

## Internet 資源搜尋系統的發展與運用 Internet Search Tools: Development and Utilization

卜小蝶 **Hsiao-Tieh Pu**

世新大學圖書資訊學系副教授

Associate Professor, Department of Library and Information Studies

Shih Hsin University

E-mail: [htpu@cc.wjtc.edu.tw](mailto:htpu@cc.wjtc.edu.tw)

### 【摘要】

Internet 上多元的搜尋系統是目前網路上最常被利用的服務之一，由於網路上的資源包羅萬象、類型繁多、又隨時變動，若不能好好利用這類系統，尋找資源將有如大海撈針般的困難。但隨著 Internet 的成長，這些資源搜尋系統也急速增加，了解這些系統的設計架構、種類特色、和未來發展，並從中選擇最適當的系統來協助檢索，已變得相當重要。同時對於圖書資訊人員而言，除了熟悉這些系統的運用外，也應開始思考資源搜尋系統在圖書資訊服務的可能應用及影響，以因應未來電子化圖書館的發展趨勢。

### 【Abstract】

One of the most popular services available on the Internet is the various search tools. With the rapid growth and the diversity of network resources, search tools have proved to be most helpful in finding out relevant information. The purpose of this paper is threefold. First, it introduces the development of network information retrieval and discusses the structural design of a typical search engine. It then reviews all types of search tools by cross comparing their generic features, such as content and search capability. Finally it explores the information service issues that arise from the nature of search tools.

關鍵詞：網際網路；網路資源搜尋；搜尋引擎；資訊服務

Keywords：Internet；Networked Information Retrieval；Search Engine；Informa-

## 一、前言

Internet 上多元的搜尋系統是目前網路上最常被利用的服務之一，由於網路上的資源包羅萬象、類型繁多、又隨時變動，若不能好好利用這類系統，尋找資源將有如大海撈針般的困難。正如同傳統圖書資訊服務人員，必須學習如何挑選適當主題的資料庫、熟悉各種查詢指令之後才能勝任資料檢索的工作。在網際網路的時代，如何正確且有效率地利用網路資訊檢索技術查詢到所需資源，也成為人人必須學習的知識。但隨著 Internet 的成長，這些資源搜尋系統也急速增加，了解這些系統的設計架構、種類特色、和未來發展，並從中選擇最適當的系統來協助檢索，已變得相當重要。同時對於圖書資訊人員而言，除了熟悉這些系統外，也應開始思考資源搜尋系統在圖書資訊服務的可能應用及影響，以因應未來電子化圖書館的發展趨勢。

以下第二節首先回顧 Internet 資源搜尋技術的發展，第三節簡述搜尋系統的設計架構，第四節以使用的觀點分析各類搜尋系統の種類及特色，第五節則說明搜尋系統的選擇與評鑑，第六節則進一步探討搜尋系統對圖書資訊服務的衝擊與影響，最後第七節是結語。

## 二、Internet 資源搜尋技術的發展

資訊檢索 (Information Retrieval) 一直是圖書資訊界的重要研究領域，但長久以來並未能在其它學門中發揮其重要性，然而藉由 Internet 的普及與網路資源的激增，以資訊檢索為基礎的網路資源搜尋技術已逐漸成為 Internet 上相當受到重視的技術之一，正如 Datapro 的調查顯示，資源搜尋系統已成為少數熱門新興發展軟體之一 (註 1)。

Internet 資源搜尋技術的發展其實很早，由於早期 Internet 上多是個人資源、且以簡易文字模式展現 (ASCII file)，整體資料量不大，加上又多在 UNIX 作業系統上發展，僅利用 UNIX 所提供的 "grep" 文字比對程式，就足以應付一般查詢需求。然而隨著資源的種類多元化，許多針對資源特性所開發出來的軟體也一一出現，如以 FTP 檔案名稱為主要查詢對象的 Archie (註 2)、以 Mailing Lists Archives 為查詢對象的 LISTSERV (註 3)、以 Telnet 連線的線上資料庫為主要瀏覽對象的 Hytelnet (註 4)，對大量的文字檔案建立索引並提供全文檢索功能的 WAIS (註 5) 等。但是隨著資料量的成長，瀏覽及傳播這些非結構化的文字檔案變得相當困難，於是 Gopher 層級式的瀏覽方式 (Hierarchical) 因為相當容易使用，一提出便大受使用者的歡迎 (註 6)。但是如何快速查詢數萬個 Gopher 站上所條列的選項資訊又成為必須克服的難題之一，因此以查詢這些選項為主的程式如 JUGHEAD、VE-

RONICA 等也被開發出來（註 7）。

Internet 資源搜尋技術發展至此看似完善，但隨著資料型式更多元、資料種類不再侷限於文字檔案，如一些聲音、圖片、影像等多媒體資料的急速成長。此外，也有很多使用者開始覺得 Gopher 層級式的瀏覽方式不夠親切自然，於是取而代之的是以超文件型式、瀏覽多媒體資訊的 WWW 系統（註 8）。然而隨著 WWW 網站的增加，找尋適合的 WWW 站也變得很困難，於是有人將這些網站分門別類的建立起主題目錄（Subject Directory），方便使用者以主題類別瀏覽網路資源，不過由於資源的成長十分驚人，這些主題目錄無法有效地應付分門別類的查詢需求，於是所謂的 Web 搜尋引擎（Search Engine）也就應運而生。這些系統主要提供 Web Page 的全文查詢，雖然以資訊檢索理論觀點來看，這些系統提供的檢索功能還相當簡單，但由於收錄資料、查詢速度、使用層面等都遠比以往的資訊檢索系統表現的更為傑出，而深受使用者重視（註 9）。如 AltaVista 宣稱收集了六十多萬個 Web Server 上五千多萬篇 Web Page、及總計超過一萬四千多個網路新聞群上的三百多萬篇文件，而且每隔幾個月就成長數倍，其使用率更高達每天一千二百多萬人次；又如 InfoSeek 每天也有七百多萬次的查詢筆數，由於從這些系統的運作所獲得的經驗，幾乎重新寫下新的資訊檢索理論，也因此建立起所謂 Real-world Search Engine

