

h-11

中文知識分類主題規範工具整建芻議

國家圖書館編目組

國家圖書館研究組

一、計畫緣起

國家圖書館（以下簡稱本館）自來就利用《中國圖書分類法》與《中文圖書標題表》等資訊組織規範工具，將採訪進館的圖書資訊進行整理，惟近年來學術發展日新月異，步入數位時代後，網際網路使得科際間知識交流蓬勃發展，現有的工具已經不敷運用，為順應時代發展的需要，正積極進行分類法與主題詞的研修工作，希望能建立一套切合目前圖書館面對資訊媒體所需的資訊組織規範工具。

本館基於修訂分類表與主題詞的職責，察覺對於圖書館建置的書目資料庫，其在使用分類表與主題詞進行標引工作後，實際運用分類表與主題詞的種種情形，有必要進行系統性的分析與探討。同時，對於歷年圖書館館員在分類標題標引工作中所討論出來的顯性或隱性標引規則，亦應進行有效的「知識管理」，以確使圖書資訊組織業務更具科學化。

建立一個動態式、配合當前學術發展的圖書資訊組織規範工具是本館現階段急欲達成的目標；致力於有關書刊文獻分類、主題標引技術的創新，也是本館奮鬥不懈的目標；研發建置書刊文獻智慧標引系統，則是我們的遠景與策略發展目標。

二、整體目標

本館肩負了國內各級圖書館圖書資訊組織規範工具的研擬修訂、說明解釋、更新維護、推廣實施等角色地位，更是國家標準書目、出版品預行編目、全國書目中心等實際執行單位，致使中文圖書原始編目的工作需求與數量為全國之最；同時，也是華文世界中正體字圖書資訊組織的主要辦理單位。面對網際網路資源大量湧現，如何改進原始編目工作的理論與實務，成為本館亟於解決與克服的

課題，人工智慧型的電腦輔助標引，是我們的努力目標，而中文知識分類主題規範工具的整建則是我們的基礎工作。因此，在目標與基礎建設之間，我們擬具了下列六項計畫目標：

（一）建置並永續發展電子式「中文文獻分類法」與「中文主題詞表」規範工具，以供本館館員配合現行的資訊組織（編目、索引）系統，進行人工與電腦輔助書刊文獻的分類、主題的標引作業。

（二）建立並永續發展「中文文獻分類法」與「中文主題詞表」中，分類類目與主題詞之間的關聯，以滿足書刊文獻資訊組織層級體系（樹狀）結構與主題詞詞間關係體系結構之雙重需求。

（三）研發改良可供目前各種資訊檢索系統或搜尋引擎共用之「檢索詞彙知識系統」，使其具備動態式建立學科專業詞彙間之知識關連，以縮短使用者檢索資料庫時能夠在較短的時間內，獲得切合所需的資訊，且不遺漏重要的相關性資料。

（四）建置並永續發展智慧型之「中文文獻分類法」與「中文主題詞表」修訂系統。經由日常可行之修訂資訊蒐集機制之建立，在電腦輔助與人工並行的作業下，系統不定期（適切之時機）提出分類號與主題詞的修訂建議，經一定的審核機制，於最短的時間內，完成修訂作業。

（五）建置並永續發展本館書目索引資料庫之自我自動偵測與檢視分類號與主題詞系統。就現行之書目索引系統（本館館藏目錄系統、本館期刊論文索引影像系統、全國博碩士論文資訊網等），進行回溯性之分類主題分析，經由電腦輔助與人工並行作業方式，確立適切合理之分類主題標引規則，並將此標引規則回饋至母系統（現行之書目索引系統，回饋母系統之架構與程序），建立書目索引系統動態式智慧型之分類主題檢測機制。

【館務報導】

(六) 建置並永續發展書刊文獻之智慧標引分類與主題系統。以智慧標引之理論與實務為基礎，融入館員歷年累積之標引規則與經驗；將現行的書目索引編目系統，具備智慧型動態標引功能，以處理曾經出現的書刊文獻；並制訂控制式分類主題體系架構的「中文文獻分類法」與「中文主題詞表」等技術規範工具，以處理不曾出現的書刊文獻，或不以先前出現書刊文獻慣用類目詞彙為主的書刊文獻。並永續發展且動態修訂此一分類主題動態標引系統。

