Part 2: Comparison task and conclusions. *Journal of Documentation*, 66(5), 668-680.

- Prebor, G. (2010). Analysis of the interdisciplinary nature of library and information science. *Journal of Information Science*, 42(4), 256–267.
- ProQuest. (2009). *LISA guide*. Retrieve 8 May, 2012 from <http://www.csa.com/factsheets/supplements/LISAguide.pdf>
- Quintarelli, E. (2005). *Folksonomies: Power to the people*. Paper presented at ISKO Italy-UniMIB Meeting. Retrieve 28 October, 2010 from <http://www.iskoi.org/doc/folksonomies.htm>
- Rafferty, P., & Hilderley, R. (2007). Flickr and democratic indexing: Dialogic approaches to indexing. *Aslib Proceedings: New Information Perspectives*, 59(4/5), 397-410.
- Ravari, M.T. (2012). A statistical study on Persian subject headings development. *International Journal of Information Science and Management*, 10(1), 73-88. Retrieved 23 May, 2012 from http://www.ricest.ac.ir/ijist/Vol10N1/ijism-V10N1_files/ijism101-73-88.pdf
- Rentsch J.R., & Hall R.J. (1994). Members of great teams think alike: A model of team effectiveness and schema similarity among team members. In *Advances in Interdisciplinary Studies of Work Teams, Vol. 1*(pp. 223-261). Greenwich, CT: JAI Press.
- Rentsch, J.R., & Klimoski, R.J. (2001). Why do great minds think alike: Antecedents of team member schema agreement. *Journal of Organizational Behavior*, 22(2), 107-120.
- Rentsch, J. R., Small, E. E., & Hanges, P. J. (2008). Cognitions in organizations and teams: What is the meaning of cognitive similarity? In B. Smith (Ed.), *The people make the place: Dynamic linkages between individuals and organizations*. LEA's organization and management series(pp. 127-155). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.
- Rolla, P.J. (2009). User tags versus subject heading: Can user-supplied data improve subject access to library collections? *Library Resource & Technical Service*, 53(3), 174-184.
- Rorissa, A. (2010). A comparative study of Flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science and Technology*, 61(11), 2230-2242.
- Rouse, W.B., Cannon-Bowers, J.A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, & Cybernetics*, 22(6), 1296-1308.

- Schaffernicht, M., & Groesser, S. (2011). A comprehensive method for comparing mental models of dynamic systems. *European Journal of Operational Research*, 210(1), 57-67.
- Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, M., & Riedl, J. (2006). Tagging, communities, vocabulary, evolution. In *CSCW '06: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (pp. 181-190). New York: ACM Press.
- Smith, G. (2008). *Tagging: People-powered metadata for the social web*. Berkeley, CA: New Riders.
- Spalding, T. (2007). When tags work and when they don't: Amazon and LibraryThing. *Thingology Blog*, Retrieved 30 December, 2011 from <http://www.librarything.com/thingology/2007/02/when-tags-works-and-when-they-dont.php>
- Spiteri, L. F. (2007). Structure and form of folksonomy tags: The road to the public library. *Webology*, 4(2). Retrieved 14 July, 2011 from <http://www.webology.org/2007/v4n2/a41.html>
- Staggers, N., & Norcio, A.F. (1993). Mental models: Concepts for human-computer interaction research. *International Journal of Man-Machine Studies*, 38(4), 587-605.
- Strader, R.C. (2009). Author-assigned keywords versus library of congress subject headings: Implications for the cataloging of electronic theses and dissertations. *Library Resources and Technical Service*, 53(4), 243-250.
- Stvilia, B., & Jørgensen, C. (2010). Member activities and quality of tags in a collection of historical photographs in Flickr. *Journal of the American Society for Information Science and Technology*, 61(12), 2477-2489.
- Sutton, S.A. (1994). The role of attorney mental models of law in case relevance determinations: An exploratory analysis. *Journal of the American Society for Information Science*, 45(3), 186-200.
- Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, MA: MIT Press.
- Thomas, M., Caudle, D.M., & Schmitz, C.M. (2009). To tag or not to tag? *Library Hi Tech*, 27(3), 411-434.
- Toker, S. (2012). *The mental model comparison of expert and novice performance improvement practitioners*. Cambridge, UK: ProQuest, UMI Dissertation Publishing.
- Vander Wal, T (2007). *Folksonomy*. Retrieved 08 May, 2012, from

