

一個社會性書籤網站之探索分析

An Exploratory Analysis of a Social Bookmarking Site

盧 能 彬

Neng-Pin Lu

長庚大學資訊管理學系助理教授

Assistant Professor, Department of Information Management,

Chang Gung University

E-mail: nplu@mail.cgu.edu.tw

黃 士 瑋

Shih-Wei Huang

長庚大學資訊管理學系碩士班研究生

MA Student, Department of Information Management,

Chang Gung University

E-mail: m9844023@stmail.cgu.edu.tw

【摘要 Abstract】

社會性書籤網站應用 Web 2.0 群體智慧，讓使用者利用書籤收藏自己有興趣的網路資源，並藉由標籤的標記功能管理書籤。關於社會性書籤網站的相關研究，目前已有相當多的國外研究成果發表，然而國內社會性書籤網站的相關分析卻仍然有限，因此本研究以台灣的 funP 推推王書籤網站為研究對象，嘗試探索台灣社會性書籤的使用概況：首先，蒐集 2010 年 7 月 27 日至 2010 年 8 月 27 日為期一個月的書籤資料為研究樣本，進行書籤之分類、書籤之標籤分佈、以及書籤標記時序等基本統計；接著，再擷取 2010 年 7 月 27 日至 2010 年 8 月 2 日為期一週的書籤資料，進行細部的使用者、標籤、網路資源等書籤三大元素的使用頻率分析與關聯分析。本研究結果發現，三大元素的使用頻率均符合齊夫定律，而三大元素的關聯關係也均符合冪次定律。換言之，使用頻率較高的少數使用者、少數標籤、以及少數網路資源主宰了多數的書籤。因此，本研究再檢視 Top 20 標籤，發現第一名的標籤為部落格，而其他則多為部落格相關詞彙。最後，再檢視 Top 20 網路資源，則確認它們多數為部落格網站。

Applying the collective intelligence of Web 2.0, social bookmarking sites allow users to bookmark network resources of personal interest and to manage their bookmarks via tagging. Although a lot of research about foreign social bookmarking sites has been published, research on the usage of social bookmarking sites in Taiwan is still limited. Therefore, in this research we attempt to reveal the usage patterns of social bookmarking

in Taiwan by investigating the social bookmarking site funP. First, we collected the bookmark data of funP from July 27, 2010 to August 27, 2010 as our research sample, and we conducted basic statistical analyses of the data, including the categorization of bookmarks, tag distribution of bookmarks, and bookmark timing of users. Then, we took the one-week bookmark data from July 27, 2010 to August 2, 2010 to perform usage-frequency and association analyses of users, tags, and URLs within bookmarks. We find that usage-frequencies of users, tags, and URLs all follow Zipf's law, while relationships between users, tags, and URLs all follow the power law. In other words, a minority of high-frequency users, tags, and URLs dominate the majority of bookmarks. Due to the discovery, we further inspected the Top 20 tags and find that the tag "blog" has the highest usage-frequency and the rest are mostly blog-related terms. We eventually inspected the Top 20 URLs, confirming that most of them direct to blog sites.

關鍵詞 Keyword

社會性書籤 使用者 網路資源 標籤 齊夫定律 冪次定律 部落格

Social bookmarking ; User ; URL ; Tag ; Zipf's law ; Power law ; Blog

壹、緒論

在 Web 1.0 的時期，只有少數網站建構者透過網際網路作為傳遞資訊的媒介，網頁瀏覽者在網際網路中多為被動地接收資訊。直至 Web 2.0 群體智慧(collective intelligence)概念的出現，網頁瀏覽者才由單純瀏覽網頁的被動者轉變為發佈內容與回應內容的主動參與者(Peters & Stock, 2007)。社會性書籤網站(social bookmarking site)是以社會性標籤系統(social tagging systems)所建置的網站，為 Web 2.0 群體智慧中，網頁瀏覽者共同參與的重要應用之一(Hoegg, Meckel, Stanoevska-Slabeva, & Martignoni, 2006; O'Reilly, 2007)。社會性標籤系統將個人的書籤使用行為透過網路共享標籤與資源，與其他使用者形成連結(Furnas et al., 2006)。使用者可以利用標籤標記有興趣的網路資源，並透過書籤的發布，達成網路資源共享及個人化網路資源管理的目的。國外的 Delicious、CiteULike、Digg 與國內的 funP 推推王、MyShare、HEMiDEMi 等，均是當今社會性書籤網站的代表。

