

網路圖像檢索需求初探

卜小蝶

摘要

本文分析大量的網路圖像與文字式網頁搜尋引擎檢索詞彙紀錄，藉以瞭解臺灣地區網路使用者的圖像檢索行為特性及需求，並進一步比較網路圖像與網頁檢索行為與需求之異同。研究結果包括網路檢索詞彙普遍簡短；網路檢索詞彙中的專有名詞比例很高，而在圖像的檢索比例則又更高、且以人名居多；網路圖像與網頁檢索行為皆有 20/80 原則的現象，即少數熱門詞彙佔有相當高的檢索比例，而且圖像檢索需求則更集中於少數的熱門主題類別；影視娛樂及動漫電玩等資訊是圖像與網頁共同的熱門檢索主題類別，但也有一些重要的圖像需求如動植物及自然天文等類，在網頁需求中並不明顯。

一、前言

圖像 (Image) 是一種視覺呈現的作品，其形式可以以相片 (Photomechanical)、圖畫 (Photographic) 或數位 (Digital) 等方式來表達。^[1] 其實圖像所包含的範圍相當廣泛，而網路圖像 (Web Images) 則指放置在網路上包括如人物風景相片、繪畫作品、新聞圖片、美工圖片、產品型錄、漫畫等公開的數位化圖像檔案，其不僅主題包羅萬象，同時檔案格式也很多元。隨著網路的蓬勃發展，網路圖像資源的成長相當驚人，據稱網路資源中有 73% 屬於視覺性

關鍵詞 (Keywords)：網路使用者研究；網路圖像檢索；檢索紀錄分析

Network User Studies ; Web Image Retrieval ; Query Log Analysis

卜小蝶：世新大學資訊傳播學系副教授；E-mail: htpu@cc.shu.edu.tw

資訊^[2]，而圖像則是其中最常見的類型。此外，在 Google 的圖像搜尋引擎中，目前就收集了超過 3 億 9 千萬張的網路圖像縮圖 (Thumbnail)。^[3]未來隨著更多收藏單位將圖像數位化後提供上網 (如圖書館、博物館、新聞社、公司、學校等)，以及更多的使用者上網查詢圖像，網路圖像檢索研究也就日形重要。

而面對如此快速成長的網路圖像資源與使用者，目前已出現一些網路圖像搜尋系統 (Web Image Retrieval System)，例如西方地區的 AltaVista、Yahoo、Excite、FAST 等、以及華人地區的 VisionNEXT 等。^[4]這些系統和一般網路搜尋引擎所提供的網頁搜尋服務類似，主要是用來協助使用者以關鍵詞方式，尋找特定圖像的位置。所不同的是，網路圖像搜尋是以圖像為搜尋對象，如檢索「海珊」指的是海珊的照片，而非其相關文字報導。這類系統通常是將圖像的檔案名稱 (Filename) 或圍繞圖像的文字 (Surrounding Texts) 建立索引，使用者在檢索時其實與檢索網頁沒有太大不同，但若是圖像沒有合適的文字描述，或是這些文字未能正確表達圖意，則容易造成查無所獲 (Zero Hits) 或錯誤組合 (False Drops) 的情形。此外，也有嘗試以圖像內容本身的資訊 (如顏色、形狀、紋理等) 來建立索引，以便進行圖像間的相似性 (Similarity) 比對，但這種作法由於缺乏語義 (Semantic) 層次上的分析，也未能充份滿足多元化的檢索需求。

相對於上述自動化的作法，其實圖書館在圖像的組織與檢索向來相當重視，而且也已累積不少經驗。^[5]一般而言，其作法與處理文字資訊類似，也就是先建

[1] 此地是根據哈佛大學視覺資源學會 (Visual Resource Association, VRA) 的定義："An image is a visual representation of a work. It can exist in photomechanical, photographic and digital formats," from <http://www.gsd.harvard.edu/~staffaw3/vra/vracore3.htm>.

[2] M.S. Lew, "Next-generation web searches for visual content," (*IEEE Computer*) 33(Nov. 2000), pp.46-53.

