

# 輔助新聞檢索之視覺化介面實作 與使用者評估

Designing and Evaluating User Interface for Assisting  
News Navigation and Retrieval

謝吉隆

**Ji-Lung Hsieh**

國立師範大學圖書資訊學研究所助理教授

Assistant Professor

Graduate Institute of Library and Information Studies

National Taiwan Normal University

沈柏辰

**Bo-Chen Shen**

國立師範大學圖書資訊學研究所

Graduate Institute of Library and Information Studies

National Taiwan Normal University

楊立偉

**Li-Wei Yang**

龍捲風科技董事總經理

Managing Director

Tornado Technologies

國立臺灣大學工商管理學系暨商學研究所兼任助理教授

Adjunct Assistant Professor

Department and Graduate Institute of Business Administration

National Taiwan University

---

投稿日期：2015年5月17日；接受日期：2015年6月30日

Email: 謝吉隆 jirlong@gmail.com; 沈柏辰 paichen.shen@gmail.com; 楊立偉 wyang@ntu.edu.tw

## 【摘要 Abstract】

網路新聞平台使得大眾易於取得多元且大量的新聞資訊，但卻因搜尋結果所傳回的資訊是依照搜尋準確率排列，無法展現所得資訊的時間性或其間的關聯性，因而使用者反而需要花更多的時間去檢索或瀏覽結果，同時也挑戰了使用者的背景知識和檢索技巧。著眼於此，本研究開發一套動態新聞視覺化瀏覽系統，企圖視覺化新聞檢索結果中的資訊關聯性；並嘗試在視覺化介面中整合並允許使用者刪修檢索歷程、擴大或縮小搜尋範圍，以降低搜尋歷程的使用者認知負荷。本研究透過主題式任務來評估本視覺化系統的使用性，經過訪談與循序序列探勘方法後獲得以下回饋：一、使用者最常用的功能是展開概念網絡中的相關關鍵字以發掘更多相關新聞，可達到觸類旁通的功效；二、使用者認為本介面所提供的三種相異的網絡視覺化繪圖方法可分別提供其逐一檢視相關關鍵字或者觀察概念群集的功能；三、使用者能夠利用歷程紀錄的功能，一方面修正檢索目標，另一方面則可刪除不必要的概念以降低參與者的認知負荷。然而，亦有使用者認為操作不便或因為介面複雜而產生新的認知負荷。

The development of online news website let the public obtain news much easier and more convenient. However, retrieved information are sorted by relevance, therefore can not display the temporal features or relations among information, which poses challenges to users' ability of information retrieval skills and their background knowledge. To assist users in browsing news, we built up a keyword visualization user interface that allows users to browse news from different aspects, to control their visit paths and to expand or filter information. We conducted interview and applied sequential analysis method to evaluate user experience in this study. The results show that 1) the function that users used the most is clicking to expand related keywords to discover more keywords and related news; 2) participants reported that the three network layouts implemented in the study can assist them to navigate keywords one by one or to read briefly potential keyword categories; 3) by visualizing visit path, participants found that they can locate themselves in information retrieval process or to alter keywords set for alleviating cognitive

load. However, the feedback of user study also shows that gathering excessive visualization design on an interface may increase cognitive load.

### 【關鍵字 Keywords】

資訊視覺化；新聞導覽；概念圖；資訊檢索歷程

Information visualization; News visualization; Information visualization evaluation

## 壹、新聞環境與檢索

資訊社會的發展讓我們具有高度理性化的資訊生產。然而，其流動性、去鑲嵌性、空間壓縮、時間壓縮、即時關係等特性，也帶來了資訊超載、無法有效整理資訊等現象（Lash, 2002）。易言之，當資訊不但日益龐大且逐步複雜化之後，我們很難從中得到一種有序的、關係化的資訊產出。以線上新聞為例，即使搜尋引擎日益進步，但它也無法避免資訊超載、錯誤資訊、或資訊失控等非理性現象。可能的原因一則是搜尋引擎能傳回大量結果，但使用者不確定其所需要的結果為何；另一種更具體的原因是，搜尋結果的排列順序是以相對於檢索詞彙的搜尋準確率而定，但這樣的資訊排列方式只能將新聞依序羅列，卻無法呈現出這些新聞間是否存在特定的關聯性。例如，以關鍵字查詢所獲的新聞，往往會過於重複、且難以被檢索；再者，現下新聞網站在每篇新聞下方或者側欄會推薦相關的新聞，但不受時間與空間限制的網路空間卻也讓過去與今日的新聞同時出現在同一個新聞版面，因而也大幅增加使用者自行判斷資訊時效性的難度。為了解決這樣的問題，過去的研究多以自動化摘要、自動化分類等協助使用者有效地擷取或再組織有用的資訊（卜小蝶、陳思穎，2007）。例如過去已有學者針對於不同網站上重複的新聞進行事件本體研究或多新聞文稿間的自動摘要（Chen, Kuo, Huang, Lin, & Wung, 2003）；鄭卜壬（2009）亦曾用本體方法在重大傳染病流行期間偵測中、外新聞所發現的實際疫情數量，避免因為新聞轉載的問題而高估疫情。

上述方法雖然可以避免資訊搜尋中的新聞重複問題，但它所帶來的是一種精簡的資訊美學，卻仍然無法在資訊搜尋間建立連結，因而呈現出資訊關聯的複雜度。基於此般需求，視覺化技術希望能提供在眾多資訊中架接橋樑的有效工具。隨著網路速度的提升，過去原本被認為因耗費大量計算資源、因而多作為個人電腦桌面程式設計的視覺

化技術，在整合 JavaScript、HTML5 Canvas、CSS、JSON 以及雲端運算等相關技術之後，專用型的視覺化桌面程式已逐漸成為輕量型的線上服務，可被鑲嵌在網頁的邊欄上，提供使用者綜覽、導覽、閱讀與搜尋等輔助功能，成為使用者閱讀或瀏覽資訊的另一種途徑。資訊視覺化提供介面讓使用者以視覺參與資料探勘，同時呈現高維度的異質性資料，以圖形展現資料與資料間的關係，或提供圖形化的摘要。資訊視覺化的價值並非提供美觀的介面，而是強化使用者對議題的認知、節省搜尋時間、突顯資料特徵、協助進行推論或者有效地傳遞資訊（Card, MacKinlay, & Shneiderman, 1999; Friedman, 2008）。而在建立資訊關聯的面向上，將搜尋結果視覺化可協助使用者理解搜尋結果（Kao, Lei, & Sun, 2008），例如將搜尋結果匯集成數個核心概念或群集。或基於資訊間的相似性提供延伸脈絡，讓使用者得以調整或者擴展其搜尋（Keim, 2002）。

建基在這樣的功能之上，歷年來視覺化工具在新聞或資訊檢索的實際應用，包括 So-net 所提供的 Blog Keyword Visualizer，能讓使用者即時觀察整個部落格平台的熱門關鍵字（簡稱 BKV，於 2007 年 1 月已經停止服務）；而 Google 所提供的 Wonder Wheel 則以關鍵字網絡輔助使用者檢索線上資訊（亦已停止提供服務）。專為瀏覽新聞所設計的視覺化介面則有 Newsmap，其利用顏色與方塊面積分別呈現新聞類別與熱門程度；而 ThemeRiver 或是 Google Trend 則以時間軸來表現議題的趨勢變化；而國內曾元顯教授亦曾為臺灣師範大學圖書館的聯合新聞資料庫，用關鍵字共現在同一則新聞的關係來建立關鍵字網絡，以協助讀者獲取相關關鍵字並修正檢索方向<sup>1</sup>。然而，上述案例均專注在視覺化時間、熱門程度、或者關鍵字間的脈絡等因子，對於讀者在視覺化新聞介面的檢索與檢索歷程，包含圖文呈現、檢索歷程調整與回溯等卻少有相應的輔助設計，但這對於探究人與介面的互動歷程，進而建立檢索資訊間的複雜關聯的現象，卻是極為重要的面向。例如 Bates（1989）即以採野莓（berrypicking）模式譬喻檢索歷程可被視為一連串的搜尋與瀏覽行為。以讀者閱讀新聞為例，讀者會經常在搜尋、瀏覽、修正方向的反覆過程偶遇一些有趣的新聞，而非目的性地找到或者尋獲他所需要的新聞。對於現在擁有大量新聞內容、且囊括諸多

---

1 國立臺灣師範大學資訊中心主任曾元顯教授所設計的新聞瀏覽器利用網絡方法繪製關鍵字與關鍵字的共現關係（<http://ir.itc.ntnu.edu.tw/udn/search.aspx>），資料涵蓋的範圍為聯合新聞資料庫 2000-2009 年間的所有紙本報紙新聞。

推薦、置頂、熱門新聞等功能的線上新聞網站而言，讀者可能會感興趣的線上新聞經常出現在超鏈結列表中，而不是以一種集中且被妥善安排的方式來提供讀者瀏覽。在這樣的情形下，使用者必須要透過搜尋或者不斷探索式地點選連結來獲取相關資訊。

因此，在上述資訊大量重複且時間性模糊的環境下，新聞讀者可能會花更多時間在瀏覽（Delphi Group, 2004; Kao et al., 2008），而不是帶有清楚目的性的搜尋活動（Cove & Walsh, 1988）。準此，本研究除了嘗試將新聞以關鍵字共現關係作為脈絡以視覺化網絡外，尚企圖在介面設計上結合使用者於檢索歷程的可能需求。對檢索歷程的需求可以在過去的研究中找到例證，例如江信昱與卜小蝶（2010）在探究使用者於資訊再尋獲的過程的檢索策略時提及，使用者在重返網頁時其多希望能夠獲得「路徑（path）」而非標的（target），因而認為「路徑」在再尋獲的過程中是一個重要的線索。除此之外，使用者有限的注意力亦輔助檢索介面設計的重要考量之一（卜小蝶、陳思穎，2007）。準此，考量「路徑」作為一種檢索歷程的線索，本研究將嘗試在新聞關鍵字視覺化的介面上同時視覺化使用者的檢索路徑以降低使用者的注意力消耗。本研究將對所設計的介面進行使用者評估，一方面藉以觀察使用者在資訊視覺化介面上的資訊檢索行為，一方面期能推廣本研究於相關性質網站上的應用。

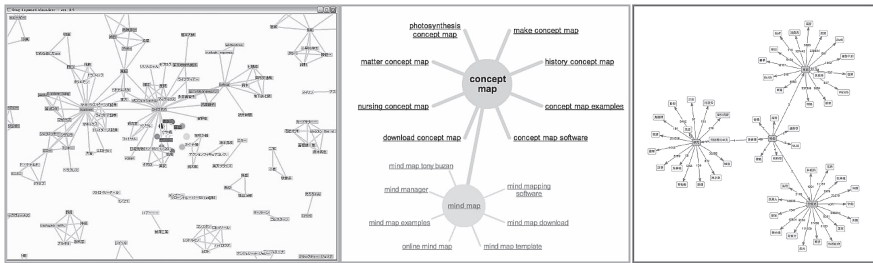
## 貳、文獻探討

### 一、新聞視覺化與視覺化技巧

過去已有不少研究與商用的搜尋引擎提供視覺化介面以輔助資訊或者新聞事件的檢索。如開發新聞輔助瀏覽介面（如 NewsMap 採用 Shneiderman 與 Wattenberg（1992）所設計的 TreeMap 概念，以顏色來區分新聞的類別、以面積來區分新聞的熱門程度），或是透過時間軸的形式展現議題發展的變化（如 ThemeRiver（Havre, Hertzler, Whitney, & Nowell, 2002）和 Google Trends 視覺化多種議題在時間軸上的變化）。除了此種視覺介面的呈現方式之外，還有利用網絡來呈現搜尋關鍵字間的關係之作法。關鍵字網絡，或曾元顯稱之為概念圖的方式（Tseng, Chang, Rundgren, & Rundgren, 2010），可以利用網絡來呈現關鍵字與關鍵字間的關係，藉此除了能夠觸類旁通找到新的關鍵字以修正檢索方向外，尚可以透過被展開的相關關鍵字來輔助理解新聞議題。如

