

台灣電子期刊使用者行為分析 ：以 Elsevier SDOS 電子期刊系統為例

Exploring the Behavior of E-Journal Users in Taiwan : An Analysis
of Web Server Log of Elsevier Science Direct On Site

戴玉旻

Yu-min Tai

交通大學資訊科學研究所碩士生

Master Student, Institute of Computer and Information Science,

National Chiao Tung University

E-mail: ymtai@cis.nctu.edu.tw

陳莉君

Li-chun Chen

交通大學資訊科學研究所碩士生

Master Student, Institute of Computer and Information Science,

National Chiao Tung University

E-mail: ls86068@cis.nctu.edu.tw

保弗·卡克拉克

Rolf Kwakkelaar

Elsevier 亞太區電腦科技顧問

Head of IT-consultancy, Elsevier Science (Asia Pacific)

E-mail: r.kwakkelaar@elsevier.com.sg

柯皓仁

Hao-ren Ke

交通大學圖書館副教授兼數位圖書資訊組組長

Associate Professor, Head of Digital Library and Information Division,

Library, National Chiao Tung University

E-mail: claven@lib.nctu.edu.tw

【摘要 Abstract】

隨著數位圖書館時代的來臨，各類電子資料逐漸普及，WWW 版電子期刊、全文資料庫、索摘資料庫、電子書已經成為教學及研究時重要的資訊來源，探討並了解使用者在此類數位圖書館系統的使用行



為，可做為數位圖書館系統的研發者在改進系統功能時的重要依據，圖書館在進行教育宣傳時也能針對使用者較常發生的錯誤以及不熟悉的功能加強訓練。本研究係由分析 Elsevier SDOS 的使用記錄檔，期能獲得國內 SDOS 使用者的使用習慣和資訊尋求行為，並提出心得與建議。

In the era of digital libraries, WWW-based electronic databases have become important resources for education and research. Investigating into and then comprehending the behaviors of users in using these systems is very valuable. Based upon the findings of user behaviors, system developers can improve system features, such as query capability and user interface. In conducting user training, librarians can easily locate those system features unfamiliar to users. This paper analyzes and explicates the web server logs of the Taiwan Elsevier SDOS E-journals system, one of the largest and most heavily used full text STM databases worldwide. Results include characteristics of user sessions, server loading, query complexity, query terms, query refinement, browsing, and full-text downloading.

關鍵詞 Keyword

數位圖書館 使用者行為 使用分析 使用記錄檔分析 電子期刊

Digital library ; User behavior ; Usage analysis ; Transaction log analysis ; Electronic journal



壹、前言

圖書資訊學的研究目標之一在於培養人的資訊素養，使人能夠有效率及有效地找尋資訊，進而運用資訊，產生知識。為了達成此一目標，友善、方便、強大的資訊檢索系統乃是不可或缺的，也因此，「針對現有資訊檢索系統進行評估，使其能日新又新、滿足人的資訊檢索需求」一直是圖書資訊學中重要的研究課題。

有關資訊檢索系統的評估標準，張嘉彬(註 1)將資訊檢索系統分為線上資訊檢索系統(包含一般資訊檢索系統及線上公用目錄)以及網站搜尋引擎，並歸納分析文獻上相關研究，整理出二十一種評估標準，包括：資料收錄範圍、檢索功能、回現率、精確率、回應時間、檢索結果展示、使用者研究、成本研究、資料新穎性、資料正確性、檢索者、錯誤率、響應程度、基於使用、輔助說明、容易連線程度、資料庫簡介、資料庫建構、資料庫大小、欄位索引情況、連結性，其中又以回應時間、精確率、檢索結果展示、回現率、資料收錄範圍、資料新穎性、檢索功能及成本研究等標準較被學者專家們提及。這些標準多是考量資訊檢索系統的設計、檢索界面、檢索技術、資料內容等方面，然而資訊檢索系統是設計來讓人使用的，因此應該要更進一步了解使用者資訊需求、檢索問題、使用者行為、使用者和系統之間的互動、檢索詞彙和檢索策略的應用、檢索後的相關判斷及滿意度等因素，藉以探討與評量資訊檢索系統，了解使用者的行為模式，除了有益於資訊檢索系統的改進之外，也能夠增進圖書館對使用者檢索模式和使用困難的了解，提升對使用者服務的品質，並促進使用者教育的實施。(註 2)

隨著數位圖書館時代的來臨，各類電子資料與數位圖書館系統逐漸普及，WWW 版電子期刊、全文資料庫、索摘資料庫、電子書已經成為教學及

研究時重要的資訊來源，我們認為對此類數位圖書館系統(註 3)進行評估及使用者研究有其重要性。本研究之目的即是藉由分析 Elsevier SDOS (Science Direct On Site)電子期刊系統的使用記錄(WWW server log)，藉以明瞭台灣的使用者在 WWW 版電子期刊系統的資訊尋求行為(Information seeking behavior)，以提供此類系統設計與改進，以及圖書館進行資訊服務時的參考。SDOS 系統內包括 Elsevier 出版的一千三百餘種電子期刊，透過 WWW，使用者可於線上檢索、瀏覽、列印以及下載所需的期刊論文。SDOS 系統是由國科會科學技術資料中心主辦的全國學術電子資訊資源共享聯盟(Consortium on Core Electronic Resources in Taiwan, CONCERT)引進台灣，於 1999 年 1 月起正式提供服務，台灣的 SDOS 使用者是透過設置在中央研究院的鏡射站(Mirror site)來使用 SDOS，該鏡射站的網址為<<http://sdos.ejournal.nccu.net>>。

貳、相關研究及文獻

WWW 除了扮演重要資訊媒體的角色之外，還提供了探索使用者資訊尋求行為的重要參考資料。WWW 網站都有所謂的使用記錄檔(WWW server log file)，儲存使用者到訪 WWW 網站的相關資訊，包含使用者要求取得的檔案(如 HTML 檔、圖檔)、啓動的執行檔(如 ASP、PHP)、使用者的來源網域(Domain)或 IP、使用的瀏覽器、到訪過程中發生的任何錯誤...等。藉由分析 WWW 網站的使用記錄檔，可以了解下列資訊：(註 4)

- 一、使用者及其分布(Users and their distribution)：使用者來自何處？使用者是否經常到訪？是否有穩定的使用群？
- 二、使用者需要的資訊與服務(Information or services needed by users)：哪些網頁或服務曾經被使用？使用的頻率如何？哪些資訊



及服務最常被使用？哪些資訊及服務對使用者較有價值？

- 三、資訊或服務的使用模式與使用者的資訊行為 (Information or services usage pattern / users information behavior)：使用者如何瀏覽與使用資訊或服務？使用者是否主動參與互動式的溝通，如討論區、留言板等機制？

國內外已有若干針對使用記錄檔進行分析研究的相關文獻。Z. Zhang (註 5) 針對 *Review of Information Science (RIS, ISSN 1431-5971)* 此一電子期刊的使用記錄檔分析使用者來源網域、索取的資訊與服務的種類 (如全文、RIS 背景資訊、檢索服務、書籍或期刊評論、留言板...等)。Z. Zhang 僅探討單一電子期刊的使用情形，分析的資訊也較為簡單，且並未提及使用者的檢索情形；而本研究則探討 Elsevier SDOS 所收錄一千三百餘種全文期刊的瀏覽、檢索、全文下載情況。

S. Jones 等學者 (註 6，以下簡稱為 Jones-2000) 是針對紐西蘭數位圖書館 (New Zealand Digital Library) 的電腦科學技術報告子系統 (Computer Science Technical Reports, CSTR) 分析其使用者來源網域、使用者檢索及瀏覽檢索結果的情形。本研究與 Jones-2000 的相似之處在於皆對使用者檢索情形進行分析 (分析的種類也頗為類似)；相異之處則有以下幾點：

- 一、CSTR 的館藏為電腦科學領域的技術報告，Elsevier SDOS 則是包含電腦、醫學、生命科學、社會科學... 等多領域的期刊論文，且截至 2000 年 7 月為止，前者收錄的技術報告總數約有 46,000 篇，而後者收錄的期刊論文已經超過 1,000,000 篇。
- 二、本研究除了檢索情形的分析外，還進一步探討使用者瀏覽 Elsevier SDOS 網站以及全

文的下載行為，Jones-2000 則未對此多加著墨。

- 三、Jones-2000 尚以人工方式分析部份記錄，本研究則無。

B. J. Jansen 等學者 (註 7，以下簡稱為 Jansen-2000) 藉由 Excite 網站搜尋引擎的交易記錄檔，分析使用者的檢索策略，以了解使用者在網站搜尋引擎上的使用行為。此外，卜小傑 (註 8、9、10，以下將註 9 簡稱為卜小傑-89) 則是分析 Dreamer、GAIS、Openfind 等三個網站搜尋引擎的檢索詞彙紀錄，探討台灣地區使用者在使用網站搜尋引擎的檢索行為及檢索詞彙特性。這兩項研究儘管都是從使用記錄檔分析使用者行為，但都是以網站搜尋引擎的使用者為分析對象，顯而易見地，此類使用者和電子期刊系統使用者之資訊尋求目的是大相逕庭，然而此二類使用者在檢索行為上是否有異同之處呢？這也是本研究想要了解的。

參、使用記錄檔資料前置處理

本研究中採用的 SDOS 使用記錄檔涵蓋了 2000 年 1 月 1 日至 2000 年 9 月 18 日的使用記錄，使用記錄檔的大小約 1.3GB。SDOS 使用記錄檔是採用 NCSA 的取用記錄格式 (Access log format)，其中較重要的欄位包括使用者網域名稱或 IP 位址、日期與時間、使用者請求 (Request)、傳回的檔案大小... 等，表一為前述這些重要欄位的範例。