由於本計畫之落實受到書目資料之持續發展、分類主題之增刪修訂、資訊組織標引規則之變易、數位資訊科技發展之日新月異等因素影響，擬採行分階段方式進行，以確保能達到本計畫所揭示的終極目標。

三、計畫內容

本計畫係將本館之書目資料庫進行分析，再將之與現有之分類與主題規範工具進行整建；相關之電子檔案、書面文件、系統概況以及應進行之工作項目說明如下：

(一) 本館館藏目錄書目資料庫分析

1. 列入分析之書目資料庫

- (1) 本館館藏目錄資料庫 (<http://lib.ncl.edu.tw/urica.html>)，約 100 萬筆。
- (2) 漢學研究中心典藏書刊目錄資料庫 (<http://ccs.ncl.edu.tw/data.html>)，約 5-6 萬筆。
- (3) 全國博碩士論文資訊網 (<http://datas.ncl.edu.tw/theabs/1/>)，約 28 萬筆。

2. 主要分析項目

- (1) 書目資料庫中分類號、主題詞（標題詞）使用現象。
- (2) 以索書號為主之分析項目。包括：①特藏號、分類號、傳主號、機關號、著者號、作品號、部冊號、版次號、年次號等；②學科主題內容、體裁（著作形式）、地域、時代、語文、用途等；③出版形式，圖書、期刊、視聽資料、電子書、拓片等；④其他可供分類類分之相關判別條件。
- (3) 以標題詞等為主之分析項目。包括：同義關係、層級關係、相關關係，單一概念

詞、單一概念複合詞、形容詞詞組與連詞詞組間關係，原標題表複分詞提升為主題詞後之適用狀況、自由詞與主題詞關係、類稱詞與專有詞關係、用與不用關係、主表與附表關係，詞頻、詞族關係（參見關係）等關係。

- (4) 各筆書目記錄在資料庫中所呈現的分類號與主題詞間關係、主題詞與題名間關係、題名與分類號間關係，以及分類號、主題詞與題名相互間之關係。

3. 工作項目與產出結果

- (1) 將館藏書目資料庫中之標題詞轉換成主題詞（標題詞轉換成主題詞之相關處理原則由本館提供；其中自然科學與應用科學之標題詞轉換成主題詞工作，已完成大部分）。書目資料庫之主題詞分析工作，必須在各筆書目記錄之標題詞全部轉換成主題詞後，才能據以進行各項分析工作。
- (2) 分析全國博碩士論文資訊網，將各筆博碩士論文書目記錄中之關鍵詞轉入館藏書目資料庫中該筆書目記錄之主題詞欄項中；未入藏之學位論文者，則將之列為待徵集圖書。
- (3) 完成書目資料庫中分類號使用現象報告。內容包括：類號使用分佈情形、使用類號與分類法類表類號之關係、各使用類號之筆數量化與類表關連、類表複分使用現象、錯誤類號之數量與原因分析、以及其他足以彰顯分類標引工作與分類表間相互關連之事項。
- (4) 完成資料庫中各類鄰近類號查驗分析報告。內容包括：相同主題鄰近類號同異間之類分判別原則，以確定分類標引工作之正確性與一致性，並對書目記錄進行差異解釋或修正。
- (5) 完成資料庫中各單一使用類號之組合報告。內容包括：組成單元項目、各項目之優先順序、組合原則、辨正與注意事項、以及其他足以彰顯該分類號組成原因之相關事項。
- (6) 完成複分表（含子項）在資料庫已使用類

