



# 網路資源搜尋引擎

## Search Engine

羅思嘉

成功大學圖書館

### 【摘要】

網路資源的搜尋一直是資訊檢索者所需獨自面對與克服的問題，尤其隨著網路資源的增加資料搜尋的問題更形顯著，不少網路搜尋系統紛紛在此情況下因運而生。過去有不少介紹臺灣網路資源搜尋引擎的相關文章，本文將就 AltaVista、Openfind、Yahoo 等國際常用網路資源搜尋引擎做一介紹。

### 【 Abstract 】

It has been a great challenge to the Network user to find the information they want on the Internet. The search engines have been played a major role in locating the information on the Net. There are several articles about the Search Engines for the Internet sources in Taiwan were published in the past. This article is focus on three wide used search engines, AltaVista、Openfind and Yahoo.

關鍵詞：搜尋引擎、網際網路、Altavista、Openfind、Yahoo

Keywords: Search Engine、Internet、Altavista、Openfind、Yahoo



## 壹、前言

在網路資源利用上資料的搜尋一直是資訊檢索者所需獨自面對與克服的問題，尤其隨著網路資源的增加資料搜尋的問題更形顯著。之所以將資訊搜尋問題稱之為資訊使用者必須獨自面對與克服的問題，其因素在於資料的範圍及多樣化，與以往資料庫的資訊檢索不同，是一個無邊界的資料搜尋。面臨這樣的問題，不同學科範圍的研究者試圖提出不同的解決方案，有以圖書館資料編目方式處理網路資源的，有另以一套分類體系收集組織網路資源的，也有提供一套簡潔著錄原則讓網頁資源提供者附加於資料上以利搜尋，或是將網路資源建成一個資料庫以網頁內容或標題為主體透過搜尋引擎查詢資料者；這些不同方式在資料收集上各有其特點。本文將就三種不同的資訊搜尋引擎做一介紹與比較，AltaVista、Openfind、Yahoo。

## 貳、AltaVista (<http://www.altavista.com>)

AltaVista 是以索引檢索的方式提供網路資源搜尋功能，AltaVista 透過網路使用者自行登錄以及利用軟體到網路上搜尋新增的網頁資料。資料的檢索可分為兩種模式，簡易檢索與進階查詢。簡易檢索部份系統提供一個單一對話方塊，可輸入檢索詞彙或詞句，點選查詢 (search) 按鍵或按 <Enter> 鍵後系統進行檢索，系統依檢

索網頁的關連性排序顯現，必要時可依語文限定所找出的資料。當所需要的資料在時間、或者是需要利用布林邏輯運算子組合不同檢索策略時，AltaVista 透過進階檢索提供不同的檢索功能。

除了檢索外，AltaVista 也提供一個分類表，以瀏覽的方式查尋所要的資料。AltaVista 共將資料分為 13 類，汽車 (Automotive)、商業經濟 (Business & Finance)、電腦與網際網路 (Computers & Internet)、醫藥保健 (Health & Fitness)、嗜好興趣 (Hobbies & Interests)、家庭 (Home & Family) 媒體與娛樂 (Media & Amusements)、人物與聊天 (People & Chat)、參考工具與教育 (Reference & Education)、購物與服務 (Shopping & Services)、社會與政治 (Society & Politics)、運動與娛樂 (Sports & Recreation)、旅遊與假期 (Travel & vacations) (圖 1)。

## 參、AltaVista 檢索特色

### 一、「單字檢索」與「語句檢索」

AltaVista 檢索系統在進階模式中提供檢索者兩種語彙檢索功能，一是「單字查詢」，可輸入單一字詞或多個字詞，系統將所輸入的每一字詞視為獨立字詞，單獨查詢，如在進階模式下鍵入 library education，系統查詢網頁包括 library 或 education 兩個詞彙 (圖 2)；另一則是「語句

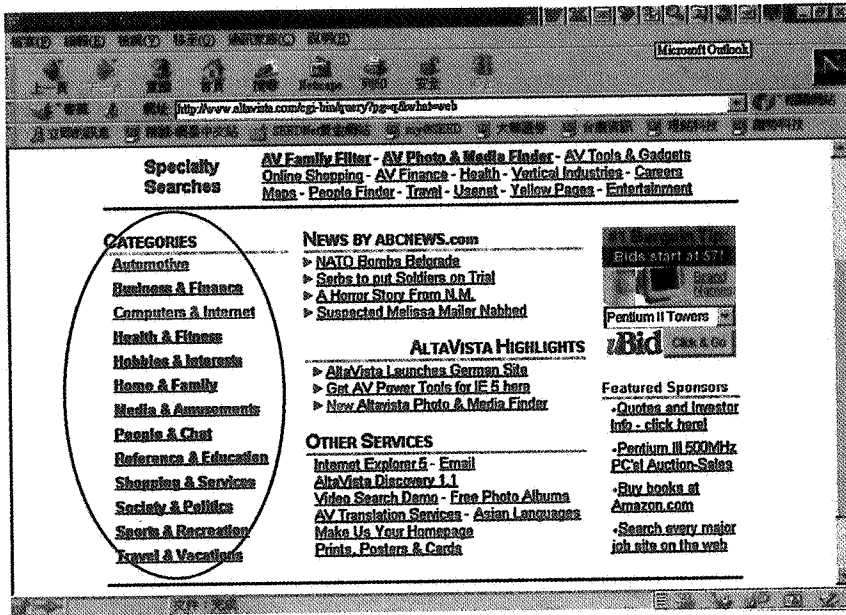


圖 1：分類瀏覽表

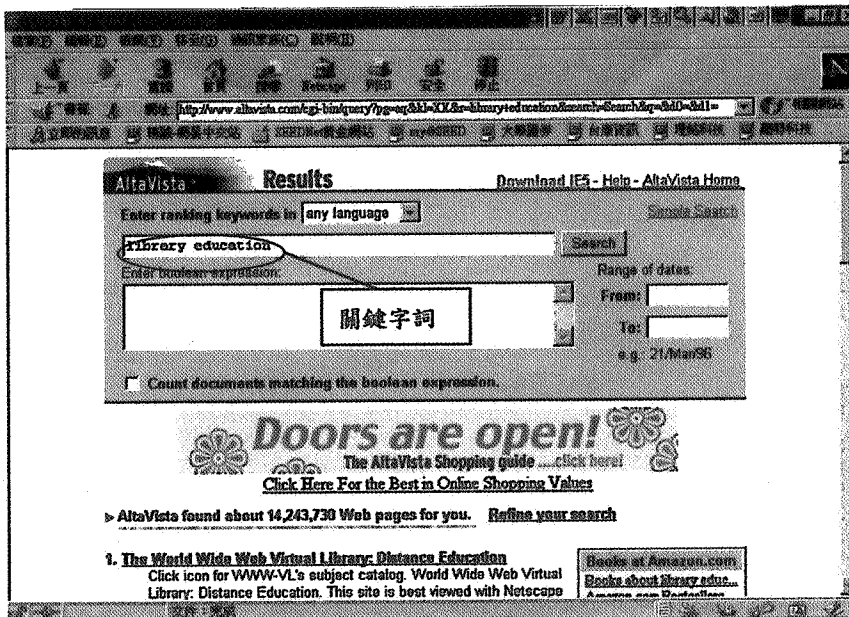


圖 2：找尋網頁中包含 library 或 education 的資料，系統找到 14,629,555 檢索結果。



檢索」，若以雙引號加在檢索詞前後，系統會將所輸入的字串視為一連結語句，依所輸入的順序與彼此間的相對位置進行查詢，如輸入 "library education"，系統會找出名稱與網頁中包含 library education 這個字串的資料，簡易模式中系統自動將檢索視為「語句檢索」（圖 3）。