Tseng 等人（2010）便曾以概念圖的形式，來視覺化同一句子或同一篇文章中的關鍵字與關鍵字的共現（co-occurring）關係（如圖 1 右圖）。此外，這種網絡視覺化的技術也曾應用於觀察社群網站的議題變化上，如 So-Net 過去曾開發 BKV 來即時監測日本部落格網站中部落客貼文中的關鍵字（如圖 1 左圖）。參考 BKV 的做法，Takama、Matsumura 與 Kajinami（2006）的研究也指出關鍵字網絡有助於讀者獲取部落格上的消費者意見，並藉其獲知文章間的關係。Google 亦曾在 2009 年發表了 Google Wonder Wheel 的關鍵字網絡視覺化功能，來輔助使用者探索新的關鍵字（如圖 1 中間），讓搜尋工具變成探索工具，不但有助於使用者修正檢索方向（Albakour et al., 2011），更可在實用層面上成為語言學習的工具（Warner & Jones, 2011）。

圖 1 關鍵字網絡視覺化範例（由左至右分別為 So-Net 所開發的 BKV、Google Wonder Wheel、與國內學者曾元顯教授所開發的聯合新聞網導覽介面）



註：BKV 介面截圖取自 <http://livedoor.blogimg.jp/yururidiet/imgs/0/7/076f9665.JPG>、Google Wonder Wheel 介面取自 <http://www.mindmaptutor.com/wp-content/uploads/2010/05/concept-map1.gif>，聯合新聞網導覽介面截自 <http://ir.itc.ntnu.edu.tw/udn/KMap.aspx>。

大體來說，這些新聞議題或檢索的視覺化輔助工具可分類為綜覽、導覽、監控與鷹架等數項功能。綜覽類服務的主要功能是以圖形與比例等提示使用者現下的熱門新聞（例如 NewsMap）。監控類的服務則著重於提供即時訊息的比較與量化，主要技術是統計當下或者某一時間區間內的關鍵字。如前述 So-Net 所開發的 BKV 即以關鍵字網絡的動態視覺化介面提供使用者從打開該服務開始，整個部落格平台所即時產製的關鍵字量、熱門程度以及關鍵字間的共現關係，讓使用者可隨時得知目前他人正在討論的議題。國內的 OpView 社群口碑分析平台也開發出類似的動態「風暴圖」，讓使用者得以分析某些議題及關鍵字間的競爭性

關係，其可探知企業與其他競爭對手間的正負面社群意見成長趨勢。鷹架類的服務則較不著重於直接讓新聞閱讀者讀取新聞標題或內容，亦不注重新聞議題的即時提供，反而較注重如何在視覺化介面上呈現被隱藏在大量新聞議題後的人際關係、關鍵字關係、事件始末等，因此通常會將多個資料維度如時間、資料量、發生頻率、關係脈絡等盡可能地綜合在一個介面上，以靜態或動態的方式來呈現，以提供對事件或議題的理解。例如 Chung、Liu、Li 與 Chen (2010) 所開發的動態社會網絡介面，即視覺化新聞中關鍵人物的互動關係。而導覽類的輔助工具如目前暫停服務的 Google Wonder Wheel，其主要概念為提供使用者從目前到下一個步驟的檢索路徑，以視覺化方法讓使用者知道他現在的查詢關鍵字與系統建議使用的相關關鍵字間的關係，而得以調整其檢索方向。前述曾元顯教授所開發的聯合新聞網視覺化輔助瀏覽介面即屬於這類型的服務。由於本研究和前人研究相較下重點在於提供對瀏覽、檢索歷程的紀錄與輔助，也屬於導覽類的服務。

## 二、搜尋與瀏覽策略與介面設計

使用者在尋找資訊時，並不是個永遠以單一方向進行的流程。為了找到滿意的目標資訊，使用者會不斷地反覆進行搜尋活動，進而提高其對搜尋目的的了解。Bates (1989) 的採野莓模式便是在描述如此一連串的活動。若搜尋活動的單位是從產生問題到利用資訊系統解決問題，使用者經常交錯進行檢索與瀏覽。易言之，在一連串的資訊搜尋行動中，除了特定資訊的搜索外，瀏覽行動亦佔據極大的比例。如 Delphi Group (2004) 的報告即指出，知識工作者每天花費超過 20% 的時間進行資訊的搜尋，而其中約 70% 的時間是花費在瀏覽資訊。此外，75% 的使用者在其 Yahoo 市場調查專案中也表達其偏好為瀏覽資訊，而非搜尋資訊。由此可見，搜尋與瀏覽行動經常是互補的，會交叉出現在使用者運用搜尋引擎來解決問題的歷程之中。相較於搜尋，瀏覽更注重於發掘資訊。因此，在搜尋目標不明確時，使用者會花費更多的時間來瀏覽資訊，以發現更切合目標的關鍵字 (Delphi Group, 2004:11)，而非直接利用關鍵字進行搜尋 (卜小蝶、陳思穎, 2007)。

然而，當搜尋與瀏覽在資訊搜尋行動中是互補之時，資訊科學領域通常卻較為著重使用者如何搜尋資訊的行為，而較少關注使用者如何進行資訊的瀏覽 (Bates, 1989)。所謂資訊的瀏覽，依據 Cove 與 Walsh (1988) 的研究，大致可分為三種瀏覽策略：(一) 搜尋瀏覽一封閉且具結構化的瀏覽活動，具有目的性的尋找資訊；(二) 一般瀏

覽一對資訊是有需求的，但很可能對大部份的資訊產生興趣，（三）偶然瀏覽—完全開放且無結構的瀏覽活動。無論是哪一種類型的檢索，瀏覽會與搜尋穿插出現在個體的資料搜尋歷程中。如當使用者帶著模糊的目標時，除了不甚明確的目標搜尋之外，經常也會出現較為發散、不具結構化的瀏覽行動。Carledge 與 Pitkow（1995）的研究便指出此種瀏覽與搜尋行動的交錯歷程。他們透過探究檢索過程中所拜訪的網頁數與拜訪次數的歷程紀錄，運用受試者的滑鼠與鍵盤使用行動發現，使用者在瀏覽資訊時，最常用滑鼠點擊頁面中的超連結（51.9%），其次是點擊「上一頁」以回溯搜尋的步驟（40.6%）。這兩個動作的使用頻率遠超過其他包含直接輸入並開啟超鏈結、點選熱門網頁列表、「下一頁」等功能（這些動作的使用頻率最多只有 2%）。藉由如此的使用動作與歷程分析，他們發現使用者最常使用的探索策略有二種：原路而回（*leaving as you've entered*）與回歸原位（*looping back*）。前者是指所有使用者都經常使用倒退鍵回溯（*backtracking*）操作歷程；而後者則是使用者在經過一段時間的探索後，利用歷程清單（例如 Chrome 瀏覽器的使用歷程）或首頁圖示回到起始點再重新出發。

當資訊搜索行動是這般搜尋與瀏覽交錯出現的複雜歷程，資料搜尋的結果也經常是種結構錯雜的資訊組成。要統整並理解這些錯雜、大量的搜尋結果，有效的分析工具就是視覺化設計。在視覺化的設計上，目前可分為「由外而內」或「由內而外」的兩種策略。如卜小蝶與陳思穎（2007）所探討的搜尋結果自動分群，即為一種協助使用者從外而內理解搜尋結果的方法。以 Vivisimo 為例指出（如下圖 2 左側），這類的服務可以幫助使用者建立搜尋結果的類別架構並建立認知框架，達到觸類旁通或產生新概念的效益（卜小蝶、陳思穎，2007）。相反的，Tseng 等人（2010）所開發的網絡視覺化方法則提供一種「由內而外」的觀察。他們的介面會自動化擷取文章或內容新聞中的關鍵字，並建立起關鍵字和關鍵字間的共現關係（兩個關鍵字同時出現在一則文章或者一個句子中時，可稱為其間有共現關係，如下圖 2 右側）。當使用者輸入關鍵字進行查詢時，便會即時擷取文章內容中的關鍵字，並描繪出關鍵字和關鍵字間的關係，甚至藉由圖形理論方法來建立關鍵字和關鍵字的群集。和前者自動化分群相較之下，兩者的相同點在於均能提供「建立類別架構」與「觸類旁通」的功能，而相異之處在於，自動化分群的方法是將搜尋結果的「文件」進行分群，而網絡方法的分群是在抽取出關鍵字後，繪製關鍵字間的共現關係；因而前者較著重自動化提供分類的功能，後者較著重觸類旁通的功能。



由此可知，在資訊搜尋行動中，除了理解資訊搜尋與瀏覽的歷程外，適當呈現搜尋後的資訊關聯或核心概念，可以幫助使用者理解搜尋結果（卜小蝶、陳思穎，2007）。如本研究將會實作的 Force-directed 演算法與 Chameleon 演算法即為能夠透過關鍵字間的共現強度而「自然」產生網絡群集的演算法（Fruchterman & Reingold, 1991）。讓使用者除了得知關鍵字和關鍵字間的共現關係外，尚能夠得知主要的觀念群集。

圖 2 視覺化輔助新聞檢索範例  
左側為 yippy.com 對搜尋結果自動化群集的結果；右側為曾元顯教授所開發之新聞查詢結果之關鍵字視覺化介面。



## 參、研究方法

### 一、資料範圍與處理流程

本研究新聞資料庫包含線上新聞發稿日期從 2011 年 1 月至 2011 年 12 月共 12 個月的文字新聞（不含影音新聞），共獲 793,522 則新聞。內容來自兩個線上新聞平台，分別為聯合新聞網（408,234 則）與中時電子報（385,288 則）。每則新聞資料欄位包含來源網站、新聞類別（該網站上的分類例如：財經、政治、生活天氣等等）、新聞發稿時間、標題、與內容等五個欄位。

本研究所設計的視覺化輔助介面核心在於呈現相關關鍵字的脈絡，讓使用者在未閱讀新聞前即可展開以探索相關概念；並結合時間軸、多元網絡繪圖演算法、以及記憶瀏覽歷程等互動功能，希望能夠提供使用者理解搜尋結果的概念範圍、框架、關聯性與時間脈絡等。因此，相較於提供有效的資訊檢索，本研究目的在於協助使用者瀏覽

搜尋結果。基於這樣的目的，當使用者鍵入檢索關鍵字後，本系統會即時對搜尋結果的前 100 篇新聞進行關鍵字擷取。關鍵字擷取方法採用楊立偉 (2003a, 2003b) 的自動擷取權威詞 (power term) 技術。該技術平均每一篇文章約取得 5 個權威詞，長文 (400 字元以上) 權威詞量最大值為 18 個，非長文 (少於 400 字元) 最大值為 11 個，最小值均為 1 個權威詞。考慮一般桌上型電腦的效能，每秒可處理 103 篇文章。擷取權威詞效能以天下雜誌網站開放全文進行評估，選擇該網站的原因在於其提供人工定義的關鍵詞 (例如 2013 年 11 月 27 日刊載的「火箭教育，打造奇蹟小學」一文 <http://www.cw.com.tw/article/article.action?id=5054132> 有 8 個人工關鍵字)。自 2010 年 1 月至 2014 年 2 月間共可取得 1,036 篇文章，含長文 313 篇、非長文 723 篇。根據所取得的權威詞，與人工定義的關鍵詞相同或互為子字串則列計為命中。命中關鍵詞個數除以天下雜誌所提供的人工關鍵詞數作為準確率 (預先篩除未在內文出現的人工定義之關鍵詞)，可得長文之平均準確率為 86.17%，非長文為 90.66%。