的研究領域（註 10）。

雖然這些搜尋引擎可以讓使用者快速地查詢到網路資源，但使用者普遍認為查詢出來的資料品質不佳，很多改進構想因此產生。諸如，美國麻州大學的 Croft 教授就提出所謂十大當前最重要的檢索技術發展方向（註 11）。除了加強傳統索引搜尋技術（Indexing and Searching）、分散式檢索技術（Distributed IR），也包括新興技術如：強調從網路中過濾不具備檢索價值的資料、並萃取重要資訊的資訊過濾技術（Information Filtering）（註 12），針對各種聲音、影像資料所發展的多媒體資訊檢索技術（Multimedia Information Retrieval）（註 13），都成為研究網路資源檢索專家們所亟思解決的問題。另外，更有文獻預測 Internet 資源搜尋技術最終渴望達到個人化資訊服務的目標，屆時所謂的智慧型資訊檢索代理人（Intelligent Information Agent）會被發展出來（註 14），如此，每個人都能根據個人需求命令系統收集、整理資訊，而不再受限於現行系統複雜的指令，以及瀏覽一些不相關的資訊。但在這類 Intelligent Agent 技術尚未成熟之前，我們仍必須了解並善用各種資源搜尋系統，如此才能快速有效地查詢到所需資訊。

### 三、資源搜尋系統架構

若想充份掌握搜尋系統的使用，實有必要先瞭解其系統設計的理念。由於將如第四節所介紹，搜尋系統的

種類相當多，限於篇幅，以下便以搜尋系統中最主要的 Web Search Engine 為例，簡單介紹其系統架構，包括資源自動收集 (Robot)、建立索引 (Indexing)、及檢索軟體 (Search Software) 三部分，及其功能與特性等，相信將有助於讀者更有效利用這些搜尋系統。一般而言，Search Engine 的系統架構和一般的光碟或線上資料庫系統並無太大不同，只是其收集的方式需要相當自動化、收集的對象更為多元大量、而且索引及搜尋的方式得更重視效率。

## 1. 資源自動收集 (Robot)

Robot 是 Search Engine 中負責資料收集的軟體，又稱為 Spider、或 Crawler (註 15)，他可以自動在設定的期限內定時自各網站收集網頁資料，而且通常是由一些預定的起始網站開始遊歷其所連結的網站，如此反覆不斷 (recursive) 的串連收集。這類軟體所關心的無非是收集的資料是否夠新、夠多，對於資料的品質就較難掌控，當然有些 Search Engine 重視資料的"量" (如 Hotbot)，有些則重視資料的"質" (如 WebCrawler)，因此 Robot 的成效不能一概而論。基本上這類軟體尚有許多需要改進之處，譬如多數 Robot 缺乏判斷網頁的相關性，往往造成同一個網頁會重複收集，同時其不斷大量的收集動作也會影響網路的流量。此外，某些型式的資料仍不易被記錄，如一般表格式資料庫中的記錄。

另外有些網站限於資料的隱私性，並不歡迎 Robot 的造訪，於是有所謂的 Robot Standard 出現，規定其擷取資料的權限。對於管理網站的網主而言，藉由使用記錄 (Transaction Log) 可判斷出那些資料或檔案曾被 Robot 擷取過，如 WebCrawler Robot 所在的主機位址為 spidey.webcrawler.com，AltaVista Robot 的名稱為 Scooter，或是檢視是否擷取過 robot.txt 這個檔案等都是很好的線索。

## 2. 建立索引 (Indexing)

雖然網路資源搜尋技術受到相當嚴苛的挑戰，但目前大部分 Search Engine 建立索引的方式和傳統資訊檢索系統仍相當類似，如逐字反轉檔 (Word-Inverted File) 還是相當常見的索引方法，建立索引時字形變化 (Stemming)、大小寫 (Capitalization) 的處理等問題，其實都算是共同的問題。只是在處理這麼龐大的索引檔時，就需要一些較複雜的技巧，如：如何有效率的切割檔案 (Partition)、如何應付頻繁的記錄異動、如何建構最佳化的檔案結構等。一般而言，索引的產生是將網頁中每個 Word 或者 Phrase 存入 Keyword 索引檔中，另外除了來自網頁內容外，網頁作者所自行定義 Meta Tag 中的 Keyword 也常被納入索引範圍。這類 Keyword 索引只能算是 Content-based Word List，並未解決語意的問題 (如同義詞、相關詞)，因此也有一些搜尋系統嘗試採用其它

方式建立索引，如 Excite 就利用所謂的 Concept Indexing 建構索引，配合字典來抽取語意。

### 3. 檢索軟體 (Search Software)

檢索軟體是決定 Search Engine 是否能普遍為人使用的關鍵要素，因為使用者多只能藉由搜尋速度、搜尋結果來判斷一個系統的好壞，而這些工作都屬於檢索軟體的範圍。Search Engine 所提供的檢索功能目前雖不如線上資訊檢索系統完整，但如以相關性排序 (Document Ranking) 檢索結果、提供相似性回饋反覆查詢 (Relevance Feedback) 等功能看，這在一般資訊檢索系統也不多見，其成效雖不一定卓著，但的確有助於使用者瀏覽與查詢網路資訊。其中 Ranking 是一項很重要的特色，也有越來越多的先進技術發展。整體而言，Search Engine 有很多方式來決定網頁的相似性 (Relevancy)，如根據字彙出現位置 (出現在 "TITLE" 或網頁的顯著位置的 Word 較重要)、出現的頻率 (出現頻率太少或太高都必須忽略)、網頁被連結的熱門程度 (愈多人連結的網頁表示愈重要)、及網頁是否被評論過 (被收錄在評論中表示愈熱門) 等，來決定網頁的相關程度，愈相關者就排在愈前面。不過有愈來愈多的駭客 (Hacker) 以破解 Ranking 方法為樂，想辦法將網頁排到最前，其中如 Spamming 的問題 (讓網頁充滿某個 Keyword，如此每次查詢時該網頁就會得到較高的

Ranking) 就經常發生，雖然有些 Search Engine 的設計者會針對這種情形予以懲罰 (Spam Penalty)，但還是有不少問題待克服。

一如 Croft 教授所預測，未來 Search Engine 最重要的發展方向無疑是朝向 Integrated Solution 改進 (註 16)，因此搜尋系統未來將持續整合其它學門技術，如資訊過濾、多媒體檢索、多語文檢索 (Cross-Language Retrieval)、Metadata、智慧代理人等，以發展出更先進的網路資源搜尋技術。

