

圖書館學與資訊科學論文叢刊 第七輯 胡述兆·盧荷生主編

多種機讀格式並存之 編目與檢索架構研究

余顯強 著



漢美圖書有限公司 印行

台北·紐約·洛杉磯

G254.364
20071

港台書

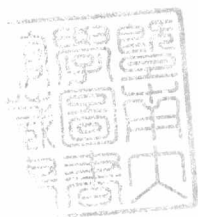
圖書館學與資訊科學論文叢刊 第七輯

胡述兆
盧荷生

主編

多種機讀格式並存之 編目與檢索架構研究

余顯強 著



漢美圖書有限公司 印行

台北・紐約・洛杉磯

多種機讀格式並存之
編目與檢索架構研究

版權所有
翻印必究

本公司法律顧問
紀鎮南律師

著作者：余顯強

出版者：漢美圖書有限公司

登記證字號：新聞局局版台業字第 3204 號

發行人：盧荷生

印行者：漢美圖書有限公司

地址：台北市重慶南路一段 99 號世界大樓 1110 室

電話：(02)383-1081；傳真：(02)389-6242

郵政劃撥帳號：07580875

本公司美國分公司：

紐約：52-25 Barnett Ave.

LIC, N.Y. 11104, U.S.A.

TEL：(718)898-9300

FAX：(718)898-9537

洛杉磯：21540 Prairie St., Unit D

Chatsworth, CA 91311, U.S.A.

TEL：(818)998-0974

FAX：(818)998-0512

1997 年 6 月初版

ISBN 957-9005-91-5 (平裝)

主編者弁言

「圖書館學與資訊科學論文叢刊」自 1990 年問世以來，已出版六集六十冊。由於這套叢刊涵蓋了台灣地區各大學圖書館與資訊學系博碩士論文的精華，資料新穎、內容充實，自出版以來，受到兩岸同道及海內外圖書館界的歡迎與好評。

第七輯徵選的十篇論文，六篇來自台大圖書館學研究所，四篇來自輔仁大學圖書資訊學研究所。在台大的六篇中，有三篇是博士論文，分別為陳昭珍博士的「古籍超文件全文資料庫模式之探討」，傅雅秀博士的「從科學傳播的觀點探討中央研究院生命科學專家的資訊尋求行為」，及王梅玲博士的「英美圖書館與資訊科學碩士教育之比較研究」。這三篇論文各具特色，陳文是為中國古籍之整理建立一套現代化的模式，傅文係從科學傳播的觀點探討生命科學家的資訊尋求行為，王文則係比較英美兩國的圖書館與資訊科學的碩士教育，皆係經過名家指導與專家考驗的力作，兼具學術與實用價值。其餘各篇探討的主題，包括：大學院校圖書館參考服務對網際網路的應用，網路資源主題檢索機制，機讀編目與檢索架構，圖書館導覽系統，企業識別系統 (CIS) 在圖書館的應用，資訊資源行銷，以及紙本書與電子書之比較，均係分別針對一個實際問題，作深入研究，故實用價值極高。

本輯之徵稿，係請王梅玲博士負責聯繫，特在此致謝。

胡述兆 盧荷生 謹識

1997 年 5 月於台北

謝辭

研究方向擬定之初，想走一段較具實務的研究方式。因此，嘗試以系統分析的方法，配合關連式資料庫系統的架構，研究異質資料並存的資料庫結構，以其能達成理論與實務結合的研究論文。

首先，非常感謝指導教授吳政叡老師，在論文撰寫的這一段期間，悉心地逐一指導與修正論文的方向與內容，使研究結果能獲得預期的成效，並使本論文能適時地完成。感謝政大圖書館館長胡歐蘭老師與陳昭珍老師，能在百忙之中，給予指正本論文的缺失，與改善的建議。

