

機讀編目格式之認識與應用

黃鴻珠

MARC: Basic Concepts of Its Structure and Application

Hong-chu Huang

Associate Director

University Library

Tamkang University

Abstract

This article introduces the general characteristics of MARC format and explicits the desired preparation when it is used in automation. Finally emphasizes on the problems of application of Chinese MARC to Chinese library automation.

前 言

「機讀編目格式」(Machine Readable Cataloging Format)的產生緣自美國國會圖書館為應用電腦處理圖書資料之需，係一九六五年集合圖書館界及電腦界的專家共同研究，於次年推出的圖書館編目資料記錄(record)結構式，當初稱之為機讀編目格式第一代(MARC I)。一九六八年該單位又宣佈該格式第二代MARC II。國會圖書館設計的機讀編目格式簡稱為LC MARC，由於設計得宜，極富彈性，推出之後，各國紛紛根據該一模式來規劃其本國的機讀編目格式，因而有英國的UK MARC，加拿大的Canadian MACR，日本的Japan MARC，澳大利亞的Australia MARC等等。國際圖書館協會鑑於電腦化的最終目的乃在於資訊的交換與共享，不同的機讀編目格式阻礙了此項目的達成，因而集合各國專家一起研究，終於在一九七七年推出「國際機讀編目格式」(Universal Machine Readable Cataloging 簡稱UNIMARC)作為各國機讀編目資料交換與共享的規範。

我國圖書館界爲促使圖書資料之處理早日進入自動化的境界，中國圖書館學會與國立中央圖書館首先成立「圖書館自動化作業規劃委員會」，以研擬全國自動化作業計劃。該委員會並於民國六十九年五月組織「中文機讀編目格式工作小組」(Chinese MARC Working Group)，歷經八個月的努力，而在民國七十年二月間完成「中文圖書機讀編目格式」的規劃工作(註一)，並於同月在中文圖書資料自動化國際研討會上公諸於世。

機讀編目格式爲圖書館資料電腦化的基石，一談到圖書館業務電腦化就離不開機讀編目格式。由於我國正在積極推展圖書館自動化，因之特就機讀編目格式的概念及其應用所面臨的問題加以探討。

機讀編目格式的特性

數據(data)、程式(program)、硬體、通訊網路(communication)，及人員爲電腦化作業的五大要素。機讀編目格式爲圖書館業務電腦化爲記載數據之需而設計的記錄(record)格式，一般具有下列特性：

一、記載範圍以編目資料及一般認定共同需要的資料爲主。

機讀編目格式顧名思義，其所記載項目係根據編目規則需求而設定，如記述編目，主題編目的項目，及一般認定共同需要的項目爲主。如要用於解決採購、流通等問題，尚得加上這些功能所需的「欄位」(field)，即使以之用於解決編目的主要工作，例如印製卡片目錄，仍須加上適應各館需要的欄位，例如識別特藏資料的欄位。因之，它是電腦化中數據的基礎，而非全部。

二、各國所公告的格式均爲交換格式(communication format)。

基於電腦資料原始的鍵入(input)耗時且費力，爲求減省人力並加快速度，交換及共享各館既已鍵入的資料是人人所殷切盼望的。爲達此目的，因而有記載項目的規劃，即前項特點的產生。依此特點產生的記錄格式稱之爲「交換格式」。各館應用時得視自身需要再加入所需的欄位，或改變記載的方式作爲「自用格式」(internal format)。交換及自用格式在內涵上是有所區別的，例如美國國會圖書館發行的機讀編目磁帶是以交換格式記載，有別於其所使用的自用格式，再如同樣採用這種磁帶作爲數據基石的OCLC或RLIN(Research Libraries Information Network)或WLN(Washington Library Network)系統，也各有其本身的自用格式。所以交換格式產生的目的係爲便於各館以一種