## 四、資源搜尋系統的種類

前述已簡要說明 Internet 資源搜尋系統的基本設計架構，基本上資源搜尋系統的區分方法非常多，有從技術區分，也有依據資源特性 (註 17)。但對一般使用者而言，以搜尋資料的特性來區分資源搜尋系統還是較合適，如有些是用來查詢 Web Page 上的字串，有些是查詢網路新聞中的訊息，有些則是用來尋找個人的 E-mail 住址等 (註 18)。為了兼具資源以及技術特性，本文提出的分類方式如圖 1 所示，首先將資源搜尋系統區分為主題目錄、搜尋引擎、Meta Search、Mega Search 等四類，原則上這四類都是廣義的搜尋引擎，只是其技術特性差距大，因此特別提出以利區分。至於其中所謂搜尋引擎，依其收錄資源特性還可再區分包括 Web、網路新聞、電子郵件、電子論壇等不同的資源搜尋系統。為了讓讀者可以對現行較重

要的搜尋系統有一清楚認識，以下則將這些系統細分如圖 1 之八類分別敘述。

### 1. 主題目錄 (Subject Directory)

在查詢網路資源時，固然可以利用 Web 搜尋引擎獲得完整的資訊，但有時所查詢的主題過於廣泛，若是利用這類系統查詢，很容易就會迷失在大量資訊中，此時就須仰賴經過人工分類組織的主題目錄。這些主題目錄根據用途又可區分為一般性 (General)、最新資源 (What's New)、最

佳資源 (Best of the Web)、及特殊資源 (Specialized Topics) 等類型。一般性主題目錄如眾所週知的 Yahoo 就是其中的經典，透過 Robot 自動收集網路資源及使用者主動提供網站登錄 (每日就高達三千多則)，再由人工分類整理，目前其大小分類已有七、八萬種，對於查詢有明顯類別區分、又希望有較高品質的資訊，利用主題目錄來查詢將有很大助益。表 1 所列為一些目前較常被使用的分類目錄。

雖然分類目錄的建置，傳統上一一直是圖書資訊學界所努力經營的專業

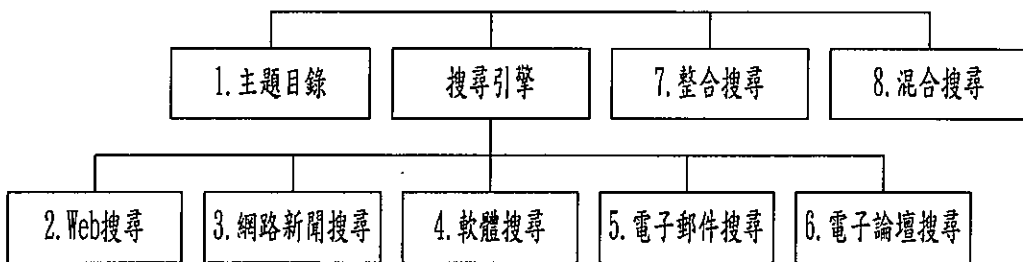


圖 1 搜尋系統的種類

表 1 一般性主題目錄舉例

系統名稱	系統位址 (URL)
Yahoo	<a href="http://www.yahoo.com/">http://www.yahoo.com/</a>
Virtual Tourist	<a href="http://www.vtourist.com/webmap/">http://www.vtourist.com/webmap/</a>
BUBL	<a href="http://bubl.ac.uk/">http://bubl.ac.uk/</a>
W3C Virtual Library	<a href="http://www.w3.org/hypertext/DataSources/bySubject/Overview.html">http://www.w3.org/hypertext/DataSources/bySubject/Overview.html</a>
Excite's NetDirectory	<a href="http://www.excite.com/Subject/">http://www.excite.com/Subject/</a>
Lycos A2Z	<a href="http://a2z.lycos.com/">http://a2z.lycos.com/</a>
Nynex Interactive Yellow Pages	<a href="http://www.niyp.com/">http://www.niyp.com/</a>
Scott Yanoff's Internet Services List	<a href="http://www.uwm.edu/Mirror/inet.services.html">http://www.uwm.edu/Mirror/inet.services.html</a>

領域，但目前網路上的分類目錄卻多由商業機構所建立，僅有少數如BUBL是由圖書資訊學界所直接建置。這一方面說明由於網路資源相當新穎，圖書資訊學界也在探索最佳的整理方法，另一方面也透露出若網路資源有商機，很自然分類整理的工作也會有商業機構代勞。

前述主題目錄提供相當方便的管道以查詢特定主題，不過一般主題目錄都未有時間性的排序，若想要查詢較新的資源，則可嘗試利用所謂的What's New的目錄，如Starting Point就針對最近幾星期內出現的新資源做一摘要條列。其實一部分的主題目錄也會提供What's New這個分類，對於想

快速瀏覽新資源的使用者而言還是相當方便。表2是這類系統的舉例。

雖然詳盡的分類整理有助於瀏覽，但同一類資源過多時，又很難找到最佳的資源，因此就有一些主題目錄將同類資源以優劣等級排序，節省使用者瀏覽判斷的時間，如Magellan就根據資源的深度（收錄範圍是否廣泛、新穎）、使用方便性（組織方式是否易於瀏覽）、吸引力（是否具創新、有趣、熱門等特點）等評鑑標準將資源分為四個等級，雖然這種評比方式相當主觀，但也節省不少使用者過濾資源的時間。表3是其中較為流行的系統。

表 2 最新資源主題目錄舉例

系統名稱	系統位址 (URL)
Starting Point	<a href="http://www.stpt.com/new.html">http://www.stpt.com/new.html</a>
What's New Too	<a href="http://newtoo.manifest.com/">http://newtoo.manifest.com/</a>
NCSA What's New	<a href="http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/whats-new.html">http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/whats-new.html</a>
Netscape What's New	<a href="http://www.netscape.com/escapes/what_new.html">http://www.netscape.com/escapes/what_new.html</a>
What's New on Yahoo	<a href="http://www.yahoo.com/new/">http://www.yahoo.com/new/</a>
Net Happenings	<a href="http://www.mid.net:80/NET/">http://www.mid.net:80/NET/</a>
Net Scout	<a href="http://wwwscout.cs.wisc.edu/scout/">http://wwwscout.cs.wisc.edu/scout/</a>

表 3 最佳資源主題目錄舉例

系統名稱	系統位址 (URL)
Magellan	<a href="http://www.mckinley.com/">http://www.mckinley.com/</a>
Lycos Top 5% Sites	<a href="http://www.pointcom.com/">http://www.pointcom.com/</a>
Yahoo Internet Life Reviews	<a href="http://www.zdnet.com/yil">http://www.zdnet.com/yil</a>
Excite Reviews	<a href="http://www.excite.com/Reviews">http://www.excite.com/Reviews</a>
NetGuide Live's Best of the Web	<a href="http://www.netguide.com/">http://www.netguide.com/</a>
C Net Best of the Web	<a href="http://www.cnet.com/Content/Reviews/Bestofweb">http://www.cnet.com/Content/Reviews/Bestofweb</a>

