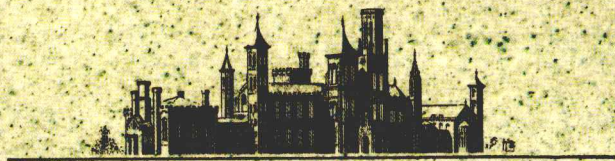


新圖書館學叢書



機讀編目格式

在都柏林核心集的應用探討

吳 政 叡

台灣學書局印行



機請繪圖格式

在圖繪環境心理學的應用設計

周 國 彪

香港大學出版社

**機讀編目格式
在都柏林核心集的應用探討**

吳政叡 著

機讀編目格式在都柏林核心集的應用探討

吳政叡著.-- 初版.-- 臺北市：臺灣學生，1998(民 87)

面；公分

含索引

ISBN 957-15-0930-2 (精裝)

ISBN 957-15-0931-0 (平裝)

1.元資料 2.資料描述格式 3.電子資料處理 4.機讀編目

312.972

87016889

機讀編目格式在都柏林核心集的應用探討

著 作 者：吳 政 叡
出 版 者：臺 灣 學 生 書 局
發 行 人：孫 善 治
發 行 所：臺 灣 學 生 書 局

臺北市和平東路一段一九八號

郵政劃撥帳號 00024668 號

電 話：(02) 23634156

傳 真：(02) 23636334

本書局登記證字號：行政院新聞局局版北市業字第玖捌壹號

印 刷 所：宏 輝 彩 色 印 刷 公 司
中和市永和路三六三巷四二號
電 話：(02) 22268853

定價：精裝新臺幣二八〇元
平裝新臺幣二一〇元

西 元 一 九 九 八 年 十 二 月 初 版

自 序

本書《機讀編目格式在都柏林核心集的應用探討》是作者所寫作有關都柏林核心集的第二本專書，相較於第一本書《都柏林核心集與元資料系統》的偏重基本和全面性介紹，本書則是深入的探討都柏林核心集在圖書館、資料處理與檢索等方面的理論與實務課題。

都柏林核心集是資源描述型元資料（Metadata）的一種，元資料主要是描述資料屬性的資訊，用來支持如指示儲存位置、資源尋找、文件紀錄、評價、過濾等的功能。以圖書館的角度來看，就其本義和功能而言，元資料可說是電子式目錄，因為編製目錄的目的，即在描述收藏資料的內容或特色，進而達成協助資料檢索的目的。因此元資料是用來揭示各類型電子文件或檔案的內容和其他特性，其典型的作業環境是電腦網路作業環境。換言之，元資料是因應現代資料處理上的二大挑戰而興起的：一是電子檔案成為資料的主流，另外一個是網路上大量文件的管理和檢索需求。

值得注意的，雖然從圖書館界的角度來看，元資料的功用和角色跟目錄相似，資源描述型元資料也是當今元資料的主流，但是元資料不等於書目描述資料，因為那祇是元資料中的一類，尚有其他種類的元資料扮演不同的角色，例如元資料中可放入與資料安全相關的資訊，像是數位簽名等。

附帶一提的，有很多人曾問作者為何要將 Metadata 翻譯成「元資

・機讀編目格式在都柏林核心集的應用探討・

料」，主要是因為字首 meta 表示「結構」的意思，而且已在很多理工的領域中被廣泛使用，作者在參閱大部分的理工辭典對 Meta-language 等的翻譯後，也將 meta 字首翻成「元」，是取其「結構」之義，而非「最初」之義。既然將其引進到圖書館學和資訊科學領域時，其使用方式和本義並無不同，為尊重學術慣例，理應加以沿用，以避免無謂的困擾，因此將 Metadata 翻譯成「元資料」。