程式即可轉入他館的資料，並將自己的資料轉成交換格式，以便供其他各館使用。

三、機讀編目格式係以定長欄 (fixed field) 及變長欄 (variable field) 混合方式記載資料。

電腦處理資料時，係以一套程式指明作業程序與企求結果，而其處理的對象為數據 (data)。為便於程式識別數據，一般採用三種方式予以記載：一為定長欄，二為變長欄，三為定長欄與變長欄混合形式。「定長欄」係每一項目固定其長度並給一名稱，程式運作時只要據其名稱，即可在所設定之位址取得所要的資料。例如一本書須記載書名、著者、版次、出版項，並劃定其書名佔四十位 (character)，著者佔二十位，版次佔十位，出版項佔三十位。要令電腦尋取版次時，它會自動將第六十一位至七十位的資料取出。這種形式的優點是位址固定，運作容易；缺點是每本書的資料限於所設定之長度內。因之，不足時只好填以空格，超出時得刪除多餘部份，較適用於有固定型態特性之數據。這種屬性與圖書資料的屬性差異很大。在編目規則中各書要依其特性加以著錄，然而著錄的項目有別，例如「聖經」需要記載劃一書名 (uniform title)。這不是每本書都需要的。又如內容註在叢書中往往需洋洋數張卡片方可完整記載該套書的內容，而一般的書則無需記載此一項目，而且即使同一項資料，它們的長度也迥然有異。因之，全部採用定長欄的形式並不適於記載圖書資料。「變長欄」即欄位之數據可依實際內涵加以記載，其間再以特定符號作為識別。記載圖書資料最理想的是有固定屬性的項目採固定欄長，無固定屬性者則採變長欄的形式。機讀編目格式即為適應此項需要而設計。一般定長欄設立的目的如下 (註二)：

- (一)說明整個記錄 (record) 的特性，例如是否為官書。
- (二)使整個記錄中變長欄所記載的消息更明確，例如有無插圖。
- (三)加入某些編目規則中所沒有的資料，例如編目用語。
- (四)顯示變長欄中某些重要項目的有無，以便日後選用，例如內容形式代碼的設計。
- (五)變長欄中的資料再以代碼資料顯示於定長欄中，以加速檢索，例如出版年代。

採用定長欄及變長欄混合形式來記載數據，為機讀編目格式之最大特性。這項特性乃依適應實際需要而定，以「中文圖書機讀編目格式」第二版 (民國七十年六月) 為例，共設定有90個欄位，其中3欄採定長欄方式，其餘87欄則採變長欄 (佔97%)，而在變長欄中又有74欄 (佔82%) 可重複記載 (見表一)。因實

際應用時，每書所需欄位均在一半以下，如全採用定長欄位方式，則空白欄位一定很多，形成浪費。

四、機讀編目格式主要規劃工作為：(一)記錄結構形式，(二)內容識別符號，(三)資料內容代碼 (coded data)。

機讀編目格式規劃的主要工作通常包括上述三項，每個記錄在磁帶上的結構形式一般依據 ISO—2709 書目資料磁帶交換格式 (documentation-format for bibliographic information interchange on magnetic tape)，分由四部份組合而成，分別名為標示 (label)、指引 (directory)、資料欄 (data fields)，及記錄終止符號 (record separator)。「標示」記載整個記錄的長度、屬性及其結構，諸如指標長度，分欄識別符號的長度及指引格局等。有人喻之如一書的書名頁。「指引」與「資料欄」相對應。在「資料欄」中每有一個欄位，則該欄的欄號 (佔 3 位)、長度 (佔 4 位) 及首字位址 (佔 5 位，第一個欄位的首字位址由 0 (零) 計起)，即記於指引中，每個欄位以 12 位址形成一個單元。因之，有人將「指引」喻如一書的目次。資料欄的形式依 ISO-2709 之規定可採四種不同方式，其區別在於指標及分欄識別界號長度的設置詳情如下：

指標位數 = 0	> 0	= 0	> 0
分欄識別符號 = 0	= 0	> 0	> 0

中文機讀編目格式、國際機讀編目格式除系統識別號的欄位外，均採用第四種方式，而且指標位數及分欄識別符號的位數均為「2」，即各佔二個字址。資料欄 (datafields) 中各欄的結構式均以指標作為前導，而後以分欄識別符號及數據記載之。分欄識別符號及數據的多寡依實際需要而訂，但各欄在數據截止時必須附以欄位終止符號，同時一個記錄結束前以記錄終止符號為界。分欄識別符號、欄位終止符號，及記錄終止符號，在各國的機讀編目格式中均有規定。