表 4 特殊主題目錄舉例

系統名稱	系統位址 (URL)
Argus Clearinghouse	<a href="http://www.clearinghouse.net/">http://www.clearinghouse.net/</a>
EEVL (Edinburgh Engineering Virtual Library)	<a href="http://eevl.icbl.hw.ac.uk/">http://eevl.icbl.hw.ac.uk/</a>
FindLaw	<a href="http://www.find-law.com/">http://www.find-law.com/</a>
HealthAtoZ	<a href="http://www.healthatoz.com/">http://www.healthatoz.com/</a>

另外還有一些主題目錄是以特殊學科主題、或特殊用途為主的目錄，前者如 EEVL 是以收錄網路上工程科學資源為主的目錄，這類目錄相當多，其實各個學科領域中的單位團體網站（如學校、研究機構）多會建立類似的主題目錄，只是收錄內容的數量及品質差異很大，如何在眾多同一學科主題目錄中找到最具參考價值的目錄，圖書資訊人員應可以提供協助。此外如 Argus 則是一種收集各類主題目錄的目錄。表 4 是這類系統的一些舉例。

## 2. Web 搜尋 (Web Search Engine)

顧名思義，Web Search Engine 主要是以 Web Page 為搜尋對象，但近來這類 Search Engine 已有涵蓋愈來愈多資料庫的趨勢，如 AltaVista、InfoSeek 等除了 Web Page 外，尚可查詢網路新聞的資料，Lycos 尚可查詢 E-mail 住址等。這些 Search Engine 和傳統線上資料庫查詢軟體有許多相似之處，如提供關鍵語、布林檢索等功能，但也有許多不同的地方，如在資料收集方面，傳統查詢系統是以人工建檔為主，而如第三節所述，Web Search Engine 多由 Robot 之類的程式，自動搜尋網

路上的資源並建立索引；又如查詢指令，傳統多需使用者學習複雜的指令語法，而 Web Search Engine 則多提供自由語句輸入方式等。

Internet 上的 Web Search Engine 種類繁多，表 5 是目前幾個較重要的系統。其中 AltaVista 是以其速度、收錄數量著名，舉例來說，其負責索引建立 (Indexing) 的伺服器所使用的記憶體據說很早就高達 6GB+210GB RAID disk，負責搜尋 (Searching) 的伺服器則有 256MB+4GB，而 Robot 的伺服器則有 1+30GB RAID disk 等，同時其收錄的 Web Page 全文也超過 110 億字 (Word) (註 19)。此外，針對向來為人詬病檢索品質不佳的問題，AltaVista 也已提供如 LiveTopics 的 Term Suggestion 服務，也就是可以根據使用者所輸入的字彙及查詢結果，即時產生其它相關詞彙供選取查詢，這項技術的提出相當受到注意，雖然其使用率及效果仍有待觀察，但已經為 Web Search Engine 開啓新的功能。此外，由於多國語文檢索是各家努力目標，如 AltaVista 就提供了二十多種語文檢索。

另外，Excite 系統則是以檢索精確度見長，其提供如概念檢索 (Concept Search)、相似性回饋 (Relevance Feed



back) 等較高品質的查詢功能, 此外, 最近也提供所謂 Personal Page 的服務, 讓使用者以個人感興趣的主題為查詢範圍, 而非以整個資料庫為查詢對象, 這種以個人化為訴求的服務, 的確有去蕪存菁的效果。還有 HotBot 是以柏克萊大學所設計的 Inktomi 為核心所發展出來的系統, 其號稱是網路上收錄最多 Web Page 的搜尋引擎, 除此之外, 其提供的 NewsBot 服務, 具備新聞媒體資訊 (Newswire) 的過濾功能, 協助使用者獲取品質更佳的資訊。再者, InfoSeek 系統則以商業資料庫檢索服務受到重視, 其也提供了 Relevance Feedback 的功能以及廠商資訊、即時新聞等資料庫, 不過也是少數需要收費的系統之一, 其中一項有趣的應用是所謂的個人化廣告, 系統會根據查詢詞彙顯示適當廣告, 而非如其它系統人人都看相同廣告而失去參考價值。

至於 Lycos 則是其中較早發展出來的系統, 所收錄的 URL 數量也相當驚人, 且其索引深度涵蓋如 Frame、Image

Map 等許多資訊, 因此可查詢到其它系統未納入索引範圍的資訊。其他著名系統還有 Open Text 的 Ranking 效果較其它系統出色, 同時在發展多國語文檢索時也相當積極。WebCrawler 是屬於早期發展的 Search Engine, 雖然收錄的資料量不如前述幾個搜尋引擎, 但其 "少而精" 的收集方式, 讓查詢資料的品質較佳, 對於查詢特定機關團體有很好的效果。綜合上述, 每個 Search Engine 都有其特色, 很難評斷那一套系統才是最佳, 而且大部份的系統都隨時在更新、進步, 一時實難分軒輊。因此使用者不宜只依賴單一系統, 一定得多嘗試幾種, 才能得到較滿意的查詢結果。

### 3. 網路新聞搜尋 (Newsgroup Articles Search)

Internet 上的全文文獻資源除了 Web Page 外, 網路新聞群的訊息也相當重要, 表 6 所列為一些可查詢這類資源的系統。其中以 DejaNews 為例, 其收錄了一萬五千多種新聞群超過 70GB

表 5 Internet 上著名的 Search Engine

系統名稱	系統位址 (URL)
AltaVista	<a href="http://www.altavista.digital.com/">http://www.altavista.digital.com/</a>
Excite	<a href="http://www.excite.com/">http://www.excite.com/</a>
HotBot	<a href="http://www.hotbot.com/">http://www.hotbot.com/</a>
InfoSeek	<a href="http://www.infoseek.com/">http://www.infoseek.com/</a>
Lycos	<a href="http://www.lycos.com/">http://www.lycos.com/</a>
Open Text	<a href="http://www.opentext.com/">http://www.opentext.com/</a>
WebCrawler	<a href="http://www.webcrawler.com/">http://www.webcrawler.com/</a>

表 6 網路新聞群全文訊息查詢系統舉例

系統名稱	系統位址 (URL)
DejaNews	<a href="http://www.dejanews.com/">http://www.dejanews.com/</a>
Reference.com	<a href="http://www.reference.com/">http://www.reference.com/</a>
AltaVista	<a href="http://altavista.digital.com/">http://altavista.digital.com/</a>
HotBot	<a href="http://www.hotbot.com/">http://www.hotbot.com/</a>
InfoSeek	<a href="http://www.infoseek.com/">http://www.infoseek.com/</a>