[3] 截至 12/31/2002，Google 所收錄的圖片數量，在網路上所公佈的數據為 390,000,000 張。網址：<http://image.google.com>。

[4] 許多入口網站都提供有圖像檢索的服務，如 AltaVista (<http://www.altavista.com/>)、Yahoo (<http://www.yahoo.com/>)、FAST (<http://www.alltheweb.com/>)、Google (<http://image.google.com/>) 等，也有專門的圖像搜尋引擎，如 VisionNEXT 所開發的 eefind 搜尋引擎 (<http://www.visionnext.com/>)。

[5] 有關圖像組織的探討，在 E. Rasmussen, "Indexing images." (*Annual Review of Information Science and Technology*) 32(1997), pp.169-196. 一文中完整介紹。而圖像檢索的研究，在 (*Library Trend Fall 1999*) 特刊中針對 "Progress in Visual Information Access and Retrieval" 主題也有一系列的探討。

立一套組織圖像資訊的分類架構，及相關的主題控制詞彙，再以人工方式對圖像的種種特徵加以分析，以便建立索引，例如Getty Research Institute 就有相當多有關藝術圖像的收集與整理計畫，同時該組織也很積極地建立圖像相關的主題分類標準。^[6]透過這種書目控制的作法來組織整理圖像之後，使用者即可以分類瀏覽或輸入關鍵詞來查詢圖像，雖然這種方式對於檢索品質的提昇有很大幫助，但考量人工分析的成本相當高，同時，在面對成千上萬、異動頻繁的網路圖像資源，這種作法也是緩不濟急。

綜合上述，不論是以自動或人工方式來組織圖像或提供檢索，都有許多待改進空間。而為了讓網路圖像的組織與檢索能更具效益，實有必要先由網路圖像的使用者著手，畢竟使用者才是這類資訊的真正需求者。換言之，瞭解圖像檢索行為是有效組織與檢索圖像的基礎。然而目前有關圖像檢索行為的研究仍相當有限^[7]，而針對網路圖像檢索行為的分析則更為缺乏。

本文主要藉由大量的網路圖像搜尋引擎檢索詞彙紀錄分析，希望能初步瞭解臺灣地區網路使用者的圖像檢索行為特性及需求，包括檢索詞彙長度、檢索詞彙頻率、輸入詞彙語言、及重要的檢索興趣等。並進一步以作者先前對文字式網頁（以下簡稱網頁）搜尋引擎的檢索詞彙紀錄分析結果為基礎^[8]，嘗試比較網路圖像與網頁檢索行為與需求之異同。研究結果一方面可提供網路圖像使用者研究的參考，同時對於發展圖像或多媒體檢索系統也將有所助益。

二、相關研究

圖像資訊的使用者來自相當多元的學科背景，包括藝術、建築、新聞、教育、出版、娛樂、廣告、法律、醫學、工程、地理等，因此有關圖像檢索行為的研究，多半是針對特定主題收藏或特定類型人員來進行分析，例如 Rorvig^[9] 分

[6] Getty Research Institute, from <http://www.getty.edu/>.

[7] A.A. Goodrum, "Image information retrieval: An overview of current research," *Information Science* 3(Feb.2000), pp.63-66.

[8] 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（1999.12），頁 49-62。

[9] M. Rorvig, et al. "The NASA image collection visual thesaurus." Proceedings of the ASIS 17th Mid-Year Meeting, Ann Arbor, MI. (1988).

析 NASA 圖像索引典的使用情形；Ornager^[10]、Markkula & Sormunen^[11] 調查記者查詢新聞圖片的行為；Keister^[12] 分析美國醫學圖書館醫學影像資料庫的檢索詞彙；Markey^[13]、Hastings^[14]、Chen^[15] 等人則針對藝術史研究人員或學生，瞭解其對藝術圖像的需求及使用情形。這些研究提供了一些有趣的檢索行為觀察結果，包括使用者利用文字較利用圖像來檢索圖像的比例高；圖像的資訊需求通常較文字性資訊需求來得專指 (Specific)^[16]，例如想查詢倫敦大橋的圖片時，會輸入「倫敦大橋」，而非較不專指的「橋」；或是使用者常常會同時利用圖像所具備的抽象概念與基本特徵來檢索，而非單純只利用概念或內容來檢索圖像。也有一些研究嘗試瞭解影響圖像檢索行為的因素，例如 Jorjenson^[17] 與 O'Connor^[18] 都曾提出檢索任務 (Retrieval Task) 對使用者檢索行為的影響，譬如在沒有檢索任務的要求下，使用者很容易以簡短敘述方式 (Narrative) 來描述圖像需求，而且其所提供的敘述，有很多是超出該圖像所要表達的內容。綜合上述，圖像檢索行為受到使用者所欲從事的任務 (Task)、學科背景 (Domain) 及圖像收藏特性 (Types) 的影響。這方面的研究目前以小規模的實證研究為主，其實驗環境多以非數位化圖像資料為主，對於圖像檢索系統的具體影響有待進一

[10] S. Ornager, "Image retrieval: Theoretical and empirical user studies on accessing information in images," Proceedings of the ASIS 60th Annual Meeting 34 (1997), pp.202-211.

