

網路資源搜尋工具(上) Web Search Tools (I)

謝寶媛

Pao-nuan Hsieh

國立台灣大學圖書資訊學系副教授

Associate Professor

Department of Library and Information Science

National Taiwan University

E-mail: pnhsieh@ccms.ntu.edu.tw

【摘要 Abstract】

網路資源搜尋工具可以分為關鍵字查詢式的搜尋引擎和整合式搜尋引擎，以及主題分類式的網站指南。本文依類分別介紹常用的網路資源搜尋工具及其查詢策略，並列表彙整分析各搜尋工具之特色，供選用與查詢時之參考。

Finding information on the web combines keyword searching in search engines and meta-search engines, with web directories to locate information on a topic via the Internet. The paper explores a variety of search engines, meta-search engines, and web directories in order to help you gain skills in conducting research on the Internet. Search strategies are discussed and comprehensive comparison among search engines and meta-search engines are tabulated.

關鍵詞 Keyword

網路資源搜尋工具 搜尋引擎 整合式搜尋引擎 分類目錄

Web search tools; Internet search tools; Search engines; Meta-search engines

壹、前言

網路資源包羅萬象，數量龐大，是研究、教學、生活的良師益友。有些人視上網找資料為最佳捷徑，有人則是非不得已不為。有人可以在網路上悠遊自在，有人是屢戰屢挫，而且很多時候挫折不是來自找不到資料，而是找到了太多資料。到底身為參考館員與資訊專家的我們該怎樣善用網路資源呢？第一步當然是熟悉網路資源搜尋工具與搜尋技巧，第二步則是將這門知識有系統地廣為傳佈，讓每位讀者都成為高效能的資訊公民。

大家一提到上網找資料，無不想到搜尋引擎。甚至很多人認為上網找資料，不就是找個搜尋引擎，打上關鍵字，就大功告成了。沒錯，只要上搜尋引擎打上關鍵字，如果沒有拚字錯誤，要找不到資料的機會還真是不太多，可是要很有效率地找到相關有用的資料，不用花很多時間過濾資料，卻不是那麼簡單。

必也正名乎！在國內一般都把所有網路資源搜尋工具統稱為搜尋引擎，這是不太恰當的。搜尋引擎只是網路資源搜尋工具之一，是指像 Alta Vista 和 Google 等，專門蒐集網路上成萬上億的網頁資料，並將全文加以分析編製成索引，以供網頁全文檢索之服務。那麼除了搜尋引擎之外，還有那些網路資源搜尋工具呢？

網路資源搜尋工具基本上是遵循 Yahoo 模式，經由主動登錄和人為篩選方式，蒐集數以萬計的網站資料，整理成主題式的分類目錄，讓使用者逐層瀏覽分類目錄，找到符合需求的網站；同時建構索引資料庫，方便不熟悉分類架構的使用者可以關鍵字找到適合的分類主題與網站。這種類型的網路資源搜尋服務是謂分類目錄(subject catalog)，Yahoo 是典型代表。

網站指南整理的是網站的資訊，通常一個網站是由數個、數十個網頁所組成，每個網頁各有標題

各有特色，因此第二種類型的網路資源搜尋服務就是以網頁為蒐集對象，提供關鍵字查詢服務，這就是大家習稱的搜尋引擎(search engines)。搜尋引擎是派遣機器人(robot, spider, worm)到網路上漫遊尋找網頁，自動蒐集網頁的連結(links)、標題(titles)、和文本，再將搜尋到的網頁文件與各個連結資料讀取回來組成資料庫，然後將資料庫中的文件加以整理析出編製成索引(Index)。當使用者輸入關鍵字進行查詢時，就是針對搜尋引擎自建的索引資料庫做查詢的工作，將使用者輸入的關鍵字去比對資料庫的索引，把符合的資料全數搜尋出來，輸出為查詢結果。搜尋引擎主要是提供網頁全文檢索，事實上很多搜尋引擎也提供分類目錄，但是這種分類目錄和前面介紹的網站分類目錄不同，是網頁分類目錄。換句話說，在各個主題類目下所排列的，不是網站名稱，而是屬於這個類別的大量網頁標題與位址。由於網頁數量相當龐大，所以網頁分類目錄很難達到分類瀏覽的功用，但是可以限定在同一類別中做關鍵字查詢，使查詢結果更為精確相關。有些搜尋引擎則是在各相關類目下提供精選網頁，提供人為篩選的高品質網頁。

由於每個搜尋引擎都是自建資料庫，因此各個搜尋引擎的資料庫不僅規模互異，內容更是各有千秋，所以想要完整搜尋網路資源時，千萬不能只利用一家搜尋引擎，而應該貨比三家，才不會有遺珠之憾。鑑此，第三種網路資源搜尋服務便應運而生，那就是整合式搜尋引擎(meta search engines)。既然使用者必須一次查詢多個搜尋引擎，每個搜尋引擎支援不同的查詢功能，如果使用者只要在整合式的查詢界面中輸入查詢策略，而由系統自動將使用者的查詢策略送到各個搜尋引擎去查詢，再將查詢結果依相關性或搜尋引擎依序條列，不是可以節省使用者很多的時間。整合式搜尋引擎所提供的正是這種服務，有些可以讓使用者自行指定欲查詢的搜尋引擎和查詢順序，是一般讀者

經常忽略的一種網路資源搜尋工具。

尋引擎，到底有哪些差異呢？請參見表一之彙整。

主題瀏覽式的分類目錄和網頁全文檢索的搜

表一：分類目錄與搜尋引擎之比較

	分類目錄	搜尋引擎
資料蒐集方式	利用登錄或電腦輔助方式蒐集網站資料，再由人工加以分類，並由人工撰寫或自動摘要方式提供網站說明。	利用機器人自動到網路上蒐集網頁資料，並加以分析整理，製成索引資料庫。
資訊品質	經由人工檢視、過濾，品質較一致。	由程式依制式準則判斷，品質良莠不一，難控制。
查詢方式	只要約略知道主題類別，就可以依分類層次找到所需資訊；也可以關鍵字尋找。	需要確知關鍵字或片語，及其間關係，才能找到精確的資料。
查詢結果	網站。 資料品質高，數量在可控制範圍，可以瀏覽挑選相關資料。	網頁。 資料品質難控制，資料數量通常多到無法控制，很難經由瀏覽挑選相關資料。
適用情況	想要對某一主題有概要性的瞭解。 定期瀏覽掌握新知，沒有確切的資訊需求。 希望找到經由人工篩選的優質網站。 例：British literature space exploration	所要尋找的資料已有明確定義，能以關鍵字或片語的組合來描述。 希望能從各種不同來源找尋廣泛的資訊。 例： Charles Dickens Mars Pathfinder
代表網站	Yahoo 是分類目錄的代表 Argus Clearinghouse 大英百科之 Internet Guide 蕃薯藤 新浪網 搜狐	Alta Vista 是網頁全文檢索的始祖 Excite Google HotBot Lycos Openfind

