

都柏林核心集使用者查詢行為調查： 以施合鄭基金會為例

吳政叡

輔仁大學圖書資訊系教授兼系所主任

【摘要】

作者用與施合鄭基金會合作所建立的網路書目資料處理系統（臺灣地區中文聯合系統，使用都柏林核心集格式）為對象，以民國90年2月起至民國90年12月止的11個月為期，利用系統來自動記錄使用者查詢的關鍵詞，及其他相關資料，加以適當的分析，來探究使用者的查詢行為和資訊需求。調查結果發現，就查詢的方式來看，即使（全部欄位）關鍵詞的查詢方式較方便，但是大部分時候使用者還是比較喜歡較精確的單一欄位方式。再以都柏林核心集的15個欄位來分析，傳統的查詢角度（欄位）如書名和作者，仍然佔有絕對的優勢。最後，調查結果顯示，單一欄位查詢方式，其每次查詢平均產生的個回覆款目，遠小於全部欄位查詢方式，因此單一欄位方式較全部欄位方式，不會帶給使用者過多的垃圾資料和資訊超載問題。不過由於本研究所使用的系統，尚屬於陸續建立館藏資料的草創階段，加上使用量並不太大，因此所得資料尚有待後續更多的研究來證實。

關 鍵 詞：都柏林核心集、查詢過程記錄分析、檢索詞彙記錄分析

一、前 言

不管是查詢過程記錄分析 (Transaction Log Analysis) ^[1] 抑或檢索詞彙記錄

[1] 張嘉彬， 臺大電子圖書館與博物館系統之評估研究：以查詢過程記錄分析法為例，〈中國

分析 (Search Term Log Analysis)^[2] 都是利用電腦來自動記錄系統使用者的檢索或是查詢過程，然後利用所得資料來分析使用者的檢索 (或是查詢) 方式、行為與資訊需求等。由於在記錄的過程中，使用者一般不會覺察到或是有意識到，因此可以說是一種非常自然和客觀的不涉入觀察方式。有關查詢過程記錄分析法的發展歷史與優缺點，在「臺大電子圖書館與博物館系統之評估研究：以查詢過程記錄分析法為例」一文中已有詳實的敘述^[3]，不在此贅述。

雖然自 1960 年代以來，查詢過程記錄分析已經有大量的研究成果和文獻存在，不過由於使用者的查詢行為，可能會因為文化、地域、語文和工具類型等因素的影響而有所不同，其研究仍然有持續進行的必要，尤其是當有全新的熱門資料類型或是檢索系統與工具出現時，更是有必要針對新的資料類型或檢索工具來進行調查。

1990 年全球資訊網 (World-Wide Web, 簡稱 WWW) 的誕生^[4]，迅速創造了一個熱門流行的新型態資料網頁。隨後基於網頁檢索的需求，產生了一個新的檢索工具搜尋引擎 (Search Engine)。臺灣網路使用者檢索詞彙分析研究一文即是針對搜尋引擎所做的初步研究^[5]，不過由於在該研究中所收集的期間較短 (一為 3 個月，一為 2 星期) 且為單一的時間片斷^[6]，因此在資料的代表性上似乎是有不足的。另一方面，受限於所取得資料的內容，其研究焦點也祇在檢索詞彙的平均長度、頻率分佈、與核心詞彙的比例^[7]，因此研究面相對也顯得有些不足，有待後續更多的研究來建立更穩固的結論。

由於搜尋引擎運作的方式是屬於全文檢索，主要以自動拆字 (或詞) 做索引的方式來建立其資料庫，做為檢索的基礎，這種方式有高回收率與低精確率的特性。因此目前在使用 WWW 上的搜尋引擎來查詢資料時，有資料回覆量太多和垃圾比率太高兩個致命傷。因而在 1990 年代中期以後，元資料 (Metadata) 逐

圖書館學會會報》，66 (2001.6)，頁 107。

[2] 卜小蝶，「臺灣網路使用者檢索詞彙分析研究」，《國家圖書館館刊》民國 89 年第 1 期 (2000.6)，頁 26。

[3] 同註 1，頁 107-109。

[4] T. Berners-Lee, L. Masinter, and M. McCahill, "Uniform Resource Locators (URL)," 1994, from <ftp://ds.internic.net/rfc/rfc1738.txt>, p.1.