[11] M. Markkula, & E. Sormunen, "End-user searching challenges indexing practices in the digital newspaper photo archive," *Information Retrieval* 1 (2000), pp.259-285.

[12] L. Keister, "User types and queries: Impact on image access systems," Proceedings of the ASIS 57th Annual Meeting 31 (1994), pp.7-22.

[13] K. Markey, "Access to iconographical research collections," *Library Trends* 37:2 (1988), pp.154-174.

[14] S. Hastings, "Query categories in a study of intellectual access to digitized art images." Proceedings of the ASIS 58th Annual Meetin 32 (1995)

[15] H. Chen, "An analysis of image queries in the field of art history," *Journal of the American Society for Information Science and Technology* 52:3 (2001), pp.260-273.

[16] P.G.B. Enser, "Progress in documentation: Pictorial information retrieval," *Journal of Documentation* 51:2 (1995), pp.126-170.

[17] J. Jorjenson, "The applicability of existing classification systems to image attributes: A selected review," *Knowledge Organization and Change* 5 (1995), pp.189-197.

[18] B. O'Connor, M. O'Connor, & J. Abbas, "Functional descriptors of image documents: user-generated captions and response statements," *Journal of the American Society for Information Science* 50:8 (1999), pp.681-697.

步研究。

過去限於實驗環境的不足，針對網路圖像檢索行為的研究相當少見。其中針對 Excite 搜尋引擎所提供的圖像檢索詞彙所進行的分析^[19]，研究者以人工方式分析了 9,855 個使用者所輸入的 33,149 個檢索詞彙，發現圖像檢索詞彙平均長度並不長（平均為 3.74 words），檢索詞彙重複的比例並不高（高頻詞彙佔總檢索次數不到 10%），同時詞彙的修改多數是將詞彙修飾的更為明確（例如將 "girls" 修改為 "pretty girls"）。但上述研究是以 28 個圖像相關詞彙（如檢索詞彙中出現 "picture" 即視為圖像檢索詞彙），間接由網頁檢索詞彙中過濾擷取出所謂的圖像檢索詞彙，因此每個詞彙基本上都增加了一些圖像限定的詞彙，這多少也解釋其統計的詞彙長度為何會比網頁檢索詞彙還長。此外，上述研究主要提供初步的統計分析，有關圖像檢索行為及其在檢索系統的應用需要更深入的探討。

三、研究方法

本研究以超過 450 萬筆臺灣地區圖像及網頁搜尋引擎檢索紀錄作為實驗分析對象，實驗記錄之內容主要為使用者輸入之檢索詞彙及其累積頻率。如表一所示，本文主要利用 VisionNEXT（新視科技）及 Dreamer（夢想家）二家公司所提供之檢索記錄為主。VisionNEXT 所開發的圖像搜尋引擎提供華人地區最大的圖像搜尋服務，透過此圖像搜尋引擎，使用者可以中文及英文查詢包括在華文地區以及西方國家的網路圖像。目前其所收錄的網路圖像已超過 500 萬張縮圖，其中有 10 萬張縮圖是經過人工分類，並給予適當的關鍵字。其所提供的檢索功能包括關鍵字查詢與主題分類瀏覽等二種檢索模式，就圖像的數量與檢索功能而言，VisionNEXT 算是相當完整，因此其也成為海峽兩岸許多重要入口網站的後端伺服器（Backend Server），包括如臺灣地區的 PCHome 及大陸地區的 Sina 新浪網等。而 Dreamer 則曾經是臺灣地區相當重要的入口網站之一，主要提供網站主題目錄瀏覽及網頁全文檢索的服務。

^[19] A. Goodrum & A. Spink, "Image searching on the Excite web search engine," *Information Processing and Management* 37:2(2001), pp.295-312.