綜言之，網路資源搜尋工具可以概分為搜尋引擎(search engines)、整合式搜尋引擎(meta-search engines)和網站指南(web directory)三大類，茲分別詳加介紹。

貳、搜尋引擎

搜尋引擎(search Engines)，亦有人稱為索引式查詢工具(searchable Indexes)主要是由三部分所組成：機器人(robot)、資料庫、和網頁分類目錄(subject Catalog)。

所謂的機器人是一種能自動在 WWW 超文件結構中讀取文件，並拜訪所有相關連結的電腦程式。由於搜尋引擎是靠站主主動登錄和機器人拜訪

的方式來蒐集網頁資訊，還是有時差，不是最即時的資訊，所以有時候透過搜尋引擎找到的資料，其連結已無效，據說 Alta Vista 的機器人必須花六個月的時間才能遊歷一趟，但是搜尋引擎是不斷地在更新其索引資料庫這是可以確定的，有可能找到的資訊是最即時的，這也是可以肯定的，至於是不是每一筆都是最即時的，那可就不一定了。還有一點需要說明的是，搜尋引擎的機器人是不能搜尋到網路上的所有網頁資訊的，因為有些網站是不歡迎機器人拜訪的，有些由資料庫產生的網頁文件機器人也無法蒐集不到，PDF 文件內容機器人也無法分析，還有些由 javascript 產生的頁框機器也是無法搜尋的。所以，儘管搜尋引擎的資料庫容量再

大，還是沒有辦法完全掌握 Internet 上的所有資訊。

搜尋引擎，所搜尋的是網頁，所查詢到的結果也是網頁資訊，是為網頁全文檢索工具。事實上很多搜尋引擎也提供分類目錄瀏覽和查詢，但是這種分類目錄和蕃薯藤、Yahoo 等的分類目錄是不同的，因為蕃薯藤和 Yahoo 等網站指南類的網路資源搜尋工具所整理分類的是網站，而搜尋引擎所整理分類的則是網頁。換句話說，搜尋引擎各個主題類目下所排列的，不是網站名稱，而是屬於這個類別的大量網頁位址。由於網頁數量相當龐大，所以搜尋引擎的分類目錄很難達到分類瀏覽的效果，但是可以限定在同一類別中做關鍵字查詢，使查詢結果更為精確相關。有些搜尋引擎則是在各相關類目下提供精選網頁，以提高資訊品質。

一、搜尋引擎之收錄範圍

每個搜尋引擎都有其收錄範圍，在使用每個搜尋引擎時，應該瞭解其收錄範圍，才能知道是不是選對了工具，可不可以找到所需要的資訊。搜尋引擎的收錄範圍，可以從地域、語文、網站類型和專業性四個方面來看，通常我們可以在 Help 中找到有關收錄範圍的說明。

(一)地域

大部分的搜尋引擎是搜尋全球網際網路的網站或網頁，尤其是大型的搜尋引擎，如 Alta Vista 以 Unicode 處理各種不同語文的文件。另有些網路資源搜尋工具則是以特定的地域範圍為對象，如來科思香港、來科思亞洲，雅虎中文、雅虎台灣、雅虎香港、雅虎大陸等。有些較小型的搜尋引擎則是專門收錄某一地區的資訊，如歐洲、北歐、台灣、香港。全球性的搜尋引擎收錄範圍雖然廣泛，卻很可能不如地域性搜尋引擎的內容豐富、深入、完備。

(二)語言

中文網站往往使用不同的語文編寫，台港地區常用的當然是 Big5 繁體中文；大陸和新加坡則是用 GB 簡體中文，因此，有些搜尋引擎只收錄繁體中文網站的資訊，特別是台灣與香港地區的搜尋引擎。有些搜尋引擎則主要收錄簡體中文網站的資訊，尤其是大陸地區的網路資源搜尋工具。當然也有些網路資源搜尋工具是繁簡體中文網站兼收的。

另外，英文網路資源搜尋工具，有些是兼收各國的資源，有些則是以美國為主，有些是偏重歐洲資訊，也有兼收亞洲、中文網站資訊的。

(三)網站類型

在網際網路的網站中，目前以 WWW 網站最為普遍，可以瀏覽網頁和提供超文件連結。此外，尚有 FTP、Gopher、BBS、新聞群組、圖片、影像等各種資料類型的網站。大型網路資源搜尋工具通常能夠搜尋各種形式的網站資訊，有些則是提供選項由使用者決定欲搜尋的網站或資訊類型。但是也有相當多的網路資源搜尋工具只整理或搜尋 WWW 網站，無法搜尋其他形式的網站資訊。

(四)專業性

一般常見的搜尋引擎是綜合性的，廣泛收錄各個學科、各個主題、各行各業的資訊。但是也有一些搜尋引擎是專門收錄特定學科、特定主題、特定行業的資訊。如醫學資訊查詢、新聞追蹤、企業查詢、人名查詢、電子郵件地址查詢等等。

專業搜尋引擎對特定資訊的查詢非常有用，我們可以利用網路資源搜尋工具的分類目錄或是整合式搜尋引擎指南，得知有那些適用的專業搜尋引擎，如 Yahoo 的 reference。

二、搜尋引擎的查詢方式

搜尋引擎查詢的方式，可以分為兩種：

(一)關鍵字搜尋：

搜尋引擎多半是根據文件的標題 (title)、MetaTag 中的描述來決定索引關鍵字；有些則是以

文件的前數百個字或是前幾行中出現頻率最高的字詞當作索引關鍵字；有些則是計算整篇文件中字詞的出現次數，找出出現次數最高的字詞來當作索引關鍵字。Alta Vista 可說是關鍵字搜尋引擎的代表。

使用者查詢關鍵字索引式的搜尋引擎時，必須自己界定所有相關的關鍵字或片語，才能找到完整、相關的資料。當使用者輸入關鍵字或片語時，搜尋引擎便根據使用者輸入的單字或片語去比對關鍵字索引，再將符合的輸出為結果。為提高查詢結果的精確度，關鍵字查詢式的搜尋引擎都提供複雜的查詢技巧和限制查詢範圍的方法。