- <http://www.vanderwal.net/folksonomy.html>
- Voorbij, H.J. (1998). Title keywords and subject descriptors: A comparison of subject entries of books in the humanities and social science. *Journal of Documentation*, 54(4), 466-476.
- Wang, P., & Soergel, D. (1998). A cognitive model of document use during a research project: Study I. Document selection. *Journal of the American Society for Information Science*, 49(2), 115-133.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.
- Winn, W. (2001). Cognitive perspectives in psychology. In D.H. Jonassen (Ed.), *Handbook of research for educational communications and technology*(pp. 79-112). Bloomington, IN: AECT. Retrieved 16 May, 2011 from <http://www.aect.org/edtech/ed1/04.pdf>
- Yang, Y., Wu, M., Cui, L. (2012). Integration of three visualization methods based on co-word analysis. *Scientometrics*, 90(2), 659-673.
- Yeung, A.K. (1998). Information organization and data structure. Retrieved 02 January, 2013 from <http://www.ncgia.ucsb.edu/giscc/units/u051/>
- Yi, K., & Chan, L.M. (2009). Linking folksonomy to Library of Congress subject headings: An exploratory study. *Journal of Documentation*, 65(6), 872-900.
- Yuan, J., Wu, Y., & Yang, M. (2007). Discovery of collocation patterns: from visual words to visual phrases. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*(pp. 1-8).
- Zhang, J., Chen, C., & Li, J. (2009). Visualizing the intellectual structure with paper-reference matrices. *IEEE Transaction on Visualization and Computer Graphics*, 15(6), 1153-1160.
- Zhang, X., & Chignell, M. (2001). Assessment of the effects of user characteristics on mental models of information retrieval systems. *Journal of the American Society for Information Science and Technology*, 52(6), 445-459.
- Zhang, Y. (2008a). The influence of mental models on undergraduate students' searching behavior on the web. *Information Processing and Management*, 44(3), 1330-1345.
- Zhang, Y. (2008b). Undergraduate students' mental models of the web as an information retrieval system. *Journal of the American Society for Information Science and Technology*, 59(13), 2087-2098.

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

附錄

附錄一 FP-tree 路徑規則中的各階層社會標記關鍵字類別

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	規則序號		
04	06	19	03				001		
	13	15					002		
		22					003		
	10	08					004		
	26	17					005		
	05						006		
	23						007		
	02						008		
	18						009		
01	04	10					010		
		19					011		
		05					012		
		06	05				013		
		11					014		
		22					015		
	15						016		
	09	04	10					017	
			17					018	
			06					019	
			23		11			020	
			11					021	
			26					022	
		12	06	19					023
				26					024
				10	23				025
			19	18					026
				19					027
				16					028
			23	17				029	
			04	16				030	

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	規則序號
			07				031
			02				032
			22				033
			05				034
			14				035
			13	17	16	21	036
		10	19				037
			02				038
		22					039
		05	02				040
			10	22	11		041
			18				042
			23				043
		06	11				044
			10	03			045
			02				046
			05				047
			03				048
			26	08			049
			18-				050
			22				051
		18					052
		13					053
		17	16				054
		02					055
		19-					056
		23					057
		03	11				058
			21				059
		08					060
		26					061
	05	26	15				062
		02					063
		23					064
	12	06	11				065
			10				066

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	規則序號
			15				067
		20					068
		02	25				069
			15				070
		05					071
		18					072
		04	10				073
			05				074
			19				075
		03					076
		16					077
		17					078
		23	26				079
		26					080
		10	22				081
			11				082
		11					083
	10	14					084
		19					085
		23					086
		22					087
	17	21					088
	02	26					089
		17					090
	08						091
	06	10					092
		02	17				093
		19					094
	18	13					095
	19	22					096
		18					097
	23	18					098
	25						099
	03						0100
	26	15					0101
		08					0102

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	規則序號	
	16						0103	
	14						0104	
	13	22					0105	
	07						0106	
	21						0107	
	11						0108	
09	10	03					0109	
		11					0110	
		19	16				0111	
		18					0112	
	16						0113	
	17						0114	
	04	23						0115
			06	05				0116
		19	23				0117	
		10					0118	
		16					0119	
		05	02				0120	
		03					0121	
	12	06	03					0122
			04					0123
		10					0124	
		05					0125	
		02	18				0126	
		18					0127	
		22					0128	
	06	10	19	02	22			0129
			18					0130
		05					0131	
		19					0132	
	23	13					0133	
		17					0134	
	19	15	07				0135	
		23	02			26	0136	
		25					0137	
	05	02					0138	