在擷取關鍵字後，統計每個關鍵字所出現的文章數，並計算每一篇文章內的兩兩關鍵字的共現次數，累積後作為關鍵字共現強度以建立關鍵字網絡。本研究由於實作了共現強度閾值的調整機制，故保留所有強度的共現關係。此外，本系統允許使用者能夠隨意的輸入關鍵字或是點擊相關關鍵字以延伸查詢其他相關的關鍵字，當事件觸發後，系統會將被點擊的相關關鍵字視作輸入關鍵字，重複執行上述相關文章檢索、關鍵字擷取、過濾、共現分析、資料格式化等步驟並重繪網絡視覺化介面。下表 1 為本平台後端關鍵字擷取的結果。

表 1

關鍵字擷取範例

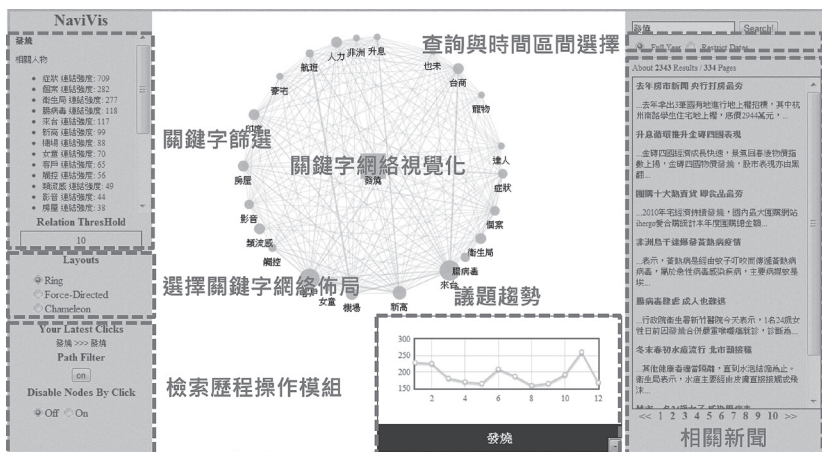
以下新聞為查詢「流感」所取出的一篇新聞文章，自動化擷取關鍵字的功能多擷取出流感疫苗與劑型兩個關鍵字。

新聞原文	相關關鍵字
<p>新北市衛生局指出，國際間流感疫情升溫，民眾出國前應先打流感疫苗，目前新北市剩餘成人劑型 (3 歲以上) 流感疫苗僅有 1 萬 5000 劑，3 歲以下幼兒劑型流感疫苗仍有 2 萬劑。為避免出國染病敗了遊興且有礙健康，民眾出國前最好能先接種流感疫苗，若到各區衛生所接種，不收費，若到合約醫療院所接種，會酌收掛號費及診察費。</p>	<p>流感：5                      流感疫苗：4                      劑型：2</p>

## 二、視覺化介面設計

本系統使用 Java 作為後端資料處理的伺服器語言，視覺化的原始資料儲存於 MySQL 資料庫，前端視覺化介面以 HTML5、CSS 與 Javascript 進行開發。為了便於提供互動功能，故關鍵字網絡視覺化則由研究者查詢相關視覺化演算法後以 Javascript 與 JQuery 自行繪圖（硬體設備為 Intel(R) Xeon(R) CPU E31220、記憶體 16GB）。透過對於視覺化介面與檢索歷程的文獻回顧，本研究歸納可輔助瀏覽過程的介面設計應支援使用者以下數項可能操作：（一）回溯檢索歷程，無論是倒退或者直接點選，以避免因注意力不足而迷失在搜尋過程中；（二）透過視覺化介面提供觸類旁通或者檢視概念類別的機會，尤其在缺乏知識背景的狀態下，受試者需要相關詞彙與類別架構來調整其檢索與瀏覽的方向。因而本研究所設計的介面<sup>2</sup>如下圖 3 可分成關鍵字網絡視覺化模組、新聞閱讀模組、檢索歷程操作模組、關鍵字篩選模組、資訊模組（含時間趨勢與相關新聞量統計）。

圖 3 系統介面綜覽。



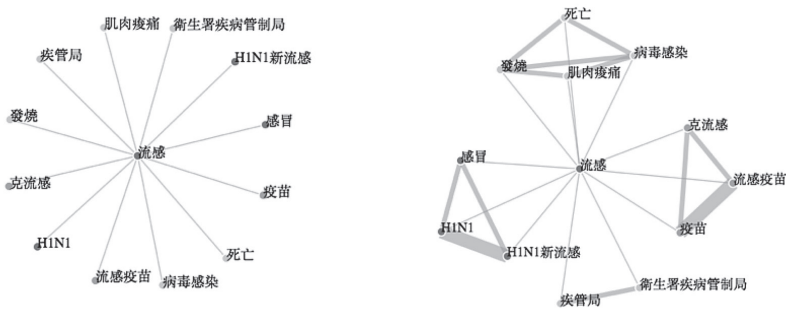
### （一）關鍵字網絡視覺化模組

過去以視覺化為輔的互動式搜尋系統，通常利用星狀圖來呈現主

2 <http://socialabservice.twgogo.org/NavigationVis/>。目前已修改原先查找中文新聞用途為查找紐約時報的相關新聞，企圖輔助非英語系的讀者瀏覽國際新聞。

要關鍵字與相關關鍵字的關係，另以關係線條的粗細來表示關係強度，例如 Tseng 等人（2010）所設計的視覺化導覽系統，或者已經停止提供服務的 Google Wonder Wheel。然而，相較於星狀圖，本研究主張進一步視覺化相關關鍵字間的關係可協助使用者判斷關鍵字類別。舉下圖 4 為例，左圖為檢索「流感」後所得到的相關關鍵字星狀網絡，讀者僅能得知其查詢的關鍵字「流感」和哪些關鍵字可能相關；但如果在繪圖的同時亦繪製了相關關鍵字的共現關係如右圖，則可得知流感相關的關鍵字間的相關性（例如「疫苗」與「流感疫苗」共現次數高）。使用者透過這樣的關鍵字網絡，獲得的不只是一個可以延伸查詢的方向，還有可能透過關鍵字網絡就預先獲得可能的資訊，而調整其檢索的大方向。比方說，從下例中，就可以得知流感相關新聞議題的討論可能可以分出主管機關、病毒與疫苗的討論、以及 H1N1 的描述等等面向。

圖 4 不同相關關鍵字視覺化結果之比較



而本介面可以透過點選相關關鍵字來延伸查詢，例如下圖 5 中的結果即為使用者查詢「流感」後點擊「疾管局」的結果，有更多的醫療院所、防疫政策、補助等專責機關的相關訊息隨後被展開。系統記錄且視覺化了使用者查詢「流感」後再點選「疾管局」的路徑。然而，這種視覺化方法和 Google Wonder Wheel 與 Tseng 等人（2010）的設計相較之下，由於為了要呈現更多的脈絡和延伸訊息，勢必產生更多的關鍵字，顯然可能會增加使用者的認知負荷（為使用者評估所欲解答的問題之一）。因而本研究設計三種網路視覺化的布局方法。圖 6 中間 Force-directed 演算法所產生的布局為預設演算法，該演算法相較於一般最佳化演算法，其模仿彈性拉力的概念以產生較為「自然」且「直觀」的網絡群集（Tikhonova & Ma, 2008; Force-directed Graph Drawing, 2015），其產生群集的因子為關鍵字和關鍵字間的共現強度，並傾向於產生考

慮全域 (global) 繪圖的最佳解而忽略區域 (local) 的群集結果 (Tseng, Chang, Rundgren, & Rundgren, 2010)。但由於本研究所設計的介面讓使用者可以調整門檻值來增減相關關鍵字量，當關鍵字量大的時候可能會有不易一一檢視關鍵字的問題，故設計了環狀布局 (圖 6 左側)，相關詞彙會等間隔地放置在環形的邊上，使用者可以逐一瀏覽所有的關鍵詞，並且迅速找到高度相關的關鍵詞。為了解決 Force-Directed 布局傾向於全域最佳解造成部分關鍵字群集被分散放置的問題，本研究尚實作了 Karypis、Han 與 Kumary (1999) 提出的 Chameleon 群集分析演算法 (如圖 6 右側)，Chameleon 演算法的主要目的在於強調關鍵字分群，因而弱化了節點與節點間距離的強度關係，以避免高度連結的網絡繪製後難以觀察群集。藉著布局的切換，本研究預期可協助使用者因應不同的需求而切換介面以觀察關鍵字，亦預期切換至適當的布局可有效降低大量關鍵字所造成的認知負荷。

圖 5 展開相關關鍵字的示意圖  
使用者在查詢「流感」後展開了「疾管局」關鍵字

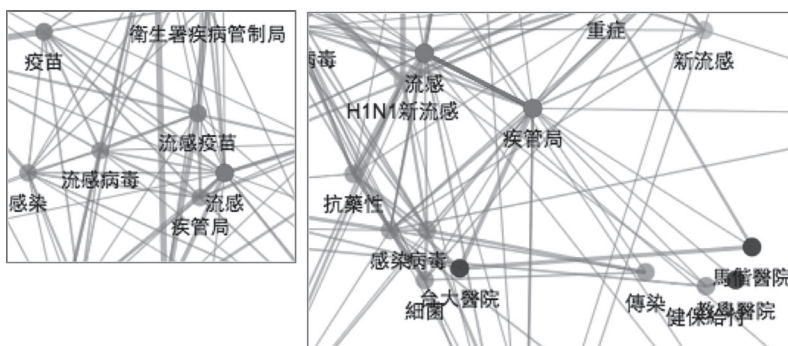
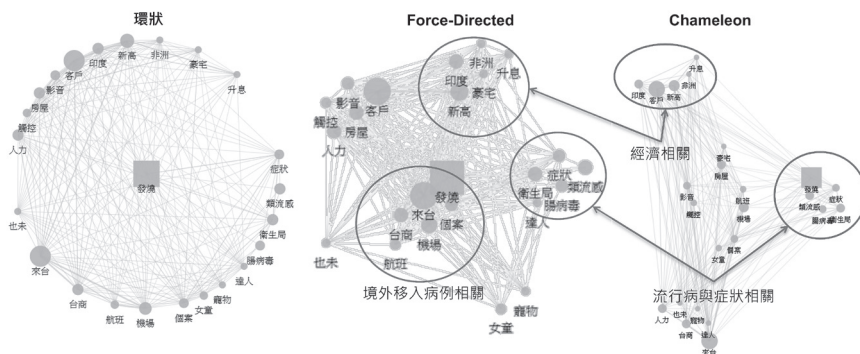


圖 6 三種不同的關鍵字網絡布局介面比較



## (二) 檢索歷程操作模組

為了記錄使用者的檢索歷程，系統會從一開始的檢索詞彙節點以不同顏色的線條或點來描繪，使用者能夠藉由節點的形狀或顏色來分辨哪些關鍵字是已點擊的狀態（如圖 5），並記錄點擊歷程，也就是檢索過程。在記錄點擊歷程的預設功能下，「歷程操作模組」允許使用者只留下最近兩次展開關鍵字的結果（開啟「Path Filter」選項），以協助使用者降低關鍵字量，或者捨棄過去的關鍵字。並允許使用者自行來關閉已點擊的節點（開啟「Disable Node By Click」選項）。但本研究預先開啟歷程的紀錄模組，因為使用者平日使用瀏覽器時，其隨時可以利用上一頁回到前頁，也是預先開啟，故採用預先開啟的做法。

## (三) 關鍵字篩選模組

由於檢索後的相關關鍵字過多可能會提高瀏覽的認知負荷，所以使用者可從介面上修改關鍵字間的共現關係強度閾值來納入或者濾去若相關的關鍵字。圖 6 所展示的是閾值為 10，也就是兩兩關鍵字曾經一起出現在 10 篇新聞以上的結果。當閾值調整越低，越容易出現較不相關的關鍵字。由於對搜尋結果進行斷字後即儲存所有的關鍵字對，不需重新抽取，故在本系統中利用閾值調整關鍵字量非常快速。