也非常感謝高錦雪老師，在研究所學習期間，與論文撰寫之初，給予相當多的協助；感謝林麗娟老師，在個別研究時的諄諄教導；感謝中圖編目組主任鄭恆雄先生與林淑芬小姐，提供書目與權威機讀目錄格式的電子檔，使研究期間，對於機讀資料格式的查核與分析有莫大的助益；感謝上海交通大學楊宗英教授，提供圖書館自動化相關的文獻資料與著作；也感謝所有在這段期間，直接或間接幫助我、教導我的所有師長。

此外，也非常感謝同學康芳菁與郭麗芳，協助修飾論文內容；好友陳俊昇提供相關的資料協助研究；還有，特別感謝好友楊惠淑同學，鼓勵我報考研究所，才有撰寫這份論文的機會。

最後，感謝我的父母，感謝他們多年來對我的照顧。無論在生活與學習上，給我非常自由的發展空間，僅藉此表達感謝，並以此論文獻給我最親愛的父母。

摘要

近年來，由於資訊的普及和網路的發達，使圖書館的地理界限，越來越模糊，也使資訊的取得越來越迅速。為求加速資料的建檔、減低原始編目的人力成本，直接轉錄網路上其他圖書館書目資料的情況將越來越普及。

由於各國所使用的機讀編目格式並不一致，因此運用 E-R 模式和 E-R 圖等關連式資料庫分析工具，分析 MARC 格式間欄號與分欄資料的實體關係，來研究多種機讀編目格式並存的資料庫檔案架構，提供兼容各種 MARC 格式的書目記錄。改善現階段不同 MARC 之間，資料轉換與交換的問題，避免圖書館自動化系統轉錄不同 MARC 格式的資料時，必須取捨無法對應的欄位，而影響資訊的透通性(transparency)。

本研究建置的檔案種類包括：儲存 MARC 資料的書目檔與權威檔；管理不同 MARC 格式的欄位，輸入與查核的編目參數檔；提供查詢與權威控制的索引檔等等。

最後，以 USMARC 和 Chinese MARC 格式的書目資料輸入測試，來確認資料欄位在資料庫各檔案的儲存，是否符合本研究的提議。並歸納研究結果，提出若干建議，提供國內圖書館界與書目自動化系統發展時的參考。

Study of Hetero-MARC Databases in the Application of Cataloguing and Retrieving

Abstract

In recent years, with the growth of the distributed system and heterogeneous DBMS as well as the repaid development of the network, people can get information more easily and quickly. In order to ease the establishment of the data and decrease the cost of manpower in original cataloguing, it will be more convenient to utilize other libraries' bibliographical data on network. However, the various Hetero-MARC formats of the bibliographical data circumvents the above mentioned scheme in the international situation.

In the study, we attempt to utilize the RDBMS, such as E-R model to analyze the relations between different fields in the MARC format. An integrated interface for cataloguing and retrieving between USMARC and UNIMARC (ex: Chinese MARC), which cover the world wide standard. Several types of tables are established, including

- (a) bibliographical and authority tables for restoring Hetero-MARC record;
- (b) cataloguing tables for inputting, checking and managing Hetero-MARC format field;
- (c) indexing tables for retrieving and authority controlling

For a pilot study, we input bibliographical data into USMARC and Chinese MARC format to verify the data's restoration in relational tables. Eventually, conclusions are made and suggestions are provided for the future development of the bibliographical automation system.

目 次

第一章 緒論	1
第一節 研究動機與目的	1
第二節 研究範圍與限制	7
第三節 研究方法與步驟	9
第四節 名詞解釋	10
註釋	16
第二章 機讀格式介紹與文獻分析	19
第一節 機讀編目格式的起源與特性	20
第二節 書目資料庫機讀編目格式並存的需求	23
第三節 機讀編目格式之間的相容性	29
第四節 國內機讀編目格式使用需求	31
註釋	34
第三章 資料庫介紹與文獻分析	37
第一節 資料庫系統	37
第二節 資料庫結構化查詢語言	43
第三節 系統分析模式	46
註釋	51
第四章 系統架構	54
第一節 編目機讀格式之架構分析	54
第二節 書目與權威記錄檔案之建立	67

