

中文圖書標題表的使用情況分析：以NBINet為例

吳政勳 (Cheng-Juei Wu)

輔仁大學圖書資訊系專任教授

E-mail: lins1022@mails.fju.edu.tw

中文摘要

由於「中文圖書標題表」可以說是臺灣地區圖書館編目人員在給與圖書標題時的主要依據，而標題是揭示書籍內容最重要的工具之一，因此對於「中文圖書標題表」的分析和探討有其必要性。作者藉由隨機抽樣方式，自「中文圖書標題表」抽取80個標題後，利用國家圖書館NBINet來進行主題檢索分析。首先，發現約有26%的標題，從未被使用過。其次，就主題複分而言，約有23%的標題並未使用到任何複分的功能。最後，就標題的整體使用情況來看，低使用頻率的標題（使用次數不超過30者）有60個，佔全體80個標題的75%（即3/4），其合計的被使用機率小於5%。相反的，極少數1/8會造成使用者資訊超載的超高使用頻率標題（資料筆數超過100以上者），卻佔全體使用機率的88%。由此可知，絕大部份的資料，集中在極少數的標題上，此部分的標題，每個平均約有775筆記錄，因此潛在的使用者資訊超載問題是非常嚴重的。

Practice analysis of Chinese Subject Headings: Using NBINet

Subject analysis is the main tool for cataloguers to expose the contents of materials such as books, and the Chinese Subject Headings is the standard for this practice in Taiwan. In this study, we randomly select 80 subjects from the Chinese Subject Headings as the searching terms on the NBINet, which is the largest bibliographic database in Taiwan, for practice analysis. Firstly, the experimental results show that 26% of randomly selected subjects are never been used on the NBINet. Secondly, 23% of randomly selected subjects do not have any forms of subdivisions. Lastly, three fourths of randomly selected subjects can be classified as low-usage subjects, which have less than 31 records. However, the appearance of all low-usage subjects is less than 5% on the NBINet in total. On the contrary, the appearance of high-usage subjects, which have more than 100 records, is about 88% in total, and there are only 13% of subjects belonging to this category. In addition, on average, each of high-usage subjects has 775 records. This indicates the potential problem of information overload is severe.

關鍵字：中文圖書標題表，主題分析，標題，資訊超載，Chinese Subject Headings，Subject Analysis，Subject Heading，Information

一、前言

主題 (subject) 或標題 (subject heading) 在黃淵泉《中文圖書分類編目學》一書定義為：「一個詞或一組字用以表示資料討論的主題。」 [註 1] 除了基本的書目資料，如書名、作者、出版社、與出版日期等，分類號和主題可以說是圖書著錄人員用來揭露資料內容的兩個主要工具。其中分類號因為肩負排架功能，祇能有一個，因此如薛理桂在「分類與編目之發展趨勢」一文中提及 [註 2]，傳統的分類法，是採用「單一分類 -- 單一位置」 (single classification--single placement) 概念。這使得分類號不若主題來的靈活，所以主題可以說是目前著錄人員用來揭露資料內容最重要的工具。

根據研究顯示，主題檢索是圖書館使用者在線上檢索時最常用的方法 [註 3-4]。然而以往圖書館界對於主題或標題的探討，大都集中在控制詞彙的應用 (即實踐主題標目的統一原則 [註 5]) 與詞彙的選用、主題標目的結構 [註 6]、主題標目的複分方式 [註 7]、主題分析的作法 [註 8-9]、標題表缺失 [註 10-11] 等，近年來雖然也有許多探討主題與檢索效益的關係，例如「中文標題檢索效益之研究--以國立臺灣大學TULIPS系統為例」 [註 12]，不過仍然是在傳統的主題標目結構下來探討。

目前臺灣地區圖書館編目人員在給與圖書標題時的主要依據為「中文圖書標題表」，此標題表是由1984年當時中央圖書館 (現今國家圖書館的前身) 使用的「中文圖書標題總目初稿」增修而來，於1993年出版。 [註 13] 基本上是以「中國圖書分類法」為基礎，參考「美國國會圖書館標題表」 (LCSH) 的詞表方式製作完成 [註 14]，是目前臺灣地區最完整的中文標題表。

至於「中文圖書標題表」的優缺點，在「評《中文圖書標題表》一兼談標題表的敘詞化改進」一文中有較詳盡的評述。 [註 15] 在優點方面，文中提到「具有結構完整、選詞精當、宜於擴充、實用性強等優點」 [註 16]；在缺點方面，作者從等同率、關聯比、和參照度等3個定量評判公式，指出入口詞過少、標題間聯性差、和交互參照不足等缺點。 [註 17]