## (四) 新聞閱讀與資訊模組(含時間趨勢與相關新聞量統計)

為了降低受試者瀏覽新聞內容的負荷故檢索結果以新聞片段呈現於搜尋列下方，並以顏色凸顯搜尋的關鍵字（如圖 3 右側）。而為了確認參與者實際有閱讀新聞，也讓受試者專注於操作本研究所提供的多種介面設計，避免受試者因為切換新聞閱讀視窗而忽略了原本在進行中的視覺化互動，故讓閱讀者於彈出式視窗中觀看新聞本文。當使用者有點選某則新聞而在彈出式視窗中觀看新聞時，則可認定參與者嘗試在理解時事議題，也可說是認真地在參與本研究。雖然本系統主要以網絡圖呈現關鍵字，但亦同時在下方呈現關鍵字的趨勢圖，以橫軸表示時間（以月為單位）；縱軸為新聞量，藉此可觀察相關新聞議題的新聞量趨勢，再藉由系統介面修改查詢列底下的時間區間以便觀察。使用者能夠利用用於控制查詢結果時間區間的小日曆，選擇新聞資料起始與結束的日期。

### 三、使用者評估—訪談法

視覺化介面的使用者評估至近年仍被認為是困難的議題 (Lam, Bertini, Isenberg, Plaisant, & Carpendale, 2011; Plaisant, 2004, Goodall 2011)，主要原因是包含不易擬定目標、設定自然情境、或因為慣採開放性任務所造成的評估困難 (Santos, 2008)，因而多採用複合方法包含操作案例分析、訪談、觀察，或者對操作歷程進行分析 (Lam et al., 2011)。Plaisant (2004) 摘要視覺化評估的方法包含以下三類：控制實驗 (controlled experiments) 以比較不同介面、設計元素、或服務的效率；使用性評估 (usability evaluation) 主要被用以快速獲取使用過程可能遭遇的問題並提供改善與優化介面時參考 (Goodall, 2011)；案例分析 (case studies) 則用以在真實使用情形下獲知使用者的應用脈絡等，以發現設計目的和使用者的實際需求的不一致 (Goodall, 2011)。由於本介面嘗試整合瀏覽歷程記憶與關鍵字網絡視覺化作為新聞檢索與瀏覽的輔助介面，並非屬於視覺化機制或檢索成效的比較，若與一般搜尋引擎檢索介面相較下，可能過度突出本介面的特色；若與其他類似目的的檢索介面相較下，由於資料集不同與其他設計不同，研究者認為使用者回饋容易過度著重在比較功能的有無，故採用使用者評估的方法，以快速獲取使用者所提出的問題、回饋與建議。參考卜小蝶與陳思穎 (2007) 的網路分群搜尋引擎評估與黃婷穎 (2011) 的網路評價視覺化系統，本研究採用「主題任務觀察法」佐以螢幕側錄程式來記錄使用者的操作歷程，並透過訪談來了解獲取使用者回饋，最後再利用循序樣式探勘方法來獲取參與者使用本介面時的操作歷程特徵，藉以檢視可能的介面設計問題。本研究共招募參與者共十位 (編號為 F01-F10)，其中包含北部公立大學八位圖書資訊學研究所學生與兩位圖文傳播研究所學生，每位學生過去均無使用過任何關鍵字網絡導覽介面的經驗，並且都有每日閱讀線上新聞的習慣。抽取圖資所學生的原因在於其曾學習過資訊檢索與使用者行為，故認為對於瀏覽過程或者使用者互動能提供比較多的回饋；額外抽取兩位圖文傳播所學生的原因是，本研究以新聞瀏覽作為議題，故找有受過大眾傳播訓練者，期望其對於新聞瀏覽與理解上能夠回饋較不同的意見。停止找更多參與者的原因為，目前對於本研究所設計的視覺化介面在關鍵字過多造成認知負荷、以及瀏覽歷程的應用與視覺化介面的優點，受試者有一致的看法。

主題任務評估方法為過去視覺化介面 (含新聞檢索視覺化介面)

常用的評估方法，多採用指定任務，但通常為了觀察自然情境下的使用情形而額外允許使用者自定任務（卜小蝶、陳思穎，2007）。為了提高參與者探索介面的可能性，任務的類別多採用開放型的任務而不給予明確的搜尋目的，這也是過去視覺化介面評估慣用的任務形式。例如重視時間趨勢的 ThemeRiver 新聞視覺化系統的作者引導二位參與者自介面中找出指定時間內的新聞主題，再請參與者指出新的新聞議題，並請參與者對視覺化的特色進行評論（Havre et al., 2002）。而本研究設計的介面嘗試綜合了主題的延伸瀏覽、檢索路徑的記錄、以及新聞篩選、時間趨勢等多種視覺化特徵，更適合採用開放性問題，請參與者透過操作介面來了解本研究提供的主题。主题則選擇臺灣社會當時較熟悉的國內主题「塑化劑」，以及臺灣社會較易生疏的國際新聞「埃及政變」與「佔領華爾街」作為主题任務。三個主题皆屬於長期發展且牽涉複雜人事物的新聞議題，為了引導參與者盡量使用各種介面的功能，也為了避免參與者將任務理解為一個「搜尋任務」，故要求參與者盡量選擇一個最生疏的主题進行瀏覽。除此之外，研究者於實驗前亦向參與者說明，在操作過程中仍可放棄此主题，而改用參與者感興趣的主题。

實驗操作前由研究者給予新聞主题任務說明單，並對參與者展示系統功能與說明任務。當參與者開始使用系統瀏覽新聞主题時，為了避免影響後續對參與者操作進行循序探勘的結果，於研究開始前告知參與者，除非參與者對介面操作提出問題，否則研究者將採不介入方式，從旁觀察並記錄參與者在操作過程中臉部表情與特殊使用行為（如使用過程的停頓、或少用、重複使用、反覆切換部分介面選項）以便進行後續訪談。為了避免漏記操作過程與後續進行操作歷程循序樣式探勘，同時利用螢幕錄影軟體 Free Screen Recorder 紀錄每位參與者的使用過程。本研究不限制參與者的使用時間，在參與者自認為其得到的資訊足夠而停止使用本系統後（10 位使用者均使用 15 分鐘以上），立即對參與者進行訪談，操作含訪談共 40 分鐘以上。訪談問題為介面中四個模組相關操作的問題，例如「於彈出式視窗閱讀新聞原文」或「切換至環形布局」。由於本研究採用使用性評估方法，故重點在於探索使用者操作上可能會遇到的問題（Plaisant, 2004）。所以訪談大綱主要針對「實驗者不瞭解受試者某些特殊操作的意義」、「受試者不斷重複使用某些功能」、及「受試者僅使用少數功能」的情形進行發問。訪談大綱與問題如下表 2 所示。由於訪談的內容為特殊的操作，故需要在受試者一完成檢索任務後立刻詢問。



表 2  
主題任務評估的訪談大綱

觀察項目	訪談問題	使用行為範例
當實驗者不理解受試者某些特殊使用情	「我剛剛看見你…為什麼？」	拖曳節點，接著切換數次網絡圖布局
受試者重複操作某項功能	「你用了…功能好幾次，為什麼？」	重複切換網絡圖的布局
受試者僅使用某項功能	「你只用了…功能…為什麼？」	大部份的操作過程均在利用新聞列表閱讀新聞文章

#### 四、使用者評估—循序樣式探勘

為了對參與者的操作過程提供量化的證據，本研究利用循序樣式探勘的分析方法，求取新聞主題任務中參與者顯著的操作歷程，包含關鍵字網絡視覺化、歷程記錄與控制、網絡視覺化布局規則切換、新聞閱讀器與時間區間控制器等諸多視覺化互動功能。本研究希望透過循序樣式探勘來得知大部分參與者究竟是如何使用本系統所設計的功能？功能與介面的切換間是否存在順序關係？而參與者間是否存在何種共同樣式的表現？

因為受試者的操作過程多在 15-20 分鐘內，且本研究刻意將閱讀新聞本文設計為原視窗內的浮動視窗，強迫受試者必須關閉新聞閱讀視窗才可以回到視覺化瀏覽介面，藉此可降低受試者因為閱讀新聞而不易回到其瀏覽過程的情形，因此本研究認為可將每一位受試者的操作過程視為連續過程。經過這樣的觀察，本研究採用 Rakesh 與 Ramakrishnan (1995) 所提出的循序樣式探勘方法分析受試者的操作序列。該演算法原始目的是用於挖掘顧客交易資料庫中的循序頻繁的交易序列（例如常見的購買樣式組合）。透過將資料庫視為一群交易集合，把每一筆交易以交易時間排序，以 AprioriAll 演算法探勘客戶最頻繁的交易序列。例如典型的尿布與啤酒的購物籃分析案例，於某大賣場中進行分析的結果發現，60% 的顧客會在購買尿布後購買啤酒，這樣的數據能夠提供決策者進行購物動線的調整（若尿布與啤酒若設置在鄰近的位置，能夠增加銷售量）。AprioriAll 演算法的主要設計是透過支持度的設定，找出所有顧客交易之大型序列。當最小支持度為 0.6 時，代表在所有使用者中必須有 60% 的受試者都有該操作序列。當每個使用者的行為序列越短且越多樣時，那麼就必須要降低最小支持度

來找尋數量上突出的序列。

下表 3 為本研究使用 AprioriAll 演算法幫助分析使用序列之實際探勘過程圖，假設資料庫中包含五位參與者的使用序列  $S_n$ ，且設定最小支持度為 0.6 ( $Sp=0.6$ )。首先將所有子序列 SubSq 於  $S_n$  中取出，並將相同之 SubSq 刪除（順序與序列包含之項目完全相同即刪除，項目相同但順序不同則視為兩獨立之 SubSq，例如 {A, B} 與 {B, A} 為兩獨立之 SubSq）。接著依照規則找出所有最大序列 LSq，若該 SubSq 之支持度小於最小支持度（0.6）則刪除；所有大於最小支持度之 LSq 不得被其他 LSq 所包含（例如，{B, D} 包含 {B}，則刪除 {B}；{A, B, D} 包含 {A, B}，則刪除 {A, B}；但是 {A, C, D} 並不包含 {D, A}），最後得以產生所有大型序列 LSq。

表 3  
AprioriAll 演算法範例（支持度  $Sp>0.6$ ）

$S_n$	Sub-Sequence of $S_n$ (SubSq)	Large k-sequence (LSq)
S1: {a, b, c, d}	{a}, {c}, {b}, {d},	{a}, $sp=0.6$
S2: {c, b, d}	{a, c}, {a, b}, {a, d}, {c, b}, {c, d}, {b},	{c, d}, $sp=0.8$
S3: {a, c, d, d, b}	d}, {b, c}, {b, a},	{b, d}, $sp=0.6$
S4: {b, c, d}	{a, c, b}, {a, b, d}, {c, b, d}, {a, c, b},	
S5: {b, a}	{a, d, d}, {a, d, b}, {b, c, d}	

本研究於新聞主題任務中，利用螢幕擷取軟體記錄所有參與者之完整使用過程，為了能夠以 AprioriAll 演算法探勘大型序列樣式，研究者以人工的方式查閱所有使用過程之影片，並將影片中參與者所有使用情況進行編碼，轉換為便於進行樣式探勘的資料格式，而編碼的方式主要依據系統所設計之功能如下表 4。並請第二人進行編碼，由於編碼紀錄的過程多為明確的介面操作，所以兩者編碼有差異處多為動作的漏編，交叉檢驗後進行後續的循序分析。