(一) 概念式搜尋

這類搜尋引擎先將資料庫中的文件加以分類整理，當使用者輸入關鍵字查詢時，搜尋引擎先根據使用者輸入的字或詞搜尋相關之概念 (concept)，決定搜尋的範圍，再將相關的結果呈現給使用者。換句話說，概念式搜尋所查詢的是字義而不是字形。Excite 是概念搜尋的典型代表。由於概念式搜尋已將相關的關鍵字組合成概念，因此不需要太複雜的查詢策略，通常只要輸入明確的關鍵字即可。

如果所要查詢的資料非常明確時，建議先利用概念式搜尋找到相關的資料，再利用關鍵字索引做廣泛的搜尋。

如果所要查詢的資料不太容易界定歸類，建議直接利用關鍵字搜尋引擎，但是最好能儘量將相關的關鍵字一一輸入查詢，才能得到較完整適用的資料。

三、搜尋引擎的個性

您是否有鍾愛的搜尋引擎，還是隨心情看天氣，每次用不用的搜尋引擎呢？您是怎樣決定使用那一個搜尋引擎的？在 About.com 的 Web Search 專欄中，Chris Sherman 說搜尋引擎也跟人一樣，

是有個性的，有優點也有缺點，有令人讚賞的特質，也有令人氣憤的缺點。您知道您經常用的搜尋引擎是什麼個性嗎？①

(一) Alta Vista 是飽學之士 (The Motley Savant)

如果您向一位博學多聞的學者請教一個問題，肯定不會失望，一定可以得到很多豐富的資訊。沒錯，Alta Vista 正是如此，因為 Alta Vista 建有超大型的網頁索引資料庫，可以滿足使用者完整搜尋的需求，所以無論您丟什麼關鍵字給 Alta Vista，都可以得到不少的結果；實在一點的說法是，大多數時候所得到的查詢結果是超乎您想像的多。

資料庫夠大是 Alta Vista 的優點，也是缺點。因為利用搜尋引擎找資料，不在找到的資料量多，而在相關度高。想想看在這麼大的資料庫中如何找到符合需要的資訊，每個搜尋引擎都有一套相關性的計算方法，Alta Vista 也不例外，而且迭有更新。最新推出的 raging search 就結合傳統的相關性分析方法和連結分析來判斷查詢結果。

總言之，Alta Vista 是完整蒐集各方面資訊的很好選擇，但是如果希望找到的是較為精準的資訊，不妨試試其他的搜尋引擎。

(二) Excite 是導遊 (The Concierge)

當您問飯店旅遊櫃台的先生，這個城市有何可看之處，他會先瞭解您有興趣的是美術館還是登山，再提供您建議。Excite 正是這樣，不只是將使用者輸入的查詢字詞到資料庫做比對，更嘗試瞭解使用者的資訊需求，分析查詢策略，做概念對映 (concept match)。所以，Excite 在查詢有關事件、地名、公司機構或是球隊等團體資訊時，最是好用。當我們需要人工智慧提供查詢建議時，Excite 確實可以發揮最高效益，若只是做一般關鍵字查詢，還是試試其他的搜尋引擎吧！

(三) Hotbot 是人氣先生 (Mr. Popularity)

HotBot 和 Alta Vista 一樣，建有大型資料庫，

不同的是 HotBot 查詢結果之前列有 Top 10 sites，這十筆不是最符合查詢策略的資料，卻是最受歡迎的十筆熱門資訊。HotBot 是利用 Direct Hit 人氣引擎(popularity engine)的技術，來找出這十個最熱門的網頁。基本假設是越受歡迎的連結其內容越相關，所以是以連結點閱的次數計算相關性。下次用 HotBot 時，別忘了進 Top 10 sites 去看一下同好們都點閱那些網頁。但是每個人的興趣需求各不相同，除了十大熱門網頁外，HotBot 更提供既深入又廣泛的資訊，是完整搜尋的好工具。所以，如果您不僅想要找相關的資訊，更想知道別人都逛那些網頁，HotBot 是不錯的選擇。

(四) Go (Infoseek) 是參考館員 (The Reference Librarian)

Go (Infoseek)就像是圖書館的參考館員，無論您問多麼困難多麼模糊的問題，她都可以提供很豐富的資訊來源，而且有很多是您壓根兒沒想到的。Go 提供兩種查詢結果，一個是人工評選的網站指南，另一個是相關網頁。Go 會將同一網站的網頁聚集，只顯示一筆，以 more result from URL 來進一步顯示，也提供結果再查詢(search within results)的功能，就像參考館員在一旁建議如何縮小查詢範圍以得到更精確的資訊。所以，如果您還沒有慣用的搜尋引擎，也許可以試試 Go。

(五) Lycos 是機器人(Part Man, Part Machine)

Lycos 是老資格的搜尋引擎之一，由傳統的機器自動編制的索引經由 Open Directory 計畫轉型為人工編輯的指南。換句話說，Lycos 的查詢結果是由編輯親自檢視分類的網頁。人工分類還是比電腦分類來得強，所以 Lycos 的查詢結果通常會比其他搜尋引擎要令人滿意。所以，如果您希望查詢結果能依照專家編輯的主題分類來呈現，Lycos 是您可以嘗試的選擇。

(六) Yahoo 則是收藏家(The Eccentric Collector)

Yahoo 是網站指南，意味著其內容完全是由編

輯人工檢視歸類的。可以很方便地沿著分類架構逐層瀏覽找到相關的網站。Yahoo 的查詢服務是比對網站目錄中有沒有符合需求的分類主題，然後按分類架構列出查詢結果。如果沒有辦法在分類目錄中找到符合的主題，Yahoo 會自動切換去查 Inktomi 資料庫(類似 HotBot 的資料庫)，可以找到符合需求的網頁資訊，而且所查到的網頁遠超過 Yahoo 所建網站目錄下的網頁資訊。但是與 HotBot 不同的事，Yahoo 的查詢結果不會先列出熱門網頁。儘管 Yahoo 非常有名，但是其資料庫還是比其他搜尋引擎的規模小多了，而且連結的更新速度也不夠快。所以，Yahoo 是想要某一特定主題有那些網站的好工具，卻不是完整資訊搜尋的好建議。