表 7 Software Search 系統舉例

系統名稱	系統位址 (URL)
Snoopic	<a href="http://www.snoopic.com/">http://www.snoopic.com/</a>
FTPsearch	<a href="http://129.241.190.13/ftpsearch/">http://129.241.190.13/ftpsearch/</a>
Shareware.com	<a href="http://www.shareware.com/">http://www.shareware.com/</a>
Jumbo	<a href="http://www.jumbo.com/Home_Page.html">http://www.jumbo.com/Home_Page.html</a>
ArchiePlex	<a href="http://flosun.salk.edu/archieplex.html">http://flosun.salk.edu/archieplex.html</a>

的全文，提供使用者以發表者、新聞群名、主題等項目查詢。這類系統除了提供一般全文查詢的功能外，也開始積極發展個人化新聞過濾服務，如 SIFT 即讓使用者挑選新聞群或關鍵語建立個人興趣檔 (Profile)，再根據此檔過濾每日大量增加的新聞訊息，擇其重要者提供給使用者瀏覽。

#### 4. 軟體搜尋 (Software Search)

Internet 上的資源除了一般文字資料外，大量且多樣的免費軟體是吸引使用者利用 Internet 的主要原因，面對成千上萬種軟體，如何從中查詢到所需軟體便成為 Software Search 這類檢索系統的重點。表 7 是一些查詢軟體資訊的系統舉例，這類系統囿於軟體資訊相當有限，因此查詢功能至多是可限制查詢範圍為檔案名、domain、主

機等。

#### 5. 電子郵件搜尋 (E-mail Address Search)

尋找個人 E-mail 住址也是 Internet 使用者經常須要查詢的需求之一，表 8 所列為一些較常使用的系統舉例。

其中以 Four11 為例，其提供全球一百多國六百五十多萬人的 E-mail 住址資料庫查詢。又如 Bigfoot 除了可查詢 E-mail 住址外，還可以查到該人的電話及住址，相當方便。

#### 6. 電子論壇目錄搜尋 (Discussion Group Listing Search)

除了上述一些主要的資源檢索系統外，Internet 上還有許多特殊資源值得查詢，如欲查詢特定主題的討論群可利用查詢 Discussion Group Listing 的

表 8 電子郵件住址搜尋系統舉例

系統名稱	系統位址 (URL)
Four11	<a href="http://www.four11.com/">http://www.four11.com/</a>
Bigfoot	<a href="http://www.bigfoot.com/">http://www.bigfoot.com/</a>
Whowhere	<a href="http://www.whowhere.com/">http://www.whowhere.com/</a>
Switchboard	<a href="http://www.switchboard.com/">http://www.switchboard.com/</a>
Internet Address Finder	<a href="http://www.iaf.net/">http://www.iaf.net/</a>
LookUp	<a href="http://www.lookup.com/lookup/search.html">http://www.lookup.com/lookup/search.html</a>

表 9 電子論壇目錄查詢系統舉例

系統名稱	系統位址 (URL)
Liszt	<a href="http://www.liszt.com/">http://www.liszt.com/</a>
Search the List of Lists	<a href="http://catalog.com/vivian/interest-group-search.html">http://catalog.com/vivian/interest-group-search.html</a>
List of Listservs	<a href="http://www.tile.net/listserv/">http://www.tile.net/listserv/</a>
Indiana University Search for Mailing Lists	<a href="http://www.ucssc.indiana.edu/mlarchive">http://www.ucssc.indiana.edu/mlarchive</a>
Directory of Schorlerly and Professional E-Conferences	<a href="http://www.n2h2.com/KOVACS/">http://www.n2h2.com/KOVACS/</a>
L-Soft Listserv List Reference	<a href="http://www.libsoft.com/lists/listref.html">http://www.libsoft.com/lists/listref.html</a>

系統，如Liszt可查詢六萬六千多種討論群，使用者只需輸入主題，即可找到相關的討論群位址及參加的方法。表9即是這類系統的舉例。若欲查詢討論群內容全文，則可利用Listserv之類的查詢軟體，以查詢個別討論群的全文訊息；此外，因多數討論群皆有對應的網路新聞群，也可以利用如DejaNews之類的網路新聞群全文檢索系統。

## 7. 整合搜尋 (Meta Search)

由於Search Engine種類繁多，查詢Web Page時經常需要使用數種Engine，以免遺漏相關訊息，這樣逐一使用性質相同的Search Engine相當浪費時間，因此就有所謂的Meta Search

系統的開發。這類系統提供使用者一次查全的服務，例如MetaCrawler就收集了九種較常使用的Search Engine（如AltaVista、Lycos、InfoSeek等），使用者僅需輸入一次查詢，其即代為查詢九種系統，之後並將結果做一整理（如刪除重複的URL、過濾有安全限制的URL等），再按相關性排序，使用上相當方便有效。表10是一些常用的Meta Search系統舉例，由於這類系統可一次查詢多種Search Engine，因此速度上會較一次查詢一種Search Engine為慢。同時其濃縮摘取結果的方式可能遺漏部分資訊。

## 8. 混合搜尋 (Mega Search)

Internet上的資源種類相當多元，

表 10 Meta Search 系統舉例

系統名稱	系統位址 (URL)
MetaCrawler	<a href="http://www.metacrawler.com/">http://www.metacrawler.com/</a>
Reference Find	<a href="http://m5.inference.com/ifind/">http://m5.inference.com/ifind/</a>
Internet Sleuth	<a href="http://www.intbc.com/sleuth/sleuth.html">http://www.intbc.com/sleuth/sleuth.html</a>
Search.Com	<a href="http://www.search.com/">http://www.search.com/</a>
SavvySearch	<a href="http://quaraldi.cs.colostate.edu:2000/">http://quaraldi.cs.colostate.edu:2000/</a>
All-in-One Search Page	<a href="http://www.albany.net/allinone/">http://www.albany.net/allinone/</a>

資料庫型式不同、查詢介面自然也有所不同，因此使用者必須了解並學習各類資料庫的查詢指令，不僅相當困難而且也頗為費時。Mega Search 這類系統即提供統一的查詢介面，讓使用者同時查詢多種異質性的資料庫，如 NlightN 即提供數百種線上資料庫如 NLM、Knight-Ridder News，其做法類似傳統線上資料庫廠商如 Dialog，提供數百種資料庫而且用同一套查詢方式查詢，所不同處在於前者不僅提供不同資料庫系統的查詢（索引、摘要、電子新聞群、Web Page 等），同時查詢指令也相當簡單，一般使用者可自行檢索，不似傳統線上資料庫需要中介者的協助。這類系統通常還提供線上訂購的功能，使用者在查到資料後，尚可線上訂購全文（若為需付費的資料）或瀏覽全文（若為免費的資料）。