表 4  
被編碼以觀察的使用者操作介面動作

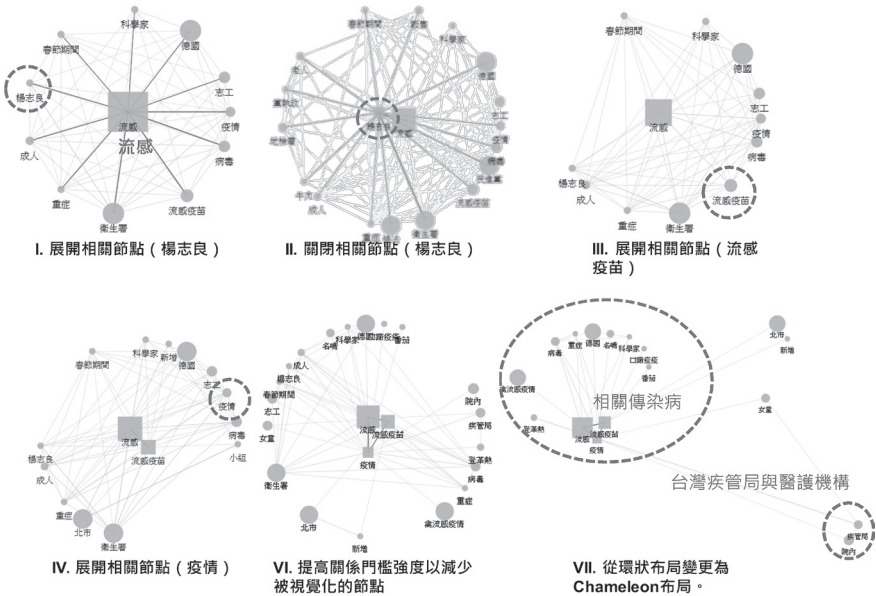
動作分類	動作與編碼
開啟功能（檢索歷程、刪除節點等）	開啟檢索歷程紀錄、關閉檢索歷程紀錄 允許使用者關閉部分節點、禁止使用者關閉部分節點、獲取當次查詢的詮釋資訊
新聞查詢與閱讀	輸入查詢關鍵字、於彈出式視窗閱讀新聞原文、點擊相關概念節點並展開、回溯曾點擊的節點
視覺化布局切換與互動	切換至環形布局、切換至 Force-Directed 布局、切換至 Chameleon 布局、放大布局視野、縮寫布局視野、拖曳節點、移動布局
過濾瀏覽範圍（包含調整共現強度與時間區間）	提高布局的關係強度門檻值、降低布局的關係強度門檻值、設定查詢日期區間、開啟關鍵字趨勢圖、縮小趨勢圖以觀看長期發展

## 肆、研究結果

### 一、介面操作範例

下圖 7 展示了某位使用者在實驗開始受研究者引導下的介面操作流程。一開始，使用者在「環狀布局」的視覺化選項下以「流感」作為關鍵字查詢。旋即點選了一個相關的詞彙「楊志良（為 2011 年的衛生署署長）」。然而，他對這個詞彙不感興趣，因此使用「允許使用者關閉部份節點的功能（Disable Node by Click）」關閉了他。之後他連著點選展開了兩個詞彙「流感疫苗」、「疫情」，再以「提高布局的關係強度門檻值」的功能刪減掉一些較不相關的字。最後利用 Chameleon 布局找出關鍵字群集。在這些操作後，使用者自行開始閱讀相關新聞。

圖 7 視覺化介面操作展示過程  
以「流感」為關鍵字進行查詢為例。



## 二、主題任務觀察與訪談

在參與者表示操作完介面後，即以訪談進行使用性評估。因為重點不在於理解新聞議題，故針對介面操作或者視覺化布局理解的內容進行詢問，訪談結果如下。

使用者進行查詢後，本介面會將被提取的新聞內容進行斷詞，抽取出重要詞彙後，依照詞彙的共現關係繪製為概念網絡。大部份的使用者在查詢後認為其能夠理解並運用關鍵字被視覺化為概念圖（或網絡）的意義（亦即了解關鍵字共現的意義），並接續點選節點以展開更多的關鍵字。在本研究中，受試者平均有五分之一的時間在展開關鍵字（20.0%），其中 F06 與 F10 的使用次數超過平均一個標準差（6.7%）以上。訪談中的受試者回饋表示，展開關鍵字有兩種時機，其一是當使用者不理解核心主題時，點選並展開更多相關關鍵字，可幫助其理解該主題與相關新聞。例如 F05A 在抱怨關鍵字網絡介面回饋太多相關關鍵字難以消化時，隨後提及「……至少讓我知道有可以有相關的東西啦……（F05A: 084）」，表示展開相關關鍵字的功能即使可能產生過多無法消化的關鍵字，仍可讓受試

者透過觀察關鍵字理解相關議題。其二是參與者也被發現會直接展開其不懂的關鍵字來理解內容，除了下述 F04 會不斷點擊看不懂的關鍵字，F02 在查詢時看到不理解的關鍵字「彭博」，其透過觸發相關關鍵字，在未經瀏覽新聞前大致已可從相關關鍵字推知「彭博」二字為（前任）紐約市長。

F04Q：我剛剛看你有一段時間不斷的點開關鍵字，那時候你想做什麼？

F04A：喔，因為我看到幾個關鍵字不知道在說什麼，我就是想點他（F04: 063）。

在展開感興趣的關鍵字以進一步查詢後，有四位參與者（F03、F04、F05、F06）均提及本系統產生了過多的關鍵字而不易消化（這也是目前最主要的負面回饋）。例如 F06 即表達「像你這些關鍵字就太多啦，讓我不知道自己到底要看什麼……。」（F06: 113）但本研究發現參與者利用本系統的其他功能，採用三種策略解決這個問題。其一是關閉一些展開過的節點以降低關鍵字量。關閉展開過的節點為本研究中最多參與者回饋的功能，共有八位參與者對該功能有正、負面意見。關閉展開過的節點的原因可能是因為參與者點擊了其誤以為感興趣的關鍵字，或者是參與者在查詢過程中對原先的關鍵字失去興趣。並有三位參與者一方面肯定本研究所設計關閉相關關鍵字功能，一方面認為應該允許被關閉的是未曾被展開過的關鍵字，而不是被展開過後的關鍵字（F04、F05、F10）。這是因為本研究設計關閉節點功能的用意是讓使用者能夠隨時改變其查詢歷程關鍵字的組合，但顯然參與者除了肯定這樣的功能外，也希望能夠關閉那些被大量展開的節點。從這樣的使用者操作中同時可歸納出，透過「歷程路徑紀錄功能」搭配「關閉節點」的功能，一方面紀錄使用者的檢索歷程，一方面又能夠允許使用者自由修剪其檢索歷程，在實用上，應可降低使用者再檢索、重新定義關鍵字的次數。但在刪除低度相關節點的設計以及預設所呈現的關鍵字量上可以再對介面進行修改。

F03A：去掉關鍵字是因為我覺得有些相關關鍵字好像不是我要的（F03: 43）。

F07A：其實是我點錯了，我想要把他擦掉，還有就是在我這條路，覺得不對了，因為你這像是一個雲嘛，是不同的關鍵字組合嘛，

那我想要重新換關鍵字，重新組合關鍵字，所以去掉我不想要的 (F07: 127)。

從以下的參與者回饋可發現三種布局的設計能夠有效提供不同目的的觀察，例如受試者 F01 即。為了降低過多關鍵字所產生的認知負荷，研究者發現參與者會切換關鍵字網絡布局，選用 Chameleon 布局的分群效果或較規則的環狀布局來進行瀏覽，這種情形下受試者所關切的並非減少關鍵字量，而是透過切換布局妥善安排這些關鍵字。參與者發現環形布局可將關鍵字排列整齊以逐一觀察；或有參與者（如 F07）發現 Chameleon 布局清楚地群集了相關關鍵字，可觀察搜尋結果的類別架構。在本研究中 F01 與 F07 為最常切換網絡視覺化布局的受試者，切換佈局的動作分別佔其整體操作序列長度的 32% 與 25%。

F05A：Ring（環狀）最好用了，因為其他的很複雜，我這麼聰明不需要他幫我分類 (F05: 70)。

F08A：我覺得，（不同的布局）都有意義，像那個（Chameleon 布局）分的比較開阿！這個（環狀）就有點像是在列出來，比較清楚 (F08: 132)。

F07A：喔，我剛剛是在看這個分群（Chameleon 布局）好像挺有道理的，譬如某一些關鍵字就是在講財經這一塊的，他會在一起；或是有一些是關於國際關係的會聚在一起 (F07: 119)。

F09A：他（Chameleon 布局）看起來有群集的概念在裡面啊，就可以一目了然 (F09: 145)。

第三種減少相關關鍵字的策略則是使用「提高布局的關係強度門檻值」的功能，當門檻值提高，與被展開的關鍵字相關的關鍵字就減少。但這是僅有三位參與者提到的功能，原因可能如兩位參與者對該功能的抱怨，本研究採用滑鼠滾輪改變門檻值的速度過慢，無法快速地減少相關的關鍵字量 (F05、F09)。

亦有部分參與者提出其對整套介面的綜合性意見。最主要的意見是認為介面過度複雜 (F04)，而 F06 認為可以藉由隱藏部分功能至滑鼠右鍵選單，得以降低整體介面在視覺上的複雜度。而 F08 則注意到系統介面設計未採用適當的隱喻，其提及「這個（開啓檢索歷程紀錄）到底是關掉還是打開？你應該做成像開關一樣 (F08: 137)」。除此之外，研究者觀察到 F05 大部份時間都在瀏覽新聞列表並展開相關新聞

出來閱讀，經過詢問後，F05 認為要理解新聞議題最主要還是靠使用者逐一閱讀感興趣的新聞，其表達不願意被視覺化的設計所引導「……我就是不喜歡這種好像我一定要照著走一樣，直接看文章比較實在（F05: 75）！」F05 的意見可理解為，即使有新式視覺化介面的設計，但使用者為了達成其檢索目的或者依照其閱讀習慣，可能有其偏好或慣用的資訊瀏覽方法。而 F05 的操作序列亦顯示出，其一開始會嘗試使用各種不同的功能，但其在使用過程的後半大量地閱讀新聞（其操作序列長度為 53，第一次閱讀新聞為 23 的位置，此後共展開了 8 次新聞閱讀視窗）。

### 三、操作歷程循序探勘分析

利用 AprioriAll 演算法（Rakesh & Ramakrishnan, 1995）對前表 4 所列的 21 項操作進行分析，使用者的操作序列最長為 62、最短為 32，平均長度為 41.5。在最小支持度為 0.6 的情形下，共獲得 45 個支持度大於等於 0.6 的序列（序列長度為 2~4）。下表 5 就本研究所設計的主要功能包含切換視覺化布局與調整視野、控制歷程、篩選檢索結果與閱讀新聞等之相關序列中，挑選支持度較大者進行說明及論述。圖 8 以狀態轉變圖列出所有支持度在 0.7 以上的循序序列狀態轉變，為了便於觀察，故拆解了長度為 3 以上的序列成為長度為 2 的成對序列。但仍有兩個動作「切換至環狀布局」與「開啟瀏覽器歷程紀錄」沒有支持度在 0.7 以上的序列，故特別針對這兩個動作，列出支持度 0.6 的操作序列，以虛線做代表。其中較特殊的是，當「允許使用者關閉部分節點」功能開啟後，使用者再進行「回溯曾點擊的節點」，會關閉其所點擊到的節點，故其實執行的是刪除部分節點的功能，以虛線做為代表。