由於本文的重點在了解標題在編目上的使用情況，乃選擇全國最大的「全國圖書資訊網路系統」 (National Bibliographic Information Network, 簡稱NBINet) 為實驗平台 [註 18]。NBINet具有國家書目資料庫的特質，起源於1987年國家圖書館推行的離線合作編目作業計畫，並於1991年結合國家圖書館和16個主要的國立大專院校圖書館，正式成立NBINet，使用Chinese CATSS系統提供線上服務，至1998年4月改用INNOPAC系統。 [註 19]

根據國家圖書館「NBINet簡介」網頁上公布的統計資料顯示 [註 20]，截止至2002年 (民國91年) 3月底止，書目記錄筆數有3,729,942 筆。以記錄類型來看，圖書資料佔絕大多數，有3,557,611 筆；記錄語文來分，絕大多數為中文資料，有2,938,320 筆。另一方面，目前參與的合作館已擴及各類型圖書館，共70所。 [註 21] 由上述的統計數據來看，NBINet可以說是國內最主要和具代表性的中文圖書書目資料庫，因此也是測試「中文圖書標題表」的最佳實驗系統。

至於NBINet書目資料庫的缺失，在「NBINet合作編目資料庫內容發展之探討」一文中已有詳盡的描述和探討。 [註 22] 而可能跟本文相關的有三個：重複記錄、未建立標題、和自訂標題，不過就此次實驗觀點來看，未建立標題和自訂標題二者，對實驗結果幾無影響，一方面是數量可能不多，另一方面是其影

響散佈在全部標題表中，而被大大稀釋掉。可能影響比較大的書目記錄重複問題，固然會影響統計數字的結果，不過對一個如此龐大的書目資料庫而言（將近373萬筆記錄），重複記錄畢竟是極少數，同時也是稀釋散佈在全部資料庫中，因此對最後分析結果的影響，應該是可以忽略的。

為了解主題（或標題）在編目上的使用情況，和主題在書目資料庫中的資料分佈狀況，作者從「中文圖書標題表」中隨機抽取80個標題 [註23]，以全國最大的「全國圖書書目資訊網」（NBINet）為實驗對象，在西元2002年4月5-7日三天進行調查，以下是調查所得的各項統計數據、表格、與結果分析。

二、標題複分的整體使用情況

首先，為了明瞭標題複分的應用情況，作者製作了抽樣標題複分次數的分配表（如下面的表1）。

表1. 抽樣標題之複分次數分配表。[註24]

標題存在形式數目	標題個數
0	21
1	18
2	7
3	6
4	2
5	5
6	3
7	2
8	2
9	2
10	1
15	2
16	1
17	1
19	1
24	1
25	1
35	1
44	1
47	1
155	1

從上面表1觀察到的第一個顯著現象，是約有26%（即1/4強）的標題，從未被參與NBINet 70個合作館的編目人員使用過。由於參與NBINet的合作館均為國內主要的圖書館，再加上書目資料庫已有近373萬筆書目記錄，因此已具備有一定的涵蓋和代表性。綜括來說，如此高比例的標題從未被使用過，隱含著某些問題值得進一步探討和深思，作者以為標題未被使用的主要原因大致如下：

- (1) 標題已過時：例如由於學術思潮的演變，有些詞彙已被遺忘或淘

汰，但是標題表尚未透過定期的檢視來去除和更新。

(2) 標題太過冷僻艱深：編目人員完全不知其意，自然避免使用。或者根本未覺察到有這些詞彙的存在。

從表1看到的第二個顯著現象，是約有23%（即1/4弱）的標題，祇有1種標題存在形式，換言之，這些標題並未有使用到任何複分的功能。然而這並不必然是因為資料筆數較少所致，例如標題「膽固醇」祇有此標題本身存在，並未使用到任何形式的複分，但其資料筆數為63，也足以讓使用者頭痛，感受到資訊超載的壓力。幸好從此次抽樣調查的資料來看，就標題存在形式為1的類別來說，資料筆數超過10的標題祇有4個，約佔18個標題的22%，屬於少數。

從表1發現的第三個現象，是每個標題的複分數目平均為9.36（以61個標題計算，扣除21個存在形式數目為0的標題）。然而，大部份標題的複分數目在1-5（含），有40個標題，約佔61個標題（扣除21個存在形式數目為0的標題）的66%。

三、標題的整體使用情況

為了解主題（或標題）在編目上的使用情況，以及主題在書目資料庫中的資料分佈狀況，作者製作了抽樣標題之記錄數量的次數分配表（如下面的表2）。

表2.抽樣標題之記錄數量的次數分配表。[註25]