第三節 編目作業參數之架構分析與檔案建立	74
第四節 查詢架構分析	84
第五節 關鍵字詞索引檔案之建立	92
第六節 權威索引檔案之建立	97
第七節 機讀編目格式轉換	105
註釋	111
第五章 資料測試	115
第一節 書目與權威資料輸入	115
第二節 查詢基本功能測試	128
註釋	135
第六章 結論與建議	136
第一節 結論	136
第二節 建議	138
參考書目	140
索引	149

圖表目次

表 1-1 : 美國機讀編目格式(USMARC)對中國機讀編目格式(Chinese MARC) 無法對應之欄號及分欄一覽表	3
表 1-2 : 中國機讀編目格式(Chinese MARC)對美國機讀編目格式(USMARC) 不必對應之欄號及分欄一覽表	4
表 1-3 : Chinese MARC 與 China MARC 欄號 100 差異比較表	5
表 2-1 : 各國國家圖書館使用 MARC 格式調查表	25
表 2-2 : 「全國圖書資訊網路」合作建檔格式一覽表	28
表 2-3 : 全國圖書資訊網路新系統需求書, 有關 Chinese MARC 與 USMARC 處理需求一覽表	32
表 2-4 : 全國圖書資訊網路新系統需求書, 有關 Chinese MARC 與 USMARC 查詢功能需求一覽表	32
表 2-5 : 上海市立圖書館自動化系統有關各 MARC 處理需求一覽表	33
表 3-1 : 表格、行、列的用語種類	42
表 4-1 : Chinese MARC、China MARC 與 USMARC 字段一覽表	55
表 4-2 : Chinese MARC、China MARC 與 USMARC 定長欄號一覽表 ...	56
表 4-3 : Chinese MARC、China MARC 與 USMARC 一般性欄號一覽表	57
表 4-4 : Chinese MARC、China MARC 與 USMARC 連結款目段欄號一覽表	58
表 4-5 : Chinese MARC、China MARC 與 USMARC 為定長分欄之欄號一覽表	58
表 4-6 : Chinese MARC 欄號 200 題名作者敘述項, 各分欄間之指定標點符號	80
表 4-7 : TULIPS 系統查詢檢索關鍵字比對範例	85

表 4-8 : 布林邏輯型態	94
表 4-9 : 不同機讀權威記錄格式所屬檢索點一覽表	99
表 4-10 : 依欄位性質轉換的設定與處理	108
表 4-11 : 依資料性質轉換的設定與處理	109
表 5-1 : 書目主檔欄位測試資料分佈情形	120
表 5-2 : 書目欄號副檔欄位測試資料分佈情形	120
表 5-3 : 書目分欄副檔欄位測試資料分佈情形	121
表 5-4 : 權威主檔欄位測試資料分佈情形	124
表 5-5 : 權威欄號副檔欄位測試資料分佈情形	124
表 5-6 : 權威分欄副檔欄位測試資料分佈情形	124
表 5-7 : 關鍵字詞主檔欄位測試資料分佈情形	125
表 5-8 : 權威索引主檔欄位測試資料分佈情形	126
表 5-9 : 反見索引檔欄位測試資料分佈情形	127
表 5-10 : 設限查詢參數檔內容範例	133
表 5-11 : 設限書目欄位參數檔內容範例	133
圖 2-1 : UNIMARC 和 USMARC 相符合的欄位段	30
圖 3-1 : 關連式資料庫系統檔案結構圖	41
圖 3-2 : 實體關連圖使用之圖形與符號	48
圖 4-1 : 定長欄號之實體關係圖	55
圖 4-2 : 欄號與一般性分欄之實體關係圖	56
圖 4-3 : 欄號與定長分欄之實體關係圖	58
圖 4-4 : 編目機讀格式欄號與分欄實體關係圖	61
圖 4-5 : 書目系統檔案架構圖	64
圖 4-6 : 書目資料實體關係圖	65
圖 4-7 : 權威資料實體關係圖	66