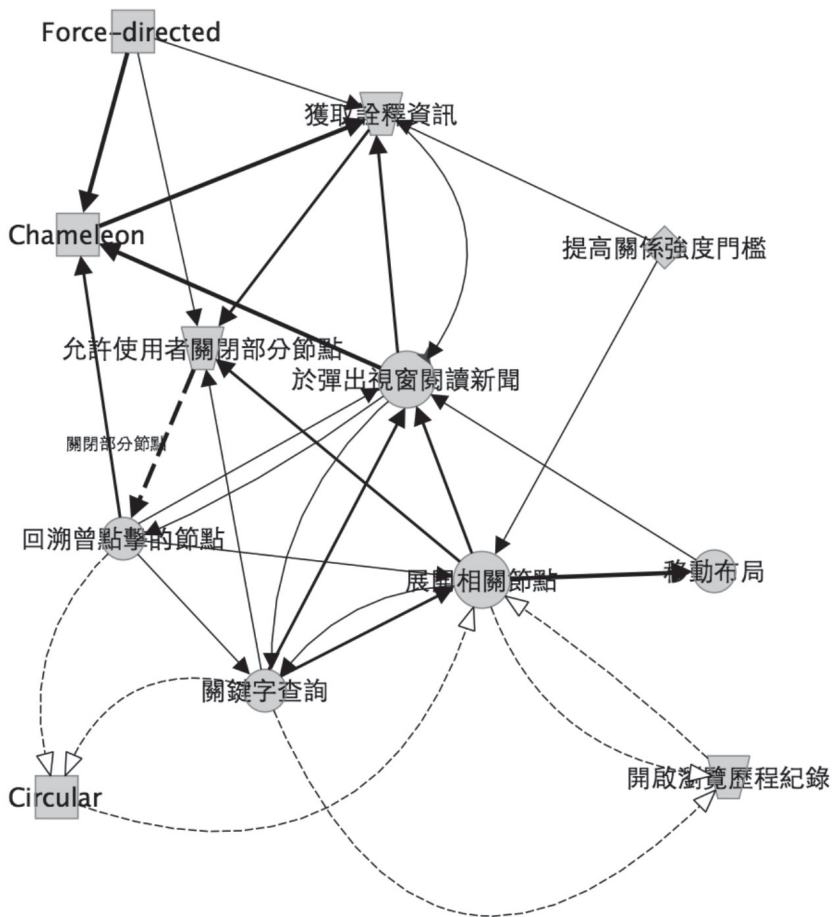
表 5  
循序探勘結果（為方便說明取功能操作上較不同的 13 個序列供參考）

相關功能	編號	歷程記錄與控制	支持度
切換視覺化布局 與調整視野	01	輸入查詢關鍵字→切換至環形布局 →點擊相關概念節點並展開	0.6
	20	切換至 Force-Directed 布局→獲取當 次查詢的詮釋資訊	0.7
	43	切換至 Force-Directed 布局→切換至 Chameleon 布局	0.9
	44	於彈出式視窗閱讀新聞原文→切換 至 Chameleon 布局→獲取當次查詢的 詮釋資訊	0.9
	02	切換至 Chameleon 布局→移動布局	0.6
	45	點擊相關概念節點並展開→移動布 局	0.9
控制歷程	08	開啟檢索歷程紀錄→點擊相關概念 節點並展開	0.6
	34	輸入查詢關鍵字→允許使用者關閉 部分節點→回溯曾點擊的節點→點 擊相關概念節點並展開	0.7
篩選檢索強度	32	提高布局的關係強度門檻值→點擊 相關概念節點並展開	0.7
	33	提高布局的關係強度門檻值→獲取 當次查詢的詮釋資訊	0.7
閱讀新聞	13	於彈出式視窗閱讀新聞原文→於彈 出式視窗閱讀新聞原文→於彈出式 視窗閱讀新聞原文	0.6
	14	於彈出式視窗閱讀新聞原文→點擊 相關概念節點並展開→於彈出式視 窗閱讀新聞原文	0.6
	42	輸入查詢關鍵字→於彈出式視窗閱 讀新聞原文→於彈出式視窗閱讀新 聞原文→獲取當次查詢的詮釋資訊	0.8



圖 8 支持度大於 0.6 的操作歷程狀態轉換圖

(實線代表支持度為 0.7 以上，從細到粗分別為支持度 0.7、0.8、0.9。虛線為支持度 0.6 的動作序列)



從表 5 與圖 8 中可觀察到以下包含開始操作、再搜尋、視覺化布局的操作、操作歷程紀錄與視覺化、閱讀新聞內容、篩選相關新聞等數個操作流程的事實：

- (一) 整個介面操作始於「關鍵字查詢」，參與者下一個動作多為「展開相關節點 (0.8)」或者「於彈出視窗閱讀新聞 (0.8)」
- (二) 會產生再搜尋的操作包含在閱讀新聞之後做再搜尋 (於彈出視窗閱讀新聞→關鍵字查詢：0.7)、或者在進一步展開節點或回溯節

點後感到不滿意而調整搜尋方向（展開相關節點→關鍵字查詢：0.7、回溯曾經點擊的節點→關鍵字查詢：0.7）。

- (三) 視覺化布局的切換為本研究的核心設計。開始進行視覺化布局的切換以進行觀察分成兩種情形，第一種情形是在查詢關鍵字後切換至環狀布局（0.6），第二種情形是在閱讀新聞後切換至 Chameleon 布局（0.9）。在「回溯曾點擊的節點」的動作後，參與者可能會繼續展開節點或閱讀新聞（均為 0.7），但卻有較高的機率去切換視覺化布局以「重新整理」視覺化的節點（切換至 Chameleon 與環狀布局分別為 0.8 與 0.6）。這是因為「回溯曾經點擊的節點」之前若為「允許使用者關閉部分節點」可能會因此刪去數個節點而導致參與者想利用切換視覺化布局的方式來進行重新整理以便觀察。切換至不同佈局的操作佔每個受試者的操作序列的平均比例為 12.8%，甚至有兩位受試者切換佈局的序列長度超過原序列的 20%（F01：32%、F07：25%）。切換布局的動作代表參與者可能覺得現有布局不利於其觀察。此外，研究者注意到切換至環狀布局的時機多為參與者輸入關鍵字或者刪除部分關鍵字後調整視覺化布局的動作，原因如訪談中所述，參與者表達切換至環狀布局具有重新整理視覺化布局的效果。
- (四) 在視覺化介面的「轉換」上，可發現在使用者切換至 Force-Directed 布局後，多會順道觀看 Chameleon 布局（切換至 Force-Directed 布局→切換至 Chameleon 布局：0.9），但可能的原因是並非「順道觀看」，而是自前述訪談結果，有參與者反應難以理解 Force-Directed 的視覺化結果。也有部分參與者會關閉部分節點以降低如訪談中所述的認知負荷（切換至 Force-Directed 布局→允許使用者關閉部分節點：0.7）。
- (五) 本研究的設計重點之一為歷程紀錄，故預先開啟歷程紀錄的選項。研究者發現，即使有告知使用者可關閉歷程紀錄，使用者鮮少關閉歷程紀錄的功能。除非受試者有點選關閉歷程紀錄的功能，才會再度開啟歷程紀錄，因此，開啟歷程紀錄如圖所示，支持度不是最高的，但若出現的話，必定是出現在再搜尋之後，重新開啟歷程紀錄。
- (六) 在新聞內容閱讀方面，由圖中可發現參與者經常由其他動作進入點選彈出視窗以閱讀新聞，代表其為了探索指定的新聞議題而有確實觀看新聞。除此之外，在觀看新聞後，使用者可能會選擇調整視覺化布局（於彈出視窗閱讀新聞→切換至 Chameleon 布局：

0.8)、或者回溯其展開節點的歷程(0.8)、或者擬定關鍵字再查詢(0.7)、或者觀看當次查詢的詮釋資訊(0.8)。

- (七)關於視覺化結果的篩選功能，在本研究中發現參與者鮮少開啟時間趨勢(時間相關的序列均未達最小支持度0.6)，可能是受到主題任務為開放型任務，而對觀察議題趨勢沒有額外的操作指示。但參與者會嘗試提高共現關係強度的門檻以減少延伸詞彙的數量，並接續展開新的節點(0.7)或者閱讀因為提高門檻而被篩選出來的高度相關新聞內容(0.7)。
- (八)參與者會反覆操作的動作主要為閱讀和視覺化布局的切換。例如就表5的序列13、14顯示6成的參與者會不斷開啟新聞內文或者透過點選相關節點來載入更多感興趣的新聞內容後再繼續閱讀。而沒有列在表列中的序列40則是有8成以上的參與者會連續展開節點三次以上。視覺化布局的操作則集中在從某個布局切換至 Force-Directed 布局與切換至 Chameleon 布局上(0.8)。

## 伍、總結

視覺化的設計主要是希望透過圖形與介面整合多維度的資訊，以新聞而言包含時間、主題、出現頻率、共現次數等，希望能夠有效協助使用者檢索(Card, MacKinlay, & Shneiderman, 1999)。然而，資訊的維度越高，採納的視覺化元素越多，亦會增加使用者的認知複雜度(Brath, 1997)，所以視覺化設計元素的增減應需考量到認知負荷的平衡。從使用者回饋中可得知本系統確實能夠協助部分參與者其發掘資訊，然而，當所設計的元素過多，或者背離參與者所宣稱地「只想好好閱讀新聞」的檢索或瀏覽目的時，反而會讓使用者產生額外的認知負荷而感到不耐煩。然而，就使用者回饋與分析也可看到，本研究嘗試直接將檢索歷程的操作整合在視覺化介面中，使用者回饋顯示有助於其探索新的概念，或者調整檢索的方向。而在相關關鍵字過多時，使用者會透過改變繪圖演算法或刪減檢索過的關鍵字的方式來降低認知負荷。

相較於過去的新聞或搜尋視覺化介面設計，本研究的主要貢獻在於嘗試將檢索歷程的外化與控制結合在視覺化介面上，一來可以直接在視覺化介面上呈現檢索歷程，二來也讓使用者得以修剪不必要的檢索歷程，以輔助使用者隨時調整、修改檢索的方向。除此之外，本研究提供了一個質、量化使用者評估範例，自訪談與循序探勘方法均顯

示出，本系統的檢索歷程設計能夠幫助使用者調整檢索方向，或者去除不必要的資訊。而透過使用者的回饋，亦得知本系統可加強在網絡視覺化的設計上，包含調整視覺化的網絡布局，或者調整關鍵字量以避免使用者產生額外的認知負荷。本系統除了本研究所展示的中文新聞檢索用途外，目前系統本身已應用至輔助使用者閱讀紐約時報的新聞上，企圖從輔助讀者檢索、瀏覽的方式，嘗試提供國人瀏覽並探索感興趣的新聞，或讓新聞傳媒得以擷取相關的新聞資訊內容。

本研究設計了一套視覺化瀏覽輔助介面，並採用使用性評估的方法與序列分析方法進行介面評估。採用使用性評估方法雖有助於快速發現介面操作上的問題 (Plaisant, 2004)，但對於要比較本介面與其他介面的在檢索上的效率與所獲得的資訊仍需使用控制實驗來進行。除此之外，目前亦無法解釋或推論使用者長期利用下的可能行為。此為本研究的研究限制一。再者，由於本研究著重在視覺化處理並非資料處理，但視覺化介面所呈現的關鍵字偶而會出現無法預期其關聯性的關鍵字 (例如圖 6 中「升息」與「發燒」確實為經常共同出現的金融類關鍵字)，其原因乃是關鍵字檢索的結果擷取到不同類別的新聞，需要進一步改善網絡群集演算法或利用新聞類別進一步群集該類的相關關鍵字，此為研究限制二。在類似新聞視覺化檢索的系統開發上，建議未來可提供服務給特殊主題例如健康資訊等，透過限制主題類別，可提供主題相關的領域知識以強化關鍵字擷取或者關係脈絡的表示法。例如，本研究亦嘗試針對健康與醫學相關領域的關鍵字進行擷取結果如 [http://web.ntnu.edu.tw/~jirlong/demo/news\\_vis/v6.html](http://web.ntnu.edu.tw/~jirlong/demo/news_vis/v6.html) 所示，透過將相關關鍵字區分為機構、病徵 (如發燒、流鼻水)、器官、病名等，可有助於使用者瀏覽相關新聞。