[5] 同註 2，頁 25-34。

[6] 同註 2，頁 27-28。

[7] 同註 2，頁 27。

漸應運而生成為研究的主流。

元資料的英文定義是 "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。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。^[8] 臺大電子圖書館與博物館系統之評估研究：以查詢過程記錄分析法為例一文^[9]，即是針對一個建基於 Metadata 的系統所進行之研究，該研究主要分成兩部分，一是就約2 星期內系統中各網頁被使用的情況進行分析，一是就約4 個月內所收集的 1,029 個查詢敘述做分析。^[10] 不過由於第一個部分進行的期間甚短，且限於網頁被使用的次數，因此其研究的面相相對來說較窄。第二個部分進行的期間較長，分析的面相較多，可能的缺陷是資料量 1,029 筆稍少，同時似乎也未完全按照隨機原理來收集，因此資料的代表性可能仍有疑慮，這都有待後續更多的相關研究來相互印證。

本研究所使用的資料，來自作者與施合鄭基金會長期合作所建立的網路書目資料處理系統「臺灣地區中文聯合系統」，此系統使用的都柏林核心集格式為元資料的一種，近年來在國際間非常受到重視。本文以 1個月期間，系統所自動記錄的使用者查詢資料為基礎，加以分析來探究使用者的查詢行為和資訊需求。

二、研究計畫與系統查詢功能簡介

如前所述，本研究所建基的都柏林核心集為元資料的一種，是 1995 年 3 月由國際圖書館電腦中心(OCLC)和 National Center for Supercomputing Applications (NCSA) 所聯合贊助的研討會下的產物，目的是希望建立一套描述網路上電子文件特色的方法，來協助資訊檢索。^[11] 都柏林核心集的 15 個基本項目有：主題和關鍵詞 (Subject) 題名 (Title) 著者 (Creator) 簡述 (Description) 出版者 (Publisher) 其他參與者 (Contributors) 出版日期 (Date) 資源類型 (Type) 資

[8] 吳政勳，從元資料看未來資料著錄的發展趨勢，《資訊傳播與圖書館學》3：2 (1997. 12)，頁 44-45。

[9] 同註 1，頁 103-125。

[10] 同註 1，頁 110。

[11] Stuart Weibel, Jean Godby, Eric Miller, and Ron Daniel, "OCLC/NCSA Metadata Workshop Report," 1995, from http://www.oclc.org:5047/oclc/research/publications/weibel/metadata/dublin_core_report.html, p.2.

料格式 (Format) 資源識別代號 (Identifier) 關聯 (Relation) 來源 (Source) 語言 (Language) 涵蓋時空 (Coverage) 版權規範 (Rights) [12][13]

作者近年來一直致力於都柏林核心集在圖書館的應用，先後發表一系列的論著 [14]，並提出一套利用都柏林核心集來處理書目資料的架構。 [15] 同時自民國 88 年 6 月起，與施合鄭民俗文化基金會合作，將此套架構付諸實踐。

施合鄭民俗文化基金會設立於民國 69 年 6 月 16 日，為一非營利的財團法人機構。原先設立之宗旨在於保存、發揚及研究與社區廟會相關之民俗曲藝，如子弟戲、布袋戲及陣頭等。自民國 80 年起將研究地域擴展及大陸，研究範圍增加宗教與儀式，研究視野進而包容整個文化現象。 [16]

自民國 88 年 6 月起，與施合鄭民俗文化基金會合作的第一期計畫，以八個月為期，建立一套系統，協助施合鄭民俗文化基金會處理其館藏。第一期計畫的目標在建立一個線上的操作系統，可以利用網路來執行著錄、查詢、管理等方面的功能，並且建立 5,000 筆書目紀錄。不管是基金會的工作人員，或是一般的使用者，皆可隨時上網查詢資料。

在第一期計畫順利完成後，接著進行第二期計畫，第二期第一年計畫自民國 89 年 2 月起至民國 90 年 1 月止，主要是針對尚未建檔的館藏進行編目。第二期計畫的第二年（從民國 90 年 2 月起至民國 91 年 1 月止）的主要完成工作項目有三：

- (一) 書籍編目：接續第一期計畫與第二期第一年計畫的編目工作，第二期計畫的第二年完成 2,273 筆新的書目紀錄。目前系統總計約有 14,000 餘筆書目資料。
- (二) 舊有書目紀錄增加主題欄位：主題 (subject) 對於揭示書籍的內容和協助使用者的查詢方面，有非常大的幫助，目前系統約有 2,700 筆主題資料。
- (三) 使用者的查詢行為和資訊需求研究：利用系統來自動記錄使用者查詢