除了上述這六個搜尋引擎外，在國外頗受好評的 Northern Light 和第二代搜尋引擎的新寵 Google，他們的個性又是如何呢？

(七) Northern Light 是學科專家

如果 Go(Infoseek)是參考館員，那麼 Northern Light 則是當之無愧的學科專家。Northern Light 會將查詢結果依主題整理放在資料夾中，使用者可以按主題打開資料夾瀏覽相關資訊，而且每個網站只列出一筆，但是找到符合的網站時可以點選 More results from this site 來查看其他相關網頁，所運用的仍然是資料夾的主題聚集觀念。

(八) Google 專業顧問

採用連結分析來計算查詢結果的相關度，查詢結果幾乎都能令使用者，甚至是剛好就是使用者想找的資訊。就像是專業顧問根據顧客的需求量身訂做一般。

四、查詢策略

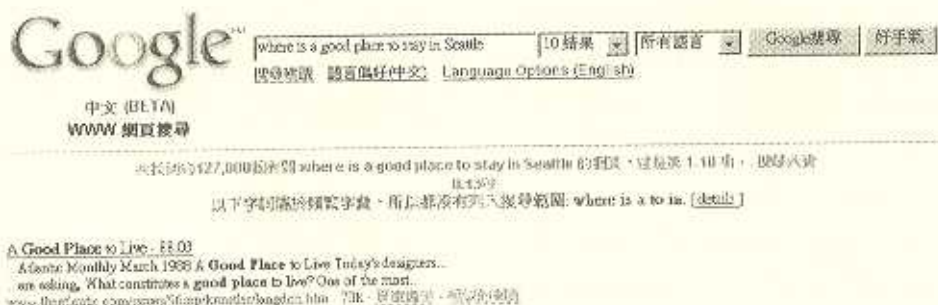
利用搜尋引擎蒐集資料的成功關鍵是明確的查詢字詞加上精準的查詢策略，接下來就根據查詢結果的精確程度，逐步介紹可資利用的查詢策略。

(一)自然語言查詢

如果上網找資料也能像與人對話一樣，直接以英文陳述句或問句來表達資訊需求，對使用者來說，應該是最直接最自然的。自然語言查詢就是利用智慧代理分析使用者輸入的問句，粹取關鍵字並判斷關鍵字間的布林關係，再送到搜尋引擎資料庫進行查詢比對，再回饋查詢結果。

目前有很多搜尋引擎都提供自然語言查詢，有些是利用 Ask Jeeves 來提供自然語言查詢。大部分搜尋引擎是以問句來區別一般查詢或自然語言查詢，所以問號是千萬不能省略！以 Alta Vista 和 Google 為例，輸入 Where is a good place to stay in Seattle? 查詢結果會顯示每個關鍵字的出現次數，以及忽略不列入搜尋的頻繁字彙 (stop words)。

word count: Seattle: 4521517; stay: 9507027; place: 42932452; good: 55607317
 ignored in: 1522806691, a: 1996443649; to: 2085805957



(二)預設 OR

有些搜尋引擎如 Alta Vista、Excite、Go (Infoseek)、GoTo、WebCrawler 等，當使用者輸入兩個以上的字詞時，是預設以 OR 來結合這兩個查詢字詞。如果輸入 knowledge management，則搜尋到的結果最前面的是在網頁中同時出現 knowledge 和 management 兩個字的，其次則是有出現 knowledge 或是有出現 management 的網頁。換句話說，直接輸入關鍵字查詢預設 OR 的搜尋引擎，所得到的查詢結果會很多，但是精確度則較低。所以，如果是利用預設 OR 的搜尋引擎查詢資料時，應善用接下來要介紹的查詢策略來提高查詢結果的精確度。但是如果輸入關鍵字的同義異詞，進行完整搜尋時，只要直接輸入關鍵字即可。

(三)預設 AND

有些搜尋引擎，如 Google、HotBot、Northern Light 等，則是預設 AND，將使用者輸入的關鍵字以 AND 組合查詢，如果輸入 knowledge management，則查詢結果中必須是 knowledge 和 management 同時出現在網頁中才算。換句話說，預設 AND 的搜尋引擎，查詢結果精確度較高，資料量相對較少。

(四)+ 包含

只要在關鍵字前加上個加號 (+)，就表示這個關鍵字一定要出現在查詢結果中。例如：
 +clinton +lewinsky，就表示找到的網頁中要有小柯也要有魯文斯基。所以，+號可以提高查詢結果的精確度。

(5) 不包含

只要在關鍵字前加上個減號(-)，就表示這個關鍵字一定不可以出現在查詢結果中。例如：
+clinton -lewinsky，就表示找到的網頁中要有小柯，但是不要有魯文斯基。所以，一號可以排除不相關的資料，提高查詢結果的精確度。

(6) 片語查詢

片語查詢(phrase searching)是一項很好用但常常被忽略的查詢策略。將兩個以上的關鍵字加上引號，當做片語來查詢可以縮小查詢範圍，找到更精確的資訊。搜尋引擎執行片語查詢時，是依照輸入字詞的文字順序，找到完全符合的字詞。例如：輸入"knowledge management"(搜尋引擎通常是以引號代表片語查詢)，只會找到 knowledge management 兩個關鍵字緊臨出現的相關資訊，而不會找 management knowledge，也不會找只有 knowledge，只有 management 的資料，更不會找到 knowledge 與 management 距離遙遠的資訊。但是並不是每個搜尋引擎都提供片語查詢的功能，請參見表二之比較分析。

有些搜尋引擎，如 Alta Vista 和 Google，為提高查詢結果的精確度，則是採用自動片語查詢，只要使用者輸入兩個以上的關鍵字，即自動去比對自建之片語字典，如果比對成功就執行片語查詢，如果比對失敗就執行一般查詢。

(7) 近似查詢

近似查詢(proximity searching)是透過限定兩個關鍵字之間的距離來縮小查詢範圍。大部份搜尋引擎是以 NEAR 來執行近似查詢，例如：輸入 knowledge NEAR management 的意思是，knowledge 和 management 中間最多可以夾若干個單字，每個搜尋引擎的預設值不同，有些是 5 個單字，有些是 10 個單字。

這項功能在執行網頁全文檢索時特別好用。因為用布林邏輯 AND 可能限制太多，找不到資訊；用 OR 又可能太鬆，找到完全沒有相關的資料。近似查詢剛好介在兩者之間，可以視關鍵字之間的出現關係，限定關鍵字之間出現的距離，查到較多、較相關的資訊。