## 誌謝

感謝龍捲風科技協助新聞搜尋與關鍵字抽取，該公司以搜尋引擎為基礎，提供企業知識管理服務、競爭性情資、網路口碑、情意運算等服務。並感謝兩位匿名審查委員提供在內文修改上的建議。

## 參考文獻

- 卜小蝶、陳思穎 (2007)。網路自動分群搜尋引擎之使用者評估研究。  
圖書資訊學研究, 2(1), 55-80。
- 江信昱、卜小蝶 (2010)。網路資訊再尋獲之檢索行為初探。圖書資

- 訊學刊，8(1)，22-57。
- 黃婷穎（2011）。網路評價視覺化系統。未出版碩士論文，國立臺灣大學資訊管理學研究所，台北市。
- 楊立偉（2003a）。中文斷詞法。中華民國發明專利，527546，2003/04/11。
- 楊立偉（2003b）。電腦新詞學習方法與系統。中華民國發明專利，517195，2003/01/11。
- 鄭卜壬（2009）。中英文媒體疫情自動分類及預警研究。（疾病管制署專題研究計畫報告 DOH98-DC-1004）。台北：衛生福利部疾病管制署。
- Albakour, M. D., Kruschwitz, U., Nanas, N., Song, D., Fasli, M., & Roeck, A. D. (2011). Exploring Ant Colony Optimisation for Adaptive Interactive Search. In *Proceedings of the 3rd International Conference on the Theory of Information Retrieval (ICTIR), Lecture Notes in Computer Science* (pp. 213–224). Bertinoro: Springer.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5), 407–424.
- Brath, R. (1997). Concept Demonstration: Metrics For Effective Information Visualization. In *Proceedings of IEEE Symposium On Information Visualization* (pp. 108-111). IEEE Service Center, Phoenix, AZ.
- Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. Morgan Kaufmann.
- Catledge, L. D., & Pitkow, J. E., (1995). Characterizing browsing strategies in the World Wide Web. *Computer Networks and ISDN Systems*, 27(6), 1065-1073..
- Collberg, C., Kobourov, S., Nagra, J., Pitts, J., & Wampler, K. (2003). A system for graph-based visualization of the evolution of software. In *Proceedings of the 2003 ACM symposium on Software visualization* (pp. 77-ff). ACM.
- Cove, J. F., & Walsh, B. C. (1988). Online text retrieval via browsing. *Information processing & management*, 24(1), 31-37.
- Chen, H. H., Kuo, J. J., Huang, S. J., Lin, C. J., & Wung, H. C. (2003). A summarization system for Chinese news from multiple sources. *Journal of American Society for Information Science and Technology*, 54(13),

1224-1236.

- Chung, H., Liu, K., Li, T., & Chen, P. (2010). Exploring dynamic social network through interactive visualization. In *Proceedings of IEEE Pacific Visualization*, Taipei, Taiwan.
- Delphi Group. (2004). Information Intelligence: Content Classification and the Enterprise Taxonomy Practice. Retrieved 2015/05/09 from <http://stratify.com/infocenter/download/DelphiResearchReport2004.pdf>
- Force-directed graph drawing. (2015, April 16). In Wikipedia, The Free Encyclopedia. Retrieved 12:45, June 9, 2015, from [http://en.wikipedia.org/w/index.php?title=Force-directed\\_graph\\_drawing&oldid=656697183](http://en.wikipedia.org/w/index.php?title=Force-directed_graph_drawing&oldid=656697183)
- Friedman, V. (2008). Data visualization and infographics. *Graphics, Monday Inspiration*, 14, 2008.
- Fruchterman, T., & Reingold, E. (1991). Graph drawing by force-directed placement. *Software - Practice and Experience*, 21(11), 1129–1164.
- Goodall, J. R. (2011). An evaluation of visual and textual network analysis tools. *Information Visualization*, 10(2), 145-157.
- Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 9-20.
- Kao, G. Y. M., Lei, P. L., & Sun, C. T. (2008). Thinking style impacts on Web search strategies. *Computers in Human Behavior*, 24(4), 1330-1341.
- Karypis, G., Han, E. H., & Kumar, V. (1999). Chameleon: Hierarchical Clustering Using Dynamic Modeling. *Computer*, 32(8), 68-75. doi: 10.1109/2.781637.
- Keim, D. A. (2002). Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphic*, 8(1), 1-8.
- Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2011). *Seven guiding scenarios for information visualization evaluation*. Research report: 2011-992-04, Canada: Department of Computer Science University of Calgary.
- Lash, S. (2002). Critique of information. London, UK: SAGE.
- Morville, P., & Callender, J., (2010). *Search Pattern*. New York, NY: O'Reilly Media.

- Plaisant, C. (2004). The challenge of information visualization evaluation. *In Proceedings of the working conference on Advanced visual interfaces* (pp. 109-116). Gallipoli, Italy: ACM.
- Rakesh A., & Ramakrishnan S. (1995). Mining sequential patterns. *In Proceedings of 1995 International Conference on Data Engineering (ICDE'95)*(pp. 3-14), Taipei, Taiwan.
- Santos, B. S. (2008). Evaluating Visualization Techniques and Tools: What Are the Main Issues. *In the 2008 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods For information Visualization (BELIV'08)*, Florence, Italy.
- Shneiderman, B., & Wattenberg, M. (1992). Tree visualization with tree-maps: A 2D space filling approach. *ACM Transactions on Graphics*, 11(1), 92-99.
- Takama, Y., Matsumura, A., & Kajinami, T. (2006). Visualization of news distribution in blog space. *In Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology* (pp. 413-416). Hong Kong: IEEE Computer Society.
- Tikhonova, A., & Ma, K. L. (2008). A scalable parallel force-directed graph layout algorithm. *In Proceedings of the 8th Eurographics conference on Parallel Graphics and Visualization* (pp. 25-32). Geneva: Eurographics Association.
- Tseng, Y. H., Chang, C. Y., Rundgren, S. N. C., & Rundgren, C. J. (2010). Mining concept maps from news stories for measuring civic scientific literacy in media. *Computers & Education*, 55(1), 165-177.
- Warner, W., & Jones, J. (2011). The wonder of words: Using technology to support vocabulary instruction. *The Agricultural Education Magazine*, 83(6), 7-9.

## Designing and Evaluating User Interface for Assisting News Navigation and Retrieval

**Ji-Lung Hsieh**

Assistant Professor  
Graduate Institute of Library and Information Studies  
National Taiwan Normal University

**Bo-Chen Shen**

Graduate Institute of Library and Information Studies  
National Taiwan Normal University

**Li-Wei Yang**

General Manager  
Tornado Technologies

Adjunct Assistant Professor  
Department of Business Administration,  
National Taiwan University

### Introduction

The internet provides news agencies an immediate, low-cost, and peer-produced publishing environment, however, it also leads to redundant and low reliability information. Several methods have been developed to assist users' news reading behaviors such as auto-clustering to reorganize news by keywords (Pu & Chen, 2007), auto-summarization of news across different news agencies (Chen, Kuo, Huang, Lin, & Wung, 2003), and building ontology models to detect different news that report the same event (Cheng, 2009). With increasing network speed, data visualization services have become more and more popular. These services not only provide elegant interfaces but also highlight the characteristics of data and scaffold



the understanding of issues, conveying a concept (Card, MacKinlay, & Shneiderman, 1999) by means which can help users to browse, search, and read information more efficiently. There have been several visualization tools for news browsing. For example, Blog Keyword Visualizer (BKV), which displays associations among blog keywords, allows users to observe hot issues in real-time. Google Wonder Wheel, like a sidebar widget, displays co-occurring relations among related terms to help users to discover related keywords. Newsmap uses area and color to represent each news topic's popularity and categories, respectively. ThemeRiver, similar to Google Trend, is designed to show and compare news trends.

Searching and browsing often appears alternatively in information retrieval process. Compared with searching, browsing focuses more on how users discover useful data by use of interactive functions of user interfaces. Therefore, while search goals are unclear, users often spend more time on browsing to discover potential keywords (Delphi Group, 2004). Therefore, users often read news for casual "encountering" of interested events, which is similar to the Bates (1989) berry-picking process. Carledge and Pitkow(1995) investigated how users relied on web browsers and found in the system log high frequency functions included clicking hyperlinks, clicking the "backtrack" button, opening URLs, opening Hotlists, and forwarding. The most frequent browsing strategies are "Leave as you've entered" and "Looping back," both of which highlighted the importance of keeping historical clicking and browsing paths. Accordingly, the study aims to build up a keyword visualization user interface that allows users to browse news from different aspects, and to provide visual clues for backtracking and memorizing users' visit history.

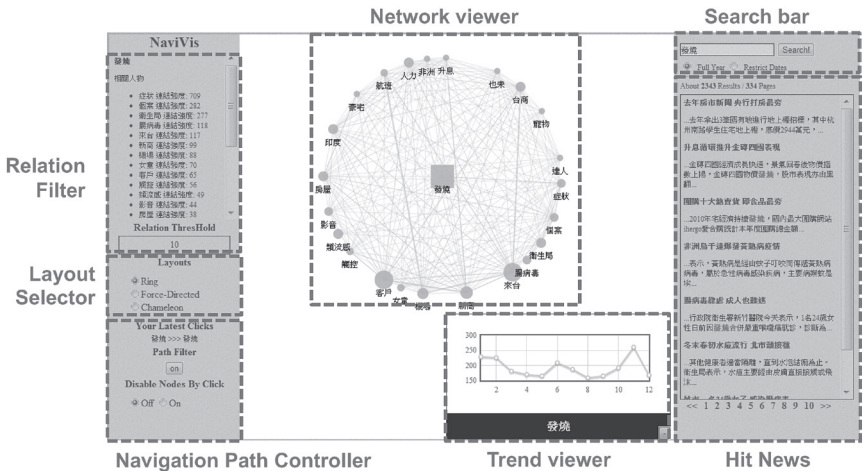
## Methods

Two Taiwanese online news sites, United Daily News and Chinatimes (total 793,522 documents from Jan 1st to Dec 31th in 2011), were crawled as the dataset. For each user query, (Yang, 2003 a, 2003 b) API was applied to retrieve the top-ranked 100 news and parse each news to obtain power terms automatically. The tool returns average five terms for each document. The efficiency and accuracy of the tool was evaluated by comparison with user-defined keywords in open accessed articles on a magazine website <http://>

www.cw.com.tw/. We compared the retrieved power terms and user-defined keywords of 1,036 articles. The accuracy rates for 313 long documents and 723 short documents are 86.17% and 90.66% respectively. —Finally, we calculated the frequency of each power term from the retrieved documents and built co-occurrence relationships for each pair of terms co-occurring in one or more documents.

To support users' information retrieval activities, the main features of our interface include a network-based keyword visualizer with three different layouts for displaying related keywords of current query, and visualized visiting path to support backtracking which aims to lessen the need for cognitive attention. Other designed features include pop-ups for displaying news trends and news reading, and a filter to decrease irrelevant keywords (shown in Figure 1). Primary modules including Network viewer and visit path controller were described as follows.

Figure 1. An overview of the network-based keyword visualizer for news browsing.