號中分佈現象報告。

- (7) 完成體裁、出版形式等在資料庫已使用類號中分佈現象報告。
- (8) 完成資料庫中每一著者及其作者號（個人與團體）之權威控制報告。
- (9) 完成資料庫中每一傳主及其傳主號（個人與團體）之權威控制報告。
- (10) 完成資料庫中分類號、類目與題名間關連性之分析報告。
- (11) 完成出各種分類標引規則匯集後所產生的不一致、相互矛盾、邏輯錯誤等問題，並列舉出實例，據以討論後決定採行的原則。
- (12) 完成資料庫中主題詞使用現象報告。內容包括：主題詞目使用分佈情形、使用主題詞與主題詞表收詞間之關係、各使用主題詞之筆數量化與詞表之關連、錯誤主題詞之數量與原因分析、以及其他足以彰顯主題標引工作與主題詞表間相互關連之事項。
- (13) 完成資料庫中各相關主題詞詞間關係之分析報告。內容包括：同義關係、層級關係、相關關係，單一概念詞、單一概念複合詞、形容詞詞組與連詞詞組間關係，自由詞與主題詞關係、類稱詞與專有詞關係、用與不用關係、主表與附表關係，詞頻、詞族關係（參見關係）等關係。
- (14) 完成資料庫中各鄰近類號之書目，在選用主題詞時之查驗分析報告。內容包括：相同主題所選用主題詞之判別原則，以確定主題標引工作之正確性與一致性，並對書目記錄進行差異解釋或修正。
- (15) 完成資料庫中各書目記錄之主題詞組成報告。內容包括：詞表詞目、附表與自由詞組之組成關係、組合原則、辨正與注意事項、以及其他足以彰顯該主題詞組成原因之相關事項。
- (16) 完成附表在資料庫已使用主題詞中分佈現象報告。
- (17) 完成原複分詞提升為主題詞後在資料庫已使用主題詞中分佈現象報告。
- (18) 完成體裁、出版形式等在資料庫已使用主題詞中分佈現象報告。

(19) 完成資料庫中每一著者（個人與團體）之主題詞權威控制報告。

(20) 完成資料庫中特定標目（會議名稱、劃一題名等）之主題詞權威控制報告。

(21) 完成資料庫中主題詞與題名間關連性之分析報告。

(22) 完成資料庫中分類號、主題詞與題名間關連性之分析報告。

(23) 檢視出各種主題標引規則匯集後所產生的不一致、相互矛盾、邏輯錯誤等問題，並列舉出實例，據以討論後決定採行的原則。

(24) 其他足以彰顯分類表、主題詞表與分類主題標引工作間相互關連事項之分析報告。

(二) 建置本館中文文獻分類法基本資料庫

1. 納入整建之電子資料

- (1) 《中國圖書分類法》增訂七版。
- (2) 《中國圖書分類法》增訂八版。
- (3) 《中國圖書分類法增訂七、八版異動類目對照表》。
- (4) 《中國圖書分類法》增訂九版。
- (5) 編目園地中新增分類類目。
- (6) 分析分類表各類號類目標引規則。
- (7) 何日章與賴永祥分類號對照表。
- (8) 書目資料庫分析所得資料。

2. 納入整建之書面資料

- (1) 本館分類標引規則。
- (2) 其他書刊文獻之分類標引規則。

3. 工作項目與產出結果

- (1) 以本館編目組目前所使用的分類表類號加上書目資料庫中所有使用過的分類類號，建構成「分類法基本資料庫」所包含的全部類號。
- (2) 將《中國圖書分類法》增訂七版、《中國圖書分類法》增訂八版、《中國圖書分類法增訂七、八版異動類目對照表》、《中國圖書分類法》增訂九版、編目園地中新增分類類目等之電子檔整合成一個檔案，並進行資料校對，修正錯誤、填補遺漏。
- (3) 將書面之各種分類標引規則，綜合分析後，條列歸納，分入各適切之欄位。

- (4) 完成分類法基本資料庫之主要欄位分析與建置。
- (5) 完成分類表基本資料庫之複分表（附表、插表、仿分等）等欄位分析與建置。
- (6) 完成上述各項整理資料之導入。
- (7) 完成檢視出各種分類類目資料、標引規則等，匯集後所產生的不一致、相互矛盾、邏輯錯誤等問題，並列舉出實例，據以討論後決定採行的原則。
- (8) 其他有助於分類法基本資料庫類目資料、標引規則等之匯集、分析、採行緣由等相關資料蒐集、討論紀錄、決行理由等資訊管理。