由於 SDOS 使用記錄檔中所記載的資料太過龐雜，為了便利分析工作的進行，本研究首先將使用記錄檔經過下列四個步驟進行前置處理。

- 一、資料清理 (Data cleaning)：去除在分析過程中不必要或不會用到的資料，例如圖檔取用記錄、HTTP 錯誤記錄、HTTP 版本資料... 等。
- 二、資料整合 (Data integration)：因為 SDOS 曾在 2000 年 4 月間更新版本，而新舊版本使



用記錄檔所記載的資訊稍有差異，本步驟整合新舊版本的使用記錄檔。

- 三、資料分類：將使用記錄依據使用者對系統提出的需求(Request)分成十四類(參見表二)，並將分類結果分別儲存。資料分類的

優點在於統計資料時，只需要取出相關的分類記錄進行分析即可。例如：當要分析使用者下載全文的情形時，只要針對 PF 及 PS 類的資料即可。

表一：SDOS 使用記錄檔範例 (僅列出部份欄位)

使用者編號 姓名 職位	日期與時間	使用者需求(Request)	傳回的檔案 大小(Bytes)
X.X.X.X	[01/Jan/2000:00:00:49 +0800]	"GET /cgi-bin/sciserv.pl/collection=journals&letter=t HTTP/1.1"	15839
X.X.X.X	[01/Jan/2000:00:01:15 +0800]	"GET /cgi-bin/sciserv.pl/collection=journals&letter=u HTTP/1.1"	9414
X.X.X.X	[01/Jan/2000:00:01:17 +0800]	"GET /images/1.gif HTTP/1.1"	.
X.X.X.X	[01/Jan/2000:00:01:28 +0800]	"GET /cgi-bin/sciserv.pl/collection=journals&letter=v HTTP/1.1"	9681
X.X.X.X	[01/Jan/2000:00:01:35 +0800]	"GET /cgi-bin/search.pl/collection=journals&search_field=xml&GetSearchResults=Search&fields=Any HTTP/1.1"	88881
X.X.X.X	[01/Jan/2000:00:02:19 +0800]	"GET /cgi-bin/sciserv.pl/collection=journals&journal=09214526&issue=v6208-2911-4&article=297_xsonmac&form=pdf&file=file.pdf HTTP/1.0"	129011
X.X.X.X	[01/Jan/2000:00:02:23 +0800]	"GET /cgi-bin/sciserv.pl/collection=journals&journal=01429418&issue=v180001&article=181_qocfo&form=pdf&file=file.pdf HTTP/1.0"	948091

表二：使用記錄檔資料分類列表

類別	內容	類別	內容
AR	閱讀文章摘要	AL	瀏覽定期期刊目次
CR	閱讀版權聲明	ES	遠端檢索畫面
HM	閱讀首頁	HP	閱讀說明檔
IL	瀏覽期刊期數列表	JL	瀏覽期刊刊名列表
PF	下載 PDF 檔案	PS	下載 Postscript 檔案
QQ	其他	SS	簡易檢索畫面
SR	執行檢索	TF	下載 TIFF 頁面

- 四、資料轉換(Data transformation)：將使用記錄檔裡的資訊轉成明瞭易懂，便於後續分析工作進行的格式。例如，將記錄檔的時間格式(如 01/Sep/2000:00:01:38)轉換成電腦較容易處理的格式(2000-09-01

00:01:38)。此外，根據使用者需求(Request)的種類進行特殊轉換處理：

- (一)Request 為檢索時：根據檢索欄位(如所有欄位、作者、標題、摘要...等)將檢索策略分別儲存。



(二)Request 為下載全文時，儲存下載全文的期刊代碼及文章代碼。

(三)Request 為瀏覽時，儲存瀏覽的期刊代碼及瀏覽頁面的 URL。

經過前置處理後，使用記錄檔的大小減為 700MB，記錄檔的筆數約五百萬筆。

肆、整體分析

因為 SDOS 採用 IP 方式控管合法使用者，使用者只要是來自合法的 IP 位址，則不需要帳號及密碼，即可使用 SDOS(註 11)，所以儲存在記錄檔中的使用者資訊就僅有連線的 IP 位址。在此取

制下，本研究是將相同的 IP 位址視為同一位使用者來分析，而經過分析的結果，共有來自 30,008 個合法 IP 位址的使用者曾連線至 SDOS。

除此之外，SDOS 並無連線區間(Session)的概念。在記錄檔中只記載使用者每一次提出需求的時間，並未記錄使用者上線及離線時間。為估算每位使用者的來訪次數及停留時間，本研究根據同一 IP 位址提出的連續二需求之間隔時間來決定連線區間，若間隔時間大於 15 分鐘，則將此二需求視為屬於不同連線區間，而在計算使用者來訪次數及停留時間時，即是以 15 分鐘時間差的設定做為計算同次來訪(同一連線區間)的評定方式。

一、使用者來訪次數

表三：使用者來訪次數分析

來訪次數	IP 數	百分比	來訪次數	IP 數	百分比
1	9611	32.03%	9	593	1.98%
2	4738	15.79%	10-19	3215	10.71%
3	2705	9.01%	20-29	1291	4.30%
4	1851	6.17%	30-39	625	2.08%
5	1439	4.80%	40-49	401	1.34%
6	1029	3.43%	50-99	599	2.00%
7	905	3.02%	100+	266	0.89%
8	740	2.47%	N/A	N/A	N/A

表三為使用者來訪次數的分析圖，由此圖中可得知大部分使用者(或是 IP)連線至 SDOS 的次數都很少，將近一半的使用者只來過一次或二次，且隨著來訪次數的增加，人數呈遞減情形。這種情形的發生至少有幾種可能性：

(一)SDOS 不符合使用者的資訊需求。

(二)在記錄檔的涵蓋時間內，大多數的使用者才剛知道有 SDOS。

(三)使用者是在已經知道某一特定的，收錄在 SDOS 的文章時才會使用 SDOS。

(四)使用者是在別的數位圖書館系統(如 CSA 或 ISI)透過串連(Linking)的機制連接到 SDOS，而非刻意連線至 SDOS。

針對(一)和(二)，我們認為圖書館需要加強對讀者的教育和宣導，讓讀者知道 SDOS 的存在，並且讓讀者更清楚 SDOS 收錄的期刊、適用的領域。至於(三)和(四)，我們認為值得進一步



探討電子期刊和紙本期刊的使用者行為是否有所差異（例如有些讀者會定期造訪圖書館，閱讀最新一期的紙本期刊，這種行為是否會轉移到電子期刊，即定期連線至電子期刊系統，閱讀最新一期的電子期刊？）。

此外，亦有少數 IP 使用 SDOS 系統超過五十次，這些 IP 可能來自：

(一) 亟需 SDOS 電子期刊的研究者所擁有之電

腦；

(二) 公用電腦；

(三) 代理伺服器。

當然，前述連線區間的定義也會影響到使用者來訪次數的判定，為此，我們針對每一個 IP 向系統提出需求(Request)的次數進行統計，結果如表四。我們可以看出，表三和表四呈現類似的傾向。

表四：使用者提出需求次數分析

提出需求次數	IP 數	百分比	提出需求次數	IP 數	百分比
1—10	9947	33.15%	81—90	651	2.17%
11—20	4209	14.03%	91—190	3249	10.83%
21—30	2638	8.79%	191—290	1309	4.36%
31—40	1829	6.10%	291—390	659	2.20%
41—50	1367	4.56%	391—490	377	1.26%
51—60	1045	3.48%	491—990	716	2.39%
61—70	891	2.97%	991—100000	409	1.36%
71—80	712	2.37%	N/A	N/A	N/A

二、使用者來訪停留時間

表五：使用者來訪停留時間分析

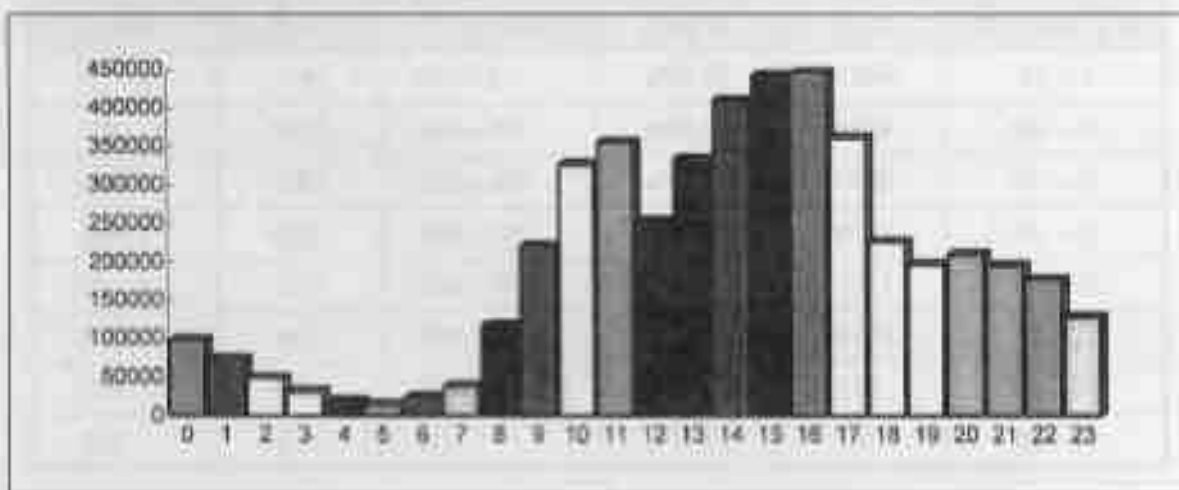
來訪停留時間	連續次數	百分比	來訪停留時間	連續次數	百分比
0—1 min	98788	30.66%	40—50 min	7904	2.45%
1—2 min	27762	8.62%	50—60 min	4928	1.53%
2—3 min	18814	5.84%	1—1.5 hr	6930	2.15%
3—4 min	14234	4.42%	1.5—2 hr	2505	0.78%
4—5 min	11729	3.64%	2—3 hr	1539	0.48%
5—10 min	41408	12.85%	3—4 hr	566	0.18%
10—20 min	47803	14.84%	4—5 hr	217	0.07%
20—30 min	23689	7.35%	5—10 hr	324	0.10%
30—40 min	12935	4.02%	10+ hr	88	0.03%

