

從網站工作記錄觀察大學圖書館網站使用情形 The Usage of a University Library Site from Observation of Web Log Files

林頌堅 **Sung-chien Lin**

世新大學圖書資訊學系助理教授

Assistant Professor, Department of Library and Information Studies, Shih-Hsin University

E-mail: scl@cc.shu.edu.tw

【摘要】

圖書館設置網站之後，為了瞭解使用情形與提昇網站的效能，有必要對網站進行評估。本論文介紹了利用網站工作記錄來作為評估資訊的方法。這種量化的方法具有客觀和經濟的優點，可以配合調查研究法得到更實際的評估。在本論文中，我們介紹了網站工作記錄的格式與從中抽取的訊息，並以一個大學圖書館的網站作為例子，實際分析在網站上的一些使用情形。從例子中，可以發現網站工作記錄上的資訊與網站使用者的特性有密切的關係，因此是相當重要的評估資料。

【Abstract】

After a library establishes its web site, evaluation is very important to understand the use of the web site and to improve the site performance. This paper introduces an evaluation method using the information extracted from web log files. This quantitative method is objective and economic and can be used in conjunction with other methods of survey research. In this paper, the format of web logs and the information that can be extracted from them are introduced. We also present an example of log files obtained from a real web site established by a university library to analyze the use of the site. Analysis of this example points to a close relation between the information carried by the log files and the use of the web site. Therefore, analysis results of the log files are important to evaluate a web site.

關鍵詞：圖書館網站評估；網站工作記錄

Keywords：Evaluation of library web sites; Web logs

壹、緒論

從民國81年教育部啓用台灣學術網路(TANet)整合校園網路與網際網路以來,大專院校便積極思考與運作,透過網際網路的便利性與穿透性使教學、研究等學術活動可以達到即時性與全球化的目標。大學圖書館也利用網際網路的優點,來提供學校師生與校外人士更多與範圍更廣的服務,如:透過網際網路的線上公眾存取目錄(Online Public Access Catalog, OPAC)與電子資料庫等等。(註1)在網際網路上,全球資訊網(World Wide Web, WWW)是一種具有非常多優點的網路系統,利用超文件傳輸協定(Hyper Text Transfer Protocol, HTTP),它可以整合多種不同網際網路服務應用,如ftp、newsgroup、gopher等等;加上圖形使用者界面的瀏覽器後的全球資訊網系統具有宜人的(user-friendly)人機介面,使用者在操作上容易學習且不易遺忘;它可以展示多媒體內容等能力,提供合適的資訊表示方式;並且具有文件間的超連結(hyperlinked),可以將分散在各處的相關文件或資源整合起來成爲一個有意義的(meaningful)資訊集合體,因此是一個愈來愈受重視的系統。目前大部分的大學圖書館都已經在全球資訊網系統上設置了網站(web sites),整合圖書館的各項服務,提供給校內外讀者方便而有效率的服務。(註2)在這些網站中,一般提供的參考

服務有:圖書館導覽、圖書館簡訊、館藏查詢系統(webpac, Web OPAC)、電子書刊、常問參考問題(FAQ, frequently asked question)、線上資料庫服務、線上服務表件申請、最新期刊目次服務、網路資源指引等等多種項目(註3)(註4);技術服務的項目則有:採訪與編目(註5);甚至有些圖書館還利用網站進行行政管理功能分析,如會議記錄、表格文件、經費管理、統計資料等,供館員使用。(註6)由此可知,大多數的大學圖書館不僅都已認知到圖書館網站設置的重要性,且利用全球資訊網提供多樣化的服務內容。(註7)

在圖書館提供網站服務之後,網站的設計者或管理者若能對這些網站的建置、架構與提供的服務常常加以評估與檢討,可以了解使用者的資訊需求,並且就全球資訊網的特性提供更高的資訊品質與更完善的服務。目前圖書館對於圖書館網站的評估方式大多利用學者和館員所發展出來的網路資源評估準則進行衡量。這些評估準則因爲甚少融入使用者的意見,很容易陷入主觀的盲點中,導致所建置的網站與使用者的需求產生極大的差距。(註8)較好的方式是組成一個由各種使用者、館員所組成的焦點團體進行需求了解與功能可靠性等議題的探討,擬出與實際使用情形較接近的評估準則做爲發展系統時的參考。(註9)因此,翁慧娟曾對五所大學的讀者與館員進行問卷調查,希望研

擬出大學圖書館的網站品質評估準則與服務項目供網站建置與改進的參考。(註10)從這項研究中，在網站品質評估準則方面，館員與讀者均重視網站資訊的更新與維護、內容的正確性以及連結的多元化、穩定度等項目；而館員與讀者均不重視的則是網站內容的畫面設計、感官享受以及一般性的消息。另外，館員在網站的規劃與設計上較重視於圖書館的任務層面；讀者則希望在網站的設計與規劃上可以一致的方式呈現組織架構，協助使用者實際取用資源時的便利。在服務項目上，館員與讀者均重視館內外資訊檢索服務，如webpac、資料庫檢索等；館員與讀者均不重視者則是一般性的資訊，如暢銷書排行榜、圖書館簡介等資訊。網站的設計者或管理者便可以依據這些評估準則分析圖書館網站的資訊內容與架構，修改網站來適合所有的使用者。然而問卷調查所得的結果是分析使用者與館員的主觀認知，問卷的受試者可能因為若干因素，如為符合社會規範、彰顯本身等等，可能影響問卷的效度。(註11)在進行問卷調查時需要對問卷設計加以注意，否則問卷所得到的結果是值得質疑的。另外問卷調查的特性之一是具有當時性(contemporary)，所以得到的分析結果只與調查當時的情況有關，當時過境遷或有新的技術或方法引入，便需要重新擬定問卷進行調查，而這需要花費很高的成本。(註12)