[12] 同註 8。

[13] 有關都柏林核心集的欄位最新資訊，請參見 <http://dublincore.org/documents/dces>。

[14] 吳政叡，《都柏林核心集與圖書著錄》(臺北：學生書局，2000.12)，頁 16-20。

[15] 同註 14，頁 15。

[16] 有關施合鄭民俗文化基金會的詳細介紹，見 <http://dimes.lins.fju.edu.tw/shc>。

的關鍵詞，及其他相關資料。

與施合鄭民俗文化基金會合作所建立的臺灣地區中文聯合館藏系統 (UCSTW，網址：<http://dimes.lins.fju.edu.tw/shc>)，提供四種不同的書目資料查詢方式：（全部欄位）關鍵字查詢、單一欄位查詢、複合欄位查詢和urn 查詢。下面圖一和圖二分別為關鍵字查詢和單一欄位查詢的畫面。

施合鄭民俗文化基金會 (UCSTW)
單一欄位查詢

- 祇須鍵入關鍵字,前後無須加任何符號.
例如: e-找出所有含字母e的資料
- 系統會自動轉換單引號(')為()。

選擇查詢類別: 施合鄭民俗文化基金會 (UCSTW) ▾

欄位名稱: 索書號 (Call Number) ▾

選擇輸出格式: 圖書館次項目格式 ▾

關鍵字:

傳送 重新輸入

圖一：關鍵字查詢的畫面

施合鄭民俗文化基金會 (UCSTW)
單一欄位查詢

- 祇須鍵入關鍵字,前後無須加任何符號.
例如: e-找出所有含字母e的資料
- 系統會自動轉換單引號(')為()。

選擇查詢類別: 施合鄭民俗文化基金會 (UCSTW) ▾

欄位名稱: 索書號 (Call Number) ▾

選擇輸出格式: 圖書館次項目格式 ▾

關鍵字:

傳送 重新輸入

圖二：單一欄位查詢的畫面

三、查詢記錄分析結果

以下作者根據系統自動記錄的使用者查詢資料，分別從以下不同的角度來分析：

- (一)查詢方式比較：四種系統書目資料查詢方式（全部欄位）關鍵字查詢、單一欄位查詢、複合欄位查詢和urn 查詢的次數比較。
- (二)都柏林核心集 15 個欄位的利用情況：從單一欄位查詢中來分析 15 個欄位的個別查詢次數。
- (三)關鍵詞查詢時的回覆款目數量：從資訊超載的角度來比較（全部欄位）關鍵字查詢和單一欄位查詢的平均回覆款目數量。

首先，本研究的調查期間，為民國90年2月起至民國90年12月止的11個月，此時期間系統的書目資料紀錄總數在12,000至14,000之間。以四種系統所提供的查詢方式來看：（全部欄位）關鍵字查詢計有212次、單一欄位查詢有1,421次、複合欄位查詢有2次、urn查詢有3,335次。

由以上的數據可以看出urn查詢最多，11個月間共有3,335筆，平均每月有303.18次查詢。其原因可能是編目人員常常直接透過urn來查詢資料所致。其次是單一欄位查詢的1,421次，而最少的複合欄位查詢祇有2次，可見布林邏輯的複合欄位查詢，立意雖好，但是較為複雜，一般極少被使用。

另一方面，（全部欄位）關鍵字查詢方式祇有212次，也遠少於單一欄位查詢的1,421次。看來雖然全部欄位的查詢方式較方便，不過其缺點是系統回覆款目過多，增加使用者過濾資料的困難。因此大部分時候使用者還是比較喜歡較精確的單一（指定）欄位方式。

以單一（指定）欄位查詢方式而言，因為都柏林核心集有15個欄位，因此下面就各欄位的利用情況來分析。根據系統資料顯示，15個欄位中祇有9個欄位有被使用，分別為：Title、Creator、Identifier、Subject、Date、Publisher、Contributor、Description、Type，其中被使用最頻繁的欄位是Title（題名或書名），而且超出其他欄位甚多，使用次數達957次，佔所有單一（指定）欄位查詢次數（共1,421次）的67.35%。其次是欄位Creator（作者），達172次，佔所有單一（指定）欄位查詢次數的12.10%。再來是Identifier（43次）和Subject（19次），其餘欄位皆不到10次。以下是9個有被使用欄位，其次數的詳細統計

如表一（按使用次數高低排列）：

表一：都柏林核心集9個有被使用欄位的次數統計表

| 欄位名稱 | 次數 |
|-------------|-----|
| Title | 957 |
| Creator | 172 |
| Identifier | 43 |
| Subject | 19 |
| Publisher | 8 |
| Date | 7 |
| Description | 7 |
| Contributor | 6 |
| Type | 1 |