表五為使用者來訪停留時間的分析結果，在總共 322,163 次的連線使用中，有 30.66% 的停留時

間在一分鐘之內，66.03% 的停留時間在十分鐘之內，這代表使用者在短暫的時間內已得到他所需要

的資訊而滿意地離開，或是因為得不到所要的資訊而失望離去。如果使用者之所以會連線 SDOS，乃是已經知道某一特定的、收錄在 SDOS 的文章或是經由串連(Linking)機制連接到 SDOS，則自然可以假設使用者的資訊需求已經獲得滿足，然而這仍有待對使用者資訊尋求的動機與行為進一步加以探討。

三、一日內各時段及一週內各日使用情形



圖一：一日內各時段使用情形分析
(橫軸：時段，縱軸：使用記錄檔之記錄筆數)

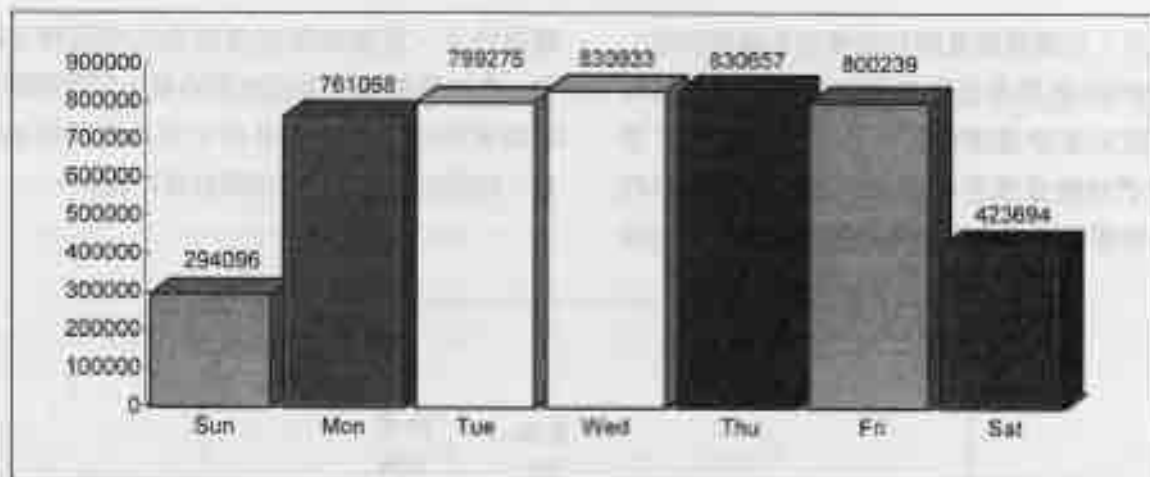
圖一為一日內各時段使用情形的分析圖，使用者利用 SDOS 的時段以上午 10 點至下午 3 點為最多，清晨 2 點至 7 點則較少人使用。圖二則為一週內每日使用情形的分析圖，週一至週五工作日的使用量幾乎不相上下，其中週一的使用情形又比一般工作日略低一些，正符合所謂憂鬱的星期一(Blue Monday)，而週末是系統使用的離峰期，週六的使用量是工作日二分之一，週日的使用量則只有工作

在表五中，有部份使用者的來訪停留時間超過三小時，甚至是十小時以上，歸究原因是由於本研究係以同一 IP 位址視為同一使用者，若使用者透過代理伺服器或公用電腦上網，那麼我們的分析就會把這些不同使用者視為同一使用者，導致可能有使用者停留時間長到超乎我們意料之外。

日的三分之一，很符合人們週末休息的心理。

了解到 SDOS 在一日內各時段、一週內每日的使用狀況後，可訂出 SDOS 系統的尖峰期與離峰期；在尖峰期間系統管理者必須多留意系統的穩定度，將系統效能保持在最佳狀態，而在離峰時間，則適合運行系統測試、備份、維護等相關作業；對使用者而言，為了能夠享有快速的系統連線，使用品質，也可以試著在離峰時間使用 SDOS。





圖二：一週內每日使用情形
(橫軸：一週內每日；縱軸：使用記錄檔之記錄筆數)

四、使用記錄檔概觀分析

我們將表二中的使用者記錄檔資料分類列表歸納為檔案取用、瀏覽、檢索和其他等四大類。其中，檔案取用包含閱讀文章摘要(AK)、閱讀首頁(HM)、閱讀版權聲明(CR)、閱讀說明檔(HP)、下載 PDF 檔案(PF)、下載 Postscript 檔案(PS)、下載 TIFF 頁面(TF)；瀏覽的範圍包含瀏覽期刊刊名列表(JL)、瀏覽期刊期數列表(IL)、瀏覽單期刊目次(AL)；檢索的範圍包含進入簡易檢索(SS)與進階檢索(ES)的畫面及執行檢索(SR)；而其他(QQ)則是不在上述範圍的使用記錄。圖三為這四大類的使用記錄分析。

由圖三得知，使用者在 SDOS 系統下的的行為以檔案取用為最多，佔了 50%；瀏覽佔了 34%，檢索的部分只佔了 16%。這樣的情形是合理的，因為使用者進入電子期刊系統的目的就是為了要取得所需的期刊文章，自然會花大半的時間在檔案取用上，諸如：閱讀文章摘要檢視是否符合需求，線上瀏覽或下載 PDF、Postscript 檔。

若使用者以取得所需期刊文章之摘要或全文檔案為其進入 SDOS 系統的最終目的，則有二種途徑可以達到此目的。第一種途徑是從 SDOS 首頁的期刊領域列表(Category list of journals)或期刊字母列表(Alphabetical list of journals)逐層瀏覽至所需文章之摘要或全文檔案；第二種途徑則是由 SDOS 首頁的快速檢索(Quick search)或經由簡易檢索、進階檢索找到所需文章之摘要或全文檔案。而圖三中瀏覽(即第一種途徑)記錄是檢索記錄(即第二種途徑)的 2 倍，似乎使用者比較偏好以瀏覽的方式取得資訊，然而經過仔細地檢視記錄檔，我們認為利用瀏覽和檢索途徑來取得所需期刊文章的比例是差不多的，原因如下：透過瀏覽方式取得文章摘要或全文時，每一次瀏覽會留下三筆記錄，包括：瀏覽期刊刊名列表(JL)、瀏覽期刊期數列表(IL)、瀏覽單期刊目次(AL)；然而檢索只會留下一至二次記錄(若直接在首頁執行快速檢索，只會留下執行檢索(SR)的記錄；若是進入簡易或進階檢索畫面後才執行檢索，則會留下進入簡易檢索(SS)或進階檢索(ES)的記錄，以及執行檢索(SR)的記錄)。



此外，在選擇瀏覽期刊領域或字母列表時，約有 76% 的使用者會選擇瀏覽字母列表，只有 24% 的使用者會選擇瀏覽期刊領域。分析其原因，我們推斷當使用者以瀏覽的方式尋找期刊時，通常都是已經確定要找的期刊刊名了，在這

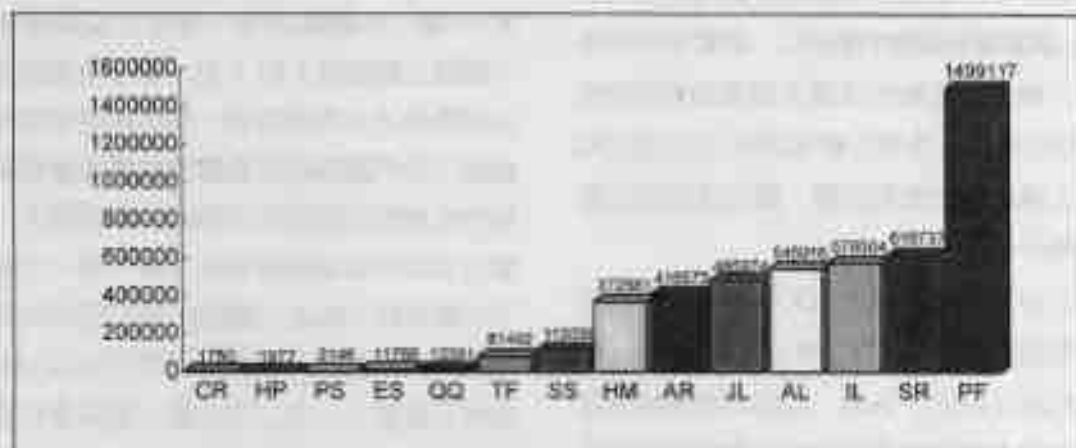
種情形下，直接按照字母列表去尋找特定的期刊，會比用領域的方式來得容易，而按照期刊領域瀏覽的話，一般都是爲了要找相關領域的期刊，相對來看，這樣的情形就會比較少。



圖三：使用記錄概觀分析

進一步檢視表列出的十四種記錄檔資料分類的情形，其結果如圖四所示，下載 PDF 檔案(PF)

約佔了記錄檔中所有記錄的 31%，由於取得文章全文是使用者進入電子期刊系統的目的，所以直覺



圖四：記錄檔分類分析
(橫軸：記錄檔需求分類；縱軸：記錄筆數)



來看，下載全文檔案會佔記錄檔中極大的比例，而圖四也證實了這一點，從圖四還可以發現，閱讀文章摘要(AR)的記錄比瀏覽類(包括 JL、IL、AL)、執行檢索(SR)、下載 PDF 檔案(PF)的記錄都來得少，亦即許多 SDOS 的使用者未曾檢視文章摘要就直接下載全文了，究竟這樣的行為模式代表什麼意義(註 12)，值得深入探究！此外，圖四裡期刊文章的比例是差不多的(SR 的記錄有 616,737 筆，而 JL、IL、AL 的平均約 536,798 筆)。