除了使檢索更精確外，「語句檢索」對一些名句的檢索有所助益，例如 "To Be Or Not To Be"，在「單字檢索」中 'to'、'be'、'or'、'not' 都屬於剔除字，在檢索上不具意義，但 "To Be Or Not To Be" 組合起來是有意義的資料檢索，透過「語句檢索」就可將包括 "To Be Or Not To Be" 這句話的網頁資料顯現出來。

## 二、布林邏輯運算

AltaVista 提供 AND、OR、NOT、NEAR 等不同布林邏輯運算子，AND 表示網頁中必需同時包括所輸入的字串，但字詞不一定要同時連接出現；OR 則是包括任一輸入字串即可；AND NOT 排除所不要的字串；NEAR 則是不單要包括所輸入的字詞，且字詞間的相隔距離不超過 10 個字詞。

(一) AND：查詢 Consortium for the Computer Interchange of Museum Information(CIMI) 這個單位利用 Dublin Core 的相關資料，可輸入這樣的檢索指令，"dublin core" and cimi（圖 4）。

(二) OR：查詢與數位圖書館，digital library 或電子圖書館，electronic library 相關的資料，可輸入這樣的檢索指令，"digital library" or "electronic library"。

(三) AND NOT：想了解 CIMI 以外的單位如何運用 Dublin Core 的資料，"dublin core" and not cimi。

(四) NEAR：收集有關電子期刊的資料，electronic near journal。

## 三、條件限制

AltaVista 搜尋功能除了布林邏輯運算與語句組合外，檢索策略也可加入不同的條件限制以使得檢索結果更符合需求。限制條件可分為兩種，一是欄位限定，一是選單限定。簡易與進階檢索兩種介面均提供欄位限定，同時欄位限定可當成單一條件檢索；不同檢索介面提供不同選單條件限制查詢，同時是配合其他檢索條件一起查詢。

### (一)欄位

AltaVista 提供 9 種不同的限制條件欄位：

-anchor，查尋附超連結的特定字串。

-applet，尋找使用特定 Java 程式的網頁。

-domain，查尋某一網域的資料，domain:uk。

-host，查尋特定網站上的資料，host:www.lib.ncku.edu.tw。

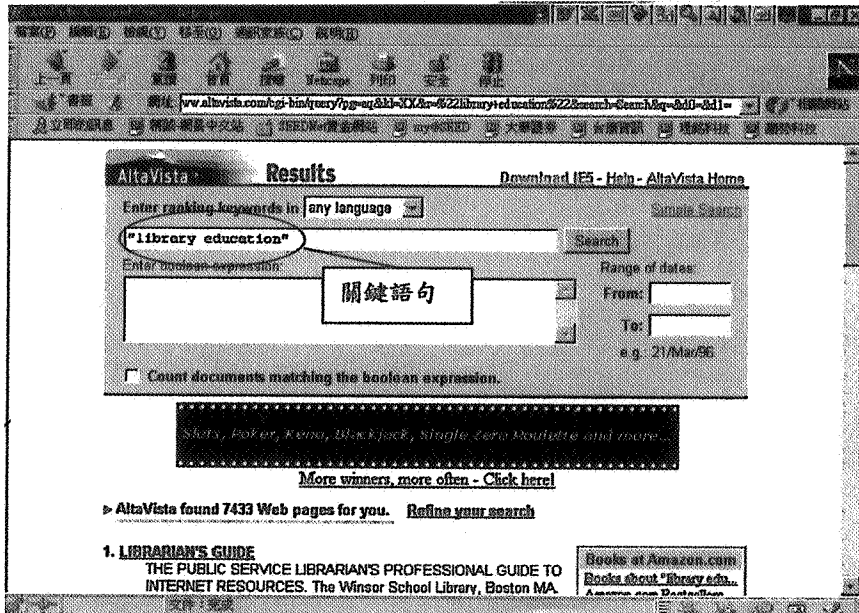


圖 3：系統將 library education 視為一體字串進行檢索，檢索出 7,677 筆資料。

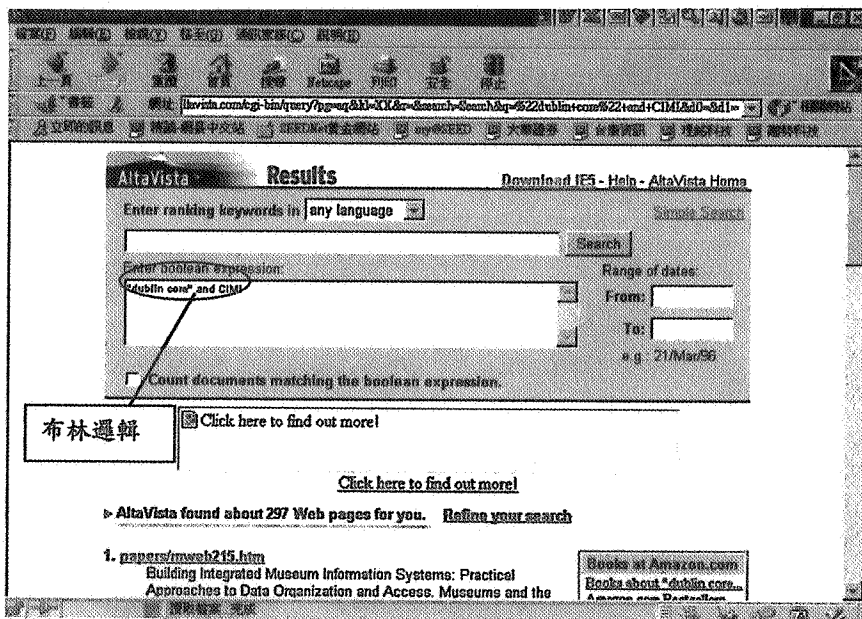


圖 4：利用布林邏輯運算查詢 CIMI 利用 Dublin Core 的相關資料



- image，以網頁影像連結為查詢依據，  
image:chopin。
- link，查檢資料連結到特定網址的網頁，  
link:www.lib.ncku.edu.tw，查尋那些網頁連結到成大圖書館的網頁。
- text，檢索網頁內容，例如 text:"information society"，找尋內文中包括「資訊社會」的資料（圖 5）。
- title，就網頁名稱欄進行資料查詢，  
title:"dublin core"，檢索網頁名稱中包括 dublin core 字串的網頁。
- url，網路網址查尋，如鍵入 url:ncku  
查尋 AltaVista 資料庫有那些可能由成大所編的網頁。