從以上的分析可以清楚得知，傳統的查詢角度（欄位）如書名和作者，仍然佔有絕對的優勢，兩者合計佔所有單一（指定）欄位查詢次數的9.45%。

最後為了觀察查詢時的回覆款目數量，系統也有記錄此方面的資訊。首先是全部欄位查詢方式，全部212次使用者的關鍵詞查詢，產生9,658個回覆款目，因此平均每次關鍵詞查詢產生45.56個回覆款目。若是扣除0個回覆款目的查詢，則查詢次數只剩下120次，因此平均每次關鍵詞查詢產生的回覆款目增加為80.48個。

其次就單一（指定）欄位查詢方式的回覆款目來分析，全部1,421次使用者的關鍵詞查詢，產生29,416個回覆款目，因此平均每次關鍵詞查詢產生17.99個回覆款目。若是扣除0個回覆款目的查詢，則查詢次數只剩下925次，因此平均每次關鍵詞查詢產生的回覆款目增加為31.80個（參見表二）。

由以上的分析清楚顯示，單一（指定）欄位查詢方式，其每次查詢平均產生的回覆款目，遠小於全部欄位查詢方式（不管是17.99對比45.56，抑或是31.80對比80.48），由此可知，單一欄位方式較全部欄位方式精確，較不會帶給使用者過多的垃圾資料和資訊超載問題。

表二：查詢方式與平均回覆款目數量的統計表

| 查詢方式 | 平均回覆款目的數量 | |
|-----------|-----------|----------|
| | 一般 | 扣除0個回覆款目 |
| (全部欄位)關鍵字 | 45.56 | 80.48 |
| 單一欄位 | 17.99 | 31.80 |

四、結語

作者以臺灣地區中文聯合系統 (UCSTW, 與施合鄭基金會合作建立的網路書目資料處理系統, 使用都柏林核心集格式) 為工具, 從2001年2月起至2001年12月止的11個月期間, 利用系統來自動記錄使用者的查詢方式和資料。調查結果發現, 就查詢的方式來看, 即便(全部欄位)關鍵詞查詢方式較方便, 但是大部分時候使用者還是比較喜歡較精確的單一欄位方式。分別是(全部欄位)關鍵詞查詢有212次和單一欄位查詢有1,421次。

上述結果和 臺大電子圖書館與博物館系統之評估研究：以查詢過程記錄分析法為例 一文中的調查結果有極大的差異, 在其研究中發現(全部欄位)關鍵詞查詢次數略多於單一欄位查詢次數^[17]; 但是在本研究中發現單一欄位查詢次數遠多於(全部欄位)關鍵詞查詢次數。

再以都柏林核心集的15個欄位來分析, 傳統的查詢角度(欄位)如書名和作者, 仍然佔有絕對的優勢, 書名(Title)欄位佔所有單一欄位查詢次數的67.35%, 而作者(Creator)欄位佔所有單一欄位查詢次數的12.10%, 兩者合計幾近80%。

最後, 調查結果顯示, 單一欄位查詢方式, 其每次查詢平均產生的個回覆款目, 遠小於全部欄位查詢方式(不管是17.99對比45.56, 抑或是31.80對比80.48, 參看表2), 由此可知, 單一欄位方式較全部欄位方式精確, 較不會帶給使用者過多的垃圾資料和資訊超載問題。

不過由於本研究所使用的系統, 尚屬於陸續建立館藏資料的草創階段, 加上使用數量並不太大, 因此所得資料尚有待後續更多的研究來證實。

[17] 同註1, 頁113。

Investigation of Users' Query on UCSTW: A Dublin Core System

Cheng-Juei Wu

Professor and Chairman, Department of Library and Information Science

Fu Jen Catholic University

【 Abstract 】

In this work, the author checks into the query log automatically recorded by the system from Feb. to Dec. 2001 to investigate users' query behaviors and information needs on UCSTW, a Dublin Core based system built under a cooperative project with the Shih Ho-Cheng Folk Culture Foundation. Firstly, the results indicate that most of the users prefer the more accurate single-field query to the all-field keyword query. Secondly, for the Dublin Core 15 fields, the title and creator fields are two dominating fields using in query. Lastly, on average, for the single-field query, there are about 18 hit items, comparing to around 45.6 hit items for the simple all-field keyword query.

Keywords: Dublin Core, Transaction Log Analysis, Search Term Log Analysis, UCSTW