在早期的 SDOS 版本中提供 PDF 和 Postscript 二種格式的全文檔案，但據 Elsevier 指出使用者比較偏好下載 PDF 格式的檔案，下載 Postscript 檔案的比例不到 1%，因此在 SDOS 較新的版本已不再提供 Postscript 格式的全文檔案了，這是為什麼圖四中下載 Postscript 檔案的記錄(PS)十分稀少。

由圖四也可看出，在記錄檔中只有 1,977 筆閱讀說明檔的記錄 (HP)，若以先前 30,008 個 IP 連線來計算，大約每 15 個 IP 中只有一個會去閱讀說明檔，此一行為傾向告訴我們幾件事：

- (一)圖書館在實施訓練課程時，必須強調並讓使用者了解各系統的檢索方法，以免導致使用者因不了解檢索方法而找不到所要的資料；
- (二)電子期刊系統的界面設計及檢索功能應該要直覺、符合常人的使用習慣，讓使用者不需要研讀太多系統說明，即可正確有效地使用系統；
- (三)電子期刊系統可利用“Today's Tip” (今日小祕訣)等管道，主動提醒或教育使用者系統

功能。

閱讀版權聲明的記錄(CR)就更少了！平均每 17 個 IP 中只有一個會去閱讀版權聲明。由於台灣使用者不時會有不當使用(Misuse)電子期刊的狀況發生，在使用者尚未養成閱讀並遵守版權聲明的習慣前，圖書館必得加強宣導合理使用電子期刊的重要性！電子期刊系統亦可採用主動式機制提醒使用者相關版權聲明。

伍、檢索行為分析

一、使用者每次來訪檢索次數 (Number of queries per session)

SDOS 的連續使用次數共計 322,163 次，其中僅 113,964 次(35%)連續有執行檢索，針對這 113,964 次的連續進行分析，每次來訪檢索次數如表六。SDOS 的使用者每次來訪檢索次數平均為 3.73 次，這結果與 Jansen-2000 針對 Excite 所統計出來的 1.6 次，Jones-2000 針對 CSTR 所統計出來的 2.04 次相去甚遠。(註 13)

表七則是比較本研究、Jansen-2000、Jones-2000 的每次來訪檢索次數，相對於 SDOS 和 CSTR 的使用者檢索超過三次以上者分別佔 30.87%和 22%來看，Excite 檢索超過三次以上者僅佔 7%，且有 67%僅檢索一次，顯示出網絡搜尋引擎使用者的每次來訪檢索次數偏低，這一點和電子期刊系統的使用者有很大的不同，之所以如此，我們認為資訊尋求的動機與目的會是主要的影響因素。



表六：使用者每次來訪檢索次數分析

每次來訪檢索次數	總筆區間數	百分比	每次來訪檢索次數	總筆區間數	百分比
1	42365	37.17%	9	1694	1.49%
2	22248	19.52%	10—19	6228	5.47%
3	14174	12.44%	20—29	1138	1.00%
4	9174	8.05%	30—39	388	0.34%
5	6225	5.46%	40—49	165	0.14%
6	4396	3.86%	50—99	149	0.13%
7	3188	2.80%	100—499	27	0.02%
8	2402	2.11%	500+	3	0.00%

表七：比較本研究、Jansen-2000、Jones-2000 的每次來訪檢索次數

每次來訪檢索次數	一次	二次	三次	三次以上
Excite (Jansen-2000)	67%	19%	7%	7%
CSTR (Jones-2000) (註 14)	44%	22%	12%	22%
SDOS (本研究)	37.17%	19.52%	12.44%	30.87%

二、檢索策略長度分析 (Length of query strategies)

所謂「檢索策略長度」係指在一個檢索策略(或簡稱檢索)中所包含的檢索詞彙個數。SDOS 系統下的檢索策略長度分布如表八所示,平均檢索策略長度為 2.27,此一結果與 Jones-2000 針對 CSTR 所分析的 2.43、Jansen-2000 針對 Excite 所分析的 2.21 都很類似(註 13),然而卻與卜小蝶-89 針對 Dreamer 與 GAIS 完全英文之檢索策略平均長度 1.10、1.22 相去甚遠。由這幾個研究結果來看,有幾項議題值得深入探討:

(一)使用者在運用 WWW-Based 網站搜尋引擎以及電子期刊系統時,是否在檢索策略長度上會有類似的傾向(如 Jansen-2000、Jones-2000、本研究有相似的結果)?

(二)東西方使用者在使用 WWW-Based 電子期刊系統時,是否在檢索策略長度上會有類似的

傾向(如 Jansen-2000、本研究有相似的結果)?

(三)東方使用者在使用 WWW-Based 網站搜尋引擎以及電子期刊系統時,在以英文進行檢索時,是否在檢索策略長度上會有不同的傾向(如本研究和卜小蝶-89 有不同的結果)?

從每次來訪檢索次數和檢索策略長度分析的結果反映出大多數使用者期望在少數幾次的簡短檢索中就能找到想要的資訊,然而,一般而言,檢索策略愈短,所包含的範圍就愈廣,檢索結果就愈多,而使用者要在大量的檢索結果中找到真正需要的資料並非易事。為了讓使用者更有效地運用數位圖書館系統,必須從二方面下手:

(一)圖書館應加強教育使用者資訊檢索的技巧(如精簡檢索策略),以較多次或較精準的檢索策略清楚地表達資訊需求;

(二)feedback、檢索擴充(Query expansion)、資料探勘(Data mining)等方法了解使用者的



與前。

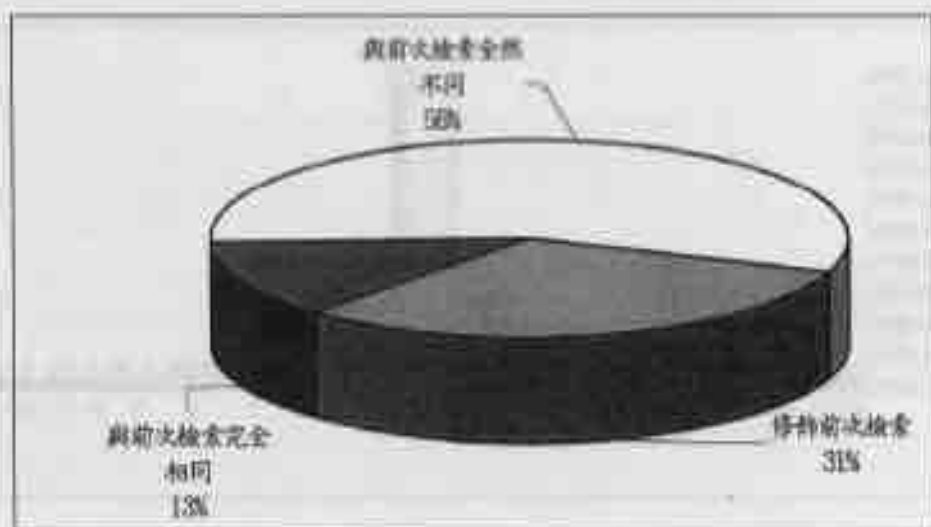
表八：檢索策略長度分析

檢索策略長度	檢索次數	百分比	檢索策略長度	檢索次數	百分比
0	596	0.14%	7	3979	0.94%
1	182754	42.97%	8	2102	0.49%
2	104095	24.48%	9	1625	0.38%
3	75268	17.70%	10	1169	0.27%
4	27900	6.36%	11—15	2589	0.61%
5	16399	3.86%	16—20	602	0.14%
6	6086	1.43%	20+	122	0.03%

三、檢索策略修飾(Query refinement)情形分析

由於任何一個資訊檢索系統的精確率(Precision)和回現率(Recall)絕不可能為 100%，在許多情形下，使用者很難單用一個檢索策略就找到

全部想要的資訊，一個有經驗的使用者應該要能根據檢索結果，修正其檢索策略，因此在本研究中擬分析 SDOS 使用者修飾檢索策略的情形，以期明瞭 SDOS 使用者是否具有檢索策略修飾的概念。



圖五：檢索策略修飾分析

如圖五所示，在 SDOS 中，56%的檢索策略是和前一次檢索策略全然不同（即與前一次檢索策略的檢索詞彙皆不相同），13%的檢索策略是和前一次檢索策略完全相同，而 31%的檢索策略

是修飾前一次檢索策略得來的（至少有一檢索詞彙與前一次檢索策略是相同的），此三種檢索策略在 Jansen-2000 中依序稱為 “Unique query”、“Identical query”、“Modified query”。由於我們無

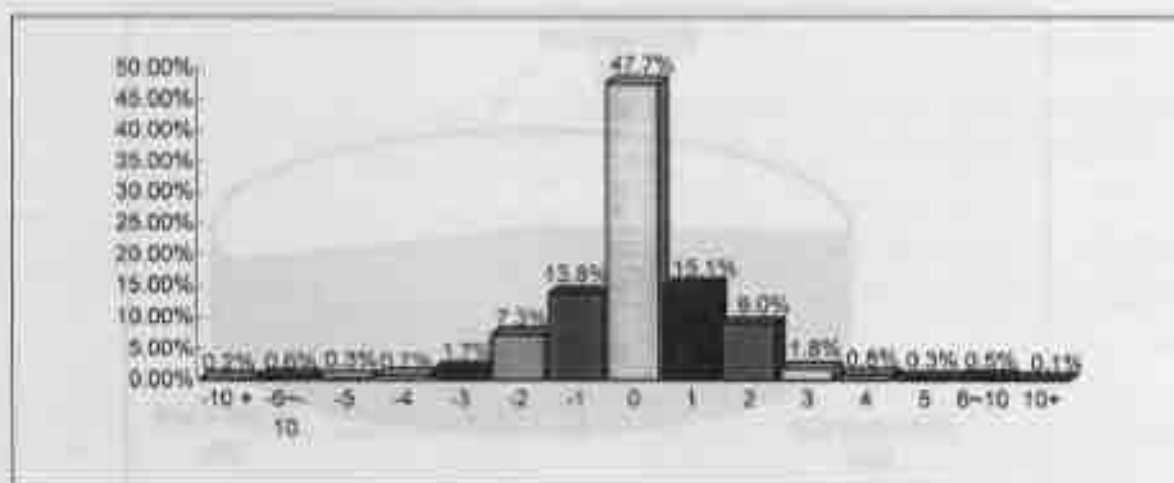


法判別產生 Identical query 的原因為何(例如因網路遲滯導致使用者不耐煩而重複按鍵)，因此以下的分析只針對 Unique query 和 Modified query。