「內容識別符號」 (content designator)，係規劃所應記載的欄位之後，每一欄位給予一個名稱及代號，每一欄位的代號英文稱為 tag，中文譯為欄號。欄號通常以三位阿拉伯數字或字母表示之。例如中文機讀編目格式以欄號 200 代表書名項，210 代表出版項，215 代表稽核項；而美國國會圖書館的機讀編目格式則分別以 245，260 及 300 來表示之。「分欄」 (subfield) 由分欄識別界號及分欄名稱組合而成。分欄識別界號一般採用元符 (\$) 代表之，各分欄名稱代號則以 a—z 及 0—9 來訂定。例如中文機讀編目格式在書名項下的正題名以 \$a 表示，平行書名以 \$d 表示。內容識別符號的設定為機讀編目格式設計的主要部份。

內容代碼的規劃係以代碼形式訂定各項目的內容。例如在中文機讀編目格式中，欄號 100，分欄 a 之下，有關官書的記載分別以 a 代表中央級機構的出版品，b 代表省級及院轄市機構的出版品……等等，其目的在便於各記錄 (record) 一再重複出現的數據以最經濟的方式來記載之。

指標 (indicator) 的內容通常均有明白的規劃。

除上述各重要特性外，一般機讀編目格式的設計均遵循 LC MARC 最初所擬定的四大原則：

一、適用於各機型電腦：機讀編目格式的設計，通常希望在各種機型電腦上均能適用，各國機讀編目格式的設計，均奉此為圭臬。

二、適用於各種類型資料：機讀編目格式的設計，一般以圖書為起點，但最終目的均希望發展成非書資料亦可適用的型式。

三、便於數據彙組：機讀編目格式中分欄的設計，主要是為適應各種需求，以便提取各自所需的數據。

四、整個機讀編目格式的代碼資料及欄位名稱如屬性相同，儘量給予一致的代碼及命名。

以上為機讀編目格式的共同特點，各國機讀編目格式均有其自身的特色，例如中文機讀編目格式除上述各特點外，允許一個記錄能以中文及外文同時記載，所採用的方法是設以特別的欄位、分欄，或以分欄記載的位置，或指標等方式，以資識別中、外文數據。它與國際機讀編目格式共同的特性有儘量採用國際標準資料，例如各國代碼採 ISO-3166。

機讀編目磁帶的產生

機讀編目格式是以電腦媒體來記載圖書館資料的架構，必須附以所屬資料的數據，且存入電腦媒體，方能成為電腦可閱讀的資料。為流傳之需，目前一般採用的電腦媒體為磁帶。它的產生，圖書館員扮演極其重要的角色，在執行此事時，圖書館員應注意下列各項：

一、識別數據記載的途徑，而後擬訂輸入格式：一個機讀編目格式的記錄，源自三個不同的途徑；一是人工給予而後鍵入，佔一個記錄的絕大部份；二為常數，可由電腦程式控制而自動轉入，例如標示 (label) 中指標長度，分欄識別符號長度，指引格局 (directory map) 中欄長位數，首字位址等均是；三是

由電腦運算或運作之後自行填入，例如：標示第一項記錄長度由每一記錄的第一位到每一記錄的終止符號，如由人工運算之後填入，不但耗時，而且易生錯誤，交由電腦處理，才真正利用到電腦的功能，又如每一記錄系統識別號 (record identifier) 一般也是交由電腦依設定方式自動填入，指引部份則是利用電腦運算每一欄位的位數，而後組合而得。因之，在準備數據時應首先判斷何者須由圖書館員準備，何者交電腦自行處理。此項工作的執行通常根據實際情況，配合電腦輸入方式，設計一格式，稱之為資料輸入表 (input worksheet)，以便圖書館人員填寫，作為鍵入資料的憑據。

二、根據編目規則、分類系統表、主題目錄及機讀編目格式準備輸入資料：輸入資料的準備，除沿用傳統的編目方法，須依據編目規則、分類系統表、主題目錄外，還得遵循機讀編目格式的需求，註明所屬欄號、指標及分欄，並得準備機讀編目格式中所要求的某些傳統編目工作未來要求著錄欄位的資料，例如前述記錄標示或編目用語或連接款目段中的各欄等。待這些資料準備妥當方能依此鍵入電腦，存放於電腦媒體中，成為機讀編目格式的目錄。