#### (二)選單

在選單限制部份，AltaVista 提供語文以及網頁日期年代的限制，語文包括中文、英文、法文等 25 種語文；日期則可輸入所希望檢索時間的區段。

#### 四、符號的運用

為符合不同的檢索需求，AltaVista 提供了一些符號運用功能來強化檢索結果，一是利用加 (+)、減 (-) 符號來確保所找出的網頁資料是否包括或剔除特定字詞，如輸入 "Library Education" -PhD，表示查詢有關「圖書館教育」方面的資料，但不要有關博士課程方面的資訊；另一則是利用星號 (\*) 來進行字串切截檢索，digital librar\*，可 digital library、

digital libraries 等資料均符合檢索策略。

#### 五、修飾檢索策略 (Refine)

為了使檢索結果更符合需求，AltaVista 會將就檢索結果再進一步分析出相關的概念，檢索者可依需求選擇所要包括或者是欲剔除的概念字詞（圖 6）。

#### 肆、Openfind (<http://www.openfind.com>)

Openfind 與其他搜尋引擎不同的是能夠讓使用者搜尋不同語文的資料，並具有相當程度的容錯能力。就 Openfind 發展者自我定位來看，Openfind 較 Yahoo 與 AltaVista 這兩個網路資源搜尋引擎不同的是，Yahoo 與 AltaVista 是以搜索以羅馬語文為基礎的網路資源，在非羅馬文字語文的網路資源上，檢索能力並不突出，Openfind 則是突破這樣的限制，使得資料檢索達到所謂的全球性資源搜尋。以查詢國內「圖書館」網頁資料為例，在 AltaVista 中應以 "library" 為檢索詞彙，並將網頁語文限定在「中文」來進行搜尋，無法直接以「圖書館」檢索，相對的在 Openfind 中可直接以「圖書館」查詢。除了資料檢索語文範圍的擴展外，Openfind 的檢索功能除精確比對外，還包括模糊比對的功能，也就是即使檢索字詞有所出入，檢索程式會儘可能找尋符合的資料（圖 7）。

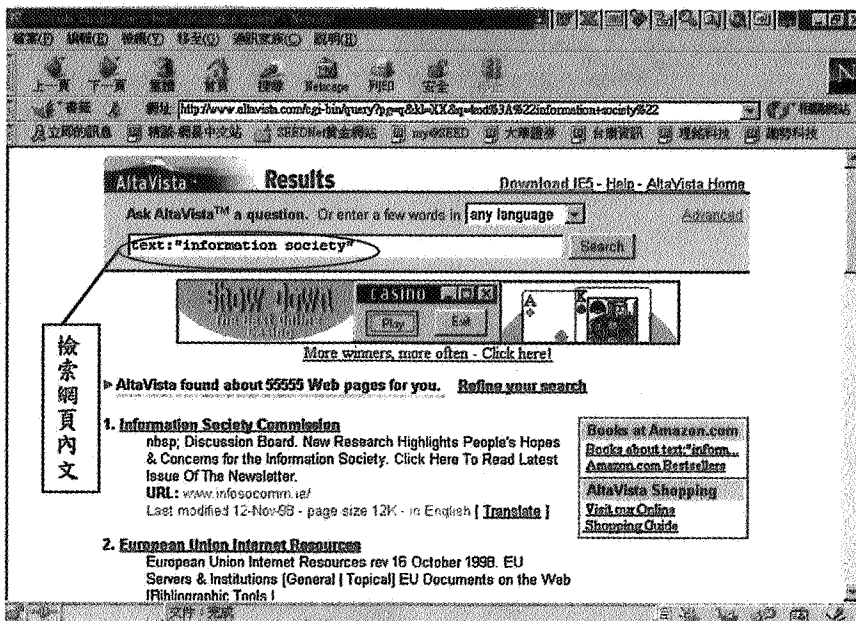


圖 5：檢索網頁內容正文包括 information society 字串的資料

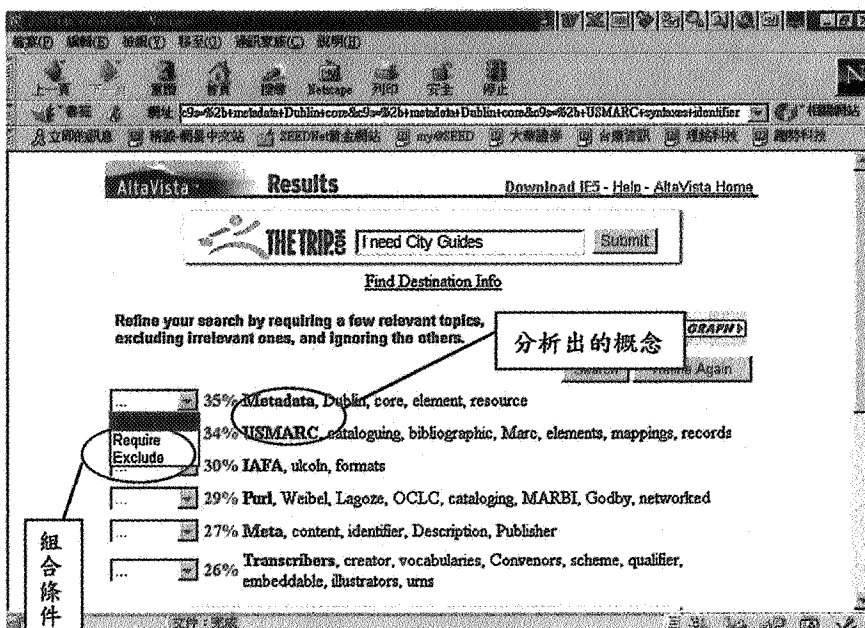


圖 6：可再利用 metadata、USMARC 等概念進一步篩選更合適的檢索結果



## 伍、Openfind 檢索功能特色

### 一、支援不同語言：

與字元比對不同，利用所謂的 Openindex 技巧，Openfind 可搜尋非字元基礎語文的資料，也因此 Openfind 提供了非羅馬文字的檢索。

### 二、模糊比對：

拼字向來在檢索的正確比對上是一大問題，在以往的比對技巧上首要條件是正確輸入檢索語彙，若輸入的資料與資料庫中現有索引資訊有所出入，檢索系統所回覆的檢索結果多是「查無此資料」，但在 Openfind 搜尋系統中的「模糊比對」功能突破這樣的限制，也就是系統會依照所輸入的字串檢索，儘可能的搜尋符合的資料（圖 8）。

### 三、網頁與網址檢索

Openfind 結合 Yahoo 網站檢索與 AltaVista 網頁檢索模式，透過 Yahoo 檢索者可找到與檢索主題相符合的網址，例如可以找出與 Digital Library 有關的網頁；AltaVista 所提供的是就所輸入的字串檢索出可能的網頁。

Openfind 的設計是先提供檢索者檢索結果符合資料的瀏覽，提供檢索結果文件資料內容符合地方的相關資料，以供查詢者判斷所獲得的資料是否符合所需，網站資訊之後提供網頁中包括所輸詞彙的網頁檢索結果。同時在資料輸出時可依網頁名