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	規則序號
		19					0139
		18					0140
		26					0141
		23					0142
	26						0143
	15						0144
	02	25					0145
		11					0146
		17					0147
	03						0148
	13						0149
	18	26					0150
	22						0151
	07						0152
	11						0153
	25						0154
	14						0155
	21						0156
	08						0157
12	04	05					0158
		02	11				0159
	06	10	26				0160
		05					0161
	15						0162
	08						0163
	26	15	17				0164
	14						0165
	05						0166
	10						0167
	03						0168
	18	13					0169
		22					0170
	11						0171
11							0172
02	13						0173
	26	13					0174

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	規則序號
	17	24					0175
13-	22						0176
23	18						0177
	15						0178
	26						0179
	13						0180
	03						0181
22						0182	
05	08						0183
	16						0184
	07						0185
	19	17					0186
	21						0187
	23						0188
	02						0189
	13						0190
25						0191	
18						0192	
19	22						0193
	16						0194
10	26	13				0195	
26	11					0196	
24						0197	
07						0198	
06	05	19	23				0199
	16						0200
	19	13					0201
	03						0202
	11						0203
03						0204	
21						0205	
16						0206	

附錄二 FP-tree 路徑規則中的各階層控制詞彙關鍵字類別

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	第八階層	規則序號
09	01	04	17					01.
			16					02.
			05	02				03.
				06	18			04.
					26			05.
				17				06.
				14	18			07.
			06	18				08.
				19				09.
				14				10.
			02	16	18			11.
			18	11				12.
			19					13.
			14	26				14.
				18	11			15.
				17	16	18		16.
					02			17.
			26					18.
09	01	12	04	14	02			19.
09	01	12	04	14	17			20.
				06	02			21.
					18			22.
					17			23.
				02				24.
				05	19	18		25.
					16			26.
				19				27.
				16	18			28.
				17	14	10		29.
			18					30.
			06	19				31.
				26				32.
				16				33.

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	第八階層	規則序號
				14	17			34.
				17	02			35.
			19	02	18			36.
				18				37.
				14	16	18		38.
					17			39.
09	01	12	19	16				40.
				26				41.
			02	18				42.
				16				43.
			05	19				44.
				16	18			45.
				02	18			46.
				06	14			47.
				14	17			48.
				18				49.
				26				50.
				17				51.
			16					52.
			14	17	18			53.
				16	25			54.
			17	16	18			55.
			10					56.
		06	19	14				57.
				16				58.
			02	26				59.
09	01	06	16					60.
			14					61.
			26					62.
		05	19					63.
			18					64.
			06					65.
			14					66.
			17					67.
			02					68.
			26					69.

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	第八階層	規則序號
		19	16					70.
			02					71.
			26					72.
			18					73.
			17	18				74.
			14					75.
		18						76.
		16						77.
		02	16					78.
		14	17	02				79.
09	01	26						80.
		17	18					81.
			14	16	18			82.
	12	04	14-	17	18			83.
			02	18				84.
			05	18				85.
				14	17			86.
				17	16			87.
				16				88.
			19	14	17			89.
			06	17	16	18		90.
				19	14	02	18	91.
				02				92.
				16				93.
			16					94.
			17	02	16			95.
		14	02					96.
			17					97.
			16					98.
		02	15					99.
09	12	05	19					100.
			06					101.
			14					102.
			18					103.
		19						104.
		06	19					105.

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	第八階層	規則序號
			16					106.
			14	17	16	18		107.
					18			108.
			02	18				109.
		16						110.
		18						111.
		17	02					112.
			16					113.
		25						114.
	04	05	14	18				115.
				17	18			116.
			19					117.
			06	16				118.
		19	02					119.
09	04		14	16	18			120.
			18					121.
		18	26					122.
		06	19					123.
			17					124.
		02						125.
		13						126.
		14	17	18				127.
		16						128.
		17	14					129.
			16					130.
		26						131.
	05	06	19					132.
			26					133.
			14	16	18	26		134.
			17					135.
		19						136.
		14	18					137.
		16						138.
09	05	17	18					139.
		17	02					140.
		02						141.