三、標點符號的處理：以電腦處理圖書資料時，為達成製作卡片目錄的目的，標點符號的使用法甚為特殊。中文機讀編目格式規定，除國際圖書館協會聯盟 (IFLA) 公告之國際標準書目著錄 (ISBD-International Standard Bibliographic Description) 所需之標點符號交由程式控制外，其餘依數據實際顯現的形式著錄。其目的在使數據具有獨立性，以適應各種不同功能之需。一般機讀編目格式，均會載明標點符號的處理方式歸程式控制或由數據準備者提供，鍵入資料時應密切注意。

四、留意數據品質：電腦處理資料的特性之一即為 G I G O (Garbage In Garbage Out)。意即電腦的忠實性很高，給什麼樣的資料即照實輸出。因之，為求得正確結果，輸入時對數據的品質管制特別重要。其控制法，一般依輸入的數據以報表印出，以便校對，此外尚可由人工或利用電腦複核。其他注意事項，如必備欄位有否遺漏，各欄可否重複使用的規定是否符合及各欄資料為文字或數字形式的屬性是否正確等等。

機讀編目資料應用前的準備工作

誠如前述，機讀編目交換格式規劃的範圍僅為編目資料及一般認定共同需要

的資料，因之，據此產生的資料僅能作為電腦化系統作業所需數據的基礎而非其全部。在應用時，圖書館從業人員還肩負下列各任務：一為述明系統的功能；二為指明各功能運作時應具有的輸入、輸出的欄位；三為說明數據與程式的關係；四為提示程式運作的程序；五為分辨作業的方式；六為決定檢索項的形式與種類；七為說明數據量的多寡、使用頻率及使用方式。茲分述如下：

一、功能敘述：

電腦在圖書館的應用已相當普遍，而且深入各種業務，舉凡採購、編目、出納、參考、管理，均能達成。唯功能的達成與數據有密不可分的關係。以機讀編目格式現有各項欄位為例，其基本功能為供作圖書館的目錄、印製各種目錄及統計資料類型等資料。以中文機讀編目格式為例，取其欄號 100，一般性資料下的分欄「a」第五項適用對象的資料作為依據，可依此選製適用不同對象的目錄，像兒童、青少年，成年或其他目錄；又如憑第六項政府出版品代碼，可印製官書目錄，如以此二項混合選用則可印製政府出版品的兒童目錄。以此類推，各種不同的目錄均可產生。

其次以機讀編目交換格式所規劃的數據為本，加入採訪作業、出納作業所需的欄位，便可解決這二方面的問題。另外，機讀目錄也可作為線上書目檢索的根據。例如美國國會圖書館的目錄可透過系統發展公司 (System Development Corp. 簡稱 SDC) 的 ORBIT 系統，洛克希德公司的 DIALOG 系統，供美國境內及境外各大圖書館線上檢索。除上述外，可用它來製作各種統計資料。例如憑欄號 100 一般性資料下，分欄 a 的第三及第四項和欄號 681 中國圖書分類法以統計各年度內各類圖書的出版量，甚而再依欄號 010 國際標準圖書號碼下的分欄 d 價格，運算出每類書的平均價格。資料一旦形成電腦可讀形式，其運用是非常靈活的，至於如何運用，唯有圖書館員較為清楚。所以在圖書館電腦化系統作業中，功能的敘述為圖書館從業人員無可旁貸的任務。

二、指明達成各種功能所需輸入、輸出的欄位及其彼此間的關係：

電腦作業特性之一是輸出係依輸入及輸入運作時所產生的結果而來。在達成預定功能時，何時需要輸入何種資料，產生何種輸出，也是圖書館員應負責說明的項目之一。例如以機讀編目資料印製館藏目錄須加入那些欄位輸出時書名卡、著者卡、標題卡、分析卡、參照卡，各是什麼樣式？應印多少張？都得詳述。圖書資料記錄的欄位多，分欄也雜，為便於程式設計員撰寫程式，輸出、輸入所需欄位最好以機讀編目格式的識別符號表示，例如正題名應標為 200\$a，即可節省程式設計員查尋的時間，加速系統完成。