有些搜尋引擎沒有提供片語查詢，而有近似查詢時，更可以近似查詢來發揮片語查詢的功能，例：solar NEAR energy。在人名查詢時也很好用，有近似查詢功能時(如 Lycos 的進階查詢)，在名與姓中間加 NEAR/2，提供更大的彈性查詢不同寫法的人名，例如：輸入 George Near/2 Bush，搜尋引擎會將各種寫法一網打盡，查詢：

George Bush

George Herbert Walker Bush

Bush, George

(8) 欄位查詢

網頁是由數個部份所組成的，包括標題、URL、文本、連結、圖像等，因此有些搜尋引擎提供欄位查詢(field searching)的功能，不僅提高查詢結果的精確度，節省查詢時間，更提供網站管理者和使用者很多寶貴的資訊。前面所介紹的各項查詢策略都是查詢網頁，如果能針對網頁標題進行查詢，精確度肯定更高，因此很多搜尋引擎便提供標題查詢的功能，例如：輸入 title:knowledge management，表示查詢網頁標題中有出現 knowledge management 這兩個關鍵字的網頁，如果輸入 title:"knowledge management"，則是表示查詢網頁標題中有出現 knowledge management 這個片語的網頁。

目前以 Alta Vista 之欄位查詢功能最為強大，茲以 AltaVista 為例，說明欄位查詢之功能。

anchor:text	查詢網頁中包含特定字或片語的超連結(hyperlink)。 例：anchor:"click here to visit gardon.com" 可以找出有提供超連結到 gardon.com 的所有網頁。利用此功能可以知道有多少網頁超連我們所建的網站。
applet:class	限定查詢含特定 Java applet 的網頁。 例：applet:morph 可以找到有利用變形技術 morph applet 的所有網頁。
domain:domainname	限定查詢特定領域名稱的網頁。 domain:tw 可以找台灣的網頁。 domain:edu 可以限定只查詢教育單位的網站，在查詢課程或入學相關資訊時非常有用。
host:name	限定查詢特定電腦主機的網頁。 例：host:www.moe.edu.tw 只找教育部主機的資料。
image:filename	限定查詢特定的影像檔。 Image:teddy bear 可以找到有泰迪熊圖片的網頁。
like:URL text	查詢與指定的 URL 相似或相關的網頁。 例：like:www.abebooks.com 可以找到和 abebooks 一樣的網路上的舊書店。 例：like:www.lib.nccu.edu.tw 可以找到和政治大學圖書館一樣的大學圖書館網站。
link:URL text	查詢連結到指定 URL 的網頁。 例：link:www.lib.nccu.edu.tw 可以找到所有連結到政治大學圖書館網站的網頁。
text:text	限定查詢含有特定詞彙的網頁，剔除影像檔的說明、連結或 URL 中出現該特定詞彙。 例：text:knowledge management 可以找到有出現 knowledge management 這個字的網頁。
title:text	限定查詢文章標題中含有特定詞彙的網頁。 例：title:knowledge management 可以找到文章標題有出現 knowledge 與 management 的網頁。 例：title:"knowledge management" 可以找到文章標題有出現 knowledge management 片語的網頁。 註：Yahoo 是以 t:來執行網頁標題檢索。
url:text	限定查詢 URL 中含有特定字詞的網頁。 例：url:knowledge 可以找到完整的 URL 中有出現 knowledge 的網頁，換句話說：可能是主機名稱，路徑名稱或是檔案名稱中有出現 knowledge 的都符合。

(九)布林查詢

當我們上網找資料時應該已經很清楚自己想要的資訊是什麼，這時可以先釐出幾個重要概念，或者是關鍵字和片語，再善加利用布林查詢 (boolean searching) 來找到更精確相關的資訊。

布林查詢是利用布林運算式 (boolean

expression) 將我們要查詢的範圍明確地加以縮小或擴大，使查詢出來的結果更加準確。換句話說，當我們上網找資料時，要打進兩個以上的關鍵字做為查詢字串時，就應該考慮是不是可以加入「AND」、「OR」、「NOT」的運算元，以增加查詢的精確度。每個搜尋引擎通常會有自己獨特的語

法或符號(如&、|)來執行布林運算,可以從 Help 中找到正確的使用方法。

1.AND,是「且」的意思。所以在兩個關鍵字中加上 AND,可以縮小檢索範圍,得到較精確的結果。例如:花木蘭 AND 迪士尼,可以找到與迪士尼電影花木蘭相關的資訊;dogs and cats,可以找到同時介紹貓和狗的資訊。當我們所要查詢的關鍵字或是概念,在不同學域或行業有不同的意義時,可以用 AND 來界定範圍。

2.OR,是「或」的意思。所以在兩個關鍵字中加上 OR,可以擴大檢索範圍,得到較多、較廣泛的結果。例如:綠島 OR 蘭嶼,可以找到與綠島有關的資訊,也可以找到與蘭嶼有關的資訊,更可以找到同時介紹綠島和蘭嶼的資訊。通常為了將相關資料一網打盡,我們會在同義詞之間加上 OR 來查詢,例:台灣 OR 福爾摩沙。而有些搜尋引擎就是預設以 OR 來執行兩個以上關鍵字的查詢,所以如果我們希望輸入的兩個關鍵字是要同時出現在網頁或索引中,應該用 AND 連接。

3.NOT,是「非」的意思。所以關鍵字之前加上 NOT,表示要剔除有關該關鍵字的資料。例如:台灣 NOT 台北,會找到台北以外的有關台灣的資料。當我們以較大範圍的關鍵字或概念進行查詢,發現有不相關的小範圍關鍵字或是概念時,可以用 NOT 來釐清,例:寵物 NOT 貓。或者是兩個關鍵字或是概念經常並列談論,但我們只想研究其中一個概念時,也可以用 NOT 來剔除沒有興趣的概念,例:狗 NOT 貓。

(十)切截查詢

若以英文關鍵字查詢,經常面臨單複數和字型變化的問題,切截查詢(truncation searching)就是

利用萬用字元(wild card symbol)來幫我們克服這樣的問題,一次查詢同一個字根不同的變形字。搜尋引擎通常是以*來執行切截查詢,例如輸入 bank* 可以查到:bank、banks、banking、bankers、bankruptcy。