假設 Unique query 的出現代表使用者的資訊需求有所變化，而 Modified query 的出現代表使用者根據檢索結果調整其檢索策略，則從 Modified query 佔所有檢索策略的比例(只考量 Unique query 和 Modified query)可大約估計使用者修飾檢索策略的傾向，比例愈高代表使用者愈有修飾檢索策略的傾向。從圖五來看，SDOS 的此一比例約為 36% (31%/87%)，與 Jansen-2000 估算的 38% 相近，但與 Jones-2000 估算的 66% 相比則偏低。在 Jansen-2000 中曾指出，在傳統資訊檢索系統中，修飾檢索策略是一個蠻普遍的行爲(亦即前述比例會比 36%、38% 來得高)，推斷之所以會如此，可能是因為傳統資訊檢索系統的

使用者多為經過專業訓練的檢索者，而 SDOS (WWW-based 數位圖書館系統)和網站搜尋引擎則多為未經過專業訓練的檢索者。此外，從前述說明可知 SDOS 使用者修飾檢索策略的傾向並不明顯，因此圖書館在進行教育訓練時可加強宣導修飾檢索策略的重要性。

進一步針對 Modified query 探討使用者如何修正其檢索策略，從圖六很明顯可以看出約 47.7% 的修正檢索是在不改變字數的情況下修正的，也就是說使用者只修改檢索策略的內容，包含了改變檢索詞彙的詞性型態變化、改變檢索詞彙等等，並未增減檢索詞彙的數目；其次才是增或減一個檢索詞彙，且比例隨著檢索詞彙數目變化量的增大而逐漸減少。此外，也可以看出增加或減少檢索詞彙趨勢是大致相同的。



圖六：Modified query 檢索詞彙長度改變情形

(橫軸：檢索詞彙增減字數；縱軸：百分比)

四、簡易檢索 VS. 進階檢索

SDOS 的檢索功能有簡易檢索與進階檢索二種，簡易檢索僅是讓使用者針對文章篇名(Article title)、摘要(abstract)、作者(Author's name)、



ISSN、PII(Publisher item identifier)、作者關鍵字(Author key words)、刊名(Journal title)任一或所有欄位輸入檢索策略進行檢索；而進階檢索則是最多有二組針對上述任一或所有欄位的檢索策略，二組檢索策略間還可選擇以 AND、OR、或 NOT 關連，此外，還可以限制要檢索的期刊領域(Journal categories)、文章類型(Article type)、語言(Language)、時間範圍(Limit dates)，並能夠選擇檢索結果的排列方式(Sort documents by)及每頁輸出筆數/Documents per page)。

在所有使用者下的檢索之中，有 91% 是採用簡易檢索，9% 採用進階檢索，這顯示儘管系統已經提供功能較為強大的進階檢索，使用者仍偏好使用直覺、簡單、好用的簡易檢索方式；然而，由於簡易檢索的功能較為簡單，無法讓檢索者有效地過濾掉無關的資訊，因此對一些進階的檢索者而言功能較為不足，所以進階檢索仍有其存在

的價值。我們認為提供多層次的檢索界面會是 WWW-based 電子期刊系統的趨勢，目前許多數位圖書館系統的檢索功能至少提供了簡易和進階兩種檢索方式讓初學者(Novice user)和進階者(Advanced user)使用，甚至有的系統還提供更進一步讓資訊檢索專家(Expert)使用的檢索方式(註 16)。

五、檢索欄位分析

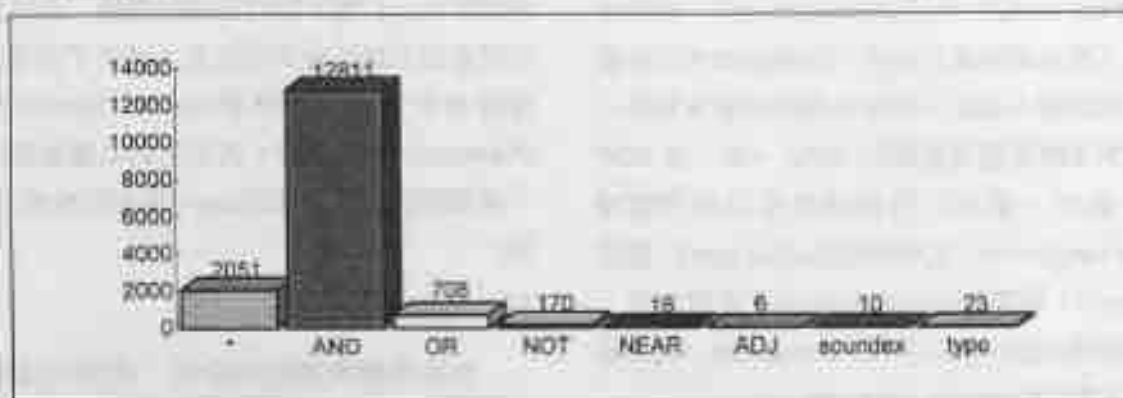
表九為檢索欄位的分析，在簡易檢索中有 84.4% 的檢索是採用系統預設的任何欄位檢索(Any)，而此一比例在進階檢索中降至 50.7%；相對地，以文章篇名(Article title)、摘要(Abstract)、作者(Author's name)為檢索點的比例在進階檢索則較簡易檢索為高。此一現象表示出進階檢索的使用者確有比較清楚的資訊尋求目標，或者懂得利用進階檢索快速有效地過濾掉無關的資訊。

表九：檢索欄位分析

欄位	簡易檢索		進階檢索	
	筆數	百分比	筆數	百分比
Any	197983	84.4%	104847	50.7%
Article title	13021	5.6%	34245	16.6%
Abstract	4711	2.0%	22004	10.6%
Author's name	16455	7.0%	28779	13.9%
Citation & abstract	2298	1.0%	NA	NA
Keyword	NA	NA	2567	1.2%
ISSN	NA	NA	0	0.0%
PII	NA	NA	192	0.1%
Text	NA	NA	662	0.3%
Journal title	NA	NA	7956	3.9%
Category	NA	NA	5447	2.7%



六、特殊功能字元使用情形



圖七：特殊功能字元使用情形
(橫軸：特殊功能字元；縱軸：記錄筆數)

SDOS 系統提供了一些特殊功能字元，用以選取多個檢索詞彙，包含：切截或萬用字元(*)、且(AND)、或(OR)、否(NOT)、鄰近(NEAR)、相鄰(ADJ)、音似(SOUNDEX)、拼字錯誤更正(TYPO)等，讓使用者得以更有彈性地將其資訊需求表達出來。從圖七的特殊功能字元使用分析情形來看，較常用的只有萬用字元(*)、且(AND)、或(OR)，推測是因為這三個特殊功能字元在網站搜尋引擎或其他數位圖書館系統中也常被拿來作特

殊功能字元使用，使用者較為熟悉且容易上手。而其他的特殊功能字元，一方面可能是因為使用者不熟悉，另一方面使用者也沒有看說明檔的習慣，因此才會較少被運用到。有鑑於運用特殊功能字元能夠更清楚地表達使用者的資訊需求或是減少輸入錯誤的產生，若圖書館在進行讀者訓練時針對特殊功能字元的使用加以說明，相信能夠讓讀者更靈活且有效地使用 SDOS 或其他數位圖書館系統。

七、檢索詞彙頻率分布分析

表十：檢索詞彙使用情形分析

使用頻率	檢索詞彙個數	百分比	使用頻率	檢索詞彙個數	百分比
1	40237	49.67%	11—20	3790	4.68%
2	13117	16.19%	21—50	2947	3.64%
3	6199	7.65%	51—100	1184	1.46%
4—5	6267	7.74%	101—200	706	0.87%
6—10	5944	7.34%	200+	623	0.77%



在 SDOS 使用記錄檔中記錄了 81,014 個不同的檢索詞彙(Unique query term)，這些檢索詞彙的出現總頻率為 887,309 次。亦即有些檢索詞彙是重複出現的。在這 81,014 個檢索詞彙中，有 40,237 個(佔 49.67%)只出現過一次。相對地，僅有 623 個(佔 0.77%)是出現過 200 次以上的(參照表十)。表十一為最常出現的前五十名檢索詞彙。這五十個檢索詞彙出現的總頻率為 159,192 次，為所有檢索詞彙出現總次數的 17.9%。