## 五、資源搜尋系統的選擇與評鑑

前面所述的資源搜尋系統種類繁多，而且這些系統經常更新，因此如何從中挑選最適合的系統來檢索資源便相當困難。其實搜尋系統的選擇與

評鑑，和傳統評鑑線上資料庫時所使用的原則頗為相似（註 20），為此以下就其中較為重要的五個方面分別說明。

### 1. 收錄範圍 (Contents and Scopes)

每種資源搜尋系統都有特定的收錄對象與方針，選擇搜尋系統必須先對以下資料特性加以瞭解。首先是資料量的大小，雖然數量多並不能確保查詢品質也能一樣好，但查詢結果可能會較完整。就收錄項目而言，許多搜尋系統不僅收集一種資源，有時還收錄多種不同資源供使用者選取查詢，一般來說，收錄項目愈多對於查詢就愈方便，因為使用同一套指令就可以查詢多種資料庫。就索引深度而言，有些系統是將全文建立索引，有些則擇其中部分資訊建立索引，索引深度愈深查詢就能愈詳盡。就新穎性及更新頻率而言，所收錄的資源是否新穎及是否經常更新，絕對會影響到查詢結果的正確性，因此若搜尋系統更新索引的頻率不高，最好就不要使用，以免查到一些老舊甚至是錯誤的資訊。就索引建立方式而言，現有的

搜尋系統多半採取自動化的索引建立方式，如此在資料維護上可較節省人力；但以目前的技術而言，想要以自動化的方式做到去蕪存菁還有一段距離，而以人工建立索引的方式在資料品質上可能較易掌握。如分類目錄這類搜尋系統就無法完全仰賴自動化索引系統，自動化的索引程式只能協助前半部的資料收集工作，之後還是得由人工逐一篩選。就處理語言而言，Internet上的資源是屬於多國語文，若搜尋系統能處理多種語文，如查詢時可以以使用者熟悉的語文輸入，系統再將相關的資訊不論語文種類一併找出，並以輸入時的語文呈現結果，則對查詢會有很大幫助。最後就提供評論而言，某些搜尋系統尚會提供資源的評論，這對使用者而言或許會有些許參考價值，雖然評論可能流於主觀，但至少應提供摘要性介紹，以免僅能由資源名稱而誤判其內容。

## 2. 檢索功能 (Search Functions)

Internet 資源搜尋系統所提供的檢索功能，其實和傳統線上資料庫大同小異，前者可能缺乏某些複雜的功能（如Set Manipulation），但也具備某些先進的功能（如Relevance Feedback）。每種搜尋系統所提供的檢索功能雖略有不同，但大致包含布林邏輯、切截、片語查詢、相近檢索、限制欄位、重複資料檢測等功能，或進一步提供相似性回饋、關鍵詞檢索、自然語句輸入等較先進功能，同時對於大

小寫、及特殊符號等處理，每個系統作法也略有差異。

## 3. 檢索結果的展示 (Display Results)

查詢結果的展現方式也是評鑑的重點之一，因為若系統呈現的方式不佳則會阻礙瀏覽的效果，例如相關性排序，Internet上資源龐大，若查詢結果只能按字母順序顯示，勢必耗費大量時間和精神去瀏覽，目前一些搜尋系統所採用的方式，即是根據資源與查詢的相關性來排序，愈相關的資料就會排在愈前面。又如有些系統允許使用者自定顯示的筆數，可節省瀏覽時間。而顯示資料的詳細程度也會影響資料的參考價值（如提供註解、列出更新日期）。此外，目前的搜尋系統在顯示查詢結果時皆會以Hyperlink方式展現，如此使用者可在查到結果後，立即連結獲取原始文件或檔案。

## 4. 檢索效率 (Search Performance)

雖然目前搜尋系統的檢索速度都相當快，但檢索效率並不一定能有相同表現。在評鑑檢索效率時可由查全率 (Recall Ratio)、查準率 (Precision Ratio)、反應時間 (Response Time)、及連線容易程度 (Accessibility) 等方向來考量。

## 5. 使用者介面 (Interface Design)

使用者介面的設計也是評鑑搜尋系統一個相當重要的因素，因為若設計不良，即使查詢功能再多、再好，

也可能無法吸引大眾使用。一般而言，搜尋系統的使用者介面是否優良，可從直觀上判斷其是否易於使用（User-friendly）；此外，其是否提供線上輔助說明（如指令的使用）；是否提供FAQ之類的文件（如介紹系統特色）；是否考慮到文字模式的使用者環境（如提供Lynx介面）等都是參考因素。

事實上，在運用這些搜尋系統時，我們會發現相同查詢需求，在不同的搜尋系統上執行，很少會有相同的查詢結果，這主要還是因為每種系統的軟體設計方法都不大相同；此外，每種查詢需求所適用的範圍也不相同，如想查詢某個單位的WWW網站位址，就不太適合利用AltaVista這類資源龐大的大站，因為很容易就找到太多資料，反而不如利用WebCrawler讓查詢範圍縮小而變得較好找。因此，所謂的最佳搜尋系統可能隨應用以及檢索技巧而異。上述的評鑑選擇標準僅能做為參考，使用者在使用時應多嘗試幾種系統，以免掛一漏萬。每種系統皆有其優缺點，在完善的系統出現之前，搜尋系統仍將朝多元化發展，使用者仍需花費心力學習。

由於Internet各類搜尋工具的數量愈來愈多，想知道最新的搜尋工具並且知道一些相關評價，則可參考一些專門收集這類資訊的目錄網站（Index of Search Tools），如表11所列。

此外，瀏覽一些以搜尋系統為主題的網站，也有助於快速的了解最新

發展現況（註21）。基本上上述各類搜尋工具多半是以美國開發的系統為主，其實各國因語文不同、資料內涵不同，往往會有一些以該國資訊為主的搜尋系統出現，如以台灣為例，中文是主要的語文，網路上的資訊又多未收錄在上述那些國際級的系統中，因此目前已發展不少以台灣網路資訊為主的搜尋系統，主題目錄如蕃薯藤Yam（<http://taiwan.csie.ntu.edu.tw/b5/yam>）、搜尋引擎如GAIS（<http://gais.cs.ccu.edu.tw/cwww2.html>）都是相當熱門的系統。