每個搜尋引擎可能用不同的語法或符號來執行切截的功能,例如:bank*、bank!、bank#、bank\$,可以利用 Help 來確定。

中文搜尋還可以用切截功能來查詢機構名稱的變異或是相關的資訊,例:輸入「*管理*研究所」,可以找到所有與管理相關的研究所,如資訊管理、企業管理、傳播管理或是管理科學。輸入「研*考*會」,則可以找到「研考會」、「研究發展考核委員會」,將團體機構的簡稱與全稱一網打盡。

(十一)大小寫相異

有些搜尋引擎是以單字的字首大寫來區辨專有名詞。例如:輸入 Bill Clinton 就只會找到有關柯林頓總統的資訊,如果輸入 bill clinton,則所找到的就不見得是柯林頓總統的資料了,還可以找到一些不相關的資訊。

(十二)模糊查詢

模糊查詢(fuzzy searching)是當使用者輸入一個關鍵字時,搜尋引擎不僅查詢所輸入的關鍵字,同時也自動去查詢與所輸入關鍵字意義相同的字詞。有些西文搜尋引擎可以自動查詢單複數或是英美不同的拼法,以及過濾掉一些頻繁字彙(stop words)再進行查詢。中文搜尋引擎的模糊查詢大概就屬同音和容錯功能的發揮。

1.同音:中文字輸入的一大困擾是,用注音輸入常會有同音的錯別字發生,還有很多時候,我們會唸但是不確定怎麼寫時,同音功能就可以幫我們解決這樣的問題。例如:我們可以試試看「沙士比亞」或「精成武」。對了,利用同音查詢外國人名或地名的翻譯,效果很不錯的。

- 2.容錯：有點類似以自然語言檢索，系統會自動過濾掉某些字，如“的”、“爲何”...，並與其他輸入字詞進行檢索。甚至有些中文搜尋引擎可以執行概念查詢，例如輸入 AIDS，不僅可以查到 AIDS，還可以查到免疫不全症候群等相關資訊。

(三)限定範圍查詢

有些搜尋引擎更進一步提供限定範圍查詢的功能，以限定日期、資料類型、語文等方式來縮小檢索範圍，以得到更精確、更符合需求的資訊。通常當我們找到的資料量太大時，就可以進一步利用限定查詢的方法來過濾掉不相關的資料類型或語文，或者所需的資訊有時間性時，亦建議採用限定查詢功能。搜尋引擎提供的限定範圍有：

- 1.語文：限查詢特定語文的資料，大部份的搜尋引擎，尤其是表單選項式的搜尋引擎都提供限定語文查詢的功能，如果我們不是要了解某一主題的全球發展趨勢，建議每次查詢都能限定語文，甚至是在個人化

時即設定查詢的語文。

- 2.日期：有些搜尋引擎提供日期範圍的選擇，可以將查尋範圍縮小到特定日期之內，如 Hotbot 可以限制查詢上星期、上兩星期、上個月、上三個月、上六個月、前一年、前兩年的資訊。這種方法可以知道搜尋引擎的資料更新頻率有多快，例如限定搜尋最近一個月的資訊，如果搜尋引擎找不到，那就表示該搜尋引擎資料庫的更新速度太慢了。但是網頁的日期資訊並不可靠，一般搜尋引擎是以最近更新日期爲日期。
- 3.分類：僅針對特定的主題分類進行查詢，大部份的搜尋引擎和網站分類指南都提供這項功能。
- 4.地域：可以針對某一地區或國家進行查詢，有些搜尋引擎則是針對不同地區或國家設計個別的搜尋引擎，如 Yahoo 就有 Local Yahoos：

Local Yahoos

Europe: [Denmark](#) - [France](#) - [Germany](#) - [Italy](#) - [Norway](#) - [Spain](#) - [Sweden](#) - [UK & Ireland](#)

Asia Pacific: [Asia](#) - [Australia & NZ](#) - [China](#) - [Chinese](#) - [HK](#) - [India](#) - [Japan](#) - [Korea](#) - [Singapore](#) - [Taiwan](#)

Americas: [Argentina](#) - [Brazil](#) - [Canada](#) - [Mexico](#) - [Spanish](#)

U.S. Cities: [Atlanta](#) - [Boston](#) - [Chicago](#) - [Dallas/FW](#) - [LA](#) - [NYC](#) - [SF Bay](#) - [Wash. DC](#) - [more...](#)

- 5.網站類型：可以限制只查詢 WWW、FTP、BBS、Gopher、Usenet 等。
- 6.領域名稱：可以僅查詢政府網站(.gov)、教育學術網站(.edu)、商業網站(.com)、非營利機構網站(.org)等。
- 7.資料類型：可以限制搜尋文本、圖形、影像、聲音等不同資料類型。

五、查詢結果的呈現

(一)相關性排序

大部分搜尋引擎是依相關性排序查詢結果，但是每個搜尋引擎判斷相關的方法略有不同，常用的方法是：

- 1.查詢字詞的出現頻率：查詢字詞在文件中出現的頻率，出現次數越多則相關性越高。
- 2.查詢字詞的出現位置：查詢字詞是否出現

在網頁標題、metatags 或文件開頭的幾個段落。查詢字詞若出現在網頁標題、metatags 和文件標題中，則相關性較高；查詢字詞若出現在文件的越前面，則相關性越高。

3. 查詢字詞的出現距離：查詢字詞在文件中出現的距離，字詞間的距離越近，相關性越高。兩個以上的字詞緊鄰出現之相關性最高。
4. 點閱次數：文件的人氣指數，其他人搜尋到此文件時，點閱的次數，當然是點閱次數愈高，相關性愈高。如 HotBot 的 Direct Hit 功能。
5. 連結分析：有多少網頁連結到此文件，被連結的次數愈高，表示相關性愈高。這是 Google 所開發的 PageRank 技術。

但是每個搜尋引擎所採用的相關性演算法不同，HotBot 是以字詞出現的頻率和位置為主要考慮因素。Infoseek 對出現在網頁標題和 metatags，給予額外的權重。Lycos 則是給予網頁標題和文件標題較高的權重，但是不考慮 metatags。Excite 甚至不索引 metatags 中的字詞。②

大部分的搜尋引擎是根據上述準則建立相關性演算公式，排序查詢結果，不管有沒有相關性評分標示，幾乎都是相關性越高的排在越前面。所以，瀏覽搜尋引擎查詢結果時，查詢筆數的多寡並不是那麼的重要，重要的是前 10 筆、前 20 筆、前 50 筆是不是符合。如果 50 筆都不符合，就表示查詢策略值得檢討。值得注意的事，有些搜尋引擎（如 Google）是採用連結分析，可能會發生一種情況，查詢結果可能不是那麼符合，但是文件中可以連結到其他非常符合的網站或網頁③。