三、說明數據與程式運作關係：

程式的運作與數據有相當密切的關係。它們的關係有些並不難瞭解，有的則非本行無從下手。例如標示 (label) 的第二分欄記錄類型，其著錄形式採代碼，以 n 代表新記錄，d 代表刪除的記錄等等。那麼當記錄中此分欄數據為 n 時，電腦應如何運作？為 d 時，又該如何執行？再如叢書項 (欄號 225) 中，指標 2 設定的情況如下：

- 0：中文要作副款目
- 1：羅馬拼音或西文要作副款目
- 2：不作副款目
- 3：中、西文均作副款目

當此指標在不同的內容下，應執行的工作差異很大，究竟副款目是什麼？如何作？不瞭解圖書館業務的人是無從理解的。

四、說明達成指定功能的運作程序：

電腦化系統的特性之一是系統功能的達成係隨使用者的意願，及鍵入的欄位而後以程式來促成，其間的程序最好圖書館員能稍加提示。例如利用中文機讀編目格式編製兒童圖書目錄時，應告訴程式設計員以欄號 100 的一般性資料，分欄 a 下的第五個代碼，位址為 17—19 的資料為憑，凡該項資料屬兒童者才能匯集而後印製。如這個目錄又要依類排比，那麼得告訴程式設計員，執行上述後，依欄號 681 中國圖書分類號排列 (sort)，再如欲同類之內依年代排比，那麼欄號 100，分欄 a 下的出版日期則是另一個查核根據。這些運作程序，圖書館員如能稍加提示，可節省程式設計員不少摸索的時間及精力。當然最便捷的運作程序也可減少電腦作業的時間，以節省財力。為配合這項工作，通常利用流程圖並輔以決策表 (decision table) 加以說明。

五、分辨達成各功能所需的作業方式——線上或離線作業：

線上或離線作業的研定，取決於對輸出結果時效的要求；有迫切需要者以採線上作業為宜。例如：以中文機讀目錄取代卡片目錄，供讀者查尋館藏資料，就非線上作業不可，相信無人肯為查尋館藏一書的有無而等待數小時，或一天，但如仍維持傳統卡片目錄方式，僅利用電腦以取代人工印製卡片式目錄則無線上作業的必要。各項功能中何者宜採線上作業？何者可用離線作業？圖書館員應參與研判。

六、檢索項 (access point) 種類及形式的決定：

所謂「檢索項」即以電腦化資料索引作為檢索所需記錄的憑據，尤其在線上作

業時，此項需求更爲明顯。究竟圖書館作業系統中，讀者會循那些線索查尋資料，也唯有圖書館員最清楚。此時所應提供者不僅是檢索項的種類、形式，甚至它是由那些欄位而來，都得說明。例如以中文機讀編目格式爲例，如欲用 ISBN 檢索資料，就得令程式設計員瞭解爲達成此功能所需建立的索引，除要取自欄號 010 分欄 a 之外，分欄 z 也不可遺漏。檢索項的設計問題較爲複雜，容後另作討論。

七、說明數據數量多寡、使用頻率及使用方式：

此項需求的說明與建檔及決定電腦的輔助儲存工具有密切的關係。系統設計之初，即應提出。

八、其他事項：

這種結構式的優點雖爲各方接受，但實際應用時仍有許多問題存在：

(一)不易獲得所有人士一致認可：機讀編目格式公告後，美國人認爲它太複雜，相反的歐洲地區却以爲太簡單（註三）。基於此一原因，各國紛紛發展成其自己的機讀編目格式，例如英國、加拿大、日本、法國等的機讀編目格式，然而各國間雖然架構相同，但實際應用時彼此資料如何交換的問題，迄今仍是各方亟待努力解決的。