稱或者網址排序（圖 9、10）。

## 四、資料檢索

### (一)單一關鍵字

輸入單一字詞，系統會找出包括輸入字詞的網頁，如 metadata。

### (二)多重關鍵字

對單一字元的語文檢索，如英文、法文等，檢索時可輸入兩個以上的檢索語彙，系統將檢索出包含且字詞順序與所輸入的檢索字詞相同的網頁，如 Digital Library，可找出 "The Magazine of Digital Library Research"，也會找出 "International Digital Electronic Access Library"，或是 "Job Search from Library Journal Digital"。

### (三)布林邏輯運算

利用布林邏輯運算「AND」(&)、「OR」(|)、「NOT」(!)，使檢索結果更為精確或是包含各種可能的資料，例如要查詢「數位圖書館」如何運用 metadata 處理電子資源，可輸入這樣的檢索策略，matadata AND digital library。

1. AND：查詢 Consortium for the Computer Interchange of Museum Information (CIMI) 這個單位利用 Dublin Core 的相關資料，可輸入這樣的檢索指令（圖 11），

dublin core and cimi

dublin core & cimi



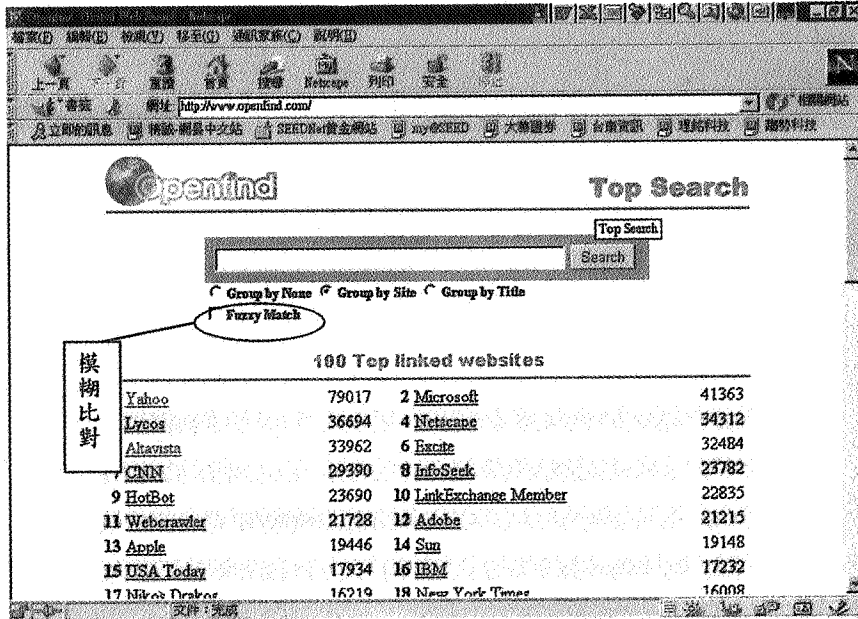


圖 7：Openfind 首頁

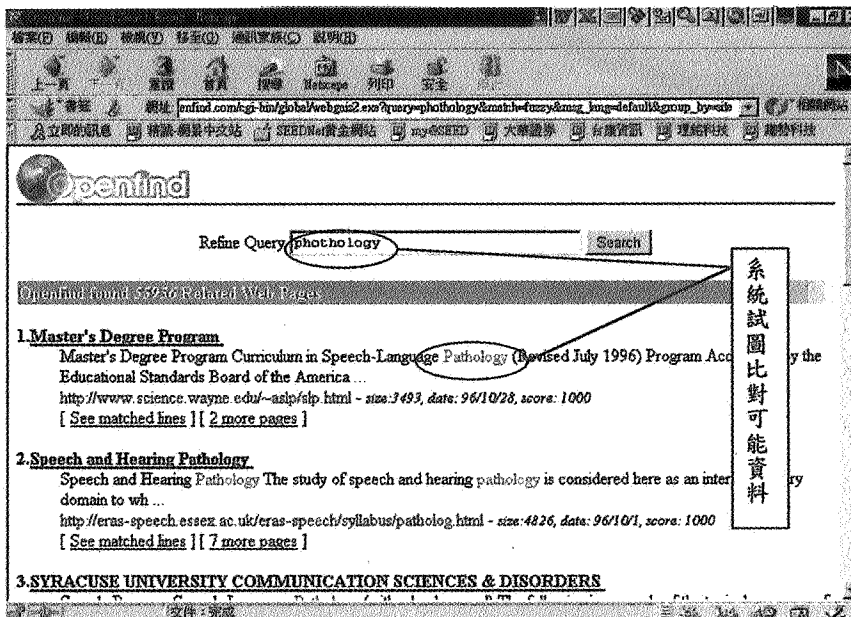


圖 8：容錯檢索，例如輸入 phothology，仍可找出 pathology 的資料

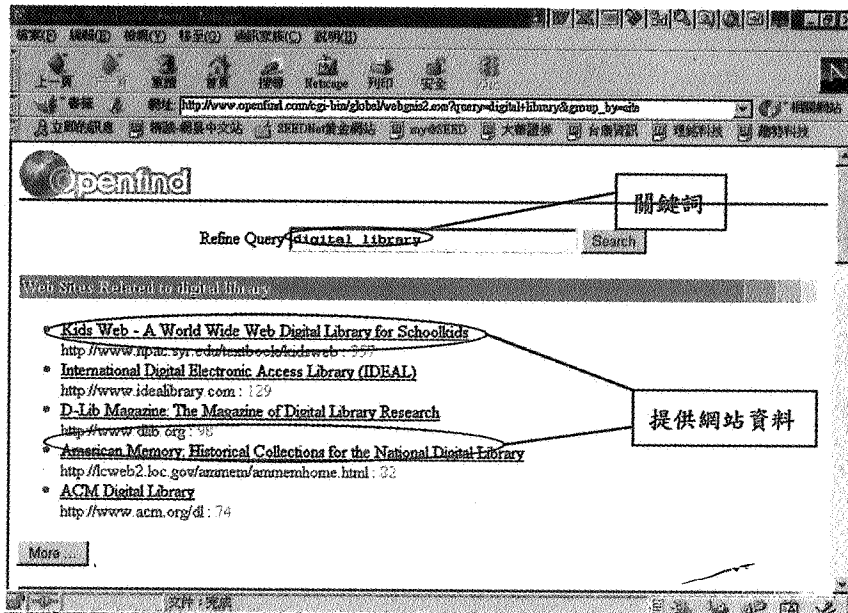


圖 9：Openfind 先整合列出與 digital library 相關的網站

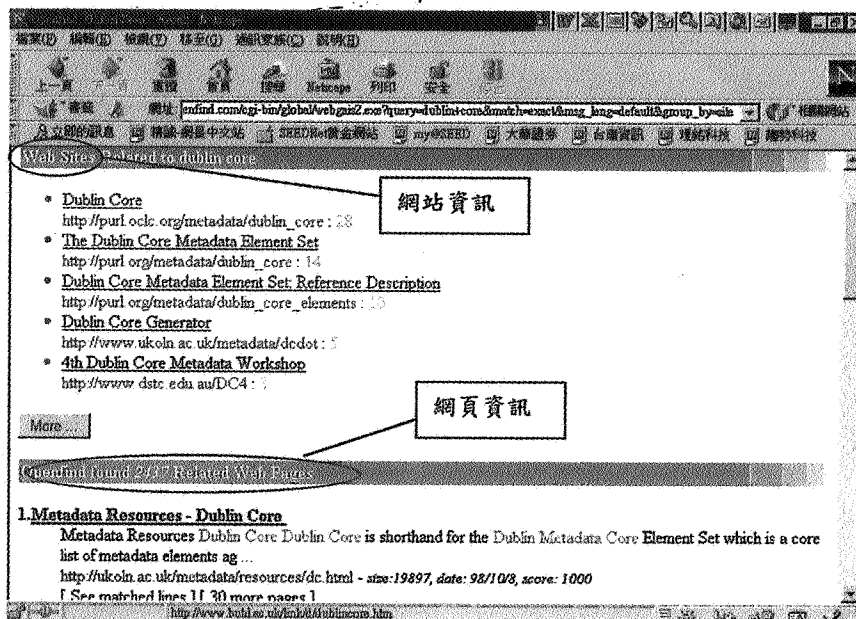


圖 10：Openfind 在網站資料後列出其他符合檢索相關網頁

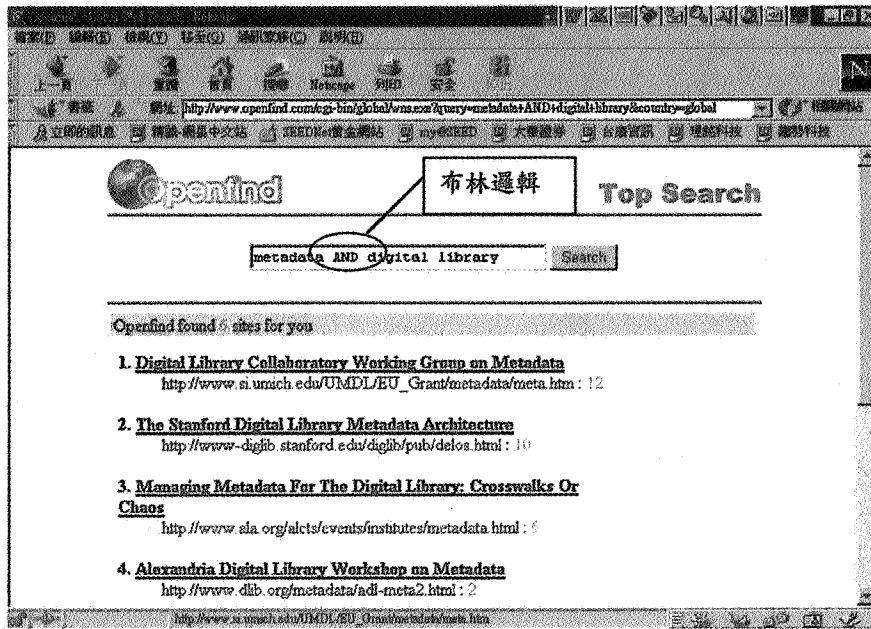