(一) 直接點閱 (direct hit/popularity ranking)

有些搜尋引擎提供直接點閱熱門或精選網站的功能。這些搜尋引擎多半是根據網頁的點閱率等

人氣指數來評選熱門網站。這項功能的基本假設是人氣指數，如果有一定數目的人執行與我們相同的搜尋，這些同好在搜尋之後，點閱率最高而且停留時間最久的網站，應該也是我們會感興趣的網站。HotBot 的 Top 10 Sites，就是根據人氣指數篩選出 10 個最相關的熱門網站。少就是美，10 個優質相關網站的價值可能遠勝於成千上萬筆的查詢結果。而 Google 的好手氣 (I'm Feeling Lucky) 則是精選網站。

(二) 直接到達 (real names results)

有些搜尋引擎提供 "Real Names" 連結，可以快速精確地找到公司機構的網站，是不錯的網站查尋服務。中文分類目錄蕃薯藤的更多搜尋中，可以選擇直接到達，執行 Real Name 查詢，例如：要找政治大學，無論輸入政大、NCCU 或政治大學都可以直接連結到政大的首頁。

(三) 群組結果 (results clustering)

您是不是曾經經歷這樣的景象，在搜尋引擎的查詢結果中，一整頁其實都來自同一個網站。有些搜尋引擎為方便瀏覽，會先將查詢結果加以群組，以精簡查詢結果，增加查詢結果的可讀性和多元性。HotBot 和 Alta Vista 是每個網站只顯示一筆查詢結果，也可以選擇某一網站，瀏覽符合的所有網頁資料。Northern Light 則提供資料夾，將查詢結果依主題、資料類型、來源或語文加以群組。

(四) 顯示筆數

大部分的搜尋引擎是預設每頁顯示十筆，如果能一次看到更多筆不是更有助於判斷嗎？大部分的搜尋引擎在進階查尋或表單選項中，都會提供顯示筆數的特定功能。

(五) 顯示方式

有些搜尋引擎可以讓使用者選擇只要顯示網頁標題，或是連摘要一起顯示。是一個網站只顯示一筆資料，還是每筆都顯示。

(六) 檔案大小

有些搜尋引擎的查詢結果提供網頁的檔案大小，方便使用者超連結時參考。

(八)顯示日期

有些搜尋引擎的查詢結果，除了網頁介紹外，還提供網頁的建置日期或更新日期。雖然網頁日期不全然可信，但是顯示日期不僅有助於判斷所查得資料的新穎性，更可以判斷搜尋引擎資料庫的更新速度。

每個搜尋引擎對網頁日期的處理方式不同，有些搜尋引擎對於沒有提供日期的網頁，一律以機器人拜訪該網頁的日期取代。有些搜尋引擎如 Northern Light 則不顯示網頁日期。

Google 並未在查詢結果中列出日期，但點選 cached，可以知道機器人拜訪該網頁的日期。

分類目錄通常都不顯示網頁日期。其實如果能加上整理或更新的日期，對使用者也是很有用的。

(九)依日期排序

聽起來很不錯，查資料總是希望能找到最新的資訊。但是日期在網路上可是個大問題，因為網頁日期應該有網頁建立日期和最近更新日期之分。事實上，根據 Go 的工程師在 1998 年的估計，只有 70% 的網站主機能夠提供正確的日期資訊，20% 則是只能回報目前日期，無法得知網頁是何時建立或何時更新的。而剩下的 10% 是根本找不到日期資訊。^④

很多專業人士還是很希望查詢到的網頁資訊能依日期排序。在眾搜尋引擎中，Go 和 Northern Light 提供日期排序的選擇，將最新的資訊排在最前面。

Go 在 Web Search 查詢結果的頂端，點選 "Sort By Date"。Northern Light 則是在進階查詢的畫面中，在選擇 Select Date Range 時，可勾選 Sort results by date。

值得一提的是，如果有興趣的是新聞話題，建議利用提供 News Search 的搜尋引擎，或乾脆利用

新聞專業搜尋引擎。

(十)庫頁備份(spidered version)

網頁更新的速度非常之快，有的時候搜尋引擎傳回來的資料，和實際連結進去看到的有所不同；還有很多時候在搜尋引擎找到一篇正合我意的文章，不幸卻已遭廢棄或不知喬遷何處。如果搜尋引擎能提供其機器人搜尋回來的版本，就不會有遺珠之憾了。

Google 就提供了這項貼心的服務，只要點選 Cached，就可以看到 Google 的機器人當初讀取回來的網頁內容，也就是索引版的網頁內容，而且還會將我們輸入的查詢字詞標示起來。在 Cached 網頁最上端會提醒 user，這可能不是最新版，若要看最新版，請點選 current page。

六、精煉查詢

(一)修正查詢策略

初步的搜尋可能產生三種結果，第一種結果是完全不符合，第二種結果是查到的資料很多，但是符合的很少，第三種結果是查詢到的結果很少，但很符合。

1. 完全不符合：如果查詢結果是零或完全不相關，先檢查一下字有沒有拼錯。沒有的話，就換個搜尋引擎，或者先利用整合式搜尋引擎查詢，找出較為適合的搜尋引擎。
2. 資料很多，但是相關的很少：資料很多，那就表示應該縮小查詢範圍，收斂查詢結果。所用的搜尋引擎若是預設為 OR，如 Alta Vista，則可以減少查詢的字詞，提高精確度。所用的搜尋引擎若是預設為 AND，則增加查詢字詞，以收斂查詢結果。搜尋引擎有沒有結果再查詢的功能，有的話，輸入新的關鍵字針對第一次查詢結果再查詢。加上限制功能，如查詢標題，或設定語文或日期範圍。利用 NOT 或減號(-)

剔除不相關的查詢字詞，如電腦 NOT 蘋果。試試看網站指南或分類目錄。

3. 資料很少，但很符合：查到的資料很少，但是都很符合，表示可能漏掉了些相關資訊，應該擴大查詢範圍。找出同義詞或其他經常並列出現的關鍵字詞，以 OR 擴大查詢範圍。利用切截，來查詢關鍵字的變形。減少用 AND 組合的查詢字詞。換大一點的主題或概念，查查看。試試其他搜尋引擎、分類目錄，或整合式搜尋引擎。查詢結果若有顯示分類架構，可以該分類架構瀏覽相關的網站或網頁。