目前已有許多國外社會性書籤網站的使用樣式分析發表，例如：Delicious 的使用樣式研究(Golder & Huberman, 2006; Cattuto et al., 2007; Chi & Mytkowicz, 2007; Krause, Hotho, & Stumme, 2008)、以及 CiteULike 的使用樣式研究(Farooq, Song, Carroll, & Giles, 2007; Helic, Trattner, Strohmaier, & Andrews, 2010)。至於國內的社會性書籤網站相關研究，則多以問卷、訪談的方式瞭解使用者利用標籤標記網路資源的動機(卜小蝶、張淇龍, 2009)，以及發展演算法與開發系統改善社會性書籤搜尋引擎的效率(鄧睿清, 2009)。然而，關於國內社會性書籤網站的使用樣式分析仍然缺乏；故本研究蒐集 funP 推推王的書籤資料為研究樣本，統計與分析其使用樣式，探索台灣地區社會

性書籤的應用概況，並比較國內與國外社會性書籤網站之使用異同。

貳、研究背景

一、社會性書籤

在 Web 2.0 尚未蓬勃發展的時期，人們對於喜愛的網頁資源只能儲存在網頁瀏覽器的「我的最愛」中，使用者無法從其他電腦取用自己收藏的網頁資源。然而，現今社會性書籤網站已漸漸取代「我的最愛」成為方便的書籤管理工具，使用者可以根據其個人喜好、興趣，創造書籤與儲存書籤於書籤網站中。這些書籤可以是公開的或是私自保存管理的，也可以分享給特定的使用者、團體或是特定網域內的使用者，而具有權限的使用者則可以藉由標籤分類或搜尋引擎瀏覽書籤。

社會性書籤系統是一個提供網路使用者組織、儲存、管理與搜尋網路資源的方法(Golder & Huberman, 2006)。Halpin, Robu, and Shepherd (2007) 定義社會性書籤系統是使用者(user)、標籤(tag)以及網路資源(URL)三大元素所構成。在三大元素的集合中，

u_1, u_2, \dots, u_i 組成使用者集合，記作

$$U = \{u_x \mid 1 \leq x \leq i, \forall x \in \mathbf{Z}^+\};$$

t_1, t_2, \dots, t_j 組成標籤集合，記作

$$T = \{t_y \mid 1 \leq y \leq j, \forall y \in \mathbf{Z}^+\};$$

r_1, r_2, \dots, r_k 組成網路資源集合，記作

$$R = \{r_z \mid 1 \leq z \leq k, \forall z \in \mathbf{Z}^+\}。$$

當使用者 u_1 將標籤 t_1 與 t_2 標記於網路資源 r_1 ，便產生了一個書籤的關聯組合 $\{u_1, \{t_1, t_2\}, r_1\}$ 。

二、社會性書籤相關研究

目前國外的社會性書籤相關研究多從使用者、標籤、網路資源三大元素的角度分析，藉以瞭解社會性書籤網站的使用行為，以及探討標籤語意問題，進而改善社會性書籤網站之使用效率與探索社會性書籤網站中的社群網絡。

關於統計分析方法的研究，Golder and Huberman (2006)分析 Delicious 熱門的網路資源樣本以及隨機抽樣的使用者樣本，發現少數使用者標記許多的標籤，但大部分使用者標記的標籤數量並不多。爾後，許多研究學者也以 Delicious 為樣本進行分析。Yanbe, Jatowt, Nakamura, and Tanaka (2007)除了觀察標籤的語意樣式之外，更透過時序分析觀察到有半數的使用者收藏的網路資源，多為 PageRank 評估為級別較高且近期發布的網頁；Chi and Mytkowicz (2007)運用條件亂度(conditional entropy)，將標籤和網路資源的關係量化成數值，再利用資訊理論(information theory)來探討社會性書籤系統的效率；Li, Guo, and Zhao (2008)開發 ISID(Internet Social Interest Discovery)系統，利用叢集分析(clustering analysis)找出以標籤為中介的使用者；Korner, Benz, Hotho, Strohmaier, and Stumme (2010)則將使用者分為分類者(categorizers)與敘述者(describers)，探討其標記標籤的動機，並從使用者標記標籤數量的多寡、標籤與網路資源的比率、每一筆書籤使用的標籤量，以及極少被使用的標籤四種統計觀點觀察使用行為。