## 六、資源搜尋系統對圖書資訊服務的衝擊

在瞭解Internet資源搜尋系統的設計與發展後，對於圖書資訊人員而言，除了熟悉這些系統的使用外，也應開始思考資源搜尋系統在圖書資訊服務的可能應用及影響。面對網路科技的挑戰，圖書資訊界對於資料的電子化（如Document Delivery）、宜人介面的設計（如Web OPAC）、資源組織整理規範的制定（如Metadata）等都投注相當心力，但對於Internet資源檢索的運用卻始終不甚積極，雖然有一些由專業圖書資訊人員所建立的搜尋系統，如NetFirst、Argus Clearinghouse、EEVL等，但其效果顯然不如一些知名系統如Yahoo、AltaVista等（註22）。在資源搜尋系統的發展上，多數圖書館的角色仍是“利用”多於“貢獻”。對圖書資訊服務而言，這些系統無疑會

表 11 一些收集搜尋工具的目錄網站

系統名稱	系統位址 (URL)
eDirectory	<a href="http://www.edirectory.com/">http://www.edirectory.com/</a>
The List of All Internet Search Engines and Online Directory	<a href="http://www.beherenow.com/spider/links.htm/">http://www.beherenow.com/spider/links.htm /</a>
Search Engines Page	<a href="http://www.vbe.com/~comsmri/searche.htm/">http://www.vbe.com/~comsmri/searche.htm/</a>
Search Engines on the WWW	<a href="http://www.iglou.com/zamboni/search.html/">http://www.iglou.com/zamboni/search.html /</a>
Search Engines, Directories, ...	<a href="http://www.exxnet.com/wsse.htm/">http://www.exxnet.com/wsse.htm/</a>
Inter-Links	<a href="http://www.nova.edu/Inter-Links/">http://www.nova.edu/Inter-Links/</a>

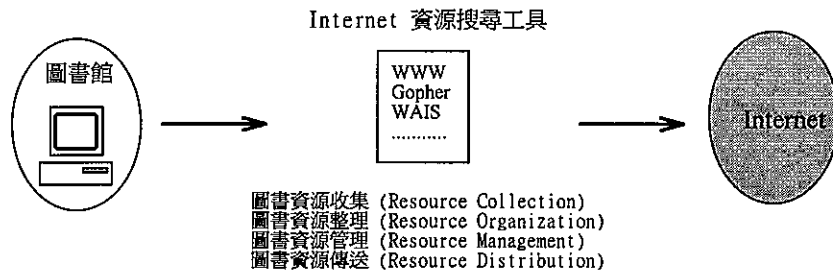


圖 2 Internet 資源搜尋系統對圖書資訊服務的衝擊

有很大的助益，不過當然也會帶來一些衝擊。這些可能的影響或許可以如圖 2 所示，從圖書資源收集 (Resource Collection)、圖書資源整理 (Resource Organization)、圖書資源管理 (Resource Management)、圖書資源傳送 (Resource Distribution) 等四個電子化圖書資訊服務的角度來加以分析：

### 1. 圖書資源收集

#### (Resource Collection)

傳統的圖書資源徵集需要相當多的人力與物力的協助，現在藉助 Internet 資源搜尋系統，收集 Internet 上的資源已相當方便。對很多圖書館而言，In-

ternet 的資源比起現有館藏資源更為豐富有趣，而且收集這些資源相當簡單經濟。譬如利用 WWW 收集資源時只須收集連結資訊 (Links)，而不需真正儲存資料本身 (Resource)。如此一來，既省空間，且資料的時效性也比較高，無須隨時注意資料是否更新。不過資料儲存在他處，也使得資料的品質與系統穩定度無法加以控制。另外透過超媒體與多媒體的整合運用，可以大大提高館內文件的品質與趣味，對讀者更具吸引力。基本上有了 Internet 資源搜尋系統，圖書館收藏的資源並不需要一車車從大門運送，很多時候只是圖書館員花點時間在網路

閱覽，以及在電腦建立一些資源目錄即可。

## 2. 圖書資源整理

### ( Resource Organization )

當圖書館在從 Internet 上收集到相當的資源後，便需要好好加以整理。這種整理在傳統上就是所謂的分類編目工作。網路上有很多資源目錄，對圖書館員而言有相當高的參考價值，另外一些圖書館的 Internet 資源分編結果也可以參閱。只是現行圖書館以傳統圖書分類法對所收藏文獻加以分類、再以 MARC 機讀格式建檔，以提供讀者利用 OPAC 查詢。若以這類方法來處理 Internet 上的資源，基本上並不太恰當，這一方面是因為現有的圖書分類法或編目規則並不適用於整理這些電子資源，因為這些資源的性質和傳統圖書相當不同；同時對使用者而言，要了解圖書館的組織分類方式還是有些困難。然而由於尚未有一國際認同的組織分類標準出現，如何讓組織分類的方法更合宜，且使用者也能感到較為熟悉易用，則是整理分類時的重點。

## 3. 圖書資源管理

### ( Resource Management )

利用 Internet 資源搜尋系統來管理所收集到的資源，在系統建構及維護上都相當容易，基本上並不需要特別再撰寫程式。此外在資料管理上，由於這些收集到的資源其實是分散在全

球，由各地一起來維護，任何時候想利用這些資訊時隨時都可以連結上去。不過因為資料是放置在他處，多少會有些不便。以利用 WWW 建立資源目錄為例，自行建立並不太難，即使是個人也可以嘗試藉此建立電子圖書館，而圖書館也可以感受到收藏全世界資訊的成就感。但是就整體典藏作業而言，因為很多資源是電子化且非本地開發，在統計與維護上需要有新的作法。

## 4. 圖書資源傳送

### ( Resource Distribution )

Internet 資源搜尋系統除了可以協助圖書館收集、整理、管理資源外，而且還可以將這些資源輕易就傳送到讀者手中。電子化圖書資源傳送具備速度快、全球性、直接性、及個人化等許多優點。譬如可以協助讀者在很短時間內取得原件全文，且可以同時比較不同單位的收集，可讓圖書館的服務品質更為提高。不過高品質的服務需要館員具備純熟的資訊技術以及豐富的學科知識。網路時代的資訊服務其難度顯然愈來愈高，館員也同樣愈來愈難為，所以一般圖書館員必須不斷充實專業知識。