記錄數量	標題個數	記錄數量	標題個數
0	21	63	
		2	
1	8	73	
		1	
2	2	82	
		1	
3	2	92	
		1	
4	2	93	
		1	
5	1	156	
		1	
6	3	191	
		1	
7	1	207	
		1	
9	2	224	
		1	

10	1	284
		1
11-15	6	445
		1
16-20	4	476
		1
21-30	7	689
		1
31-40	2	1984
		1
	1	
	3096	
52	1	
58		
1		

從表2觀察到的標題整體使用情況，以標題為單位，可整理如下：

- (1) 從未被使用過的標題：一如表1所示，有21個，約26%（即1/4強，以全體80個標題計算）。
- (2) 使用次數在1-10（含）的標題：有22個，若以全體80個標題來計算，約佔28%。若以61個標題計算（扣除21個存在形式數目為0的標題）來計算，約佔36%。
- (3) 使用次數在11-20（含）的標題：有10個，若以全體80個標題來計算，約佔13%。若以61個標題計算（扣除21個存在形式數目為0的標題）來計算，約佔16%。
- (4) 使用次數在21-30（含）的標題：有7個，若以全體80個標題來計算，約佔9%。若以61個標題計算（扣除21個存在形式數目為0的標題）來計算，約佔11%。
- (5) 使用次數在31-100（含）的標題：有10個，若以全體80個標題來計算，約佔13%。若以61個標題計算（扣除21個存在形式數目為0的標題）來計算，約佔16%。
- (5) 使用次數超過100以上的標題：有10個，若以全體80個標題來計算，約佔13%。若以61個標題計算（扣除21個存在形式數目為0的標題）來計算，約佔16%。

綜合來說，全體80個標題的使用次數（即資料筆數）是8804，平均的使用次數（資料筆數）約是110。然而，低使用頻率的標題（使用次數不超過30者，即上述第1-4項合計），有60個，佔全體80個標題的75%（即3/4）。另一方面，這3/4低使用頻率的標題（含其複分），其總資料筆數為407，約佔全體資料筆數8804的4.6%。換言之，75%的標題，其合計的被使用機率小於5%。

相反的，會造成使用者資訊超載之超高使用頻率的標題（資料筆數超過100以上者，即上述第6項），有10個，約佔全體80個標題的13%（即1/8）。但其總資料筆數為7752，約佔全體資料筆數8804的88%。換言之，極少數1/8的標題，卻佔全體使用機率的88%。由此簡單的 analysis 可知，絕大部份的資料，集中在極少數的標題（主題）上，此部分的標題，每個平均約有775筆記錄，因此潛在的使用者資訊超載問題是非常嚴重的。

四、結語

主題分析可以說是目前圖書著錄人員用來揭露資料內容的主要工具，為了探討目前主題分析作業的效益和主題檢索的便利性，作者在「模糊邏輯在主題分析的應用：標題權值的計算方式」一文中曾利用國家圖書館的NBINet來進行簡易的實驗，結果發現資訊超載和資料分佈不平均的情況嚴重，同時目前圖書館主題分析作業中所採行的主題複分，並不足以解決資訊超載的問題。[註26]後來作者在「光碟資料庫中主題分佈頻率的初步調查」一文中 [註27]，以圖書資訊界經常使用的library Literature和Lisa兩種光碟資料庫為對象，進行同樣的調查，結果顯示光碟資料庫上主題分佈不平均現象亦存在，並足以造成使用者的資訊負荷超載。

本文是上述研究的延伸，本文的重點在了解主題（或標題）在編目上的使用情況，和主題在書目資料庫中的資料分佈狀況。由於「中文圖書標題表」可以說是臺灣地區圖書館編目人員在給與圖書標題時的主要依據，而標題是揭示書籍內容最重要的工具之一，因此對於「中文圖書標題表」的分析和探討有其必要性。

此次作者使用嚴謹的方式，藉由隨機抽樣方法，自「中文圖書標題表」抽取80個標題後，以全國最大的「全國圖書書目資訊網」（NBINet）為實驗對象，在西元2002年4月5-7日三天進行調查。

首先，發現約有26%（即1/4強）的標題，從未被參與NBINet 70個合作館的編目人員使用過。由於參與NBINet的合作館均為國內主要的圖書館，再加上書目資料庫已有近373萬筆書目記錄。因此1/4比例的標題從未被使用過似乎過高，意謂著某些標題可能已過時，或是太過冷僻艱深，有重新審核標題的需要。

其次，就標題複分的整體使用情況而言（參考表1），約有23%（即1/4弱）的標題並未有使用到任何複分的功能。然而這並不必然是因為資料筆數較少所致，例如標題「膽固醇」祇有此標題本身存在，並未使用到任何形式的複分，但其資料筆數為63，也足以讓使用者頭痛，感受到資訊超載的壓力。幸好從此次抽樣調查的資料來看，就此類的標題來說，資料筆數超過10的標題祇有4個，約佔18個標題的22%，屬於少數。