除了以 Delicious 作為分析樣本之外，CiteULike、Connotea 同樣是研究學者探討社會性書籤特性的研究對象。Farooq et al. (2007)統計使用者與標籤的新增數量以觀察兩者的成長趨勢，並從標籤的重複使用次數的統計中發現，CiteULike 是屬於標籤重複使用率較低的網站；Santos-Neto,

Ripeanu, and Iamnitchi (2007)分別討論使用者與網路資源以及使用者與標籤的交集程度，並設定一個門檻值 t ，判斷出使用者之間的相似度。關於網絡分析方法的研究，Halpin et al. (2007)則根據標籤與標籤之間的相對關係，發展出衡量標籤之間的權重距離公式。以標籤為節點，權重距離為邊，繪製出具有視覺化的標籤關聯圖(inter-tag correlation graph)，藉以檢視標籤的共生網絡(co-occurrence network)。

至於國內的社會性書籤相關研究則多以問卷、訪談的方式瞭解使用者利用標籤標記網路資源的動機，以及發展演算法與開發系統改善社會性書籤搜尋引擎的效率。Liu and Chang (2008)以 funP 推推王社會性書籤網站為例，探討學習理論與社會性書籤服務整合的可行性；卜小蝶與張淇龍(2009)採三階段的內容分析、Q 方法以及訪談，探索使用者使用社會性書籤網站的動機，結果發現使用者共用標籤的頻率頗高、且多屬於意義廣泛的主題類別詞彙。發展演算法與系統開發的研究方面：戴瑋(2008)開發一套搜尋系統改善現有搜尋引擎的語意問題；陳崇正(2009)開發相似度演算法改善使用者搜尋結果，協助使用者透過相似興趣或專業來集結社群；曾姿婷(2009)利用時間序列分群演算法計算不同時間區間所形成之群聚的相似度，檢視社會性標籤隨時間的變化趨勢；鄧睿清(2009)則利用協同式過濾(collaborative filtering)的鄰居排列法結合派系篩檢(clique filtering)法的重疊特性，發掘出使用者網絡的群聚現象。本研究則嘗試統計與分析 funP 推推王書籤網站的資料，探索台灣地區社會性書籤的應用概況，並比較國內外社會性書籤網站之使用異同。

三、齊夫定律與冪次定律

齊夫定律(Zipf's Law)是語言學家 Zipf(1949)

觀察語料使用頻率時所發現。在 Brown 語料庫中，the 是最常見的單字，它在語料庫中出現的機率大約為 7%，排名第二位的單字 of 的機率為 3.5%，而排名第三位的單字 and 的機率為 2.7%。由語料出現的機率可以發現，排名第二的出現頻率約是排名首位出現頻率的 1/2，排名第三的出現頻率約是排名首位出現頻率的 1/3，以此類推，排名第 n 的出現頻率是排名首位出現頻率的 $1/n$ 。Zipf 發現單字出現的頻率與它在語料庫裡的排名成反比，定義第 n 名語料出現的機率：

$$P_n \sim \frac{1}{n^a} \quad (1)$$

若將齊夫定律中的排名屬性改用其他屬性代替，則可一般化為冪次定律(Power Law)。冪次定律說明物件或事件中屬性 x 的出現或發生機率 $\rho(x)$ 會以 $-r$ 的冪次比例關係存在(Newman, 2005)：

$$\rho(x) = Ax^{-r} \quad (2)$$

其中指數 r 描述其分配如何以冪次函數的方式改變。判斷冪次定律時，可將 $\rho(x)$ 取對數，若是符合冪次定律的情況下，所呈現的關係具有線性特徵，其方程式如下：