由於都柏林核心集 (Dublin Core) 自作者的第一本書《都柏林核心集與元資料系統》出版後，又有不小的變化，因此本書中很多都柏林核心集的相關部份，已重新改寫或者修訂過。現在將本書的章節介紹如下：第一章是元資料的興起背景與簡介，使讀者對元資料興起的背景能有充分的了解，進而對未來資料處理與檢索方面的發展趨勢，能有更清晰的認知。第二章都柏林核心集，是有關於都柏林核心集的詳細介紹。

第三章機讀編目格式轉換是本書兩個主要探討的課題之一，作者創造了幾個轉換的原則，不但順利的將較複雜的機讀編目格式轉換成較簡單的都柏林核心集，還同時達成了兩個主要的目標：一是不損傷都柏林核心集原有的架構和格式；一是資訊的損失非常低，幾乎是沒什麼資訊的流失。因此這幾個轉換的原則，不但充分發揮了都柏林核心集的彈性，也提示了將來傳統印刷媒體資料和電子印刷媒體資料合併處理的可能性。第四章是介紹都柏林核心集到機讀編目格式的格式轉換，主要是都柏林核心集轉換到國際機讀編目格式 (UNIMARC)、美國機讀編目格式 (USMARC)、中國機讀編目格式 (Chinese MARC) 的分析。

第五章分散式元資料系統 (DIMES)，由於作者所寫的一個都柏

林核心集實驗系統，已經過大幅度的改寫和換名，所以特別加以詳盡介紹，也使讀者對第六章所提的實驗，其實作環境能有充分的認識。第六章是都柏林核心集與檢索失誤率，雖然祇進行了一個簡單的先導式實驗，但是透過作者親自設計的實驗方法，其實驗結果清楚的證明了都柏林核心集在資訊檢索上所能扮演的角色，和可達到的功效。更重要的，爲了能明白闡釋實驗的數據與結果，作者創造了一個新的檢索系統衡量標準——「檢索失誤率」，不但可以有效克服傳統衡量標準—回收率和精確率的種種缺失，更可以用來作爲各種檢索系統（如圖書館自動化系統中的線上公用目錄查詢系統、光碟資料庫、WWW的搜尋引擎）的共通衡量標準，對未來檢索系統的評估與設計，有深遠和重大的影響。

作者撰寫本書時雖然力求完善，然而一己的能力畢竟有限，疏漏在所難免，尚祈各界先進和同業不吝指正。最後本書的完成，要感謝家人和同儕的協助，也是本書完成不可或缺的助力，在此表達最深的謝意。

吳政叡 謹識

民國 87 年 11 月

於輔仁大學圖書資訊系

都柏林核心集與元資料系統

一書中的自序

自小由於喜歡閱讀各種書籍，因此對資料的整理和分析也頗感興趣，在就讀輔仁大學圖書館系時，就曾以 DBASE II 套裝軟體寫了一個小型的資料庫系統來整理自己的書籍和資料。退伍後幸賴父母資助得以出國攻讀電腦學位，雖然在美攻讀電腦博士的研究領域，主要是類神經元網路（Neural Networks）和模糊邏輯（Fuzzy Logic），但卻也在攻讀博士學位的最後半年，因緣際會的接觸到電子圖書館方面的文獻，並開始收集相關的資料。學成返國後回到輔仁大學圖書資訊系任教，除了繼續電子圖書館方面的研究外，也在偶然機會下接觸到元資料並對其著迷，二年多前那時正是搜尋引擎（Search Engine）勢力如日中天之際，雖然國外對元資料的研究已在開始推動，但是國內不論是圖書館界或電腦界，對元資料卻還是相當的陌生，因此連翻譯名詞也找不到，作者在參閱大部分的理工辭典對 Meta-language 等的翻譯後，也將 meta 字首翻成「元」，因此將 Metadata 翻譯成「元資料」。

元資料（Metadata）最常見的英文定義是 "data about data"，可直譯為描述資料的資料，主要是描述資料屬性的資訊，用來支持如指示儲存位置、資源尋找、文件紀錄、評價、過濾等的功能。以圖書館的