進一步分析檢索詞彙與頻率的關係，圖八說明了出現頻率排名前 1000 名的檢索詞彙約佔總頻率的 0.016%，第 3000 名約佔 0.00494%，第

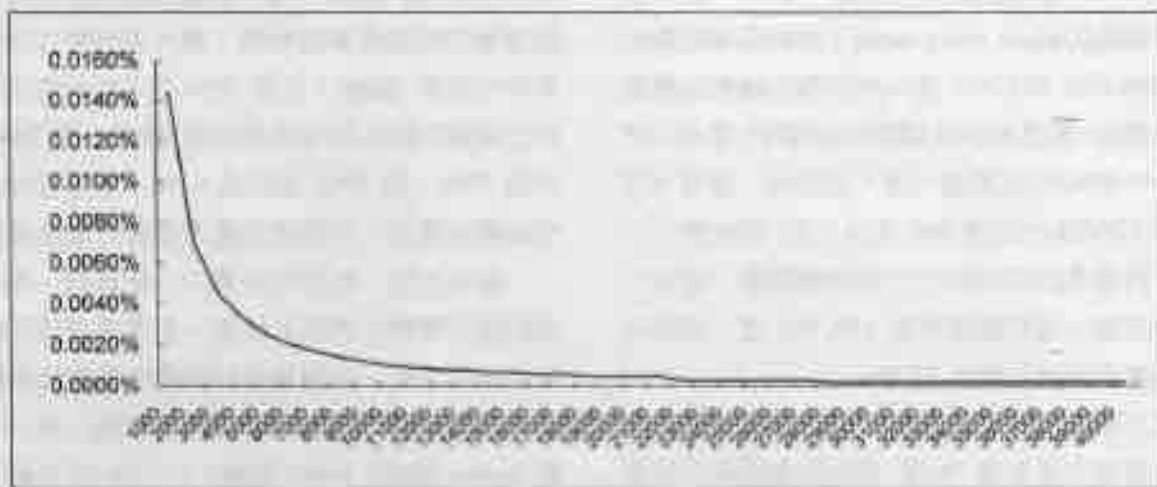
5000 名約佔 0.00272%，而隨著出現頻率遞減，則遞減的速度愈來愈緩慢。圖九則是將出現頻率累計的結果，說明了在前 1000 名中的檢索詞彙累計出現頻率約佔所有檢索詞彙 38%，前 3000 名約佔 59%，前 5000 名約佔 61%，隨著檢索詞彙出現頻率遞減，則遞增的速度也會愈來愈緩慢。

綜合而言，我們可以看出一個傾向，亦即高頻詞彙的筆數比例雖然較低，但其累計出現的頻率卻十分之高。相對地低頻詞彙的筆數比例雖然較高，但其累計出現的頻率卻非常低。此一傾向與 Jansen-2000、Jones-2000、卜小樸-89 等研究都十分類似。

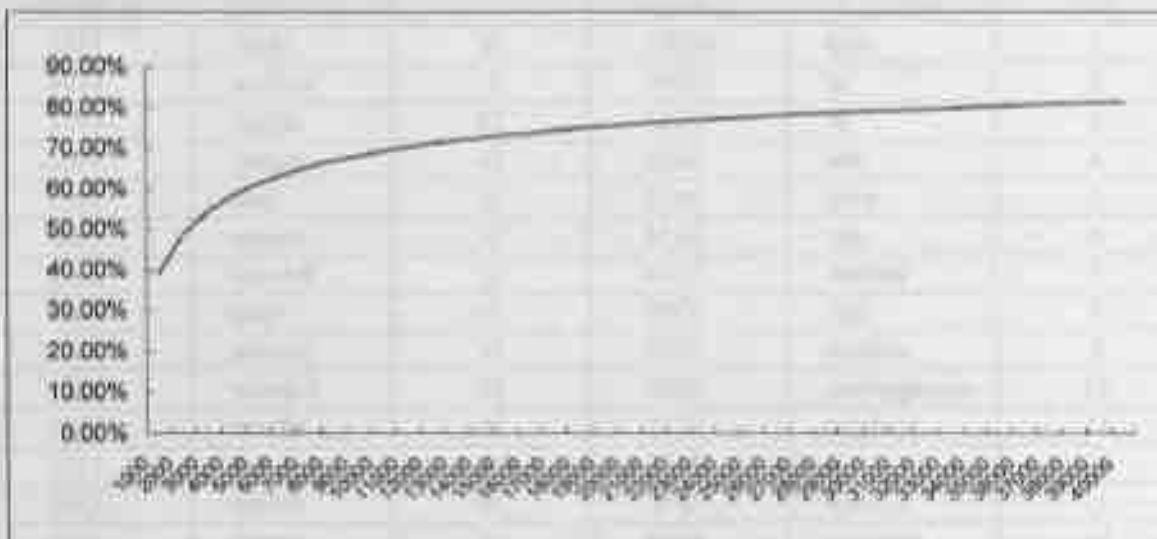
表十一：最常出現的前 50 名檢索詞彙，粗斜體字是停用字(Stopword)

名次	檢索詞彙	出現頻率	名次	檢索詞彙	出現頻率
1	<i>and</i>	62203	26	chain	1556
2	<i>of</i>	13707	27	surface	1496
3	<i>in</i>	4655	28	design	1418
4	<i>the</i>	3658	29	supply	1402
5	acid	3422	30	heat	1395
6	<i>for</i>	2854	31	copper	1366
7	review	2760	32	thermal	1362
8	cell	2388	33	data	1322
9	carbon	2355	34	plasma	1313
10	management	2289	35	transfer	1269
11	analysis	2240	36	silicon	1262
12	<i>a</i>	2152	37	network	1235
13	control	2013	38	sensor	1205
14	polymer	1979	39	neural	1198
15	protein	1927	40	DNA	1191
16	system	1893	41	stress	1186
17	fuzzy	1831	42	phase	1174
18	film	1800	43	diamond	1166
19	<i>&</i>	1727	44	thin	1146
20	model	1725	45	plant	1144
21	<i>on</i>	1708	46	flow	1135
22	<i>or</i>	1699	47	information	1114
23	water	1653	48	membrane	1104
24	oxide	1616	49	process	1095
25	metal	1574	50	GaN	1090





圖八：最常出現的檢索詞彙與出現頻率分析
(橫軸：第 N 個最常出現的檢索詞彙；縱軸：出現頻率百分比)



圖九：最常出現的檢索詞彙與出現頻率累計分析
(橫軸：前 N 個最常出現的檢索詞彙；縱軸：出現頻率累計百分比)

另一點值得注意的是，表十一中顯示出最常被使用的前四名檢索詞彙皆是停用字 (Stopword)，儘管停用字並不至於影響檢索結果的正確性，也不會對系統產生太大的處理負擔，然而若使用者具備停用字的觀念，在擬定檢索策略

時可專注於資訊需求的主題，也更容易擬定符合資訊需求的檢索策略。此外，“or”、“and”、“&”等檢索詞常被網站檢索引擎或其他數位圖書館系統視為特殊功能字元，如：“or”等同於 SDOS 的“OR”，“and”及“&”等同於 SDOS 的“AND”，這表



示使用者有可能把在其他系統的使用習慣帶過來，以致於會有誤用的情形發生。甚至會導致檢索結果和使用者原先的資訊需求差異極大。(註 17)

陸、其他分析

一、全文檔案下載分析

在先前曾經提及，全文檔案的下載(包含 PDF 與 Postscript 檔案)佔記錄檔中極大的比例。進一步分析全文檔案的下載情形，如表十二所示。在 SDOS 系統中總計約 1,000,000 篇的全文檔案中，約有 350,000 篇全文檔案曾被下載過，全文檔案

的下載總次數為 1,503,418 次。而在曾被下載過的全文檔案中，49.53% 只被下載過一次，90.25% 只被下載過五次以內。

SDOS 系統管理者可以根據此一分析來調整系統儲存空間，將儲存空間、價格、效能等因素最佳化。例如採用階層式儲存設備管理(Hierarchical Storage Management, HSM)的方式將不曾或較少被下載的全文檔案放在較慢的儲存設備(如磁帶櫃)，較常被下載的全文檔案放在較快的儲存設備(如 Fiber-Channel 磁碟機或磁碟陣列)。HSM 系統的磁碟緩存區(Cache)亦可以根據全文檔案下載的頻率趨勢來做最佳化的調配。

表十二：全文檔案下載情形分析

文章被下載次數	篇數	百分比	文章被下載次數	篇數	百分比
100+	19	0.01%	11	1667	0.48%
51—100	117	0.03%	10	2260	0.65%
21—50	1595	0.46%	9	3144	0.90%
20	255	0.07%	8	4153	1.19%
19	269	0.08%	7	6219	1.78%
18	357	0.10%	6	9177	2.63%
17	409	0.12%	5	13326	3.82%
16	533	0.15%	4	22314	6.40%
15	665	0.19%	3	40249	11.54%
14	806	0.23%	2	66121	18.96%
13	1055	0.30%	1	172744	49.53%
12	1310	0.38%	NA	NA	NA

表十三為使用者下載全文檔案的數量分布情形，在 30,008 位使用者中有 22,282 位曾經下載過全文(比例為 74%)。綜合而言，下載五篇或以下的使用者人數佔曾下載過全文的使用者之 38.40%，下載六至十篇者佔 14.40%，十一至五十篇者佔

29.89%，五十一至一百篇者佔 7.33%，一百篇以上者佔 8.91%。從曾經下載過全文的使用者比例以及下載數量的統計來看，我們可以說 SDOS 還蠻符合使用者需要的。



表十三：使用者下載全文檔案的數量分布

下載全文檔案數量	使用者人數	百分比	下載全文檔案數量	使用者人數	百分比
1	3078	13.81%	21—30	1652	7.41%
2	1909	8.57%	31—40	1082	4.86%
3	1563	7.01%	41—50	798	3.58%
4	1122	5.04%	51—60	580	2.60%
5	885	3.97%	61—70	446	2.00%
6	916	4.11%	71—80	347	1.56%
7	667	2.99%	81—90	261	1.17%
8	586	2.63%	91—100	236	1.06%
9	569	2.55%	101—200	1069	4.80%
10	471	2.11%	201—300	367	1.65%
11—20	3129	14.04%	>300	549	2.46%

如前所述，SDOS 系統內共有 Elsevier 出版的一千三百餘種期刊，然而絕大多數的 SDOS 訂戶都是只有訂購其中的部份期刊。在 2000 年，Elsevier 對古體的訂戶採取開放政策，讓每一個訂戶瀏覽及下載 SDOS 內的所有電子期刊（不論訂購與否）。在本研究中也探討了 SDOS 訂戶在下載訂購與未訂購之期刊全文的狀況。在 2000 年 1 月至 7 月間，使用者下載 SDOS 全文檔案的總次數為 1,149,974 次，其中下載非訂購期刊全文的總次數為 735,810 次，佔總下載次數的 64%。進一步分析台灣在 2000 年時五十一個 SDOS 訂戶的使用者下載非訂購與訂購期刊全文的狀況，我們以方程式一計算每個訂戶的使用者下載非訂購期刊之全文檔案次數的比例，在五十一個訂戶中，最小的比例為 26.15%，最大的比例為 99.18%，圖為這五十一個訂戶的非訂購期刊全文下載比例分布圖。從圖十可知，下載