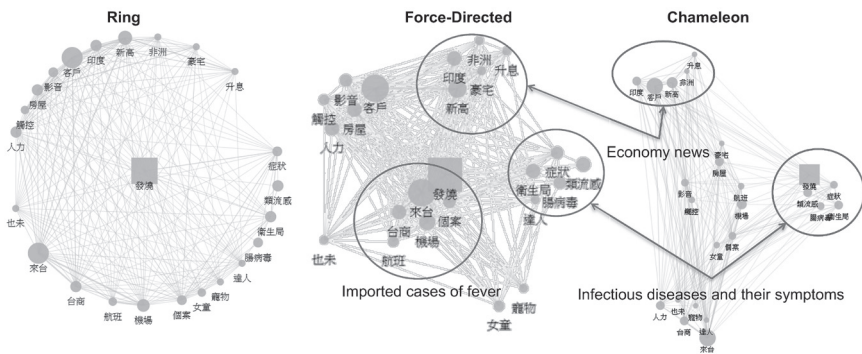


1. Network viewer. The co-occurrence relationship among keywords was visualized by network. Users can click nodes to expand the keyword network and explore more news, and disable – keywords of no interest. We implemented circular, force-directed, and Chameleon layouts for different purposes. Circular layout provides users a clear

view of the related terms located with proper spacing on the circle edge, making it easy for users to find strongly-related terms. Unlike the circular layout, force-directed layout locates each pair of terms with proper and reasonable distances according to the strength of their association (Tikhonova & Ma, 2008), which allows users to discover strong-related terms and highly-condensed subgroupings of terms. The last algorithm, Chameleon (Karypis, Han, & Kumary, 1999), is based on agglomerative hierarchical clustering method, and focuses on allocating clusters with proper distances between them, rather than allocating individual concepts directly as the force-directed layout does (Collberg, Kobourov, Nagra, Pitts, & Wampler, 2003). In this case, users can see clearer clusters than those in the force-directed method.

2. Visit (click) path controller. Users can interact with the keyword network by clicking to expand them. The clicked keyword and original query combine by the “OR” operator to be treated as a query to retrieve more news and power terms. User visit paths (highlighted by different colors) are represented by the sequence of the clicked terms and associations between them. Besides visualizing visit paths, the tool allows users to go backward by disabling nodes one by one. Once a clicked node is disabled, all of its related nodes, except for those also related to other clicked nodes, will be filtered out.

Figure 2. Three different network layouts implemented in the study.



## Evaluation

We recruited 10 university students in northern Taiwan who often read news daily to evaluate the usability, pros and cons of our tool. Each participant was asked to explore one un-familiar topic such as “Occupying Wall Street” and “2013 Egyptian coup d’état.” We used a free screen streaming software Free Screen Recorder to record their browsing behaviors. Finally, they were interviewed and asked to share how they feel about our tool and what they discovered through adjustable and interactive network layouts. Besides task observation, we applied a sequential pattern mining method, AprioriAll proposed by Rakesh and Ramakrishnan (1995), to obtain high supported successive operations for our tools.

## Results

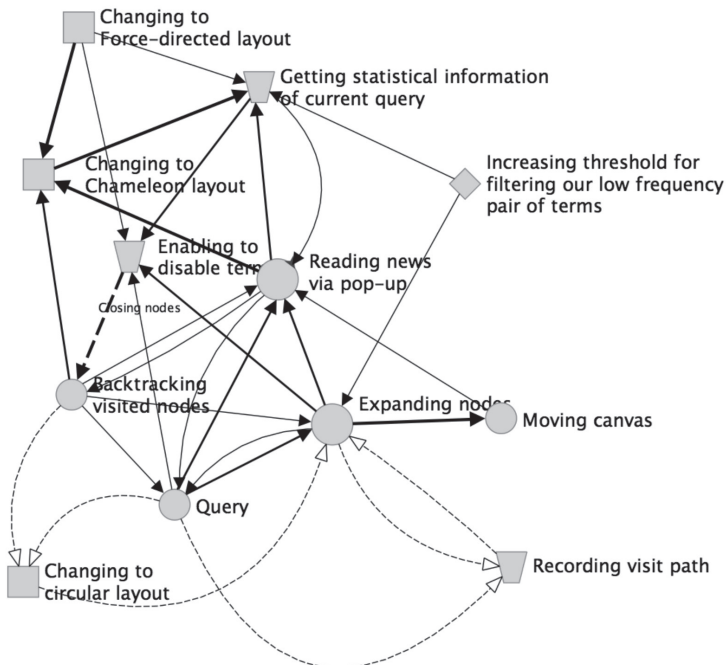
We encoded a total of 21 operations in the interface. 20.0% of all participant operations were expanding related keywords. Participants F06 and F10 expanded more keywords than others. They indicated that, instead of opening news and reading them one by one, expanding keywords can help them to understand unfamiliar topics in advance. However, some study subjects (F03, F04, F05, F06) also indicated that the keyword network generates too many keywords to consume, which means that the generated keywords may bring heavy cognitive load for users. Nevertheless, we found our participants used several strategies to overcome this problem. One such strategy is to use the function for “disabling keywords of no interest” to decrease the amount of keywords shown in the network. Another strategy is to switch network layout repeatedly to obtain a clear view by circular layout or observe concept groups by Chameleon layout. Compared with the first method, this strategy does not reduce the amount of information but organizes them according to user needs. Participants F01 and F07 were found to switch layouts more than others. The layout switching operations account for 32% and 25% of their behavior sequences, respectively.

Among the 10 participants, the longest sequence length was 62 while the shortest was 32 (average 41.5 operations). By setting minimal support to 0.6, we obtained 45 sequences with lengths from two to four. Sequences with more than two operations were segmented into pair of operations and

Figure 3 shows the transition diagram of 12 operations occurring in the 45 high-supported sequences. The thickness of directed edge represents the highest support degree of the pair. Several high-supported subsequences were described as follows.

1. After each query, next operations are often “expanding related keywords (support=0.8)” or reading news via pop-up (0.8).
2. We hoped to lower the query times of users. However, 80% of participants made more than one query (average 2.1 queries). The participants were found to re-send a query only when they read news (0.7) or accessed visit paths including expanding keywords (0.7) and backtracking their operations (0.7).
3. Participants were found to switch network layout after making a query (0.6), reading news (0.9), or backtracking several nodes (0.7). After switching to Force-directed layout, participants often switched to Chameleon layout to have more understanding on presented keyword networks.

Figure 3. Transition diagram of high-supported subsequences.



## Conclusion

The study designed a network-based keywords visualizer for assisting users to interact with news keywords to find out potential topics of interest or to get an early understanding before reading news content. A user study was conducted and showed that keyword expansion can help users to understand content of news in advance, and visit path recording and controlling can help users to trace their information retrieval process forward and backward. However, participants also reported several unsuitable designs including too large number of keywords to consume or unsuitable metaphors in visual designs, improvements to which may be considered in future work. The primary conclusion is that, although visual designs can provide more information and understanding of original data, the cognitive limitation of users' attention still needs to be taken into consideration.

## Reference

- Albakour, M. D., Kruschwitz, U., Nanas, N., Song, D., Fasli, M., & Roeck, A. D. (2011). Exploring Ant Colony Optimisation for Adaptive Interactive Search. In *Proceedings of the 3rd International Conference on the Theory of Information Retrieval (ICTIR), Lecture Notes in Computer Science* (pp. 213–224). Bertinoro: Springer.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5), 407–424.
- Brath, R. (1997). Concept Demonstration: Metrics For Effective Information Visualization. In *Proceedings of IEEE Symposium On Information Visualization* (pp. 108-111). IEEE Service Center, Phoenix, AZ.
- Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. Morgan Kaufmann.
- Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World Wide Web. *Computer Networks and ISDN Systems*, 27(6), 1065-1073.
- Collberg, C., Kobourov, S., Nagra, J., Pitts, J., & Wampler, K. (2003). A system for graph-based visualization of the evolution of software. In *Proceedings of the 2003 ACM symposium on Software visualization* (pp.

- 77-ff). ACM.
- Cove, J. F., & Walsh, B. C. (1988). Online text retrieval via browsing. *Information processing & management*, 24(1), 31-37.
- Chen, H. H., Kuo, J. J., Huang, S. J., Lin, C. J., & Wung, H. C. (2003). A summarization system for Chinese news from multiple sources. *Journal of American Society for Information Science and Technology*, 54(13), 1224-1236.
- Chung, H., Liu, K., Li, T., & Chen, P. (2010). Exploring dynamic social network through interactive visualization. In *Proceedings of IEEE Pacific Visualization*, Taipei, Taiwan.
- Delphi Group. (2004). Information Intelligence: Content Classification and the Enterprise Taxonomy Practice. Retrieved 2015/05/09 from <http://stratify.com/infocenter/download/DelphiResearchReport2004.pdf>
- Force-directed graph drawing. (2015, April 16). In Wikipedia, The Free Encyclopedia. Retrieved 12:45, June 9, 2015, from [http://en.wikipedia.org/w/index.php?title=Force-directed\\_graph\\_drawing&oldid=656697183](http://en.wikipedia.org/w/index.php?title=Force-directed_graph_drawing&oldid=656697183)
- Friedman, V. (2008). Data visualization and infographics. *Graphics, Monday Inspiration*, 14, 2008.
- Fruchterman, T., & Reingold, E. (1991). Graph drawing by force-directed placement. *Software - Practice and Experience*, 21(11), 1129-1164.
- Goodall, J. R. (2011). An evaluation of visual and textual network analysis tools. *Information Visualization*, 10(2), 145-157.
- Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 9-20.
- Huang, T. Y. (2011). *On the Visualization of Online Reputation*, Unpublished master's thesis, National Taiwan University, Taipei, Taiwan.
- Jiang, X. Y., & Pu, H. T. (2010). An Exploratory Study on the Re-finding Behavior on the Web. *Journal of Library and Information Studies*, 8(1), 29-57.
- Kao, G. Y. M., Lei, P. L., & Sun, C. T. (2008). Thinking style impacts on Web search strategies. *Computers in Human Behavior*, 24(4), 1330-1341.
- Karypis, G., Han, E. H., & Kumar, V. (1999). Chameleon: Hierarchical

- Clustering Using Dynamic Modeling. *Computer*, 32(8), 68-75. doi: 10.1109/2.781637.
- Keim, D. A. (2002). Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1-8.
- Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2011). *Seven Guiding Scenarios for Information Visualization Evaluation*. Research report: 2011-992-04, Canada: Department of Computer Science University of Calgary
- Lash, S. (2002). *Critique of information*. London, UK: SAGE.
- Morville, P., & Callender, J., (2010). *Search Pattern*. New York, NY: O'Reilly Media.
- Plaisant, C. (2004). The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces* (pp. 109-116). Gallipoli, Italy: ACM.
- Pu, H. T., & Chen S. Y. (2007). User Evaluation of Web Clustering Search Engines. *Journal of Library and Information Science*, 2(1), 55-80.
- Rakesh A., & Ramakrishnan S. (1995). Mining sequential patterns. In *Proceedings of 1995 International Conference on Data Engineering (ICDE'95)*, Taipei, Taiwan, pp. 3-14.
- Santos, B. S. (2008). Evaluating Visualization Techniques and Tools: What Are the Main Issues. In *the 2008 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods For information Visualization (BELIV'08)*, Florence, Italy.
- Shneiderman, B., & Wattenberg, M. (1992). Tree visualization with tree-maps: A 2D space filling approach. *ACM Transactions on Graphics*, 11(1), 92-99.
- Takama, Y., Matsumura, A., & Kajinami, T. (2006). Visualization of news distribution in blog space. In *Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology* (pp. 413-416). Hong Kong: IEEE Computer Society.
- Tikhonova, A., & Ma, K. L. (2008). A scalable parallel force-directed graph layout algorithm. In *Proceedings of the 8th Eurographics conference on Parallel Graphics and Visualization* (pp. 25-32). Eurographics Association.
- Tseng, Y. H., Chang, C. Y., Rundgren, S. N. C., & Rundgren, C. J. (2010).



Mining concept maps from news stories for measuring civic scientific literacy in media. *Computers & Education*, 55(1), 165-177.

Warner, W., & Jones, J. (2011). The wonder of words: Using technology to support vocabulary instruction. *The Agricultural Education Magazine*, 83(6), 7-9.

Yang, L. W. (2003a). Dian nao xin ci xue xi fang fa yu xi tong. *T.W. Patent No. 517195*.

Yang, L. W. (2003a). Zhong wen duan ci fa. *T.W. Patent No. 527546*.