在本論文中我們將探討從網站伺服器中自動記錄的『網站工作記錄』(web log files)來發掘使用者的使用情形，做為在網站品質評估準則之外的另一資訊來源。在電子商務或其他網站相關的評估中，已經有許多研究採取利用網站工作記錄來進行網站的評估，一方面了解使用者常存取的網站資料，提供系統管理者在改善網站架構與組織上的參考，讓資料的存取更便利與迅速，另一方面則可以分析進站的使用者之特質，了解目前用戶的來源，以開拓更廣大的顧客與市場。(註13)(註14)因此，我們認為可以利用網站工作記錄得到讀者與館員在實際使用上的資訊，這些資訊較為客觀，且分析較為容易，較為經濟，可以經常施行，得到較為貼切當時的使用情形，應可以與利用問卷所得到的網站品質評估準則和服務項目相輔相成。

本論文中，我們將首先詳細介紹網站工作記錄的內容，與一些網站工作記錄分析工具，並對若干從網站工作記錄中獲得的原始資料，如：取用記錄和到訪記錄等加以定義；接著探討可以從分析網站工作記錄的原始資料所延伸出來的網站使用者資訊與具有的限制；然後，則實際以世新大學圖書館的網站工作記錄做為例子，觀察使用者的行為。

貳、網站工作記錄的介紹

在本節中，我們首先介紹網站工

作記錄；接著我們從實際使用網站的範例中來看使用者在取用網站資源時所留下的網站工作記錄；從分析這些工作記錄來介紹如何獲取使用者在存取資源時的存取記錄（access）和上站記錄（visit）。最後介紹用來獲取這些資料的網站工作記錄分析工具。

一、網站工作記錄

當使用者利用瀏覽器輸入一個網頁的網址或對網頁中的連結進行點選（hit）等存取網頁資源的行為時，便是對於網站伺服器（web server）發出需求（註 15），因為所存取的網頁中可能嵌入（embedding）多項多媒體資訊，比方說一個網頁中即有包括一些相關的靜態圖形（still graphs）檔案，甚至還有音訊（audio）或動態圖形（moving graphs）等檔案，所以通常一次使用者取用網頁的行為便包含了對多個網站伺服器所提出的需求。網站伺服器將會對於每一個需求留下一些工作記錄，這些工作記錄可分為『交易記錄』（transfer logs）和『錯誤記錄』（error logs）。（註 16）交易記錄是記錄使用者對於網站伺服器發出每一個需求的相關資訊，包括時間、用戶端主機位址和存取資源等等。錯誤記錄則是當網站伺服器對使用者的

需求傳回錯誤訊息時，保留在伺服器中的記錄，這些錯誤通常是使用者所要求的資料不存在或是使用者的權限無法存取所需求的資料。另外，網站伺服器中還有兩種較次要的工作記錄：『參考記錄』（referred logs）記載這些需求是從哪些網址（URL, Uniform Resource Locator）產生的，如由搜尋引擎（search engines）或是其他連結到所需求資源的網頁；與『代理程式記錄』（agent logs）記載哪一種代理程式（agent）對網站伺服器發出需求，這類代理程式以瀏覽器，極小部分是搜尋引擎的自動網頁擷取程式（robot）。基本上，在每一類的網站伺服器中都有記錄這些工作記錄，這些工作記錄都以 ASCII 形式的一般文字（plain text）來記錄，常見的工作記錄格式如 W3C 格式、NCSA 格式等。

由於交易記錄記錄了使用者對網站伺服器發出需求的資料，我們在分析使用者取用網頁行為的時候，便可以利用交易記錄來獲取相當重要的資訊。下面我們便以一個例子詳細地介紹網站交易記錄中的欄位資訊。圖 1 便是一筆 W3C 格式的交易記錄。

我們以圖 1 的例子來介紹交易記錄中的各個欄位：

1. 日期（date）、時間（time）；需求

```
# Fields : date time c-ip cs-method cs-uri-stem sc-status sc-bytes cs-bytes time-taken
1999-10-05 00:04:28 192.192.151.53 GET /navigator_ie4.htm 200 8767 371 156
```

圖 1 一筆 W3C 格式的交易記錄

發生的日期與時間，以上面的例子來說這個需求是發生在 1999 年 10 月 5 日的零時四分二十八秒。

2. 用戶端的網路 IP 位址 (**c-ip**)：發出需求的主機之網路 IP 位址，如上面的需求是由位址為 192.192.151.53 的主機所發出的。
3. HTTP 的存取方法 (**cs-method**)：根據 HTTP，用戶端程式對伺服器發出需求來存取資源的方法 (**method**) 有 GET (將資料從伺服器中取得) 和 POST (將資料由用戶端上傳至伺服器) 兩種，由於全球資訊網上資料流 (**data flow**) 的特性是由伺服器流向用戶端的情形較多，因此工作記錄中大部分的交易記錄都是在 GET，如上面的例子中用戶端即是 GET 的方式向伺服器要求某項資料。
4. 存取資源 (**cs-uri-stem**)：用戶端向網站伺服器要求存取的資源，上面的例子中是在根目錄 (**root directory**, /) 下名為 `navigator_ie4.htm` 的超文件。
5. 伺服器所傳回之資源存取狀態 (**sc-status**)：在 HTTP 中，提供資料的伺服器會在回應用戶端需求之時，告知用戶端結果的代碼，常見的代碼有 200 (沒有錯誤，要求成功)、403 (伺服器不接受這個需求，原因通常是沒有通過驗證) 和 404 (伺服器沒有用戶端所要求的資料)。在例子中，交易記錄關於這個欄位的記載是 200，代表這個需求是成功的。

6. 傳送資源的大小 (**sc-bytes**)：從伺服器傳送到用戶端資料的位元組數，上面的例子中，這個需求共 8767 個位元組。
7. 接收資源的大小 (**cs-bytes**)：伺服器從用戶端接收到資料的位元組數，上面的例子中，這個需求共 371 個位元組。
8. 花費時間 (**time-taken**)：從用戶端發出需求到接收到資料共花費的時間，在這個例子中共花費 156 毫秒。

二、存取記錄與上站記錄

交易記錄是對用戶端對伺服器每一筆存取需求的記錄，即是使用者從伺服器中取得一個網頁超文件、靜態圖形或甚至音訊、動態圖形等資料都會在工作記錄中留下一筆記錄。從上面對於交易記錄各個欄位的分析，如果我們利用簡單的統計去分析某一部網站伺服器在一段期間的工作記錄，便可以獲知在這段期間內，在這部伺服器中各項存取的資訊，如一段時間 (月、週、日或小時) 中伺服器中的資訊存取次數、用戶端的主機發出資料需求的次數、伺服器中每一項資料被存取的次數、資料要求成功或失敗的次數等等。這些統計資訊是以存取次數為基礎的，當工作記錄中的資料量足夠的情形下，我們可以將這些資訊再延伸，以獲得更詳盡的資料。