表一：網路圖像及網頁檢索記錄實驗資料

實驗資料	來源 (收集期間)	服務類型	總詞彙筆數	總檢索次數
V-2002	VisionNEXT 圖像檢索 記錄(2002 年 7-9 月間)	網路圖像分類目錄及關 鍵字搜尋服務	325,812	2,401,628
D-1998	Dreamer 網頁檢索記錄 (1998 年 3 個月份)	網站分類目錄及網頁全 文搜尋服務	228,502	2,183,506

有鑑於華文地區幾乎不見有關網路檢索詞彙的大規模統計分析，因此本研究首先建立瞭解各項檢索詞彙統計分析的基本工具與環境，進行電腦統計分析，包括檢索詞彙的字串長度分析、檢索頻率的分佈情形，檢索詞彙的語文分佈等。而有關檢索紀錄的主題內容分析，主要是利用作者先前研究所發展出的檢索詞彙自動分類方法^[20]，將高頻熱門檢索詞彙依事先訂定之類別進行分類，以觀察各類檢索主題的分佈情形。此外，為瞭解網路圖像與網頁檢索詞彙的異同，本研究也嘗試將上述分析結果與先前研究作一比較分析，例如簡短檢索詞彙 (Short Queries) 是否為一普遍現象？高頻熱門檢索詞彙是否皆佔有相當高的比例？網路使用者的熱門檢索主題為何？網路圖像與網頁檢索主題有何異同？等等。

四、網路圖像與網頁檢索行為特性分析

(一) 網路檢索詞彙普遍簡短

根據國外相關研究^[21]及本文分析結果，不論是查詢圖像或文字式網頁資訊，網路檢索詞彙一般而言皆相當簡短。以本文所收集之 VisionNEXT 圖像搜尋引擎 (V-2002) 及 Dreamer 網頁搜尋引擎 (D-1998) 所提供共超過 450 萬筆檢索詞彙記錄統計分析為例，如表二所列，網頁中文檢索詞彙平均為 3.18 字 (Characters)，而圖像中文檢索詞彙平均為 3.08 字，顯示多數中文檢索詞彙接近為一複合名詞 (Compound Nouns)，因為根據一般中文詞彙統計分析，平均中文

^[20] Hsiao-Tieh Pu, Shui-lung Chuang, & Chyan Yang, "Subject categorization of query terms for exploring Web users' search interests," *Journal of the American Society for Information Science & Technology* 53:8(2002), pp.617-630.

^[21] B.J. Jansen, & U. Pooch, "A review of web searching studies and a framework for future research," *Journal of the American Society for Information Science* 52:3(2001), pp.235-246.

詞長約為 1.5 ~ 1.6 Characters , 3.18 Characters 約為雙詞 (Word Bigram or Word Pair) , 這與國外所統計的平均長度 2.35 words 相近。然而, 與國外不同之處, 臺灣網路使用者的西文檢索詞彙平均長度普遍低於國外統計 (網頁及圖像的西文檢索詞彙分別平均只有 1.10 Words 及 1.40 Words) , 且多半屬於一些簡短的專有名詞, 如 "Microsoft"、"Spiderman" 等。

表二：網路圖像及網頁檢索詞彙平均字串長度

檢索詞彙 \ 網路圖像	V-2002	D-1998
中文檢索詞彙	3.08	3.18
西文檢索詞彙	1.40	1.10

本文進一步統計中文檢索詞彙的平均字數, 如表三所列, 圖像檢索詞彙中有較多的一字詞, 其中有不少名稱, 如「狗」、「貓」、「魚」等皆屬於高頻的熱門詞彙, 多少影響了整體檢索詞彙的平均長度, 但一般而言, 不論圖像或網頁的檢索詞彙皆相當簡短, 並不易瞭解其資訊需求或檢索動機, 因此, 使用者經常無法順利檢索出滿足其需求的相關資訊, 而這也是長期以來發展資訊檢索系統所持續努力改善的方向之一。為了克服這類簡短檢索詞彙問題 (Short Queries Problem) , 一些研究如利用相似性回饋 (Relevance Feedback) 或擴大詞彙 (Query Expansion) 等技術^{[22] [23]}, 都是嘗試以增加系統與使用者互動的機會, 來產生較佳的檢索結果。

表三：網路圖像與網頁中文檢索詞彙平均字數統計

檢索詞彙 \ 網路圖像	一字詞	二字詞	三字詞	四字詞	五字詞以上
V-2002	4.33%	33.01%	34.69%	17.26%	10.44%
D-1998	1.67%	40.68%	19.91%	25.13%	12.61%

[22] Donna. Harman, "Relevance feedback revisited," *ACM SIGIR '92*(1992), pp.1-10.