(二)數據來源所據之編目規則迄今仍無全球一致的標準，即使有，因編目者對規則的闡釋不一，所得結果亦有差異。

(三)語文問題：語文問題來自二方面，一爲資料本身的語文，例如書名、著者的用語如英文、中文、俄文等，此項問題的解決與硬體密不可分。另一來自編目國採用的語文，有些國家以二種以上的語文作爲通行的官方用語，例如加拿大及比利時，或爲適應國內、外的需要，因而得採用二種以上的語文來記載有關欄位，例如稽核項、附記項、標題，爲此需要，資料的記載採用的方式很多，有些採取：1.新設欄位，2.重複欄位，3.重複欄位並增設指標 (indicator)，4.重複欄位並增設分欄 (subfield) 來處理。其中以第四種方式最宜，原因爲 1.：不增設欄位及指標易與其他格式維持一致性，2.保持既有欄位，便於日後擴張。

(四)資料著錄的一致性：機讀編目格式著錄內容含源自編目規則的記述編目 (descriptive cataloging) 資料，源自分類系統表及主題表 (subject heading list) 的主題目錄及其他識別資料如時代代碼、國別代碼、資料類型代碼、適用對象代碼等。因之，光格式結構的規劃仍不足，它需上述各項資料配合方能記載完整、一致的資料。

(五)編目時所依據的資料：據原書而著錄編目資料與據卡片目錄著錄編目資料

N)、系統識別號 (System Identification No.) 等係以原內涵的數據建立索引檔以爲檢索外，如採用書名，著者或書名與著者組合者，主題等，其形式的設計始終是學者專家研究的課題。在西文系統中，採書名、著者或書名與著者作爲查尋根據，常見形式有如下數種：

(一)採全銜形式，即以所屬欄位全部數據作爲檢索根據。

(二)採節縮形式，即依所屬欄位中前數個字的前幾個字母組成新的單元作爲檢索的依據，稱之爲查尋鍵 (search key) 例如採4,1,1,1者，即以第一個字的前四個字母，及第二至四個字的第一個字母組合而成。節縮式的形式很多，有前述4,1,1,1，有4,2,2,2，亦有3,2,2,1等方式。何種爲佳，須視各種因素而決定之。

(三)採節縮或全銜形式，但介係詞及冠詞不計。

(四)採節縮或全銜形式，加上非標準資料 (non-standard information) 項目，共同檢索。此法爲Cooper先生率先倡導。其所謂非標準資料在他的實驗中歸納出二十四種 (註六)；諸如出版日期，資料形態或頁數等等。

(五)奇數字母節縮法，即採用節縮方式，但偶數字母不計，例如用3,2,2,1形式，其書名爲 *Introductory Probability and Statistics* 即形成ito, po, ad, s的查尋鍵。

(六)採語音代碼 (phonetic-code) 形式，即將字母數字化以編成查尋鍵。例如Soundex系統規劃爲A, E, I, O, U, L, H, Y等字母不計，另將發音相似的字母歸成1—6的數值。其規定方式如下 (註七)：

字母	數值
B, F, P, V,	1
C, G, J, K, Q, S, X, Z	2
D, T	3
L	4
M, N	5
R	6

(七)採要語法 (key-word)，用於書名及主題的查尋，即將某些冠詞及介繫詞除外，以書名中每一用字自成一單元，作檢索項。採此方式者通常配以布式邏輯的功能，以便組合問題，以檢索真正需要的記錄。

上述爲西方拼字系統下檢索項的設計方式，由於西文結構與中文差異甚巨，西方形式並不能全適用於中文。中文圖書資料庫的設計誠屬首創，而檢索項形式

的設計問題也屬第一遭。目前中文機讀編目格式小組所採用的為查尋鍵的方式，書名、著者均取前面三個字(註八)。筆者為試驗此種形式的效果，特就「中華民國出版圖書目錄彙編」第三冊，隨意取每頁上、下各一書共2,746種作為樣本，結果發現中文書名的平均用字數為5.6字，而且採前三字作為查尋鍵時，能完全表達出書名用字者佔9.79% (詳表二)，同時取至前三字作為查尋鍵，其重複的情況佔11.54% (詳表三)。由於中國字大都以辭的形式出現，因之，在2,746種書名中統計辭的用字數，發現以二字形成一辭者有1,664次，以三字形成一辭者有392次，以四字形成一辭者有257次，以五字者有45次，平均每一辭的用字數為2.4字。就此項結果，試以奇數字及偶數字各取前三字之查尋鍵，其重複結果均降至1.7%以下 (詳表四，表五)，較取前三字者理想很多。