・機讀編目格式在都柏林核心集的應用探討・

角度來看，就其本義和功能而言，元資料可說是電子式目錄，因為編製目錄的目的，即在描述收藏資料的內容或特色，進而達成協助資料檢索的目的。因此元資料是用來揭示各類型電子文件或檔案的內容和其他特性，其典型的作業環境是電腦網路作業環境。換言之，元資料是因應現代資料處理上的二大挑戰而興起的：一是電子檔案成為資料的主流，另外一個是網路上大量文件的管理和檢索需求。

從另外一個角度來看，元資料的興起跟 WWW 與搜尋引擎的盛行頗有關連，WWW 盛行後，為因應檢索網頁內容的需要而有搜尋引擎的產生，搜尋引擎運作的方式，基本上是屬於全文檢索，主要是透過自動抓取程式在網際網路上抓取網頁，然後以自動拆字（或詞）作索引的方式來建立其資料庫，做為檢索的基礎，這種操作方式的特點是高運作效率和一網打盡，因此有高回收率與低精確率的特性，這個低精確率的缺點，隨著 WWW 網頁數量的急遽膨脹，成為無法忍受的致命傷。於是大家體會到對資料加以適當描述的重要性，這跟圖書館製作目錄的動機是一致的，這個古老的經驗又得到再一次的肯定。

這本書的主題是都柏林核心集（Dublin Core），它是眾多元資料種類中的一種，是 1995 年 3 月由 OCLC 和 NCSA 所聯合贊助的研討會，在邀請五十二位來自圖書館、電腦、網路方面的學者和專家，共同研討下的產物，是一個簡單有彈性，且非圖書館專業人員也可輕易了解和使用的資料描述格式。這種簡單有彈性和適合各種專業人員的特性，正是它在國外越來越受到歡迎的主要因素，也是作者特別青睞都柏林核心集的原因，這是因為作者同時具有圖書館學和電腦的背景，了解到在現階段，一種適合各專業人士的簡易元資料的必要性；一方面傳統的機讀編目格式過於繁瑣，也繼承太多的傳統包袱，同時

傳統圖書館的著錄方式並不適合非圖書館專業的人。另一方面以作者對目前人工智慧、類神經元網路、模糊邏輯等相關學科的了解，知道創造一個具有現今一般圖書館員智慧的自動化系統，在現階段仍是一個遙不可及的夢想，因為至今我們連模仿一個三歲小孩說和聽故事的智力都有困難，更別說是模仿一個成年的專業人士。所以綜合來說，在現階段資料的描述仍需以人工作業為主，同時以今日網際網路上資料膨脹的速度來看，光靠圖書館員來處理是不夠的，由（眾多專業的）文件或資料的創造者本身來自行加以描述，已是必然的趨勢，這正是類似都柏林核心集這種元資料受重視的原因。

二年多來作者雖然發表一系列的文章來介紹元資料，特別是都柏林核心集，但仍感到過於分散，缺乏完整性和系統性，同時國內外有關元資料和都柏林核心集的專書也甚為稀少，乃將二年多來的研究心得加以整理寫成此書，希望能為國內在元資料和都柏林核心集的研究上，起一些推波助瀾的效果。

本書的章節安排如下：第一章是時代背景介紹，使讀者對現在資訊檢索方面的新趨勢有一些基本認識，以了解元資料興起的背景。第二章是元資料概論，闡述元資料的定義和內涵，介紹現有的元資料的種類，最後選擇三個具代表性的元資料，來加以仔細的介紹，使讀者對元資料能有一個較清楚的整體認知。第三章是都柏林核心集的詳細介紹和分析，首先是都柏林核心集至今的整個發展歷程，其次是每個欄位的詳細剖析，這是有興趣使用都柏林核心集的讀者必須詳讀的章節，再來是修飾詞的介紹，修飾詞是都柏林核心集最新的發展方向，雖然整體的內容尚未有定論，但對有心深入都柏林核心集的讀者，也是必須詳讀的章節。第四節是作者針對都柏林核心集的重要特色與制