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	第八階層	規則序號
	06	19	17	04				142.
			02					143.
		17	14	16	18			144.
		03						145.
		14	16					146.
			17					147.
	19	14						148.
		02						149.
		17	16					150.
			14	16				151.
	18							152.
	02	16	18					153.
	17	14	18					154.
		16						155.
		02						156.
	14	02						157.
	16							158.
09	13							159.
01	04	06	14					160.
		05	26					161.
		19	17					162.
		17						163.
		14	17					164.
			16	18				165.
		02						166.
		16						167.
		26						168.
	05	16						169.
		06	14					170.
		18						171.
		14	18					172.
	12	14	18					173.
			17					174.
		04	19	02	18			175.
			05	06	02			176.
				19	17			177.

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	第八階層	規則序號
01	12			14	17			178.
				02				179.
		04	14					180.
		17						181.
		19						182.
		05	19					183.
			06					184.
			17	02				185.
		16						186.
		06	17					187.
			02					188.
			18					189.
	06	13						190.
		02						191.
		19						192.
		16						193.
	19							194.
	18							195.
	17	02						196.
01	02	26						197.
	26							198.
04	14	17	02					199.
			18	26				200.
		02						201.
	05	06						202.
		18						203.
		17						204.
		16						205.
	06	14	17					206.
12	04	16						207.
		02						208.
		06	19	14				209.
			17					210.
		05						211.
		14	18					212.
		17						213.

第一階層	第二階層	第三階層	第四階層	第五階層	第六階層	第七階層	第八階層	規則序號
	06							214.
	05	14	17					215.
	17							216.
05	16	18						217.
	19	17						218.
	06							219.
06	19							220.
	02	11						221.
17	02							222.

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

附錄三 以社會標記群組組成的路徑式規則階層及其關鍵

字類別與規則

序號	路徑式規則的階層及其組成的關鍵字類別	規則號碼	數量
001	Content	192, 205, 206	3
002	Content→Content	194	1
003	Content→Function	176, 193	2
004	Function	182	1
005	Function→Content	177-178, 180	3
006	Function→Other	179	1
007	Function→Title	181	1
008	Other	191, 197	2
009	Other→Topic	196	1
010	Title	198, 204	2
011	Title→Content	9, 16, 103-104, 107, 173, 184, 187, 190, 200	10
012	Title→Content→Content	2, 88, 95, 97, 186, 201	6
013	Title→Content→Function	3, 96, 105	3
014	Title→Content→Other	175	1
015	Title→Function	7, 188	2
016	Title→Function→Content	98	1
017	Title→Other	99	1
018	Title→Other→Content	5, 101, 174	3
019	Title→Other→Title	102	1
020	Title→Title	6, 8, 91, 100, 106, 183, 185, 189, 202	9
021	Title→Title→Content	11, 72,, 90, 94	4
022	Title→Title→Content→Function	199	1
023	Title→Title→Content→Title	1	1
024	Title→Title→Function	15, 64	2
025	Title→Title→Other	89	1
026	Title→Title→Other→Content	62	1
027	Title→Title→Topic	10, 14, 92,	3
028	Title→Title→Title	12, 63,	2
029	Title→Title→Title→Content	93	1
030	Title→Title→Title→Title	13,	1
031	Title→Topic	108, 203	2