某些以使用者上站造訪為基礎的資訊，如某些用戶端主機在某段期間內上站的次數、以伺服器中的那一個

網頁為進站網 (entry page) 或離站網頁 (exit page) 的次數 (註17) 等等，或是較複雜的資訊，如使用者在上站時存取資源順序的形式 (pattern)，這些造訪記錄資訊在分析使用者行為時是相當有用的資訊。比方說，如果大部分上站的使用者都不是由首頁進入，也就是以首頁作為進站網頁的比例過低，設計者或管理者便可仔細思考網站的架構是否不佳，而應重新設計網站。再者，例如多數的上站中存在某些常見的存取資源順序的形式，而這些順序相當長，設計者或管理者也可以加以思考是否重新組織網站架構，方便使用者迅速存取相關資料。但若是工作記錄中只有以存取記錄為基礎的資訊，沒有前後相關的資料，這些資訊便無法得知。

對於這問題，我們可以觀察工作記錄中時間相接近的幾筆具有相同用戶端網路IP位址的交易記錄，如圖2所示是一段從實際的圖書館網站 (世新大學圖書館，網址 <http://lib.shu.edu.tw>) 工作記錄中所取出的部分交易記錄，但為了說明清楚起見，我們去除了一些其他網路IP位址的交易記錄與較不重要的需求，如網頁中嵌入的靜態圖檔和『串接樣式表』 (cascading style sheet, css)。(註18)

從這段1999年11月9日的交易記錄清單中 (記錄前的標號是為了說明起見而加上的)，我們可以看到從IP位址為192.192.159.32的用戶端主機連上圖書館網站伺服器的使用者在接近

十三分鐘之內所進行的動作。從記錄 (a)，我們可以觀察到使用者 (註19) 在 06:07:43 向網站伺服器提出下載圖書館網站首頁 `navigator_ie4.htm` 的需求，因為在這筆記錄之前，並沒有相同用戶端主機IP位址的記錄，我們可以推斷這筆記錄是使用者此次上站對網站伺服器所發出的第一個需求，而圖書館網站首頁 `navigator_ie4.htm` 則是用戶端此次上站的進站網頁。之後的記錄 (b) 和 (c) 顯示這位使用者在 06:08:06 和 06:08:17 兩度向網站伺服器提出在目錄 `webpac2` 下的 ASP (active server page) 網頁 `webpac22.asp` 的需求，希望使用網站中的 `webpac` 功能；接著使用者在 06:08:34 從伺服器下載 `msearch.dll`，將要進行書目查詢，並於 06:09:24 使用這項的功能 (注意前一筆記錄 (d) 的 HTTP 的存取方法為 GET，而後一筆 (e) 則是 POST，這顯示前一次用戶端先下載 `msearch.dll`，後一次才是使用者進行查詢功能的記錄)；之後以 `booklist.dll` 將書目查詢的結果列出，從 06:09:31 到 06:17:39 使用者都在利用 `booklist.dll` 瀏覽查詢的結果 (記錄 (f) 到 (l))；而後 06:20:16 時使用者再度利用 `webpac`，使用 `Lib.dll` 功能 (記錄 (m) 和 (n))；在 06:20:53 時，使用者進入在目錄 `shulib/Service` 下的『文山區三校圖書館聯盟』的網頁 `allunion.htm` (記錄 (o) 和 (p))，而這個網頁是以 `frame` 的方式構成，裡面包括 `pagetitle.htm` (記錄 (q)) 和 `union.htm` (記錄 (r)) 兩個子網頁。

由於之後的交易記錄清單之中已經沒有關於這個用戶端主機IP位址的記錄（註 20），所以我們推論使用者可能在這個網頁中斷與網站伺服器的連線，進入其他兩個學校的圖書館網站。

因此，利用這樣的經驗法則（heuristics），我們可以網站伺服器中的工作記錄分析與發掘使用者上站的使用行為，如使用者上站所花費的時間、連上伺服器的網路主機、使用者在站上使用的網頁或功能、使用者常用的進站網頁和離站網頁、甚至這些使用網頁或功能的順序等等都可以分析得知。以上面的例子來說，此次上站使用者共花費13分14秒，由192.192.159.32上站，進站和離站的網頁分別是/navigator_ie4.htm 和/shulib/Service/allunion.htm，此次上站使用圖書館網站webpac中的書目查詢功能和文山區三校圖書館聯盟。根據這些資料，網