表一 中文機讀編目格式欄位統計表

欄段	該欄段之數	定長欄	變長欄	欄不可重複	欄可重複	指標	分欄
(0)	5	1	4	1	4	0	13
(1)	5	2	3	5		0	12
(2)	6		6	3	3	4	40
(3)	19		19	1	18	1	37
(4)	14		14		14	14	14
(5)	9		9		9	9	70
(6)	12		12		12	3	68
(7)	18		18	6	12	18	117
(8)	2		2		2	1	10
總計	90	3	87	16	74	50	380
百分比		3%	97%	18%	82%	28%	

* 與已設定之欄位相比

** 本表資料取自「中文圖書機讀編目格式(第二版)」圖書館自動化作業規劃委員會 中文機讀編目格式工作小組 臺北：國立中央圖書館 民國七十年六月
頁八三~一〇二

中文查尋鍵形式很多，上述僅是方法之一，他如電信局電話查尋所採用的語音方式，又如以字根組合的形式來設計，均為可行之道。唯設計時應掌握下列原則：

(一)查尋鍵愈孤一 (unique) 愈好：即查尋鍵彼此間的重複率少，便於一次

(二)回復時間愈快愈好：基於經濟及心理因素的要求，回復時間快，意謂電腦作業的時間短，當然所需費用少。另從使用者心理因素來看回復時間快，則操作人員不必久等，因之，可以提高工作興趣，反之，不但影響工作速率，同時人們也沒有久等的耐性，結果會引起厭惡使用電腦的心理。

(三)便於人與電腦交互溝通：線上檢索時，首先由人鍵入數據交由電腦處理。查尋鍵設計應便於人工的鍵入，最好鍵入的數據少，且易於轉譯(encode)。因人工鍵入數據多，則所需時間也多，同時人為錯誤率也較高，而電腦是在GIGO原則下作業，如鍵入的查尋鍵為誤的話，當然不會獲得正確結果，所有作業均屬枉然。

(四)應富於彈性：要記住一本書的正確書名或著者，說來容易，但實際上並不簡單。在工作經驗中往往碰上須經不十分正確的書名或著者去查尋所要的資料。因之，憑查尋鍵檢索時，最好能有竄字檢索(truncate)及其他功能，以便於僅憑蛛絲馬跡亦能取得結果。

(五)應顧及未來資料的量：圖書資料電腦化處理的特性之一為儲存時間久，且數量增加快。一個系統中達到百萬筆以上記錄，屢見不鮮。因之，查尋鍵的設計應顧到未來的收藏量，不能以眼前者為限。

查尋鍵的設計，在中文圖書電腦化工作中確屬首遭。本文所取樣本2,746種嫌少，將此項結果公佈，目的僅在拋磚引玉。盼各位先進、同道能重視此一問題，共同依中文特性，設計出便於國人使用的查尋鍵。

圖書資料庫中採查尋鍵方式作為檢索依據，主要是用於特定書名或特定著者的查尋，即查出某一本書的書目資料或某一著者的某一或某些作品。檢索時必須有書名或著者的線索。這種檢索項對依主題來查尋資料並不適當。因之，西方成立較晚的資訊網如RLIN或WLN，其資料庫檢索法多數改用要語檢索法，其優點是不僅可滿足特定書名或特定著者的查尋，同時可作為主題線索的查尋方法之一，即查出某一主題所屬的所有資料。要語檢索項的設計，要應用到中文圖書資料庫最大的困擾是如何由電腦自動判斷要語。因西文中字與字間有空格隔開，可作為電腦取字的根據，同時西文多數由字可代表完整的語意，但中文字與字間沒有空格，而且最大的問題是中文多數是辭的組合，如何由電腦自動判斷要語是目前亟待研究的問題。有關電腦自動判斷要語的研究，最近謝清俊先生等人的研究有極傑出的成果(註九)。此課題的研究須各方專家學者共同參與，尤需語言學家協助。