圖 11：布林邏輯運算 AND

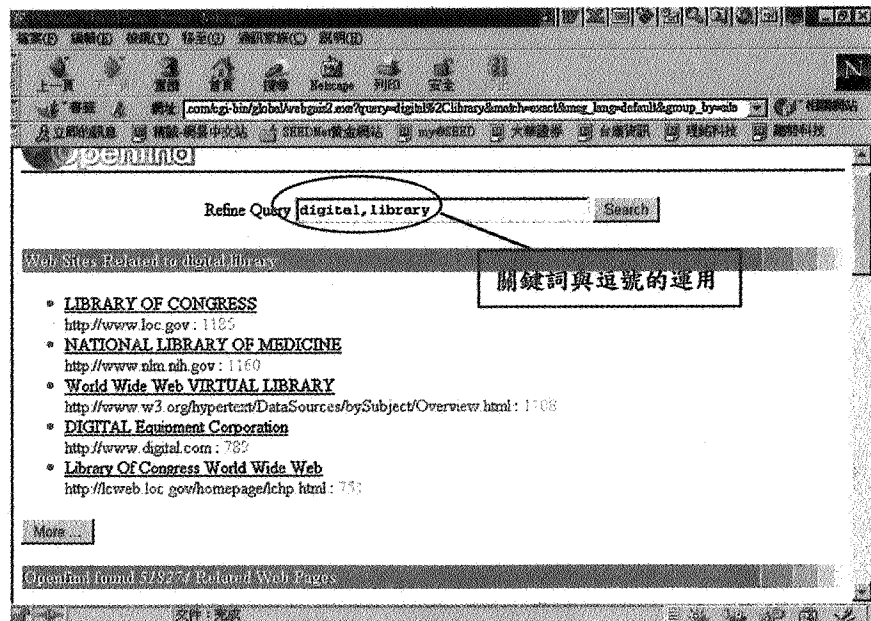


圖 12：逗號檢索的運用，digital, library



2. OR：查詢與數位圖書館，digital library 或電子圖書館，electronic library 相關的資料，可輸入這樣的檢索指令，

digital library or electronic library  
digital library | electronic library

3. AND NOT：想了解 CIMI 以外的單位如何運用 Dublin Core 的資料，

dublin core not cimi  
dublin core ! cimi

#### (四)符號的運用

1. 多重關鍵字計分查詢，逗號分割查詢法

若想要放寬字詞順序限制，增加檢索回收，可輸入數個關鍵字，各關鍵詞以逗號分隔，系統會就所輸入的關鍵字進行檢索，再依照文件中所包含的字詞數，給予權重分數，字詞間的相關位置不受到策略輸入順序的限制（圖 12）。

關鍵語句、布林邏輯運算與逗號三種不同檢索模式在利用上是可彼此互換的，運用上可考慮檢索結果範圍，以不同方式

放寬或縮緊檢索條件。例如 "digital library" 所要查詢的是名稱或網頁內容包括 "digital library" 這個字串的資料；"digital and library" 則是搜尋包括這兩個關鍵字的資料，其間的順序可不予以考慮；"digital, library" 則是以兩者其一即可的概念來搜尋網路資源。（表 1）

#### 2. 加減號的運用

利用加減符號的運用，搜尋或剔除包括輸入字詞的網頁。例如要查尋有關都柏林大學 (University College Dublin) 或都柏林 (Dublin City) 這個城市的資料，若單純輸入 "Dublin"，系統除了回覆 "Dublin" 的資料外，也會搜尋出有關 "Dublin Core" 的資料。若要剔除有關於 "Dublin Core" 的網頁資料，可以輸入 "+dublin, -dublin core" 這樣的檢索策略，系統會考慮保留有關 "dublin" 的資料，但刪除 "Dublin Core" 的資料（圖 13）。