站伺服器的管理者便可根據這些資料來瞭解使用者的興趣與行為，對使用者提供更方便而有效率的網站架構和網頁內容。

三、網站工作記錄分析工具

雖然由以上的分析，我們可以自交易記錄中大約猜測出使用者上站的行為與資源存取的順序，但是實際上工作記錄中所包含的交易記錄十分鉅量，而且因為伺服器是一個多工的作業環境，換言之，在一段很短暫的期間內可能有非常多的用戶端使用伺服器，在工作記錄中有非常多的用戶端交易記錄，實在很難以人工的方式直接觀察，得到詳密的結論。因此需要借助網站工作記錄分析工具，以自動化的方式分析工作記錄提供一些統計訊息，作為進一步分析的依據。常見的網站工作記錄分析工具有 Analog

```
(a) 1999-11-09 06:07:43 192.192.159.32 GET /navigator_ie4.htm 200 8841 324 2781
(b) 1999-11-09 06:08:06 192.192.159.32 GET /webpac2/webpac22.asp 200 5036 397 16
(c) 1999-11-09 06:08:17 192.192.159.32 GET /webpac2/webpac22.asp 200 4969 452 31
(d) 1999-11-09 06:08:34 192.192.159.32 GET /Webpac2/msearch.dll 200 7234 452 32
(e) 1999-11-09 06:09:24 192.192.159.32 POST /Webpac2/msearch.dll 200 370 650 1218
(f) 1999-11-09 06:09:31 192.192.159.32 GET /Webpac2/booklist.dll 200 7713 458 657
(g) 1999-11-09 06:10:20 192.192.159.32 GET /Webpac2/booklist.dll 200 4699 464 797
(h) 1999-11-09 06:10:40 192.192.159.32 GET /Webpac2/booklist.dll 200 7713 463 19515
(i) 1999-11-09 06:17:18 192.192.159.32 GET /Webpac2/booklist.dll 200 4699 463 766
(j) 1999-11-09 06:17:26 192.192.159.32 GET /Webpac2/booklist.dll 200 7713 463 9437
(k) 1999-11-09 06:17:28 192.192.159.32 GET /Webpac2/booklist.dll 200 7713 463 1141
(l) 1999-11-09 06:17:39 192.192.159.32 GET /Webpac2/booklist.dll 200 4699 463 11016
(m) 1999-11-09 06:20:16 192.192.159.32 GET /webpac2/webpac22.asp 200 955 457 16
(n) 1999-11-09 06:20:36 192.192.159.32 GET /Webpac2/Lib.dll 200 1863 448 1172
(o) 1999-11-09 06:20:53 192.192.159.32 GET /shulib/Service/allunion.htm 200 731 439 15625
(p) 1999-11-09 06:20:53 192.192.159.32 GET /shulib/Service/allunion.htm 200 731 439 797
(q) 1999-11-09 06:20:54 192.192.159.32 GET /pagetitle.htm 200 2856 452 391
(r) 1999-11-09 06:20:57 192.192.159.32 GET /shulib/Service/union.htm 200 6663 463 359
```

圖 2 一段相同用戶端主機 IP 位址的工作記錄清單

(註 21)、wwwstat (註 22)、Web Trends (註 23) 等等。

一般網站工作記錄分析工具所提供的功能有分析各項統計項目，如存取資源、用戶端主機和網域、進站網頁、離站網頁、資源存取的順序等等，並可將分析結果組合成具有較高可讀性的表格與報告。這些統計訊息的表格與報告便可利用來作為進一步分析的依據。一些較精緻的分析工具也可以將分析結果以各類統計圖表顯示，可以用更具體的圖像提供更清楚的展示。某些分析工具並具有過濾某些存取資源或主機位址的記錄，方便進行分析。甚至商業用途的網站工作記錄分析工具還可將初步分析結果存成SQL形式的關連式資料庫 (relational databases)，將分析過的資料儲存起來，可以進行『資料倉儲』 (data warehousing) 的處理，加入一些人工智慧的方法可以提供更精確且更豐富的網站使用資訊。(註 24) (註 25)

參、網站工作記錄的限制

雖然利用網站工作記錄分析工具得到網站的資源取用記錄與用戶端上站記錄，可以了解網站中使用者的行為。(註 26) 但是由於一些 HTTP 的特性與網站工作記錄的限制，這些資訊並非絕對的正確。首先上站記錄的資料並非是精確無誤的，有些使用者以瀏覽器進入網站，取得網頁資料後，可能因為某些事件的耽擱而必須多次暫時離開，等事件結束後再返

回。目前一般網站工作記錄分析工具多以兩次相同用戶端主機IP位址的記錄所相隔的時間長度，作為判斷使用者是否離站的經驗法則。如果間隔時間太長，此時分析工具可能會將此次上站統計成多次到訪記錄，影響上站記錄資料的正確性。在圖書館網站的分析上，這個問題更加嚴重。因為圖書館提供許多開放的電腦做為連線的用戶端主機供使用者在館中使用webpac或查詢其他資訊。在這種狀況下，如果不同使用者使用的間隔時間太短，或甚至部分使用者在離開後沒有關閉瀏覽器，下位使用者便直接使用已經連上圖書館網站的瀏覽器之情形下，網站工作記錄分析工具會將這兩次不同使用者的上站做為一次到訪記錄，破壞上站記錄的分析。

其次，網站工作記錄所記錄的是用戶端的主機IP位址而非實際使用者的識別資訊，當嘗試將交易記錄中用戶端的取用和到訪記錄視為某一使用者的行為時，事實上並非完全正確；由於目前在網際網路上，並非每一部電腦可以擁有一個專屬的IP位址，尤其是以撥接上網的使用者來說，他們的IP位址是動態設定的，所以等於說，許多人將會共用一個IP位址，而使用者在每次連線時所設定的IP位址也不相同。(註 27) 而即使是透過區域網路直接連上網際網路的用戶端主機，也極有可能是多人共用。因此我們無法將交易記錄中某個用戶端的存取和上站記錄視為某一特定使用者的

行爲。此外，會產生同樣困擾的是網路上的『代理伺服器』(proxy server)，在設定代理伺服器之後，用戶端向網站伺服器提出下載網頁資料的需求，會先向代理伺服器提出需求，而由代理伺服器傳回網頁資料回應該次需求，如果網頁資料不存在代理伺服器，無法回應該次需求，代理伺服器便會先向網站伺服器提出需求，將網頁資料下載，再回應用戶端，用戶端並不會直接連線至網站伺服器，在 TANET 上許多主機都設定了代理伺服器以符合教育部管制網路連線流量的規定，在工作記錄中我們可以觀察到許多代理伺服器所提出的需求，但無法確實知道是從哪一用戶端所發出的。

最後，正如我們在前面所提到的工作記錄中的資料相當龐大，需要借助網站工作記錄分析工具的輔助才能獲得大概的資訊。但是每一套網站工作記錄分析工具用來進行統計分析的演算法與經驗法則都有所不同，Backman 和 Rubin 曾經以當時最著名的幾套網站工作記錄分析工具進行測試，他們發現雖然這些工具所得到的統計趨勢大致相同，但是在實際的數量上卻有一些差異。(註 28) 所以網站工作記錄分析工具所得到的結果雖仍可提供網站管理者做為參考，但不是十分精確。