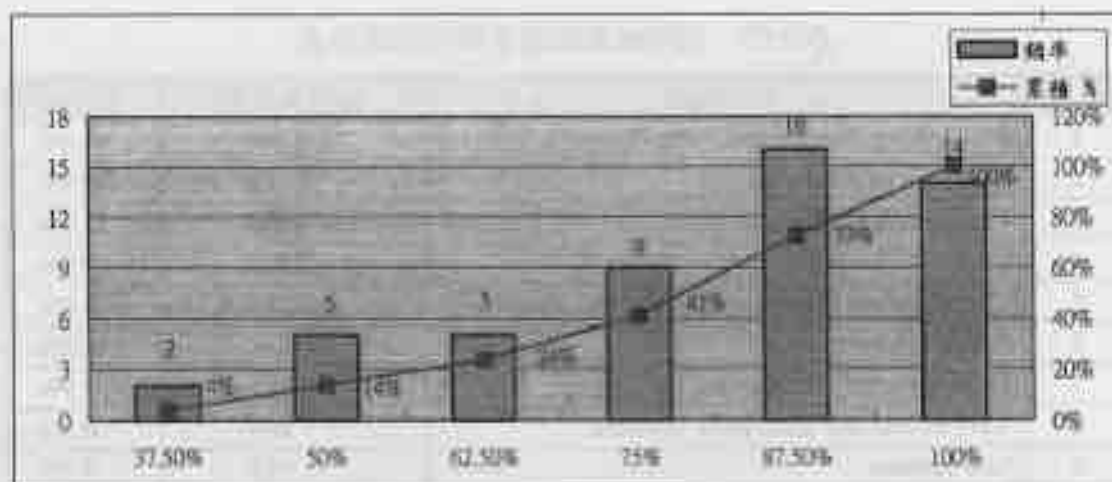
非訂購期刊全文檔案的比例高於 62.3% 的訂戶居然多達三十九部。根據此一統計，有二點值得進一步探討：

- (一) 訂戶所訂購的期刊是否真正合乎其使用者的需求，是否需將本統計的數據納入各訂戶的期刊館藏增刪政策中？
- (二) 全文期刊系統廠商是否應體認此一統計呈現的現象，檢討收費政策？事實上，Elsevier 已經針對前述現象調整其 2001 年的收費政策，期望訂戶在少量的額外花費下能取用更多資訊。

方程式一

$$\text{Ratio} = \frac{\text{非訂購期刊全文檔案下載次數}}{\text{所有期刊全文檔案下載次數}}$$





圖十：SDOS 訂戶非訂購期刊全文下載比例分布圖
(橫軸：方程式一的比例值分界)

解讀方式舉例如后：下載非訂購期刊全文檔案的比例介於 62.5% 和 75% 之間的訂戶共有九個；下載非訂購期刊全文檔案的比例小於 75% 的訂戶佔所有訂戶的 41%

二、期刊領域選取排行

在進階檢索時，使用者可依需要選擇欲檢索的期刊領域，而瀏覽時也可選擇用期刊領域來瀏覽，表十四與表十五分別為檢索與瀏覽時被選取的期刊領域排行。在進階檢索時最常被選擇的期刊領域前三名分別為工程/能源/技術類 (Engineering, Energy and Technology)、材料科學類 (Materials Science)、物理/天文學 (Physics and

Astronomy) 類，而最常被瀏覽的期刊領域則為化學/化工類 (Chemistry and Chemical Engineering)、材料科學類 (Materials Science)、生命科學類 (Life Science)，唯一有重疊的僅是材料科學類，這說明了不同領域期刊的使用者所習慣使用的方法是不大相同的，若單就進階檢索和瀏覽來比較，則使用者在瀏覽時比在進階檢索時選擇期刊領域的次數來得高。



表十四：進階檢索時選取期刊領域排名

名次	期刊領域	被選取次數	百分比
1	Engineering, Energy and Technology	1788	14.4%
2	Materials Science	1482	11.9%
3	Physics and Astronomy	1386	11.1%
4	Computer Science	1351	10.9%
5	Chemistry and Chemical Engineering	1344	10.8%
6	Social Sciences	1154	9.3%
7	Environmental Science and Technology	983	7.9%
8	Agricultural and Biological Sciences	803	6.5%
9	Life Science	710	5.7%
10	Mathematics	607	4.9%
11	Clinical Medicine	579	4.7%
12	Earth and Planetary Science	255	2.0%
	總計	12442	100%

表十五：經常被瀏覽的期刊領域排名

名次	期刊領域	被瀏覽次數	百分比
1	Chemistry and Chemical Engineering	13767	21.0%
2	Materials Science	9047	13.8%
3	Life Science	6881	10.5%
4	Engineering, Energy and Technology	5820	8.9%
5	Environmental Science and Technology	5752	8.8%
6	Computer Science	5648	8.6%
7	Social Sciences	4955	7.6%
8	Physics and Astronomy	3452	5.3%
9	Clinical Medicine	2802	4.3%
10	Agricultural and Biological Sciences	2739	4.2%
11	Earth and Planetary Science	2661	4.1%
12	Mathematics	1921	2.9%
	總計	65445	100%

三、熱門期刊排行

表十六與表十七各是檔案下載、瀏覽的熱門期

刊排行，斜體字的部分則是二級排行中有重複出現的期刊。



表十六：最常被下載檔案的前二十名期刊

名次	期刊名	下載次數
1	<i>Synthetic Metals</i>	28424
2	<i>Thin Solid Films</i>	26574
3	<i>Tetrahedron Letters</i>	23658
4	<i>Sensors and Actuators A: Physical</i>	21576
5	<i>Journal of Power Sources</i>	17867
6	<i>Atmospheric Environment</i>	16332
7	<i>Journal of Chromatography A</i>	15549
8	<i>Polymer</i>	15407
9	<i>Analytica Chimica Acta</i>	13845
10	<i>Journal of Crystal Growth</i>	12389
11	European Journal of Operational Research	11695
12	Surface and Coatings Technology	11198
13	<i>Fuzzy Sets and Systems</i>	11171
14	Materials Science and Engineering: A	11038
15	Journal of Materials Processing Technology	10561
16	Sensors and Actuators B: Chemical	10233
17	Journal of Alloys and Compounds	9749
18	Fuel and Energy Abstracts	9595
19	<i>Electrochimica Acta</i>	9488
20	<i>Tetrahedron</i>	9435

表十七：最常被瀏覽的前二十名期刊

名次	期刊名	瀏覽次數
1	<i>Tetrahedron Letters</i>	22322
2	<i>Journal of Chromatography A</i>	11530
3	<i>Atmospheric Environment</i>	10975
4	<i>Polymer</i>	9267
5	<i>Thin Solid Films</i>	9065
6	<i>Synthetic Metals</i>	8488
7	FEBS Letters	7697
8	Trends in Biochemical Sciences	7567
9	Trends in Biotechnology	6892
10	<i>Tetrahedron</i>	6878
11	<i>Journal of Power Sources</i>	6874
12	Biomaterials	6227
13	<i>Journal of Crystal Growth</i>	6198
14	International Journal of Heat and Mass	6106
15	<i>Analytica Chimica Acta</i>	6053
16	<i>Electrochimica Acta</i>	5988
17	Journal of Membrane Science	5705
18	<i>Fuzzy Sets and Systems</i>	5657
19	<i>Sensors and Actuators A: Physical</i>	5524
20	Trends in Cell Biology	5512



柒、結論與建議

隨著數位圖書館時代的來臨，各類電子資料逐漸普及，WWW 版電子期刊、全文資料庫、索摘資料庫、電子書已經成為教學及研究時重要的資訊來源。由於此類數位圖書館系統的使用者大多未接受過圖書資訊學的專業訓練，若對他們的系統使用及檢索行為有深入的了解，可做為數位圖書館系統的研發者在改進檢索功能、使用者界面...時的重要依據，圖書館在進行教育宣導時也能針對使用者最常發生的錯誤以及不熟悉的功能加強訓練。

本研究係藉由分析 Elsevier SDOS 的使用記錄檔，初步獲得國內 SDOS 使用者的使用習慣和資訊尋求行為，所得成果或許亦能應用在其他類似的數位圖書館系統上。綜而言之，使用者多以工作日及上班時間為主來使用系統，且停留時間很短，使用者多偏好使用簡易檢索，並以簡短的檢索策略尋找資料，檢索點常設定為任何欄位(Any)，檢索策略修飾的情況約佔所有檢索的三分之一，而修飾檢索策略時，每次修正的詞彙仍是以小幅度的改變居多。

根據本研究的發現，茲對圖書館、系統研發者與出版商、系統管理者提出以下與電子期刊系統相關的建議：

一、圖書館

- (一)針對每個電子期刊系統獨特的檢索與瀏覽功能多加宣導，彌補使用者不常閱讀使用檔的行為傾向。
- (二)針對電子期刊系統的合理使用多加宣導，彌補使用者不常閱讀版權聲明的行為傾向。
- (三)針對檢索策略修飾的概念多加宣導。
- (四)由下載全文的情況檢討期刊館藏之增購政策。
- (五)針對每個電子期刊系統特定的特殊功能字

元多加宣導。

二、全文電子期刊系統研發者與出版商

- (一)利用主動式機制將系統特色與功能告知使用者，彌補使用者不常閱讀使用檔的行為傾向。
- (二)利用主動式機制將相關版權聲明告知使用者，彌補使用者不常閱讀版權聲明的行為傾向。
- (三)大多數使用者較習慣使用直覺、簡單的檢索功能，但也應提供多層次的檢索功能，適量考慮初學、進階、專家使用者的檢索需求。
- (四)可由下載非訂購期刊全文的趨勢檢討收費政策。
- (五)利用相關性回饋(Relevance feedback)、檢索擴充(Query expansion)等技術來增加系統與使用者互動的機會，進而產生較符合使用者需求的檢索策略。
- (六)運用資料探勘(Data mining)的技術了解使用者的檢索與瀏覽行為傾向。