· 機讀編目格式在都柏林核心集的應用探討 ·

定原則，加以詳細的剖析，最後是都柏林核心集轉換到國際機讀編目格式（UNIMARC）、美國機讀編目格式（USMARC）、中國機讀編目格式（Chinese MARC）的分析。

第四章是介紹都柏林核心集具代表性的國外研究計畫，分別是北歐元資料計畫和分散式系統技術中心（DSTC），這兩個計畫各具特色，非常值得參考，讀者會因此對都柏林核心集的功用刮目相看。第五章是介紹作者所發展的一套整合式都柏林核心集系統--元資料實驗系統，相當實用和具有特色，同時也開放給有興趣的使用者著錄和檢索，是一個很好的實驗系統，來支援都柏林核心集和元資料的相關研究與實驗，歡迎大家使用與不吝指教。

作者撰寫本書時雖然力求完善，然而一己的能力畢竟有限，寫作期間又常須分身照顧四歲的小女虹熠（Grace），疏漏在所難免，尚祈各界先進和同業不吝指正。最後本書的完成，要感謝盧荷生老師平日不斷的督促與鼓勵，並且慨允為本書作序，同時家人和同儕的協助，也是本書完成不可或缺的助力，在此表達最深的謝意。

吳政叡 謹識

民國 87 年 2 月

於輔仁大學圖書資訊系

目 次

自 序	I
都柏林核心集與元資料系統一書中的自序	V
目 次	IX
圖表目次	XIII
第一章 元資料的興起背景與簡介	1
第一節 時代背景	3
第二節 元資料的興起	13
第三節 元資料的定義與類型	18
第二章 都柏林核心集	25
第一節 發展沿革	26
第二節 重要特色與制定原則	43
第三節 基本欄位	46
第四節 修飾詞 (QUALIFIER)	55

・機讀編目格式在都柏林核心集的應用探討・

語言修飾詞 (LANG)	55
架構修飾詞 (SCHEME)	59
次項目修飾詞 (SUBELEMENT)	63
第三章 機讀編目格式轉換	73
第一節 0 段欄號	76
第二節 1 段欄號	81
第三節 2 段欄號	108
第四節 3 段欄號	116
第五節 5 段欄號	118
第六節 6 段欄號	136
第七節 7 段欄號	143
第八節 8 段欄號	150
第四章 都柏林核心集格式轉換	157
第一節 國際機讀編目格式 (UNIMARC)	158
第二節 美國機讀編目格式 (USMARC)	163
第三節 中國機讀編目格式 (CHINESE MARC)	168
第五章 分散式元資料系統 (DIMES)	175
第一節 系統特色簡介	177
第二節 註冊子系統	179
第三節 都柏林核心集子系統	182
著錄次系統	183

查詢次系統	186
第四節 查詢子系統.....	189
第五節 URL 子系統.....	192
第六節 評價認證 (SOAP) 子系統.....	195
著錄次系統	201
查詢次系統	203
第六章 都柏林核心集與檢索失誤率	207
第一節 檢索效率的衡量方法與檢索失誤率	208
第二節 研究動機與實驗設計	211
第三節 實驗結果整理與分析	213
第四節 結 語	216
附錄一 NISO Z39.53 的語言代碼	219
中文索引	227
英文索引	231

圖表目次

圖 1-1.	WWW 的使用者介面範例	5
圖 1-2.	一個利用微軟 Frontpage 97 所建立測試網頁的 原始檔案內容	6
圖 2-1.	電子郵件中夾帶特殊格式檔案 (微軟 Internet 郵件軟體)	33
圖 2-2.	一個 CGI 程式範例(使用 PERL 語言)	36
圖 2-3.	Netscape 軟體中「一般設定」功能項的畫面	37
圖 2-4.	微軟「檔案總管」中「檢視－資料夾選項」 功能項的畫面	37
圖 2-5.	微軟 Excel 中提供的日期格式的部分畫面.....	51
圖 2-6.	微軟 Word 軟體製作 HTML 文件時 使用 ISO 639 語言代碼.....	60
圖 5-1.	分散式元資料系統首頁	176
圖 5-2.	註冊子系統首頁	180
圖 5-3.	註冊子系統(新增)註冊畫面.....	180
圖 5-4.	註冊子系統修改註冊資料畫面.....	181
圖 5-5.	註冊子系統查詢畫面	181
圖 5-6.	都柏林核心集子系統首頁	182
圖 5-7.	都柏林核心集子系統全部項目著錄部分畫面	183