四、可用的資訊

從上面的分析，可以知道網站工作記錄所能提供資訊的限制在於無法

提供十分精確的使用者網站使用行爲，但我們還是可以利用交易記錄與網站工作記錄分析工具得到不少有用的資料。交易記錄所能提供的資訊有上站的日期時間、使用的網頁資料或功能、與用戶端的主機位址等等。這些資訊可以分成時間、使用者與資源等三個方面。所以，最基本我們可以知道使用網站的時機(如：月份、日期、假期與平日等等)、可能的使用者來源(如：某一特定位址的用戶端主機、校內主機或校外主機、館內主機與非館內主機等等)與所使用的資源與功能，並可以綜合分析這些資訊而得到不同使用者來源利用網站的時機、不同使用者來源所利用的網站資源與功能、某一網站資源與功能被使用的時機等等，這些資訊有助於我們瞭解使用者利用網站的目的，並由此來增益網站的效能。下面我們便以實際的網站工作記錄來了解分析圖書館網站的使用情形。

肆、實地觀測的結果

以下以世新大學圖書館網站做為例子，介紹利用網站工作記錄所觀察到的一些使用行爲。在本論文中，所使用的是主機名稱為 lib.shu.edu.tw 的網站伺服器從 1999 年 5 月 4 日到 2000 年 5 月 5 日約一年期間的工作記錄。值得注意的是由於這些圖書館網站的資訊內容並非都在這個伺服器上，有些資料或服務為了較有效率起見，被放置在其他的服務器上，如光碟資料

庫檢索，所以關於這些資料或服務需求的工作記錄，不存在於網站伺服器 lib.shu.edu.tw 之中。因此在本論文中所探討的使用行為只侷限於一部份圖書館網站的服務功能，如 webpac、館際合作、電子期刊、網際資源、參考服務和圖書館服務等等。未來我們將會嘗試收集所有相關網站伺服器的工作記錄，進行更大規模與更詳盡的探討。

另外，此次觀察中，在分析工具的參數設定上，我們將判斷使用者的離站時間長度經驗值設為 5 分鐘，也就是當用戶端主機在 5 分鐘之內沒有繼續對網站伺服器發出需求的話，下次再發出的需求將視為另一次的上站記錄。此外，我們也忽略某些型態的存取資源，如 JPEG 和 GIF 靜態圖形、串接樣式表等，因為這些資源的統計與本論文的研究較不相關。

在本節中，我們首先觀察在一年的期間內各項的統計資料，如資源存取次數與上站次數、使用者常用的網站資源、用戶端主機與它們所屬的網

域以及常見的被存取資源的順序；其次，我們探討在歷時一年期間的各項統計數值的變化。最後，我們依照用戶端主機的網域將使用者分成數群，來推論不同類別使用者的特性。

一、一年內的各項統計資料

表 1 是這一年內網站的資源存取次數、上站次數與上站的用戶端主機 IP 數的統計表，我們可以看到在一年之中已經有超過一百七十萬次的資源存取次數，而上站次數也已經達到十四萬次，對照世新大學的學生及教師人數（如表 2）來說，以一個才開始提供啓用 webpac 不久的圖書館網站來說，這已經是一個相當好的結果，表示使用者對網站有很高的使用意願。在另一方面，我們也可以看到曾經上站的用戶端主機 IP 數超過一萬三千個，這也說明了使用者上站的主機來源相當多，不僅只是來自校內的主機，同時還有非常多的部分是來自於校外。除了利用校園內的電腦使用圖

表 1 世新大學圖書館各項資源存取與上站記錄（1999/5/4~2000/5/5）

資源存取次數	上站次數	上站的用戶端 主機 IP 數
1748347	143435	13461

表 2 世新大學學生與教師人數統計（取自網頁 <http://shu.edu.tw>）

學生人數總計	大學部	第二部	研究生	專任教師
8247 人	6114 人	2057 人	76 人	235 人

圖 3 一年內圖書館網站中前六項常用的資源與功能以及它們佔全部存取次數的比例

書館網站所提供的各項服務之外，使用者還可以利用其他網際網路上的連線主機或是家中的個人電腦撥接上網來使用。所以，綜合以上的資料，爲了提供更方便的服務與服務更廣大的使用者，大學圖書館確實有必要建置網站。

接著我們來看一年間使用者常使用的網站資源與功能，圖 3 所顯示的是使用者最常使用的前六項網站資源與功能以及它們佔全部存取次數的百分比。從圖中我們可以觀察到前六項網站資源與功能總共佔了全部存取次數的 78%，顯示使用者在利用圖書館網站時，大部分集中於這些資源或功能。另外，我們也可以觀察到在前六項的資源與功能中，只有少數是網

頁，如第三項是圖書館首頁，其餘的都是 webpac 中的功能，前兩項分別是館藏查詢與書目查詢列表，其次第四項是館藏資料，第五項則是個人資料查詢。可見這幾項 webpac 功能對於圖書館網站使用者的重要性，而且館藏查詢、書目查詢列表與館藏資料更是重要，因爲這三項的總和幾乎佔了全部存取次數的一半以上，顯示使用者在利用圖書館網頁時大半的用途是進行館藏查詢。

接下來我們來分析常上站的用戶端主機及網域。首先，表 3 中表示最常上站的前二十個用戶端主機位址，從表中，可以觀察到前二十個最常上站的用戶端主機位址大部分來自網域爲 192.192.158.xxx 的用戶端主機，只

表3 一年內圖書館網站中前二十個上站次數最多的主機位址以及它們佔全部存取次數與上站次數的比例

排名	網站地址	上站次數	%	存取次數	%
1	192.192.158.132	6,500	4.5	118,445	6.8
2	192.192.158.131	6,009	4.2	134,185	7.7
3	192.192.158.133	5,517	3.9	111,633	6.4
4	192.192.158.138	5,178	3.6	134,735	7.7
5	192.192.158.137	5,160	3.6	125,753	7.2
6	cc.shu.edu.tw	5,118	3.6	37,509	2.2
7	192.192.158.154	3,698	2.6	33,650	1.9
8	192.192.158.130	3,586	2.5	83,454	4.8
9	gate.tp2rc.edu.tw	3,500	2.4	20,230	1.2
10	192.192.158.150	3,260	2.3	28,017	1.6
11	192.192.158.152	3,029	2.1	25,541	1.5
12	192.192.158.156	3,021	2.1	19,439	1.1
13	192.192.158.157	2,965	2.1	19,495	1.1
14	192.192.158.158	2,753	1.9	16,850	1
15	192.192.148.108	2,566	1.8	39,046	2.2
16	i192.192.158.140	2,447	1.7	29,908	1.7
17	192.192.158.134	2,421	1.7	41,883	2.4
18	192.192.158.141	2,330	1.6	29,898	1.7
19	192.192.158.151	2,229	1.6	21,009	1.2
20	192.192.158.155	2,206	1.5	16,133	0.9
總和		73,493	51	1,086,813	62