#### (五)自然語彙檢索 (natural language)

以口語化方式輸入檢索問題，系統依照所輸入的資料截取關鍵字查詢可能符合

表 1：以三種不同檢索策略查詢 digital library 相關資料的檢索結果

結果 \ 策略	digital library	digital and library	digital, library
網站網頁數	740(5)	740(5)	4,993(5)
相關網頁	2,793	7,960	518,274
網站結果	786	808	51,353



的網頁資料，但系統所回覆的結果，可能因分析上的誤差而產生與預期相差甚遠的檢索結果。例如以「網際網路對圖書館的影響」為檢索語彙，系統回應 4707 篇的檢索結果，但其中不少與所希望檢索到的資料符合程度偏低，例如「西方思潮對近代中國圖書館事業的影響」、「名人談網：愛心與網路」。另一個值得注意的是，Openfind 目前在檢索上還未能辨識「剔除字」在「自然語彙」檢索中可能的意義，例如 "To Be Or Not To Be" 若單獨檢索，或者是與其他語詞組合時是不具有太多的意義，但是當其視為一個語句時，在檢索上是有不同含意的。雖說在檢索上「自然語彙」檢索還有相當大的成長空間，但在各個搜尋引擎設計，與對格式化檢索方式不熟悉的使用者來說，以「口語化」方式檢索資料，Openfind 確也提供了另一個資料檢索的思考方式。

### 陸、Yahoo (<http://www.yahoo.com>)

Yahoo 大概是國內網路使用者最早利用的國外網路資源搜尋引擎，以分類目錄的型式整理網路資源。與 AltaVista 相類似，Yahoo 同時提供一類表供檢索者以瀏覽的方式搜尋各類下的網路資料，目前 Yahoo 提供 25,000 大、小類的資料，同時為方便檢索者找到所需的網路資源，Yahoo 也提供關鍵字查詢功能，讓使用者

直接由網頁資料中找尋所需要的資訊。Yahoo 與前面所介紹的 AltaVista 及 Openfind 不同的是，Yahoo 資源資料庫的建立，除了程式的資料搜尋外，Yahoo 投資了相當的人力在資料選取與分類上。就資料量上來說，以相同策略檢索結果量來看，Yahoo 索引的資料量較前兩者少，但就另一方面來看其符合精確程度可能較前兩者高，同時網站所提供的資料品質可能較為平整。另一個較不同的是 Yahoo 在檢索結果的安排上是分別以類別、網站及網頁為單元，各別顯現（圖 14）。

利用 Yahoo 時，檢索者可就有興趣的主題，以點選方式瀏覽每一類下相關的資源，或者是利用搜尋功能檢索特定網頁、特定主題的相關網站。Yahoo 所提供的檢索功能可針對 Yahoo 完整資料庫資訊進行搜尋，或者是只針對某一主題類別下進行搜尋。Yahoo 對檢索結果的處理方式是先將類表中符合的類別整理出來，查詢者可就類別，瀏覽各所屬網站；檢索者也可直接點選類別後的各個網站選項。Yahoo 就所檢索到網頁，依其檢索詞彙出現的頻率、詞彙符合以及檢索詞彙出現所在，為每一網頁內容相關性評分，再將檢索結果依照相關性排序。

### 柒、Yahoo 檢索特色

#### 一、檢索環境設定

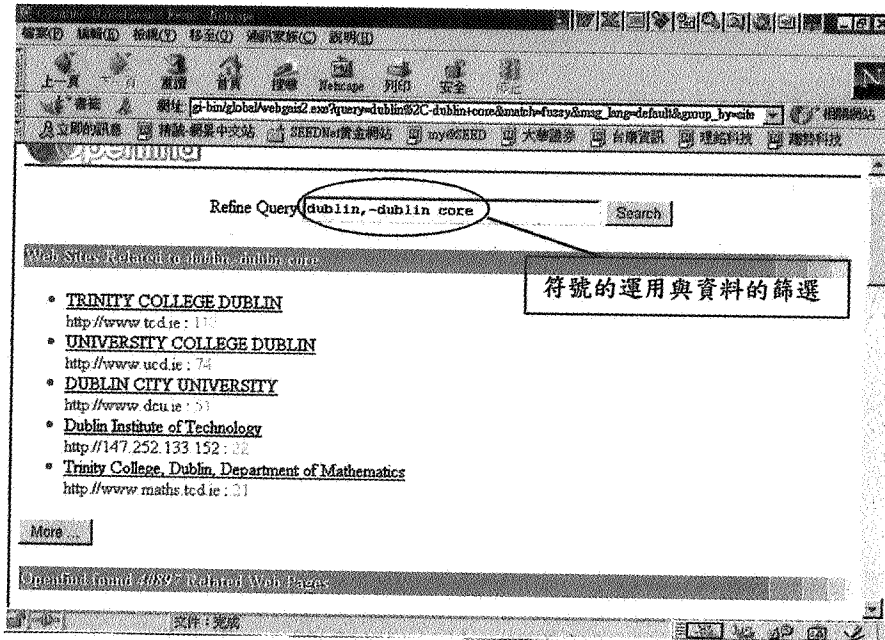


圖 13：利用減號提高檢索結果的精確率

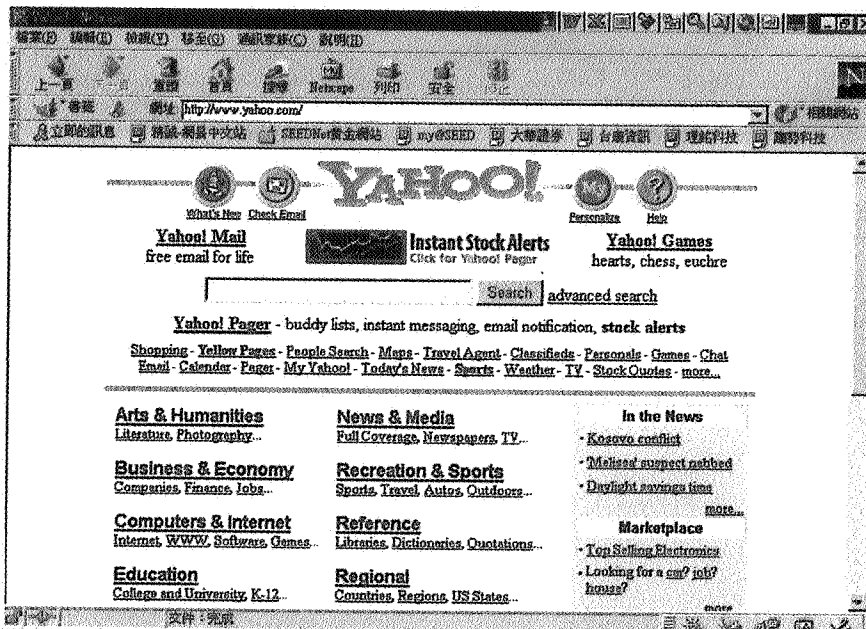


圖 14：yahoo 網站首頁



爲了能夠符合不同檢索需求，Yahoo 設計了一個檢索環境設定的功能，查詢者可依所想要檢索網路資源的範圍，如網頁、討論群，檢索比對方式、類別與網頁檢索、網頁架構時間，等做不同的設定。若只想要所搜尋到的網頁包括所有輸入的詞彙，可選擇 "matches on all words (AND)" 的比對方式 (search method)；若僅想搜尋近 1 個月內所架設的網頁資料，可由有 7 個選項的時間表中選取適當的時間範圍 (圖 15)。