三、系統管理者

- (一)了解電子期刊系統在一日內各時段、一週內每日的使用狀況，為系統訂出尖峰期與離峰期；在尖峰期間系統管理者必須多留意系統的穩定度，將系統效能保持在最佳狀態，而在離峰時間，則適合進行系統測試、備份、維護等相關作業。
- (二)根據使用者下載全文檔案的特性調整系統儲存空間，將儲存空間、價格、效能等因素最佳化。

張嘉彬一文(註 18)中曾說明系統評估之方法主要可分為三類，分別為檢索過程記錄(即使用記錄檔)分析、一般調查研究(調查、問卷及訪問法)



及觀察與實驗控制法等三種，這三種方法皆有其要缺點，而使用記錄檔分析的主要優點包含：

1. 研究人員可以類似無感式(Unobtrusively)的觀察方式來觀察使用者的行爲，使其在檢索過程中不受打擾；
2. 可提供一些實際的檢索記錄，了解使用者真正進行了那些檢索活動；
3. 記錄下來的檢索過程資料可以重複觀察讀者與系統之間的互動關係。

其主要缺點則有：

1. 無法掌握個別使用者於檢索進行時產生的不同變數，如使用者特性；
2. 無法真正代表使用者的檢索意向，檢索結果是否滿意及檢索過程中發生的困難或問題；
3. 檢索過程記錄分析是一種相當量化的研究方法，因此有學者批評其不夠深入。

我們認為使用記錄檔分析的方法和其他二種方法有互補的效果，例如可先透過使用記錄檔分析

發掘出現象，再運用其他二種方法了解之所以會發生這種現象的原因。舉例而言，從本研究的發現中，有以下幾點值得用其他二種方法再深入探究：

- (一) 電子期刊和紙本期刊的使用者行爲有何異同之處？
- (二) 許多 SDOS 使用者為何未曾閱讀文章摘要便直接下載全文了？
- (三) 網站搜尋引擎和數位圖書館使用者資訊尋求的動機、目的、行爲有何異同之處？
- (四) 電子期刊和索摘資料庫系統間串連的機制對使用者資訊尋求行爲有何影響？

本研究係以統計的觀念來處理使用記錄檔，未來我們將試著利用資料探勘(Data mining)來分析資料，可進行研究的方向之一是從使用記錄檔分析全文檔案間的關連性，以作為使用者下載全文檔案時的建議。在本研究進行之中，我們就曾實驗性地以一套資料探勘工具—DBMiner2.0<<http://www.dbminer.com/>>—來發掘 SDOS 全文檔案間的關連

ID	Name	Value	Value	Value
1	PK = (123295)	PK = (123295) AND PK = (123295) AND PK = (123295)	0.214	0.1428
2	PK = (123295)	PK = (123295) AND PK = (123295) AND PK = (123295)	0.214	0.1428
3	PK = (123295)	PK = (123295) AND PK = (123295) AND PK = (123295)	0.214	0.1428
4	PK = (123295)	PK = (123295) AND PK = (123295) AND PK = (123295)	0.214	0.1428
5	PK = (123295)	PK = (123295) AND PK = (123295) AND PK = (123295)	0.214	0.1428

圖十一：利用 DBMiner2.0 作資料探勘後檔案關連分析結果



性，DBMiner 的功能包括將資料分類、依照資料相似程度將資料分群、以及尋找資料間的關連性等。我們的實驗只針對被下載超過三十次以上的全文檔案作分析，而最後分析結果得到七十七筆相關法則(Association rules)，其中幾筆相關法則如圖十一。

由圖十一中第一筆資料來看，若某使用者下載過檔案識別碼 pid=123245 的檔案，則其也下載過 pid=106888、pid=18163、pid=84373 及 pid=88453 四個檔案的可能性是 51.429% (稱之為可信度，Confidence)，而這樣的情況出現在所有使用紀錄裡的百分比是 0.214% (稱為支持度，Support)。

藉由資料探勘找出的相關法則，可在使用者瀏覽或檢索時提供建議。比方說，當某使用者找到一

個需要的檔案時，系統可在同一頁面顯示出先前下載過此檔案的其他使用者亦曾下載過的其他檔案；如此，系統能更主動且有智慧地幫助使用者找到所需的資訊，滿足使用者的資訊需求。

此外，運用資料探勘的技術，還能分析檢索詞彙與全文檔案之間的關連性，或是更深入地分析使用者檢索詞彙間的關連性，並且將分析後的結果融入於系統中，以便在使用者檢索時，適時地提供檢索及全文檔案下載建議，提升系統的服務品質和使用者的滿意度。(收稿日期：2001 年 8 月 23 日)

致謝

感謝交通大學資訊科學研究所陳意芬、楊明洋、王悠伊、黃榮璋同學在資料探勘方面的技術支援，在此一併致謝。

註 釋：

- 註 1：張嘉樹，「系統評估研究之文獻探討：從電子圖書館角度考量」，《大學圖書館》，3：2 (民國 88 年 4 月)，頁 69-94。
- 註 2：張美美，「中文資訊檢索系統的應用和使用者檢索問題探討」，《資訊傳播與圖書館學》，6：2 (民國 88 年 12 月)，頁 23-38。
- 註 3：本文所指之數位圖書館系統僅限於 WWW 版電子期刊、全文資料庫、索網資料庫、電子書。
- 註 4：Z. Zhang "Evaluating Electronic Journals and Monitoring their Usage by Means of WWW Server Log Analysis," *Yang* 11(1999), pp. 37-42.
- 註 5：Ibid.
- 註 6：S. Jones, S. J. Cunningham, R. McNea, and S. Boddie "A Transaction Log Analysis of a Digital Library," *International Journal on Digital Libraries* 3 (2000), pp. 152-169.
- 註 7：B. J. Jansen, A. Spink, and T. Saracevic "Real Life, Real Users, and Real Needs: a Study and Analysis of User Queries on the web," *Information Processing and Management* 36(2000), pp.207-277.
- 註 8：卜小蝶，「台灣網路使用者檢索行為探析」，《大學圖書館》，4：2 (民國 89 年 9 月)，頁 23-37。
- 註 9：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》，89 年第 1 期 (民國 89 年 6 月)，頁 25-34。
- 註 10：卜小蝶，「網路使用者查詢語彙記錄之分析與應用初探」，《資訊傳播與圖書館學》，6：2 (民國 88 年 12 月)，頁 49-62。
- 註 11：SDOS 系統於 2000 年下半年起亦提供 Username/Password 認證方式，但在本研究的記錄檔涵蓋時間內，以 Username/Password 認證方式進入者並不多，故不納入本研究考量範圍。
- 註 12：之所以會有這種現象，至少有三種可能性：(1)因為電腦和網路的發達，使用者不在乎是否真正需要某一篇文章



章，先下載了再說，反正總量很大，網路很快；(2)使用者已經在其他索檢資料庫中看過這篇文章的摘要，認為確實對他有用，所以無需在 SDOS 系統中再看一次；(3)使用者是在閱讀其他文章時發現有引用這篇文章，認為確實對他有用，所以無需在 SDOS 系統閱讀摘要。

註 13：Jones-2000 計算每次來訪平均檢索次數時，把每次來訪檢索次數為零者(約佔所有連線次數的 21.51%)也計入，本研究和 Jansen-2000 則無，因為 SDOS 使用者中有 65%並未執行檢索(因每次來訪檢索次數為零，推測這些使用者應是以瀏覽方式取得資訊，若是把這 64%也納入計算，將使每次來訪的平均檢索次數大幅下降至 1.58。由於 SDOS 具備瀏覽和檢索二種方式取得資訊，不像 Jones-2000 針對的 CSTR 僅提供檢索功能，是以我們覺得在計算 SDOS 每次來訪平均檢索次數時應該要把檢索次數為零者排除。此外，本研究針對 SDOS 的連線器間定義亦會影響到每次來訪平均檢索次數的計算。

註 14：已將每次來訪檢索次數為零者排除。

註 15：在 Jansen-2000 中曾提及，許多針對傳統資訊檢索系統的研究發現，平均檢索策略長度約在 7—15，這個數字與 Jansen-2000、Jones-2000 以及本研究的發現相去甚遠。這個差異可能導因於傳統資訊檢索系統的使用者以經過圖書資訊學訓練的圖書館館員為主，而 WWW-based 系統(無論是網站搜尋引擎或是數位圖書館系統)則是以未受過圖書資訊學訓練的一般使用者為主。

註 16：SDOS 系統在 4.0 版的檢索功能中已提供簡易檢索(Simple search)、進階檢索(Expanded search)和專家檢索(Expert search)等三層次的檢索功能。

註 17：在較先前版本的 SDOS 中，“or”與“OR”、以及“and”、“&”與“AND”的意義並不相同，SDOS 會把“or”、“and”視為停用字而非特殊功能字元，亦即若使用者下“digital or communication”這個檢索策略，SDOS 會把其中的“or”去掉，而等同於使用者是以“digital communication”進行檢索，而此檢索策略的真正意義為“digital AND communication”。因此若使用者是把“or”當成“OR”來使用，則檢索的結果將會與使用者原先的資訊檢索需求有極大的差異(對“and”則是沒有影響)。在新版的 SDOS 中，已經把“or”與“OR”、“and”與“AND”的意義混為相同的，至於“&”，問題就更大了，“&”在 SDOS 為非停用字，因此若使用者以“digital & communication”進行檢索，則此檢索等同於“digital AND & AND communication”，就我們在本文寫作之時在 SDOS 執行此檢索將找不到任何符合的文章，而“digital AND communication”則有 372 篇符合的文章！

註 18：同註 1。