有少數從cc.shu.edu.tw (192.192.150.1, 世新大學的主機) 或是 gate.tp2rc.edu.tw (163.28.32.100, TANET台北區二中心) 等上站, 經查證的結果, 這些網域為 192.192.158.xxx 的用戶端主機都

是位於圖書館內提供讀者利用的開放主機, 可見得雖然連上網站的用戶端主機超過一萬三千個, 使用者可以在任何可與網際網路連線的地點使用, 但是還有非常多的使用情形是發生在

圖 4 各類主機的存取次數比例

圖書館內的開放主機。我們可以再詳細的分析這些在圖書館內開放主機的地理位置，前三個最常上站的主機位址是擺放在圖書館入口處供讀者利用 opac 功能的電腦，這三部電腦與第四、五部都是最靠近書庫的開放主機，而且這五部主機的存取次數總和佔全部存取次數的三分之一以上（35.73%），顯然許多使用者有在離資源最近的地方使用網站服務的傾向。這也表示圖書館在提供網站服務之後，仍然應重視館內開放電腦的設置與維護，才能提供使用者更好的服務品質。

以網域來區分這些主機，我們將它們分成(1)館內提供讀者利用的開放式電腦，稱為館內讀者主機（IP位址為 192.192.158.xxx）(2)館員用來執行圖

書館業務的電腦，稱為館員主機（IP位址為 192.192.148.xxx）(3)校園內其他辦公室與電腦教室的主機，稱為校內主機(4)IP位址來自於世新大學校園以外的地方者，包括TANET和其他網際網路上的主機，這些都歸類為校外主機。圖 4 是各類主機的存取次數比例。從圖中，我們可以清楚的看到大部分的存取來自於館內讀者主機，這也再度地闡明了上面的論點。其次，來自校外主機的存取也有 15%是相當值得重視的。

最後我們來看看常見的存取資源順序。常見的存取資源順序代表大部分使用者利用網站所希望完成的工作，如果順序的長度較長，可能是網站的架構結構不佳，使用者需要多次點選才能完成他們的工作。因此，網

表 4 一年內最常見的五種存取資源順序與其存取次數

排 名	存取資源的順序	存取次數
1	/navigator_ie4.htm /shulib/netshare/netfoot.htm /pagetitle.htm /shulib/netshare/net.htm	2989
2	/webpac2/msearch.dll /webpac2/booklist.dll /webpac2/store.dll	1176
3	/navigator_ie4.htm /shulib/cdromintro/cdtrial.htm /pagetitle.htm /shulib/database/trial.htm	802
4	/navigator_ie4.htm /webpac2/webpac22.asp /webpac2/counter.dll /webpac2/person.dll	720
5	/webpac2/webpac22.asp /webpac2/counter.dll /navigator_ie4.htm	586

站管理者可以根據這項資訊來判斷實際使用的情形，做為網站改版時的參考。依據我們使用的網站工作記錄分析軟體所得到的前五個常見的存取資源順序如表 4 所示。排名第一的是由圖書館網站首頁 (/navigator_ie4.htm) 進入『網際資源』 (/shulib/netshare/netfoot.htm)，而這個網頁是一個包含了兩個子網頁 (/pagetitle.htm 和 /shulib/netshare/net.htm) 的框架式網頁；第二是『館藏查詢』 (/webpac2/msearch.dll)，『列出查詢結果』 (/webpac2/booklist.dll) 後，使用者進一步點選『館藏資料』 (/webpac2/store.dll)；

第三是使用者從圖書館網站首頁進入『光碟資料庫試用』網頁 (/shulib/cdromintro/cdtrial.htm) (註 29)，這個網頁也是一個框架式網頁，包含兩個子網頁 (/pagetitle.htm 和 /shulib/database/trial.htm)；排名第四則是使用者從圖書館網站首頁進入『線上公用目錄查詢系統』網頁 (/webpac2/webpac22.asp) (/webpac2/counter.dll 是網頁中的計數器程式)，再進行『個人資料查詢』 (/webpac2/person.dll)；最後是由『線上公用目錄查詢系統』網頁回到首頁。從表中，我們可以看到網站的使用者大多重視各種館內外的資訊檢

圖 5 一年中網站使用者上站次數與資源的存取次數的變化

索服務，這與翁慧娟所做的調查是不謀而合的。（註 30）另一方面這些順序的長度都在 4 以內，顯示網站的架構良好，使用者不需太多次點選，即可順利完成他們的工作。

二、一年內各項統計數值的變化

首先我們觀察在一年中網站使用者的上站次數與資源的存取次數的變化，如圖 5 所示。從圖中，這些使用者的上站次數與資源的存取次數與學校的行事曆有很明顯的關係：在假期期間，也就是圖中的 1999 年 7 月到 9 月和 2000 年 2 月，使用者的上站次數與資源的存取次數都比其餘的月份來的低，甚至在學期間有較多假期的月份的使用者上站次數與資源存取次數

也較其餘月份來得低，如 2000 年 1 月和 4 月。這樣的結果與圖書館的讀者結構有密切的關係，世新大學圖書館的讀者以大學部學生居多，在假期中大學部學生較沒有功課的壓力，利用圖書館網站的機會便也較少。因此網站使用者的上站次數與資源的存取次數的變化較受學校假期的影響。另外，在過去一年內，每一學期間各月的使用者上站次數和存取資源數目也有越來越高的趨勢，顯示使用者逐漸熟悉利用圖書館網頁的服務。

接著，我們來看看各月上站的主機 IP 位址數目與累積 IP 位址數，各月上站的主機 IP 位址數目是指當月所有曾上站的主機 IP 位址數目，每月累積 IP 位址數則是指從 1999 年 5 月到當月