## 七、結語

善用 Internet 資源搜尋系統，對於掌握與利用 Internet 資源有很大的幫助。本文以使用的觀點介紹了很多種資源搜尋系統，也探討其系統架構，



希望對讀者能有些許助益。學習Internet資源搜尋確實較學習傳統線上資訊檢索困難，Internet的資源來源龐雜、缺乏組織、且又經常變動，不似傳統線上資料庫的產生、組織、傳播都有一定的規則可循，因此不論是提供資訊服務的人員或是使用者本身，都需要持續學習及了解各種搜尋系統的使用與發展，如此才能充份掌握Internet資源，而這些技能都需要長時間的經驗累積，無法一蹴可及。

#### 註釋

- 註 1 : Robert R. Mueller, "Information Retrieval Tools for the Internet, " In Datapro Management of Applications Software. Delran, N.J. : Datapro Research Corp., 1997.
- 註 2 : Peter Deutsch, "Resource Discovery in an Internet Environment - the Archie Approach," Electronic Networking: Research, Applications and Policy 2 (Spring 1992): 45-51.
- 註 3 : Knapp Leslie, "Communicating with Lists: A Beginner's Guide to Listservs and ListProcs," The Serials Librarian 25 (1995): 33.
- 註 4 : Scott Peter, "HYTELNET as Software for Accessing the Internet: a Personal Perspective on the Development of Hytelnet," Electronic Networking : Research, Applications and Policy 2 (Spring 1992): 38-44.
- 註 5 : Natalie Oakes Sturr, "WAIS: an Internet Tool for Full-text Indexing," Computers in Libraries (June 1995): 52-54.
- 註 6 : 卜小蝶，「Internet 資源搜尋系統與圖書館資訊服務-以Gopher為例」，中國圖書館學會會報 53期(民83年12月)，頁83-109。
- 註 7 : Louise McGillis, "Gopher Searching Using VERONICA," The Reference Librarian 41/42 (1994): 25-36.
- 註 8 : 吳毅成，「WWW全球資訊系統之介紹及其展望」，資訊與教育雜誌(民83年8月)，頁2-11。
- 註 9 : Michael Lesk, "The Seven Ages of Information Retrieval,"  
<<http://community.bellcore.com/lesk/ages/ages.html>>.
- 註 10 : ACM SIGIR'97 會議今年即增加以 Real-world Searching Engines 為題的討論會，與會者的反應相當熱烈。
- 註 11 : W. Bruce Croft 教授在 D-Lib Magazine (Nov. 1995) 以 "What Do People Want from Information Retrieval" 為題，提出所謂十大當前最重要的檢索技術發展方向，依序包括 Integrated Solutions, Distributed IR, Efficient Indexing and Retrieval, "Magic", Interfaces and Browsing, Routing and Filtering, Effective Retrieval, Multimedia Retrieval, Information Extraction, Relevance Feedback 等<<http://www.dlib.org/dlib/november95/11croft.html>>.
- 註 12 : Nicholas J. Belkin and W. Bruce Croft, "Information Filtering and Informa-

- tion Retrieval: Two Sides of the Same Coin?" Communications of the ACM 35 (Dec. 1992): 29-38 ; 卜小蝶, 「網路資訊過濾技術與個人化資訊服務」, 在 21 世紀資訊科學與技術國際學術研討會, (台北市: 世界新聞傳播學院圖書資訊學系, 民 85 年 11 月 7~9 日), 頁 339-350。
- 註 13: 曾元顯, 「多媒體資訊檢索技術之探討」, 在 21 世紀資訊科學與技術國際學術研討會, (台北市: 世界新聞傳播學院圖書資訊學系, 民 85 年 11 月 7~9 日), 頁 281-298。
- 註 14: Mark Nisson et al., "Intelligent Agents: A Technology and Business Application Analysis," Nov. 1995, <<http://www.mines.unancy.fr/~gueniffe/ueniffe/CoursEMN/I31/heilmann/heilmann.html>>.
- 註 15: Martijn Koster, "World Wide Web Robots, Wanderers, and Spiders," <<http://www.webcrawler.com/mak/projects/robots/robots.html>>.
- 註 16: 同註 11。
- 註 17: Alan Poulter, "The Design of World Wide Web Search Engines: a Critical Review," Program 31(April 1997): 131-144.
- 註 18: Internet 資源搜尋系統的一般性介紹可參考以下文獻: Gus Venditto, "Search Engine Showdown," Internet World 5(May 1996): 79-86 ; R. Scoville, "Find It on the Net," PC World 14 (Jan 1996): 125-130. <<http://www.pcworld.com/reprints/lycos.htm>> ; L. Webster and K. Paul, "Beyond Surfing: Tools and Techniques for Searching the Web," Feliciter 42 (Jan 1996): 48-54. <<http://mgi.com/melick/it96jan.htm>> ; I. Winship, "World Wide Web Searching Tools," Vine 99 (1995): 49-54. <<http://www.bubl.bath.ac.uk/BUBL/IWinship.html>> ; Ross Tyner, "Sink or Swim: Internet Search Tools & Techniques," <<http://www.sci.ouc.bc.ca/libr/connect96/search.htm>> ; J.Liu, "Understanding WWW Search Tools," <<http://www.indiana.edu/ibrcsd/search/>> ; J.Brimingham, "Internet Search Engines," <<http://www.stark.k12.oh.us/Docs/search/>> ; R.Brownlee, Web4Lib Archive: Search Engines - Reference List. April 9, 1996. <<http://www.lib.berkeley.edu/Web4Lib/archive/9604/003.html>>.
- 註 19: 簡立峰等, 「透視 Internet 資源發掘與搜尋」, Online-Today: 電子資料庫簡訊 12 期 (民 85 年 6 月), 頁 2-11。
- 註 20: Internet 資源搜尋系統的評鑑可參考以下文獻: Susan Feldman, "Just the Answers, Please: Choosing a Web Search Service," <<http://www.infotoday.com/searcher/may/story3.htm>> ; M. P. Courtois et al., "Cool Tools for Searching the Web: an Update," Online 20 (May/June 1996): 29-36. ; K.Campbell, "Understanding

and Comparing Search Engines,"  
<<http://www.hamline.edu/library/links/comparisons.html>> ; "Evaluation of Selected Internet Search Tools,"  
<<http://www.library.nwu.edu/resources/internet/search/evaluate.html>>;  
陳秀美，「檢索工具的選擇與比較，

資訊傳播與圖書館學 2 卷 4 期（85 年 6 月），頁 41-63。

註 21：A Webmaster's Guide To Search Engines.

<<http://searchenginewatch.com>>;  
Search Insider.

<<http://www.searchinsider.com>>.

註 22：同註 17。