圖 5-8. 都柏林核心集子系統單一項目著錄（一般項目） 部分畫面	184
圖 5-9. 都柏林核心集子系統單一項目著錄（關連值項目） 部分畫面	184
圖 5-10. 都柏林核心集子系統更新部分畫面	185
圖 5-11. 都柏林核心集子系統刪除部分畫面	186
圖 5-12. 都柏林核心集子系統單一欄位查詢（一般項目）畫面	187
圖 5-13. 都柏林核心集子系統單一欄位查詢（關連值項目） 部分畫面	187
圖 5-14. 都柏林核心集子系統複合欄位查詢部分畫面	188
圖 5-15. 都柏林核心集子系統單一文件資料查詢部分畫面	189
圖 5-16. 查詢子系統首頁	190
圖 5-17. 查詢子系統關鍵字查詢畫面	190
圖 5-18. 查詢子系統單一欄位查詢畫面	191
圖 5-19. 查詢子系統 URN 查詢畫面	191
圖 5-20. URL 子系統首頁	192
圖 5-21. URL 子系統中的 URL 著錄畫面	192
圖 5-22. URL 子系統中的 URL 更新畫面	193
圖 5-23. URL 子系統中的 URL 查詢畫面	194
圖 5-24. URL 子系統中的 URN 對映 URL 查詢畫面	194
圖 5-25. 評價認證（SOAP）子系統首頁	195
圖 5-26. 評價認證子系統增加與刪除評審人首頁的部分畫面	196
圖 5-27. 評價認證子系統增加評審人畫面	197
圖 5-28. SOAP 的共用查詢子系統首頁畫面	197

圖 5-29. 評價認證子系統評論資料關鍵字查詢畫面.....	198
圖 5-30. 評價認證子系統單一文件資料畫面.....	198
圖 5-31. 評價認證子系統評審人查詢畫面.....	199
圖 5-32. 評價認證子系統 SOAP 種類查詢結果畫面	199
圖 5-33. 假想評價認證子系統——輔大中文評價認證的首頁	200
圖 5-34. 輔大中文評價認證的來源文件描述首頁.....	201
圖 5-35. 輔大中文評價認證的評論部份畫面.....	202
圖 5-36. 輔大中文評價認證的更新畫面.....	203
圖 5-37. 輔大中文評價認證的查詢次系統的首頁畫面.....	204
圖 5-38. 輔大中文評價認證的查詢次系統的評論資料 單一欄位查詢畫面	205
表 3-1. 中國機讀編目格式第 0 段欄號的對照表.....	76
表 3-2. 中國機讀編目格式第 1 段欄號的對照表.....	81
表 3-3. 中國機讀編目格式第 2 段欄號的對照表.....	108
表 3-4. 中國機讀編目格式第 3 段欄號的對照表.....	116
表 3-5. 中國機讀編目格式第 5 段欄號的對照表.....	118
表 3-6. 中國機讀編目格式第 6 段欄號的對照表.....	136
表 3-7. 中國機讀編目格式第 7 段欄號的對照表.....	143
表 3-8. 中國機讀編目格式第 8 段欄號的對照表.....	150
表 4-1. 都柏林核心集對映到國際機讀編目格式的對照表.....	159
表 4-2. 都柏林核心集對映到美國機讀編目格式的對照表.....	164
表 4-3. 都柏林核心集對映到中國機讀編目格式的對照表.....	168
表 6-1. S. E. Robertson 的 2 x 2 表格	209
表 6-2. 實驗結果	214