所有曾上站的主機IP位址數目，圖6表示其間的關係，如1999年11月的上站的主機IP位址數目為2002，而累積IP數則為5804。我們可以看到，除了假期的幾個月份之外，每月大約都有1000到2500個IP位址左右的用戶端主機連上圖書館網站；而且每月的累積IP位址數也成長的很快。這表示圖書館網站的設立確實可以服務相當廣泛的使用者，並且不斷有新的使用者進入。

接下來我們來觀察各類使用者主機在各月中存取網站資源的比例，圖7是這些使用者主機的存取數目佔各月總存取數目比例的堆疊圖。舉例而言，在1999年11月，從館內讀者主機存取網站資源的次數約佔當月總存取數的70%，校外主機存取數目的比例為20%，而館員主機和校內主機則分別佔總存取數目的4%和6%。圖中最明顯之處是在假期時（1999年7到9月和2000年2月），從館內讀者主機存取網站資源的比例較低，而此時校外主機與館員主機的比例反而升高。我們可以推論可能是如同前面所討論的圖書館讀者的結構中大部分是大學部學生，除了假期中較沒有功課壓力外，大學生在假期中也較少到校利用圖書館，而在家中以撥接上網的方式連結進入圖書館網站，因此館內讀者主機存取的比例降低，但校外主機的比例則升高。而館員主機比例升高的原因則是在假期中館員仍須利用電腦主機連結網頁執行日常業務，所

以當其他主機的存取次數減少時，便顯得館員主機比例升高。

三、各類使用者所利用的網站資源和功能

最後，我們來看看不同類型的使用者來源在使用網站資源和功能時的差異性。我們將各類型使用者來源最常使用的前五項資源和功能取出來討論，各類型使用者來源最常使用前五項資源和功能的項目是相同的，雖然存取資源和功能的比例和排名不同。這表示這五項資源和功能確實是各類型使用者都常利用的。在館內讀者主機的部分，因為其中的存取數目在所有類型的使用者佔絕大多數，所以可以說館內讀者主機的前五項存取資源和功能的排名和全體使用者方面相同，且佔全部存取的比例相近。而校外主機則是其中最特異的，使用者在利用校外主機連線進入圖書館網頁時，最常用的功能是『個人資料查詢』（15.4%），而其次才是『館藏查詢』（15.1%）、『書目查詢列表』（13.0%）、『圖書館首頁』（7.9%）和『館藏資料』（7.0%），所以可能使用者在利用校外主機上站時較常使用『個人資料查詢』功能，查詢自己的借閱狀況。同樣的，雖然校內主機的使用者仍是最常使用『館藏查詢』和『書目查詢列表』功能（22.5%和14.2%），但是『個人資料查詢』功能也是佔有相當多的比例（13.3%）。

圖 6 各月上站的主機 IP 位址數目與累積 IP 位址數

圖 7 一年中各月各類使用者來源的比例

圖 8 各類使用者最常存取資源的前五項及其佔該類使用者存取次數的比例

伍、結論

圖書館在設置網站之後，爲了瞭解使用情形與提昇網站的效能，有必要對網站進行評估。以往常用的評估準則是以問卷調查的方式找出網站的設計與規劃上需要注意之處。但問卷調查容易受到受試者的心態影響，且擬定問卷的成本很高，不易經常修訂，所以可能與網站的實際使用情形有所差距。本論文介紹了利用網站工作記錄來作爲評估資訊的方法，從使用者在利用網站時記錄在網站工作記錄利用量化的方法分析，抽取出與使用情形相關的資訊。這種量化的方法

具有客觀和經濟的優點，可以配合調查研究法得到更實際的評估，幫助網站的設計者與管理者進一步提昇網站的利用效能。在本論文中，我們介紹了網站工作記錄的格式與從中抽取的訊息，並以世新大學圖書館的網站作爲例子，實際分析在網站上的一年的工作記錄。在對圖書館網站的工作記錄所做的分析中，可以觀察到使用者所利用的網站資源以webpac內的功能和資訊檢索爲主，尤其是館藏查詢是非常重要的功能。其次，一年內曾連上網站的主機數目超過一萬三千個，但大部分的上站記錄是由館內主機產生，顯示館內開放主機的重要性。另

外，使用者的上站次數與存取資源的次數和學校的行事曆間相關，這反應了讀者的結構將影響網站的使用情形。由以上的討論，可以知道網站工作記錄上的資訊與網站使用者的特性與網站架構等因素有密切的關係，因此是相當重要的評估資料。

本論文初步嘗試利用網站工作記錄來分析圖書館網站的使用情形，由以上的討論中，可以知道還有許多值得進一步研究之處。未來的研究方向包括蒐集更完整的網站工作記錄做為

進一步研究的材料，包括蒐集更長期的資料與納入所有與圖書館網站相關的網站伺服器。並且嘗試利用精緻的統計方法來分析初步的結果，以達到對網站利用情形更精確的解讀。

陸、致謝

本論文之完成，作者要感謝世新大學圖書館賴鼎銘館長和黃子芳小姐慷慨提供圖書館網站工作記錄，並於研究開始時給予寶貴的意見。

註釋

註1：彭慰、宋玉，「圖書館自動化與資訊網路」，在第三次中華民國圖書館年鑑（台北市：國家圖書館，民國88年），頁237-278。

註2：同註1。

註3：陳昭珍，「網路化的參考服務」，在圖書館自動化與網路論集，中華圖書資訊學教育學會編（台北市：文華圖書館管理公司，民國87年），頁167-180。

註4：朱碧靜、鄭雪玫，「全球資訊網在台灣地區圖書館應用現況之探討」，在圖書館自動化與網路論集，中華圖書資訊學教育學會編（台北市：文華圖書館管理公司，民國87年），頁63-88。

註5：同註4。

註6：同註4。

註7：薛理桂，「圖書館在網際網路提供的讀者服務」，圖書與資訊學刊No. 24，頁35-43。

註8：翁慧娟，「我國大學圖書館網站品質評估」（碩士論文，國立台灣大學圖書資訊學研究所，民國87年），頁3。

註9：A. Koopman, "Library Web Implementation: a Tale of Two Sites," <<http://www.curtin.edu.au:82>>.