最後，就標題的整體使用情況來看，低使用頻率的標題（使用次數不超過30者）有60個，佔全體80個標題的75%（即3/4），其合計的被使用機率祇約為4.6%。相反的，會造成使用者資訊超載之超高使用頻率的標題（資料筆數超過100以上者）有10個，約佔全體80個標題的13%（即1/8）。但其總資料筆數為7752，約佔全體資料筆數8804的88%。換言之，極少數1/8的標題，卻佔全體使用機率的88%。由此可知，絕大部份的資料，集中在極少數的標題上，此部分的標題，每個平均約有775筆記錄，因此潛在的使用者資訊超載問題是非常嚴重的。

註釋

- 註 1：黃淵泉，中文圖書分類編目學（台北市：學生書局，民 85 年 4 月），頁 11。
- 註 2：薛理桂，「分類與編目之發展趨勢」，國立成功大學圖書館館刊第 1 期（民 87 年 4 月），頁 36-48。
- 註 3：盧秀菊，「中文主題標目與標題表」，中國圖書館學會會報 59 期（民 86 年 12 月），頁 26。
- 註 4：曾繁絜，「中文標題檢索效益之研究--以國立臺灣大學 TULIPS 系統為例」，大學圖書館第 2 卷第 1 期（民 87 年 1 月），頁 101。
- 註 5：陳麥麟屏和林國強，美國國會圖書館主題標目（台北市：三民書局，民 78 年 12 月），頁 27-29。
- 註 6：同註 5，頁 41-48。
- 註 7：同註 5，頁 49-73。
- 註 8：陳昭珍，「主題檢索理論之探討—主題分析（上）」，書農 9 期（民 81 年 12 月），頁 11-27。
- 註 9：陳佳君，「從知識結構探討主題分析」，書府 16 期（民 84 年 6 月），頁 38-47。
- 註 10：同註 3，頁 31-38。
- 註 11：侯漢清，「評《中文圖書標題表》—兼談標題表的敘詞化改進」，圖書與資訊學刊 31 期（民 88 年 11 月），頁 17-23。
- 註 12：同註 4，頁 100-123。
- 註 13：同註 3，頁 34。
- 註 14：同註 11，頁 18。
- 註 15：同註 11。
- 註 16：同註 11，頁 17。
- 註 17：同註 11，頁 18-19。
- 註 18：文獻上發現 NBINet 的中文名稱有二個—「全國圖書資訊網路系統」和「全國圖書書目資訊網」，在國家圖書館有關 NBINet 簡介的網頁中（http://NBINet.ncl.edu.tw/screens/libinfo_chia.html），這二種名稱都有使用，不過在正式的「全國圖書資訊網路系統合作編目要點」中，是使用「全國圖書資訊網路系統」一詞，因此本文也使用此詞。
- 註 19：林淑芬，「NBINet 合作編目資料庫內容發展之探討」，國家圖書館館刊 88 年 2 期（民 88 年 12 月），頁 4-5。
- 註 20：國家圖書館「NBINet 簡介」的網頁網址為 <http://NBINet.ncl.edu.tw/screens/>

libinfo_chia.html。

註21：參與的70個合作圖書館名單，請見國家圖書館的網頁，網址為<http://NBINet.ncl.edu.tw/screens/coopwww.html>，或是許靜芬，「書目網路合作模式探討—NBINet 90年度合作編目業務諮詢會議的省思與展望」，國家圖書館館訊90年3期（民90年8月），頁10。

註22：同註19，頁12-16。

註23：實驗所使用的「中文圖書標題表」，乃是直接取自國家圖書館編目組網站中編目規範標準之「中文圖書標題表」，網址是<http://192.83.186.17/catweb/2-1-4.htm>。

註24：請注意表1中「標題存在形式數目」的用法，標題存在形式數目0表示標題在書目資料庫中不存在，標題存在形式數目1表示書目資料庫中祇有標題本身，未有任何其他的複分形式存在。

註25：NBINet的書目紀錄有重複情形，換言之，並未將各合作圖書館相同的書目紀錄合併，因此這裡的紀錄數量有偏高趨勢。

註26：在此項簡易的實驗中，作者祇是隨意使用四個通用的詞彙--天主教、宗教、心理學、電腦，因此其立論的基礎並不穩固，尚需使用更嚴謹的方式來進一步查核。吳政叡，「模糊邏輯在主題分析的應用：標題權值的計算方式」，圖書與資訊學刊40期（民91年2月），頁10-17。

註27：吳政叡，「光碟資料庫中主題分佈頻率的初步調查」，審查中。