[23] Bruce R. Schatz, et al. "Interactive term suggestion for users of digital libraries: Using subject thesauri and co-occurrence lists for information retrieval," *Digital Library* (1996), from <http://dli.grainger.uiuc.edu/schatzDL96.htm>.

(二) 網路檢索詞彙內容呈現多樣性

根據表四所列檢索詞彙的頻率分佈統計結果，網路圖像與網頁檢索行為皆有 20/80 原則的現象，即少數熱門詞彙佔有相當高的檢索比例。如頻率超過 3 次以上的檢索詞彙在二者分佔了 85.17% 及 88.14% 的總頻率，顯示少部分詞彙是使用者熱門的檢索詞彙。然而圖像的熱門檢索詞彙似乎更為集中，如其前 1,000 大詞彙就佔了超過 80% 的總頻率，此外，其前 25 大詞彙雖然僅佔所有詞彙的 0.00027%，但其頻率則佔了 14.61%。雖然網路檢索詞彙有集中於熱門詞彙的情形，但由低頻詞彙不論在圖像或網頁檢索中都佔有相當比例來看（如頻率小於 3 者，網頁檢索詞彙為 11.86%，而圖像檢索詞彙為 14.84%），網路檢索詞彙的內容其實相當多元，也顯示其檢索需求的多樣性。

表四：網路圖像及網頁檢索詞彙頻率統計

網路圖像 檢索詞彙	頻率 =1	頻率 =2	頻率 >3
V-2002	9.00%	3.74%	85.17%
D-1998	5.48%	3.74%	88.14%

(三) 網路檢索詞彙主要以中文表達

根據檢索詞彙的語文分佈統計結果（表五），很明顯地網路圖像與網頁檢索詞彙多以中文來表達，圖像有近 91% 為中文檢索詞彙，而網頁則有近 81% 的比例。有趣的是，許多專有西文名詞，使用者有時並未使用西文、反而是以中文來輸入，如 "Yahoo" 以「雅虎」來輸入，顯示使用者仍較習慣以中文來表達其檢索需求。這種現象在圖像檢索詞彙中則更為明顯，由於專指度高的檢索詞彙相當多（特別是人名），其中又有很高比例是查詢國外明星，如 "Britney Spears" 以「小甜甜」或「布蘭尼」輸入，因此許多檢索需求其實是以國外資源為目標，只是臺灣的網路使用者較不習慣或不擅於以西文來表達其需求。此外，西文檢索詞彙的比例並不高（圖像有 9.08% 為西文檢索詞彙，而網頁則為 19.18%），且根據抽樣觀察，這些詞彙多數為單詞、且屬於專有名詞，其專指度 (Specificity) 都相當高，如圖像檢索詞彙 "spiderman"、網頁檢索詞彙 "kimo" 等，顯示西文資訊的需求實際上並不複雜。