註10：同註8，頁5。

註11：林振春，社會調查，三版（台北市：五南，民國81年），頁94-98。

註12：同註11，頁5。

註13：R. Stout, Web Site Stats: Tracking Hits and Analyzing Traffic (Berkeley:Osborne/McGraw-Hill, 1997), 92-93.

註 14：D. Nicholas et al., "Cracking the Code: Web Log Analysis," Online & CD-ROM Review 23: 5 (1999): 263-269.

註 15：在全球資訊網中，由於具有超連結的功能與網際網路開放，許多術語的定義都是虛擬的，並非有實體存在，而是要由使用者或設計者的觀點來解釋。例如『網頁』（web pages）、『網站』（web sites）、『網站伺服器』（web servers）等。網頁一般都是指設計者設計提供使用者利用瀏覽器所看到的一個畫面，然而它並不等於是一個固定或動態產生的 HTML 檔，一個網頁經常在 HTML 檔內嵌了許多靜態圖檔、動態圖檔和音訊檔等多媒體檔案，而且許多網頁是框架式的（framed）網頁，一個網頁中可以由主要 HTML 檔以 HTML 語法的框架包含多個 HTML 檔，這些內嵌的多媒體檔案或框架內的 HTML 檔與主要 HTML 檔並非由同一個人或機構所提供，甚至這些資源與檔案可能在同一網站伺服器中或在多個伺服器中。所以當使用者在閱讀一個網頁時，事實上他同時取用了多個網站伺服器內或外的資源與檔案。網站的概念也是相當寬鬆，有些研究將網站定義為經由網站的設計者依據他的知識結構，將一組相關資源或檔案組織起來的網頁集合，網站中的資源與檔案可能在同一網站伺服器中或在多個伺服器中。此外，一個網站伺服器中可以包含多個網站，一個網站中的資源與檔案也可能放置在多個伺服器中。

註 16：同註 13，頁 6-10。

註 17：進站網頁是指使用者進入網站的第一個網頁；離站網頁是指使用者在中斷與網站連線前的最後一個網頁。進站網頁與一般所熟知的網站『首頁』（homepage）有所不同，首頁通常是指網站的設計者或管理者為一個網站中相關網頁所規劃的門面（entry），也就是他們希望使用者進入網站的進站網頁。但是實際上，網站本身是一個超文件系統，使用者可以從網站中的任何一個網頁進入系統中。雖然在大部分的情形下，使用者可能會依據設計者或管理者的規劃，由首頁進入，但是在一些情形中使用者還是會自其他網頁進入網站，這些情形可能代表首頁的設計不佳（例如國內網頁常見在首頁中只有一個很大的圖形檔案，沒有任何相關資料或指引進入網站中的其他網頁）或是熟練的使用者只需網站中的某些資料，而將這部分的網頁利用『網頁書籤』（bookmarks）儲存。

註 18：串接樣式表是用來提供網站中一組相關網頁的段落內文或標題等樣式有相同的字型（font）的參考檔案，詳細資料可參見 W3C 的文件 <http://www.w3.org/TR/CSS-access>。

註 19：在此處，我們先假設一個用戶端主機 IP 位址對應到一位使用者的理想狀況，也就是我們可以將一個用戶端主機 IP 位址的存取記錄與造訪記錄等同為一位使用者的網站使用行為。在下面章節中，我們將說明實際上使用者與用戶端間的對應關係。

註 20：事實上，之後的交易記錄清單中沒有相同的用戶端主機 IP 位址，是一個特例。在現實的使用行為之中，使用者在一天之中可能會有多次上站的記錄，甚至具有同一用戶

端主機 IP 位址的多位使用者在同一天中也會在不同的時間內上站，所以工作記錄中會留下同一用戶端主機 IP 位址多次上站的記錄，我們無法以用戶端主機 IP 位址的最後一個記錄作為使用者離開網站的憑據。在下面我們將介紹的網站工作記錄分析工具通常是以一個經驗法則，就是如果用戶端主機在一段時間之內沒有繼續對網站伺服器發出需求的話，則視為已經離站。易言之，在某一筆記錄之後一段時間內若沒有與這筆記錄相同用戶端主機 IP 位址的交易記錄，這筆記錄便當作是這個用戶端主機此次的離站記錄，而下一筆相同用戶端主機 IP 位址的交易記錄則是視為下次上站記錄的進站記錄。在一些較精緻的分析工具是可以由使用者根據他們對網站使用行為的瞭解，自行設定這段離站時間長度的經驗值。而這段時間長度與網站所記載的資訊內容與網頁的設計和編排方式有相當密切的關係。

註 21：Analog 網站首頁 <http://www.statslab.cam.ac.uk/~sret1/analog/>。

註 22：wwwstat 網站首頁 <http://www.ics.uci.edu/pub/websoft/wwwstat/>。

註 23：WebTrends 網站首頁 <http://www.Webtrends.com/>。

註 24：D. Backman and J. Rubin, "Web Log Analysis: Finding A Recipe For Success," Network Computing (15 June 1997): 87-93.

註 25：K. Bauer, "Who Goes There? Measuring Library Web Site Usage," Online (January 2000): 25-31.

註 26：同註 25。

註 27：一些研究與廠商宣稱，解決動態 IP 設定的問題可以憑藉網站在使用者端建立的 cookies。在使用者造訪網站之後，網站便利用程式在用戶端電腦中留下一小段稱為 cookies 的資訊，這些資訊可以區辨不同的使用者，再使交易記錄中心增加這個資訊，便可以達到分析出各個使用者的目的。然而這個問題主要是針對一般商業網站的使用者利用個人電腦撥接連上網的行為。但如果是如下面我們將提到的使用者利用開放設置的電腦來與網站連線的情形，便無法解決上述問題。而這種情形在圖書館網站的使用十分常見，因此，在本論文的研究中，我們仍然不使用 cookies 技術。

註 28：同註 24。

註 29：目前這個網頁已經移到 <http://lib.shu.edu.tw/shulib/DBtrial/cdtrial.htm>。

註 30：同註 8，頁 123-124。