（三）建置本館中文主題詞基本資料庫

1. 納入整建之電子資料
 - (1) 《中文圖書標題表》。
 - (2) 《中文主題詞表》（完整版）。
 - (3) 編目園地中新增主題標目。
 - (4) 書目資料庫分析所得資料。
2. 納入整建之書面資料
 - (1) 本館主題標引規則。
 - (2) 本館中文主題標引規則。
 - (3) 其他書刊文獻之主題標引規則。
3. 工作項目與產出結果
 - (1) 以本館所編訂的主題表詞目加上書目資料庫中所有使用過的主題詞（以標題詞轉換成主題詞為主）的詞目，建構成「主題詞基本資料庫」所包含的全部詞目。
 - (2) 將《中文圖書標題表》、《中文主題詞表》等之電子檔轉換至新系統。並進行資料校對，修正錯誤、填補遺漏。
 - (3) 將書面之各種主題標引規則，綜合分析後，條列歸納，分入各適切之欄位。
 - (4) 完成主題詞基本資料庫之主要欄位分析與建置。
 - (5) 完成主題詞基本資料庫之附表等欄位分析與建置。
 - (6) 完成上述各項整理資料之導入。
 - (7) 完成檢視出各種主題詞詞目、標引規則等，匯集後所產生的不一致、相互矛盾、邏輯錯誤等問題，並列舉出實例，據以討

論後決定採行的原則。

- (8) 其他有助於主題詞基本資料庫詞目資料、標引規則等之匯集、分析、採行緣由等相關資料蒐集、討論紀錄、決行理由等資訊管理。

（四）建置本館分類主題標準書目資料庫

1. 標準書目資料庫之資料來源
 - (1) 本館館藏目錄資料庫（<http://lib.ncl.edu.tw/urica.html>），約 100 萬筆。
 - (2) 漢學研究中心典藏書刊目錄資料庫（<http://ccs.ncl.edu.tw/data.html>），約 5-6 萬筆。
 - (3) 全國博碩士論文資訊網（<http://datas.ncl.edu.tw/theabs/1/>），約 28 萬筆。
2. 工作項目與產出結果
 - (1) 將本計畫所轉出之館藏目錄資料庫另行建置成一獨立之資料庫，惟新資料庫仍可以與原來之資料庫產生對應關連，以便於日後進行資料之比對、修正等相關事宜。
 - (2) 新建置之資料庫在完成分析、比對等相關作業後，應將所得正確之類目、類號、主題詞、分類主題標引規則等資訊，進行資料庫更新作業，使成為標準書目資料庫。
 - (3) 資料庫仍以原始轉出之 CMARC 格式建置，並可單筆、多筆；分類、分段欄等批次下載，亦可轉成 EXCEL 檔案及 XML 檔案格式（CMARC/XML，ISO2709 格式轉換程式由國圖提供）格式。
 - (4) 標準書目資料庫將可以成為「分類法基本資料庫」與「主題詞基本資料庫」之書例連結來源。

（五）建置本館分類主題規範工具簡易查詢系統

1. 簡易查詢系統之資料庫來源
 - (1) 本館中文文獻分類法基本資料庫。
 - (2) 本館中文主題詞基本資料庫。
 - (3) 本館分類主題標準書目資料庫。
2. 系統功能需求
 - (1) 分類、主題、分類主題間基本功能。具有：瀏覽、查詢（檢索技巧）、新增、修改、輸出列印、上傳下載、統計等功能。
 - (2) 可由分類、主題再查詢到「本館分類主題標準書目資料庫」。



(六) 研擬本館中文知識分類主題智慧標引系統功能需求書及相關報告書

1. 建置系統規劃需求書。包括：規範系統、標引系統、修訂系統、分析資料導入、系統維護、系統軟硬體環境。
2. 國際現行分類主題系統之評估與發展趨勢報告。包括：國際上較著名之分類主題系統之理論實務、研究成果、系統功能、系統評估、發展趨勢等。
3. 主題詞之詞性、學科類屬、詞間網絡關係之研究報告。包括：國內外之研究現況、研究成果、相關技術規範、發展趨勢等。