(二) 收斂查詢(sub-searching)

當第一次查詢結果不甚滿意時，不要氣餒，因為大部份的搜尋引擎都提供收斂查詢的功能，可以根據第一次的查詢結果加以精練，常見的方法有：

1. 網頁分類目錄

很多搜尋引擎也提供分類目錄的功能，有些搜尋引擎可以限定在某一分類主題內搜尋，可以大大地提高查詢結果的精確度。有些提供分類目錄服務的搜尋引擎，查詢結果會自動回覆符合查詢字詞的分類目錄，供使用者瀏覽相關網頁，如 Web Crawler。Go.com 則是列出相關的分類目錄主題 (Web directory topics) 供參考，也可以點選 search with directory 查詢分類目錄。有些則是提供選擇 Find Results in Directories 來收斂第一次的查詢結果。

2. 相關查詢(related searches)

有些搜尋引擎會在查詢結果的頁首列出建議的相關查詢，多半是取材自網頁分類目錄之主題，可以幫助使用者精練查詢結果。

3. 相似網頁(find similar)

有時在查詢結果中有一兩筆資料是深得吾心，完全符合查詢需求，這時會很希望能找到類似的網頁。查找相似網頁就在滿足這項需求。

4. 結果再查詢(search within)

利用搜尋引擎查詢資料時，常常輸入查詢字詞後得到的結果多得讓人驚訝！這時怎麼辦呢？重新查詢當然是最直接的方法，不過有些搜尋引擎提供再查詢(search within)的功能，可以輸入其他查詢字詞或限制查詢範圍，來精練查詢結果。

七、綜合分析

何時該用搜尋引擎呢？

1. 當我們想要研究的主題或觀念很明確（狹窄）或很模糊時。
2. 當我們想要找某一個特定的網站時。
3. 當我們想要查詢數百萬頁的全文網頁時。
4. 當我們想要取得某一主題的大量資料時。
5. 當我們想要查詢特定類型的文件、檔案、來源 (source location)、語文和最近更新日期時。
6. 當我們想要善用概念查詢、熱門排行、連結排行等最新檢索科技時。

表二根據前述介紹之查詢功能、結果呈現和精煉查詢等項目，比較分析常用的七個西文搜尋引擎，做為選用搜尋引擎和查詢時之參考。同時表二也可以做為一個便覽表格，隨時追蹤搜尋引擎功能之變化，不斷加以更新，做為參考諮詢之一項知識資產。（待續）

（收稿日期：2000年8月24日）

表二：搜尋引擎功能之比較分析

	Alta Vista	Excite	Go (Infoseek)	Google	GoTo	HotBot	Lycos	Northern Light	WebCrawler
資料庫	350M	250M	75M	1,060M 560M		110M	50M	265M	
進階查詢界面	Power Search, Advanced Search, Raging Search	Power Search	Advanced Search	Advanced Search		Advanced Search	Advanced Search	Power Search	
自然語言查詢	✓			✓					
查詢功能									
包含	+	+	+	+	+	+	+	+	+
不包含	-	-	-	-	-	-	-	-	-
片語查詢	自動 引號("")	引號("")	引號("")	半自動 引號("")	引號("")	引號("")	引號("")	引號("")	引號("")
完全符合 (Match All)				自動		自動	自動	自動	
部份符合 (Match Any)	自動	自動	自動	自動	自動				自動
加權關鍵字				✓					

	Alta Vista	Excite	Go (Infoseek)	Google	GoTo	HotBot	Lycos	Northern Light	WebCrawler
布林邏輯									
預設運算元	OR	OR	or	AND	or	AND	and	and	or
其他運算元	AND, AND NOT, NEAR	AND, AND NOT	and, not		AND, OR, ()	OR, NOT	or, not, adj, near, before, far	or, not	
切截(truncation)	*	概念查詢				?字中 *字尾		自動 *	
字根(stemming)			✓				✓	✓	
大小寫相異	✓	*	✓			✓			
日期限制									
RealName 連結	✓		✓	✓					
欄位查詢									
標題 (Title)	title:		title:		title:	title:	menu	title:	
網站/主機 (Site)	host:		site:		domain:	domain:	menu:		
URL	url:		url:				menu:	url:	
連結 (Link)	link:		link:	link:	linkdomain:	linkdomain:			

	Alta Vista	Excite	Go (Infoseek)	Google	GoTo	HotBot	Lycos	Northern Light	WebCrawler
結果顯示									
網站群組 (clustering)	✓		✓	✓	✓	✓		✓	
顯示日期	✓		✓			✓		✓	
依日期排序			✓					✓	
增加查詢筆數		✓	✓	✓		✓	✓		✓
庫頁備份 (spider version)				✓					
精煉查詢結果 (sub-searching)									
熱門網頁 (direct hit/ popularity ranking)						✓	✓		
查詢分類目錄	✓	✓	✓			✓	✓	✓	
相關查詢 (related searches)	✓		✓		✓	✓			
類似網頁 (find similar)			✓	✓					
結果再查詢 (search within)			✓			✓	✓		
其他特色		Concept-based	How to Search	Link Analysis	Let GoTo Help	Search Tips	Help	Folder Category	Help & Information
查詢輔助	Help: Search http://doc.altavista.com/help/search_help.htm	Search Help http://www.excite.com/info/searching.html	How to Search http://infoseek.com/Help/pg=SearchTips.html	How to Google http://www.google.com/hcjp.html	Let GoTo Help http://www.gotome.com/d/about/help	Search Tips http://hotbot.lycos.com/help/tips/default.asp	Help http://www.lycos.com/help/	Search Help http://www.northernlight.com/docs/search_help_optimize.html	Help & Information http://www.webcrawler.com/info/

註釋：

註①：Chris Sherman, "If Search Engines Were People,"

<<http://websearch.about.com/internet/websearch/library/weekly/aa052199.htm>> (10 Aug. 2000).

註②：Martin P. Courtois and Michael W. Berry, "Results Ranking in Web Search Engines," *Online*, May 1999,

<<http://onlineinc.com/onlinemag/OL1999/courtois5.html>> (13 Aug. 2000).

註③：Google 在其首頁號稱其資料庫索引了十億的網頁文件，實際上應該有一半是連結分析的成果。

註④：Danny Sullivan, "Search Assistance Features,"

<<http://www.searchenginewatch.com/facts/assistance.html>> (22 Aug. 2000).