## 二、「單一關鍵字」與「關鍵語句」

除了可就設定選項處理各輸入關鍵字外，必要時可利用雙引號“ ”，進行關鍵語句的檢索。例如所要查詢的資料是有關於「圖書館教育」的資料，若單純輸入 library education，兩者間關係爲「且」，可由 Yahoo 資料庫中找到 1,839 個網頁，其中不單是找到與「圖書館教育」有關的資料，還會找到許多「教育圖書館」(education library) 的資訊，並不符合檢索需求 (圖 16)；若配合雙引號的利用，則獲得符合程度較高的結果可能性爲大。(表 2)

表 2：「關鍵字」與「關鍵語」的檢索

結果 策略	Library education	"library education"
類別符合結果	6	1
網站符合結果	1839	12

## 三、符號的運用

### (一)加減號的運用

爲了確保字詞的出現與否，與 AltaVista 及 Openfind 相同的，Yahoo 以提供利用加減符號與關鍵字的組合使檢索結果更符合需求。例如 dublin-dublin core，可查詢到與 Dublin 有關的資料，但 Yahoo 會自動刪除與 Dublin Core 有關的網頁；而 +Dublin+core 的檢索策略則可確定所獲得資料包括 Dublin 與 Core 這兩個詞彙 (圖 17)。

### (二)切截字元

\* 與字根的配合，可擴展檢索結果所包括的範圍，以 cap\* 查詢，不單可找到 cap 的資料，也可找到 cape、capital 等相關網頁。以 cap 查詢，查詢 Yahoo 回覆 1,140 個網站資料；以 cap\* 檢索，Yahoo 回覆 11,445 個網站。

## 四、欄位限定

在檢索欄位限定上，檢索者可單就文獻名稱或者是網址部份進行查詢。

### (一)文獻名稱

t:dublin core，Yahoo 類表中有一符合類別、5 個網站、以及 264 個網頁名稱包括 dublin core。

### (二)文獻位址

u:dubline core，Yahoo 資料庫中僅有兩個網站符合，  
 • Dublin Core Metadata-a 15-element

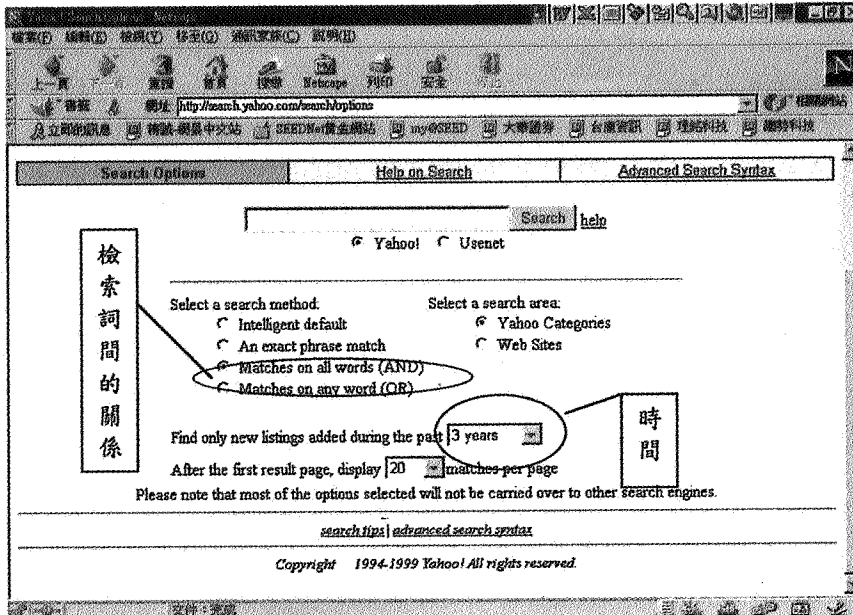


圖 15：檢索環境設定

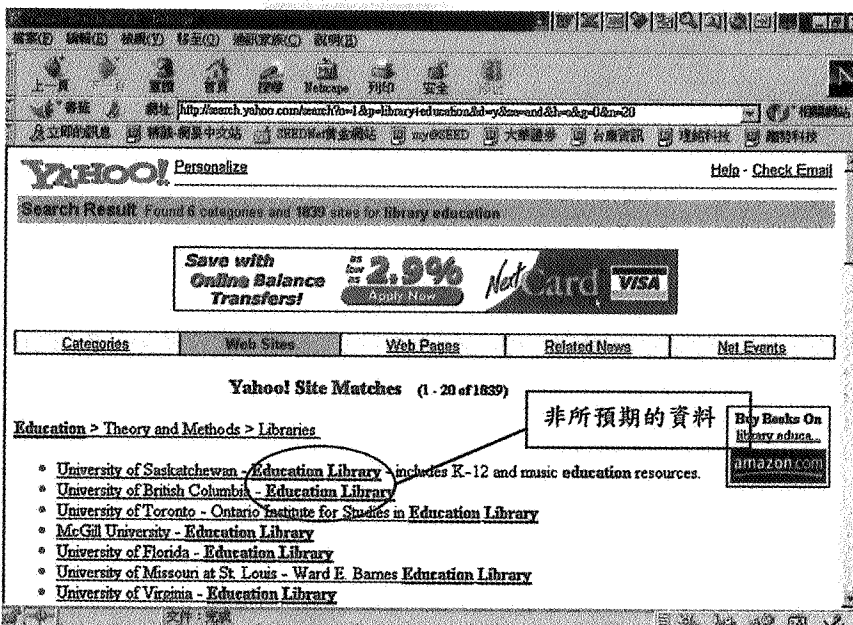


圖 16：關鍵字檢索 - library education





metadata element set intended to facilitate the discovery of electronic resources.