四、預期成果

本整建計畫預期將完成館藏目錄資料庫分析、分類與主題基本資料庫的建置、簡易的分類法與主題詞表的檢索系統、分類主題標準書目資料庫的建置等五項主要工作，以及系統需求書及相關的研究案，茲分別說明如下：

(一) 完成本館館藏書目資料庫分析

館藏書目資料庫是歷年來本館編目組同人辛勤的智慧結晶，其中經歷了分類表的各種版本，蘊藏著各種「下類號」、「給主題詞」的標引規則，經由書目資料庫的分析，將使分類號實際使用的情形、標（主）題詞實際運用的情形，得以完整的掌握，有助於對於分類法與主題詞進行全面、整體檢討；各類的分類主題標引規則，也有了科學化的整理。

(二) 完成本館中文文獻分類法基本資料庫之建置

目前大家所使用的分類法是以學術理論為主的規範工具，完成本項計畫後，所編製成的分類法基本資料庫，將是一部確實以「文獻保障」為主的分類法，其中並融入大量的分類標引規則，將對現有分類法的意涵推向新的境界。

(三) 完成本館中文主題詞基本資料庫之建置

本館目前所採行的《中文圖書標題表》及新修訂完成的《中文主題詞表》，仍屬於以學術理論為主的規範工具，完成本項計畫後，所編製成的主題詞基本資料庫，將是一部確實以「文獻保障」為主的主題詞表，其中並融入大量的主題標引規則，將對現有主題詞表的意涵推向新的境界。

(四) 完成本館分類主題標準書目資料庫之建置

為達成電腦輔助標引的發展趨勢，對館藏目錄中分類主題等資料重行檢視，訂正後建置一標準書目資料庫，將可成為有效的「練習文件」，這項文件將可成為發展電腦在資訊組織智慧標引工作上寶貴的「人工智慧」，而且是屬於動態成長式的人工智慧。

(五) 完成本館分類主題規範工具簡易查詢系統建置

在分類法與主題詞表基本資料庫完成後，於探討分類與主題間各種相互關連關係尚未完成之際，僅需要一簡易的分類法、主題詞及分類主題間對應的查詢系統，待日後發展成真正符合圖書館智慧標引所需的多功能系統。

(六) 完成本館中文知識分類主題智慧標引系統功能需求書及相關報告書

經由本項專案的研究，將會提出建置智慧標引系統的需求書。另外，經由對國際上現行分類主題系統發展現況與趨勢的研究，也有助於系統需求規範的訂定，而主題詞詞間關係的釐清，也有助於日後技術規範理論的探討、實務的應用。

五、展望未來

面對數位時代，圖書館館藏資源在類型與數量上有著極大的變化，傳統以分類、主題等方式所進行的資訊組織，無論在速度上、精準度上、完整性上都無法滿足館員與讀者的需求，圖書館員一直在尋求一個可以超越目前困境的替代方式。我們也知道適當的替代方式不會憑空出現，而是在既有的方法上尋求突破。

本項計畫只是我們對於現有分類法與主題詞的重新思索，希望藉由實際於圖書館館藏資訊組織上運用情形的分析，瞭解分類法與主題詞在現有結構、功能與資訊分析上是否能改善空間，能否更邏輯化、科學化與人工智慧化。智慧標引、數值型標引、規則型標引等都是我們尋求努力方向。

本計畫由編目組李莉菡主任、研究組江綉瑛主任率同編目組黃莉玲編輯、阮靜玲助理編輯、陳友民組員，研究組嚴鼎忠編輯、賴珍蘭小姐，閱覽組陳立原管理師，漢學研究中心資料組錢月蓮助理編輯等人共同討論擬訂，其中並獲得閱覽組俞小明主任、林淑芬編輯等人頗多協助。