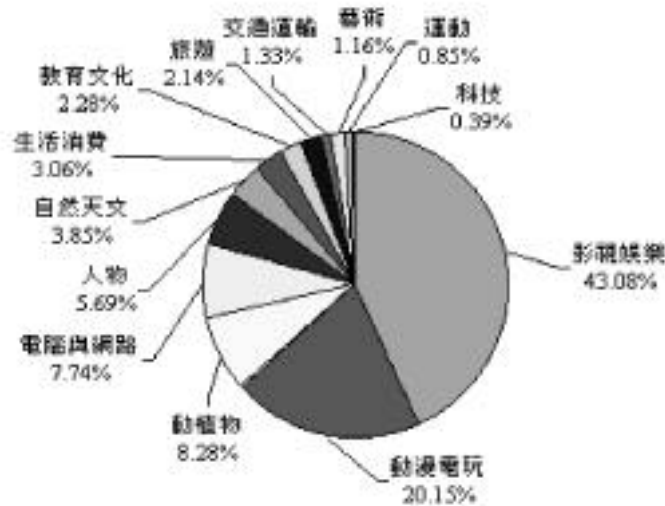
表五：網路圖像及網頁檢索詞彙語文分佈

檢索詞彙 \ 網路圖像	V-2002	D-1998
中文檢索詞彙	90.92%	80.86%
西文檢索詞彙	9.08%	19.18%

五、網路圖像與網頁熱門檢索主題分析

(一) 網路圖像與網頁熱門檢索主題類別分佈

為觀察比較圖像與網頁檢索主題的異同，作者利用先前研究所提出之檢索詞彙分類方法，將圖像及網頁的前1,000大熱門檢索詞彙進行適當分類後，所得結果如圖一及圖二所示。由於實驗中所使用之圖像搜尋引擎具有過濾色情詞彙的功能，因此在本文並未列入色情類資訊的統計。在圖像檢索中，查詢影視娛樂類者最多，佔有43.08%，依次為動漫電玩20.15%、動植物8.28%、電腦與網路7.74%、人物5.69%、自然天文3.85%、生活消費3.06%、教育文化2.28%、旅遊2.14%、交通運輸1.33%、藝術1.16%、運動0.85%、科技0.39%等。而在網頁檢索中，查詢電腦與網路類者最多，佔有28.22%，依次為影視娛樂12.83%、育樂交友11.71%、生活消費9.68%、動漫電玩8.18%、旅遊6.56%、教育學習5.89%、財經工商5.61%、媒體出版3.73%、政治社會3.57%、人文藝術2.25%、醫療保健1.08%、及學術資訊0.68%等。由二圖比較很容易看出，影視娛樂、動漫電玩類的資訊在圖像及網頁的需求上都佔有相當高的比例；而有一些在圖像需求中相當重要、但在網頁檢索中並不明顯的主題類別，例如動植物及自然天文等類；也有一些在網頁中相當重要、但在圖像檢索中並不明顯的主題類別，例如財經工商及政治社會等類。此外，圖像檢索的前三大類總檢索次數佔有65.8%、網頁為51%，顯示少數主題類別是網路使用者的熱門檢索需求，但值得注意的是，圖像檢索需求相較網頁而言有更集中於少數類別的現象。究其原因，可能是網頁資訊需求較為廣泛，圖像資訊需求基本上屬於整體網路資訊需求的一環，因此其所檢索的主題內容相對而言也就不是那麼多元。



圖一：網路圖像熱門檢索主題類別分佈



圖二：網頁熱門檢索主題類別分佈

(二) 網路圖像與網頁的熱門檢索主題

除了上述的主題類別分佈之外，本文進一步檢視一些熱門的檢索詞彙，來瞭解圖像與網頁需求內容的差異，如表六分別列出圖像及網頁的前 20 大檢索詞彙，二者所呈現的需求明顯不同。由於這兩種紀錄是不同時期收集，其熱門檢索詞彙會有所差異。對網頁檢索需求而言，多半是集中在一般性 (Generic) 的主題

檢索，如 MP3、小說等；然而對圖像而言，查詢專指度高 (Specific) 的人名卻有相當高的比例，換言之，圖像在查詢人名的需求上明顯偏高，這與網頁檢索需求有所不同。

表六：網路圖像及網頁前 20 大熱門檢索詞彙及比例

D-1998 前 20 檢索詞彙	檢索次數	比例	V-2002 前 20 檢索詞彙	檢索次數	比例
MP3	42561	1.95%	薛楷莉	41590	1.73%
模擬器	15071	0.69%	S.H.E	17351	0.72%
Icq	13899	0.64%	林熙蕾	15018	0.63%
桌面	12092	0.55%	章小蕙	14813	0.62%
桌面王	11680	0.53%	河莉秀	14492	0.60%
蕃薯藤	10000	0.46%	背景	14445	0.60%
奇摩	9328	0.43%	周杰倫	13581	0.57%
Bbs	8613	0.39%	劉佳佳	11669	0.49%
Kimo	8166	0.37%	陳寶蓮	10931	0.46%
104	7943	0.36%	遊戲王	10059	0.42%
小說	7456	0.34%	張娜拉	9629	0.40%
歌詞	7217	0.33%	賤兔	9519	0.40%
圖片	7163	0.33%	卡通	8819	0.37%
統一發票	6998	0.32%	布蘭妮	8811	0.37%
笑話	6376	0.29%	李玟	8067	0.34%
遊戲	6296	0.29%	Christina Aguilera	7619	0.32%
星座	6141	0.28%	風景	7518	0.31%
泡麵	5981	0.27%	凱莉米洛	6991	0.29%
Kitty	5640	0.26%	哈姆太郎	6867	0.29%
Game	5490	0.25%	開朗少女成功記	6845	0.29%