圖 4-8 : 書目資料庫系統整體架構圖	66
圖 4-9 : 書目記錄檔案結構關係圖	72
圖 4-10 : 權威記錄檔案結構關係圖	73
圖 4-11 : 編目類型檔案結構關係圖	75
圖 4-12 : 編目作業參數檔案結構關係圖	83
圖 4-13 : 資料庫組織方式	86
圖 4-14 : 書目資料索引組織方式	87
圖 4-15 : 書目資料索引組織實體關係圖	87
圖 4-16 : 索引項目檔案結構圖	90
圖 4-17 : 索引項目包含之索引點檔案內容範例	91
圖 4-18 : 關鍵字詞索引檔案結構關係圖	92
圖 4-19 : 查尋時資料處理過程	94
圖 4-20 : 同義詞檔	95
圖 4-21 : 權威索引檔案結構關係圖	103
圖 4-22 : 權威參見索引檔	103
圖 4-23 : 權威反見索引檔	104
圖 4-24 : 機讀格式轉換方式	106
圖 4-25 : 資料庫資料輸出機讀格式轉換過程	107
圖 4-26 : 機讀編目格式轉換對照表檔案結構	110
圖 5-1 : (a) 書目機讀編目交換格式 (b) 欄位與欄位結構	117
圖 5-2 : 「指引」格式略圖	118
圖 5-3 : 資料檔案模型測試資料內容	118
圖 5-4 : 西文測試資料以 USMARC格式顯示內容	119
圖 5-5 : 中文測試資料以 Chinese MARC 格式顯示內容	120
圖 5-6 : 中國機讀權威記錄格式範例	124
圖 5-7 : 權威「反見款目」顯示範例	131

圖 5-8 : 設限檢索流程圖	132
圖 5-9 : 設限檢索檔案結構關係圖	133

第一章 緒論

第一節 研究動機與目的

機讀編目格式(Machine Readable Cataloguing, 簡稱 MARC), 規範了書目、記述和權威格式, 及其相互之間的連結關係, 因此, 成為影響記述編目(descriptive cataloguing)優劣的重要因素之一(註 1)。由於 MARC 可促進編目自動化的便利性, 許多國家依使用上的需求, 而制定了適合各國國情的 MARC 格式, 以作為線上編目的依據。

為了使國際上各種 MARC 格式, 能彼此之間相互轉換, 國際圖書館協會聯盟(International Federation of Library Associations and Institutes, 簡稱 IFLA)於1977年出版了國際機讀編目格式(Universal MARC format, 簡稱 UNIMARC), 作為各使用單位之間, 機讀編目格式轉換的依據。但由於許多國家彼此間所使用的格式並不完全一致, 而且也有許多國家並未採用 UNIMARC 作為交換上的標準。因此, 不同機讀編目格式之間的交換, 仍存在不少的困擾。

為推行中文圖書資料的電腦建檔及自動化, 我國台灣地區也於民國70年1月出版了「中文圖書機讀編目格式」第一版。以「國際機讀編目格式」(UNIMARC-1980)為藍本, 先後參考「美國國會圖書館書目機讀編目格式」(MARC Formats for Bibliographic Data-1980)及其他各國之機讀格式, 如: 英國的 United Kingdom MARC (簡稱 UKMARC), 加拿大的 Canadian MARC, 法國的 Inter-MARC 及澳洲的 Australian MARC 等。磁帶上之書目著錄則採用「國際標準組織第2709號標準」(簡稱 ISO2709)格式。中外文書目著錄格式, 則分別依據「中國編目規則」及美國的「英美編目規則第二版」(Anglo-American Cataloguing Rules, 2nd Edition, 簡稱 AACR 2)為標

準。「中文圖書機讀編目格式」於同年七月修訂出版第二版，民國71年修訂再版時改稱為「中國機讀編目格式」第一版，至今已出版至第三版(註2，3)。

依據黃世雄教授的調查，台灣地區自動化編目採用之 MARC 格式，中文以 Chinese MARC 為主，西文則多以 LCMARC 為主(註 4)。爲了國內使用及交換上的需求，教育部委託中國圖書館學會，研訂 Chinese MARC 與 USMARC 轉換對照表，業已於民國82年出版，轉換程式雖已具雛型，但仍待改善(註 5)。