結 語

中文機讀編目格式設計只是為中文圖書資料電腦化奠下根基，至於整個中文圖書資料電腦化系統的完成仍有待前述各項研究工作的配合。因之，如要促使此系統早日臻於理想境界，對各項研究工作應給予充分資助，俾便早日完成，同時應成立常設機構負負策劃，協調各方的意見及促進全面發展。

後 記

本文書名查尋鍵之統計資料，承淡江大學教育資料科學學系鄭美珠老師，及永林清、呂盛隆、沈英杰、陳泰維、鄭陽明與陳夢玲諸同學協助。謹此致謝。

附 註

註一 中文圖書機讀編目格式 圖書館自動化作業規畫委員會 中文機讀編目格式工作小組 台北 國立中央圖書館發行 p.111

註二 Library of Congress, "Fixed Fields-- General Information" in Books: Verification, preliminary edition, 1980, Washington D C: Library of Congress p.2

註三 Lucia J. Rather, "Exchange of Bibliographic Information in Machine Readable Form", *Library Trend*, 1973 p.632

註四 Henriette D. Avram, "The Library of Congress View on Its Relationship to ALA MARC Advisory Committee", *Journal of Library Automation*, 7(2),1974, pp. 124-125

註五 Ching-chun Hsieh et.al., "The Design and Application of the Chinese Character Code for Information Interchange (CCCI)", p.14 (尚未出版，為民國七十年二月於台北舉行之中文圖書資料自動化國際研討會中的論文)

C.C. Hsieh, C. T. Chang and S. S. Tseng, "Chinese Character Database: Its Design, Implementation and Application" *Symposium on Computer Processing of Chinese Library Materials and Computer-Assisted Chinese Language Instruction at ASIS-82*, Columbus, Ohio, 1982, pp.7-9

註六 Llinas James, *A Method for Evaluating Search Key Performance Dissertation*, State University of New York at Buffalo, 1970, p.70

註七 Robert M. Hayes and Joseph Becker, *Handbook of Data Processing for Libraries*, 2/e Los Angeles: Melville Publishing Co. 1974, p. 227

註八 Chinese MARC Working Group, "Chinese MARC Format and Bibliographic Database", in *Symposium on Computer Processing of Chinese Library Materials and Computer-Assisted Chinese Language Instructions at ASIS-82*, Columbus, Ohio Taipei: American Society for Information Science, Taipei Chapter, 1982, pp. 5-26

註九 Hsieh, Ching-chun et. al., "Experiments on the Search and Recognition of Words from a Chinese Text" 此文發表於亞太地區第一屆圖書館學研討會 台北 三月十三日至十九日 民國七十二年

參 考 書 目

- 1 中文圖書機讀編目格式 第一版及第二版 圖書館自動化作業規畫委員會中文機讀編目格式工作小組 台北 國立中央圖書館發行 民國七十年
- 2 Agenbroad, James E., "Personal Names and Chinese Character Code for Information Interchange." 民國七十年二月於台北舉行之「中文圖書資料自動化研討會」之論文
- 3 American National Standards Institute, *American National Standard Format for Bibliographic Information Interchange on Magnetic Tape*. New York 1973, 1979 (ANSI Z39.2)
- 4 BALLOTS Center, *BALLOTS Data Elements Dictionary*. 2nd Standford Standard University, 1978
- 5 *Data Preparation Manual-Books*. Olympia, Washington: Washington Library Network, 1980
- 6 IFLA UNIMARC: *Universal MARC Format*, 1st and 2nd ed. IFLA International Office of UBC, 1977 and 1980
- 7 International Organization for Standardization, *Documentation-Format for Bibliographic Information Interchange on Magnetic Tape*, ISO-2709, 1973, 1980
- 8 James, Llinas *A Method for Evaluating Search Key Performance Dissertation*, State University of New York at Buffalo, 1970
- 9 Lee, Te-chu et. al., "Chinese MARC : Its Present Status and Future Deve lopment." 民國七十年二月於台北舉行之「中文圖書資料自動化研討會」之論文
- 10 Library of Congress, *The MARC II Format*, 4th ed. Washington D.C.: Library of Congress, 1972