註：網頁總檢索次數為 2,183,506；圖像總檢索次數為 2,401,628。

綜合上述，有關檢索行為特性的分析結果，包括二者在檢索詞彙的長度、頻率、及語文的使用狀況等，分析結果包括網路檢索詞彙普遍簡短；而二者的高頻詞彙檢索比例都很高，但圖像有更明顯的20/80 原則現象，即少數檢索詞彙佔有多數的檢索頻率，除此，二者其實也存在不少低頻詞彙，顯示檢索內容仍相當多樣 (Diverse)；此外，二者所使用的語言固然以中文為主，但有不少圖像檢索詞彙其實是以非中文的檢索內容為目標，只是臺灣網路使用者較習慣以中文來表達其檢索需求；最後，網路檢索詞彙中的專有名詞比例很高，而在圖像的檢索比例則又更高、且以人名居多。除上述檢索行為的特性分析外，本研究也初步觀察及比較圖像與網頁的熱門檢索需求類別，一般而言，影視娛樂類資訊在圖像及網頁的檢索需求上皆佔有相當高的比例，只是圖像在這方面的需求遠高於網頁，然而在網頁檢索需求中最重要的電腦與網路類資訊在圖像需求中就不是那麼明顯；此外，值得注意的是圖像存在一些相當重要、但在網頁需求中並不明顯的主題類別，例如動植物及自然天文等類。

六、結語

圖像檢索需求的研究在國外向來受到相當重視，有關網路圖像檢索需求的研究也正方興未艾，鑑於國內目前相關研究仍相當缺乏，上述研究結果提供網路圖像與網頁檢索行為的一般特性及檢索主題類別的分析比較。為了更進一步瞭解使用者對網路圖像的檢索需求，有必要針對上述結果再進行更深入的分析與探討，至少可包括三個方向：首先，檢索紀錄的分析範圍除了檢索詞彙 (Terms) 外，檢索歷程 (Sessions) 也提供了許多觀察檢索行為的線索，例如檢索詞彙與瀏覽或點選圖像之間的關聯，而除了利用電腦的檢索紀錄進行分析外，若能配合針對不同主題領域、任務、或身份的網路使用者進行實體訪查，則較能完整深入瞭解圖像檢索需求的內容，同時也有助於瞭解現行圖像檢索系統在資源及功能上有何限制。其次，若能更進一步探討現有圖像檢索需求 (Image Query) 的分析模式 (Analytic Model)，並針對網路圖像檢索需求發展出一套分析模式，將有助於在大規模網路圖像檢索環境中，以系統化方法來瞭解網路圖像與網頁檢索需求的差異，甚至對多媒體資訊需求的瞭解也有所幫助。最後，如何將上述對使用者檢索需求的瞭解，具體應用於圖像檢索系統也是重要的努力方向，例如改善瀏覽介面、協助檢索結果的呈現、提昇檢索功能、及輔助檢索系統的評鑑等。未來隨著

多媒體資訊的日益增多，瞭解使用者的多媒體資訊需求與檢索行為特性，將是提升資訊服務品質的重要任務，同時也是改善多媒體資訊檢索系統的基礎，本文只是初步研究，未來值得更多研究的投入。

誌 謝

感謝 VisionNEXT 及 Dreamer 公司分別提供大量的網路圖像及網頁檢索詞彙紀錄檔案、中研院資訊所莊水龍先生的技術支援、及國科會專題研究計畫之補助 (NSC91-2413-H-128-003)

An Exploratory Study on Web Image Queries

Hsiao-tieh Pu

Abstract

This study analyzed a large number of Web image and textual queries from Web image and Web page search engine logs. The main purpose was to better understand the image requests and search behavior of networked users in Taiwan. Further, the study also provides a comparison between image and text queries in terms of query length, query frequency, language used in query, and popular subject categories searched. The results show that networked users tend to input short queries; the content of the queries is rather diverse; personal name searches are obvious within image requests; entertainment and games are both important requests in image and textual queries, while image requests contain other important categories like animal/plant and nature categories, which not so obvious in the textual requests.

Keywords (關鍵詞): Network User Studies ; Web Image Retrieval ; Query Log Analysis
網路使用者研究 ; 網路圖像檢索 ; 檢索紀錄分析

Hsiao-Tieh Pu : Associate Professor, Department of Information and Communications, Shih Hsin University ; E-mail: htpu@cc.shu.edu.tw