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

序號	路徑式規則的階層及其組成的關鍵字類別	規則號碼	數量
032	Title→Topic→Content	52-53, 56, 68, 77-78, 84-85	8
033	Title→Topic→Content→Content	54	1
034	Title→Topic→Function	39, 57, 86-87	4
035	Title→Topic→Function→Other	79	1
036	Title→Topic→Other	61, 80	2
037	Title→Topic→Title	4, 55, 60, 71, 76	5
038	Title→Topic→Title→Content	18, 50, 59, 67, 70, 75	6
039	Title→Topic→Title→Function	43, 51	2
040	Title→Topic→Title→Function→Topic	20	1
041	Title→Topic→Title→Other	22, 69,	2
042	Title→Topic→Title→Other→Title	49	1
043	Title→Topic→Title→Title	19, 40, 46-48, 74,	6
044	Title→Topic→Title→Topic	17, 21, 44, 58, 65-66, 73	7
045	Title→Topic→Title→Topic→Title	45,	1
046	Title→Topic→Title→Topic→Function→Topic	41	1
047	Title→Topic→Topic	83	1
048	Title→Topic→Topic→Content	27, 28, 35, 37, 42	5
049	Title→Topic→Topic→Content→Content→ Content→Content	36	1
050	Title→Topic→Topic→Function	33, 81	2
051	Title→Topic→Topic→Function→Content	29	1
052	Title→Topic→Topic→Other	24	1
053	Title→Topic→Topic→Title	31-32, 34, 38	4
054	Title→Topic→Topic→Title→Content	23, 30	2
055	Title→Topic→Topic→Topic	82	1
056	Title→Topic→Topic→Topic→Content	26	1
057	Title→Topic→Topic→Topic→Function	25	1
058	Topic	172	1
059	Topic→Content	113-114, 144, 149, 155-156, 162, 165	8
060	Topic→Content→Content	169	1
061	Topic→Content→Content→Title	135	1
062	Topic→Content→Other	137, 150	2
063	Topic→Content→Function	170	1
064	Topic→Content→Function→Title→Other	136	1
065	Topic→Function	151	1
066	Topic→Function→Content	133-134	2

序號	路徑式規則的階層及其組成的關鍵字類別	規則號碼	數量
067	Topic→Other	143, 154	2
068	Topic→Other→Content	195	1
069	Topic→Other→Content→Content	164	1
070	Topic→Title	148, 152, 157, 163, 166, 168	6
071	Topic→Title→Content	119, 132, 139-140, 147	5
072	Topic→Title→Content→Function	117	1
073	Topic→Title→Function	115, 142	2
074	Topic→Title→Other	141, 145	2
075	Topic→Title→Title	121, 131, 138, 158, 161	5
076	Topic→Title→Title→Title	116, 120	2
077	Topic→Title→Title→Topic	159	1
078	Topic→Title→Topic	118, 146	2
079	Topic→Title→Topic→Content	130	1
080	Topic → Title → Topic → Content → Title → Function	129	1
081	Topic→Title→Topic→Other	160	1
082	Topic→Topic	153, 167, 171	3
083	Topic→Topic→Content	112, 127	2
084	Topic→Topic→Content→Content	111	1
085	Topic→Topic→Function	128	1
086	Topic→Topic→Title	109, 123, 125	3
087	Topic→Topic→Title→Content	126	1
088	Topic→Topic→Title→Title	122	1
089	Topic→Topic→Topic	110, 124	2

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

附錄四 以控制詞彙群組組成的路徑式規則階層及其關鍵

字類別與規則

序號	路徑式規則的階層及其組成的關鍵字類別	規則號碼	數量
001	Content→Title	222	1
002	Title→Content	194, 195, 220	3
003	Title→Content→Content	217, 218	2
004	Title→Content→Content→Content→Other	200	1
005	Title→Content→Content→Title	199	1
006	Title→Content→Title	196, 201	2
007	Title→Other	198	1
008	Title→Title	219	1
009	Title→Title→Content	163, 167, 169, 171, 190, 192, 193, 203, 204, 205	10
010	Title→Title→Content→Content	162, 164, 172, 206	4
011	Title→Title→Content→Content→Content	165	1
012	Title→Title→Other	168, 197	2
013	Title→Title→Title	166, 191, 202	3
014	Title→Title→Title→Content	160, 170	2
015	Title→Title→Title→Other	161	1
016	Title→Title→Topic	221	1
017	Title→Topic→Content	181, 182, 186	3
018	Title→Topic→Content→Content	173, 174	2
019	Title→Topic→Title→Content	180, 183, 187, 189	4
020	Title→Topic→Title→Content→Title	185	1
021	Title→Topic→Title→Content→Title→Content	175	1
022	Title→Topic→Title→Title	184, 188	2
023	Title→Topic→Title→Title→Content→Content	177, 178	2
024	Title→Topic→Title→Title→Title	179	1
025	Title→Topic→Title→Title→Title→Title	176	1
026	Topic→Content	152, 158, 159, 216	4
027	Topic→Content→Content	148, 155	2
028	Topic→Content→Content→Content	150, 154	2
029	Topic→Content→Content→Content→Content	151	1
030	Topic→Content→Title	149, 156, 157	3
031	Topic→Title	214	1