[http://purl.oclc.org/metadata/dublin\\_core/](http://purl.oclc.org/metadata/dublin_core/)

- OCLC/NCSA Metadata Workshop report

[http://www.oclc.org:5046/conferences/metadata/dublin\\_core\\_report.html](http://www.oclc.org:5046/conferences/metadata/dublin_core_report.html)

### 五、檢索結果的整理

若對檢索環境未多加限制，Yahoo 會就類別與網站本身進行符合比對，結果顯示則分為三部分，一是類別、一是網站以及網頁。以搜尋 "Dublin Core" 為例，所獲得的結果包括了在 Yahoo 25,000 不同類別中的 "Dublin Core" 專欄、列於專欄下的 5 個網站，如 Dublin Core Metadata、以及 2,557 個網頁（圖 18）。

### 捌、綜合比較

AltaVista、Openfind、Yahoo 三個搜尋引擎各有其特色，檢索與使用上都可協助資料搜尋者找到一些相關的資訊。以下就資料量、檢索功能、結果處理等做一分析比較。

#### 一、檢索資料量

這三個資源搜尋網站並未就所收錄的資料量提供一詳細數據，因此資料量的比較僅能就實際檢索範例結果，查詢有關

Dublin Core 的相關資料，做一簡單的比較說明。在三個搜尋站中都是以關鍵句型進行檢索，也就是說兩個字必須以輸入順序與格式同時存在網頁中。（表 3）

表 3：檢索結果摘要

結果	altavista	openfind	yahoo
網 站		67(935)	5
網 頁	6,716	2,437	2,557

#### (一) AltaVista

在 AltaVista 索引中共有 5,214 網頁包含 Dublin Core 這個字串，AltaVista 可將 Dublin Core 視為一關鍵語句進行檢索，但由於 AltaVista 未做歸類與整合的整理，5,214 個檢索結果的重覆率相當高。

#### (二) Openfind

Openfind 視 Dublin Core 為一關鍵句，先就網站與網頁名稱部份進行搜尋，列出系統判斷最相關的檢索結果，以及各網站或名稱下的網頁文件數；

進一步可查看更多的檢索結果，Openfind 共找到 67 個網站，按相關網頁數排序，67 個網站下共 935 個網頁，其他網頁 2,437。

#### (三) Yahoo

Yahoo 的檢索結果，依分類表找到一相關類別，電腦與網際網路 (Computer and Internet)，資訊與文件 (Information

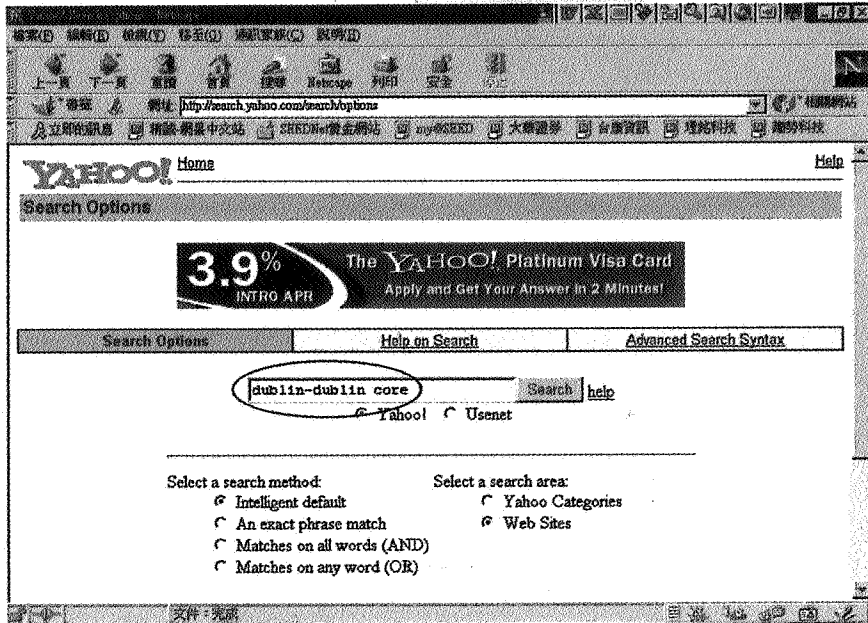


圖 17：減號的運用

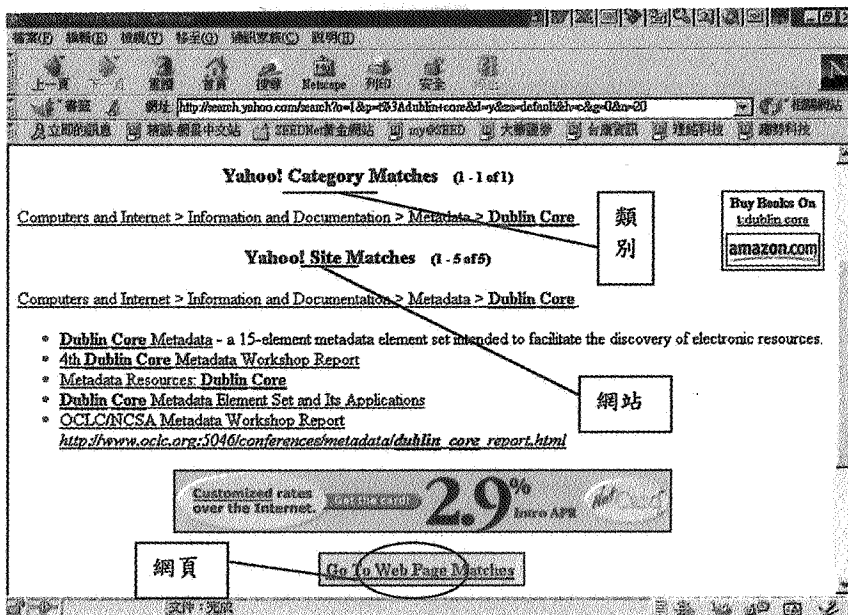


圖 18：檢索結果顯示



and Documentation) ，元資料 (Metadata) 下；Yahoo 資料庫中有 5 個相關的網站，2,557 網頁名稱或網址符合檢索條件。

## 二、功能分析

Altavista 、 Openfind 、 Yahoo 這 3 個搜尋引擎在功能設計上各有其特色，可就資料的整理方式、關鍵字詞的處理、運算子與符號的運用等不同層面來看。就資料整理方面來看，除了檢索外，Altavista 與 Yahoo 還提供檢索者以類表的方式瀏覽不同主題的網路資源；字詞處理方面，3 個搜尋引擎都可處理單一關鍵字、多重關鍵字以及關鍵語詞的查詢模式。為強化檢索的精準性與資料的回收率，布林邏輯運算子及符號運用，在 Altavista 、 Openfind 與 Yahoo 的檢索策略中均可整合運用；除此外 Altavista 與 Yahoo 還提供以

限制欄位的方式進行查詢。相對來說 Openfind 雖不強調這些運算方式的利用，但 Openfind 在跨語文以及模糊比對檢索上確實是 3 者間最特殊的。至於檢索範圍與結果的處理，Altavista 僅就網頁部份查詢、Openfind 區分網站與網頁、Yahoo 則是就類別、網站與網頁等 3 個不同層次做處理。(表 4)

## 玖、結語

搜尋引擎的發展有助於網路資源的搜尋與利用，雖說在檢索的精準度上各個搜尋引擎網站還有相當大的發展空間，但從發展之初到現在，搜尋引擎的幫助與自身功能設計的成長，在二次資訊爆炸時代的今天，確實是掌握網路資源上不可或缺的利器。

表 4：功能比較簡表

檢索功能	Altavista	Openfind	Yahoo
分類	13 類	No	14 類
搜尋	Yes	Yes	Yes
關鍵字	單一、多重	單一、多重	單一、多重
語句檢索	Yes	Yes	Yes
布林邏輯運算	And、or、not、near	&、 、!	and、or
符號	+、-、*	+、-、,	+、-、*
條件限制	9 種	no	2 種
模糊比對	No	Yes	no
自然語言	No	Yes	no
範圍與結果	網頁	網站與網頁	類別、網站與網頁