Chinese MARC 格式中，對於善本、拓片等特殊資料，有詳盡的架構與定義，但 USMARC 則無(註 6)。如表 1-1 及表 1-2 所示，Chinese MARC 與 USMARC 在轉換上，仍有部份無法對應而必須捨棄之欄位或分欄。

表 1-1：美國機讀編目格式(USMARC)對中國機讀編目格式(Chinese MARC)

無法對應之欄號及分欄一覽表

欄號	欄位 註記	指標		分欄	分欄 備註	欄、指標、欄位名稱	理由
		1	2				
007	R	00=h (Microform) 007/09 2=other z --> ?					無法轉
007	R	00=M (Motion picture) 007/14 n,u,2 --> ?					無法轉
007	R	00=S (Sound recording) 007/10 n,u --> ?					無法轉
007	R	00=S (Sound recording) 007/13 e --> ?					無法轉
008	R	06=b (No dates given; B.C. date involved) b --> ?					無法轉
008	NR	06=i (Inclusive dates of collection) i --> ?					無法轉
008	NR	19=b not applicable			b --> ?		無法轉
008	NR	18-19 (Music)			ca --> ? nn --> ? uu --> ?		無法轉
008	NR	24-29 (Music)			b --> ? s --> ?		無法轉
008	NR	18-20 (Visual materials) nnn --> ? --- --> ?					無法轉
008	NR	23-27 (Visual materials) b --> ?					無法轉
008	NR	20 (Serials) ISDS Center b--> ? 5 --> ? u --> ?					無法轉
008	NR	32 (Serials) Cumulative index availability u --> ?					無法轉
008	NR	33 (Serials) Original alphabet or script of title					無法轉
033	R	b				Postal registration no. b-->?	無法轉
035	R	b	b	z	R	Cancelled/invalid system control number =z-->?	無法轉
041	NR	b	b			Undefined bb-->?	無法轉
600	R			q	NR	Fuller form of name =q-->?	無法轉
611	R			q	NR	Name of meeting following jurisdiction name =q-->?	無法轉

資料來源：中國圖書館學會圖書館自動化規劃委員會研訂，美國機讀編目格式轉至第三版中國機讀編目格式對照表（台北市：教育部出版，民82），頁107。

表 1-2：中國機讀編目格式(Chinese MARC)對美國機讀編目格式(USMARC)

不必對應之欄號及分欄一覽表

欄號	欄位 註記	指標		分欄	分欄 備註	欄、指標、欄位名稱	理由
		1	2				
002	NR					Subrecord Map of Drectory	用不到
003	NR					Subrecord Relationship	用不到
004	NR					Related Record Drirectory	用不到
006	NR	位址 00-17				Fixed-length data elements additions materials	用不到
007	R	00=a,d (Map, Globe)				007/02 之位址不轉	用不到
007	R	00=a,d (Map, Globe)				007/05 之位址不轉	用不到
007	R	00=g (Projected, graphic)				007/02 之位址不轉	用不到
007	R	00=h (Microform)				007/02 之位址不轉	用不到
007	R	00-k (Nonprojected graphic)				007/02 之位址不轉	用不到
007	R	00=m (Motion graphic)				007/02 之位址不轉	用不到
007	R	00=s (Sound recording)				007/02 之位址不轉	用不到
007	R	00=v (Video recording)				007/02 之位址不轉	用不到
008	NR	39 (Cataloging source)					用不到
008	NR	32 (Main entry not in body of entry)					用不到
008	NR	32 (Music)(Main entry not in body of entry)					用不到
008	NR	32 (Visual material)(Main entry not in body of entry)					用不到
008	NR	33 (Visual material) type of material					在執行代 碼06位址 表示
008	NR	22 (Serial) Form of original item					用不到
008	NR	34 (Serial) Successive/latest entry					用不到

資料來源：中國圖書館學會圖書館自動化規劃委員會研訂，美國機讀編目格式轉至第三版中國
機讀編目格式對照表（台北市：教育部出版，民82），頁108。