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

序號	路徑式規則的階層及其組成的關鍵字類別	規則號碼	數量
032	Topic→Title→Content	76, 77, 126, 128, 136, 138, 207, 213	8
033	Topic→Title→Content→Content	70, 73, 75, 81, 121, 129, 130, 137, 139, 146, 147, 153, 212, 215	14
034	Topic→Title→Content→Content→Content	74, 127	2
035	Topic→Title→Content→Content→Content→Content	82, 120, 144	3
036	Topic→Title→Content→Content→Title	79, 142	2
037	Topic→Title→Content→Other	72, 122	2
038	Topic→Title→Content→Title	71, 119, 140, 143	4
039	Topic→Title→Other	80, 131	2
040	Topic→Title→Title	125, 141, 145, 208, 211	5
041	Topic→Title→Title→Content	1, 2, 13, 60, 61, 63, 64, 66, 67, 78, 117, 123, 124, 132, 135, 210	16
042	Topic→Title→Title→Content→Content	57, 58, 115, 209	4
043	Topic→Title→Title→Content→Content→Content	116	1
044	Topic→Title→Title→Content→Content→Content→Content	16	1
045	Topic→Title→Title→Content→Content→Content→Other	134	1
046	Topic→Title→Title→Content→Content→Title	17	1
047	Topic→Title→Title→Content→Content→Topic	15	1
048	Topic→Title→Title→Content→Other	14	1
049	Topic→Title→Title→Content→Topic	12	1
050	Topic→Title→Title→Other	18, 62, 69, 133	4
051	Topic→Title→Title→Title	65, 68	2
052	Topic→Title→Title→Title→Content	6, 8, 9, 10, 118	5
053	Topic→Title→Title→Title→Content→Content	7, 11	2
054	Topic→Title→Title→Title→Other	59	1
055	Topic→Title→Title→Title→Title	3	1
056	Topic→Title→Title→Title→Title→Content	4	1
057	Topic→Title→Title→Title→Title→Other	5	1
058	Topic→Title→Topic→Content	30, 52	2
059	Topic→Title→Topic→Content→Content	37, 40	2
060	Topic→Title→Topic→Content→Content→Content	39, 53, 55	3

序號	路徑式規則的階層及其組成的關鍵字類別	規則號碼	數量
	Content		
061	Topic → Title → Topic → Content → Content → Content → Content	38	1
062	Topic → Title → Topic → Content → Content → Other	54	1
063	Topic → Title → Topic → Content → Other	41	1
064	Topic → Title → Topic → Content → Title → Content	36	1
065	Topic → Title → Topic → Title → Content	27, 31, 33, 42, 43, 44, 49, 51	8
066	Topic → Title → Topic → Title → Content → Content	20, 28, 34, 45, 48	5
067	Topic → Title → Topic → Title → Content → Content → Topic	29	1
068	Topic → Title → Topic → Title → Content → Title	19, 35	2
069	Topic → Title → Topic → Title → Other	32, 50	2
070	Topic → Title → Topic → Title → Title	24	1
071	Topic → Title → Topic → Title → Title → Content	22, 23, 26, 46, 47	5
072	Topic → Title → Topic → Title → Title → Content → Content	25	1
073	Topic → Title → Topic → Title → Title → Title	21	1
074	Topic → Title → Topic → Topic	56	1
075	Topic → Topic → Content	104, 110, 111	3
076	Topic → Topic → Content → Content	97, 98, 113	3
077	Topic → Topic → Content → Title	96, 112	2
078	Topic → Topic → Other	114	1
079	Topic → Topic → Title → Content	94, 99, 100, 102, 103, 105, 106	7
080	Topic → Topic → Title → Content → Content → Content	83, 89, 108	3
081	Topic → Topic → Title → Content → Content → Content → Content	107	1
082	Topic → Topic → Title → Content → Title → Content	95	1
083	Topic → Topic → Title → Title	101	1
084	Topic → Topic → Title → Title → Content	84, 85, 88, 93, 109	5
085	Topic → Topic → Title → Title → Content → Content	86, 87	2

以關鍵字使用分析探討社會標記者與索引專家的文獻標引心智模式

序號	路徑式規則的階層及其組成的關鍵字類別	規則號碼	數量
086	Topic → Topic → Title → Title → Content → Content→Content	90	1
087	Topic → Topic → Title → Title → Content → Content→Title→Content	91	1
088	Topic→Topic→Title→Title→Title	92	1