$$\log(\rho(x)) = -r\log(x) + \log(A) \quad (3)$$

冪次定律有兩項特性：(1)極大值發生於接近原點的地方，隨後持續下降至無窮遠處。(2)衰減的速率，比常態分配緩和得多，使得產生偏離值的可能性高出許多。冪次定律經常出現在自然以及人工的現象中，包含語言中最常出現的字詞、月球表面坑洞的大小等。至於社會性書籤的分析上也常見齊夫定律與冪次定律的出現：在 Delicious 的相關研究中，Heymann and Garcia-Molina (2006)發現愈熱門的標籤，被使用者標記的次數愈多；Halpin et al. (2007)與 Li et al. (2008)均發現，多數使用者只

收藏較少的網路資源，標記較少的標籤。另外，Santos-Neto et al. (2007)與 Helic et al. (2010)也發現 CiteULike、Connotea、以及 BibSonomy 三個屬於文獻資料管理的社會性書籤網站也符合冪次定律：大多數的使用者重複使用少數相同的標籤。

參、研究方法

一、研究對象

根據 Alexa (2010)網路資訊公司的網路流量統計，截至 2010 年 08 月 10 日，funP 推推王在國內網站的排名為 205，HEMiDEMi 的排名為 1742，MyShare 的排名為 307。2010 年 5 月至 2010 年 8 月的到站率統計數據中，funP 推推王為 0.00786，HEMiDEMi 為 0.00207，MyShare 為 0.00542。funP 推推王在 2009 年至 2010 年之間，一直是國內使用率最高的社會性書籤網站，因此本研究選擇 funP 推推王作為研究母體。

funP 推推王採用類似 Digg 的營運模式(Lerman, 2007)，應用眾多網站的服務混搭(service mashup)，並結合 RSS 與詮釋資料搜尋(metadata search)功能，再加上標籤以及投票機制作為資料索引基礎來建置此社會性書籤網站(funP 推推王官方網站，2011)。funP 網站中會根據使用者收藏的書籤建議分類，計有 13 個主類別和 84 個子類別共 97 個為內建分類標籤，如表 1 所示。在 funP 推推王中，書籤若使用子類別標籤後，即會自動帶出主類別標籤。

二、樣本挑選

由於 funP 推推王的書籤資料龐大，受限於分析系統的軟硬體限制，難以完整蒐集，因此本研究參考 Marlow, Naaman, Boyd, and Davis (2006)，Lerman (2007)，Zollers (2007)等研究的方法，先蒐集 2010 年 7 月 27 日至 8 月 27 日為期一個月的書

籤資料進行基本統計；接著，再從一個月的書籤資料中擷取 2010 年 7 月 27 日至 8 月 2 日一週的資料分析 funP 網站中使用者、標籤、網路資源等三大元素之間的關聯。

三、資料收集

本研究利用 funP 推推王提供的 RSS 訂閱功能，蒐集 2010 年 7 月 27 日至 8 月 27 日為期一個月的書籤資料。透過最新書籤的 RSS feed，可以收

集到的資料有：(1) 新增書籤時間；(2) funP 推推王書籤內部網址。由於 RSS feed 所提供的資料有限，為了收集到更詳細的書籤資料，本研究另外撰寫 PHP 網頁程式語言來自動化擷取完整的書籤資料。所有收集到的書籤資料則利用 EXCEL 進行資料的分類彙整與前置處理後，儲存於 MySQL 資料庫中，以供後續 SPSS 的統計與分析。

表 1
funP 推推王之內建分類標籤

主類別	子類別
美食	食記、食譜、小吃、優惠、團購
旅行	日本、亞洲、中國、台灣、澳洲、歐洲、美國、遊記
科技	軟體、電腦、網路、手機、筆電、相機、汽車、生態、科學、超自然
動漫	圖文、動畫、漫畫、遊戲
運動	大聯盟、中華職棒、籃球、足球、網球、棒球、賽車、單車、活動、運動員
娛樂	藝人、音樂、電影、電視、偶像劇、八卦
政治	政府、政黨、選舉、人物、國際、評論、政策、時事
藝文	設計、音樂、電影、攝影、閱讀、繪畫、表演、活動
生活	職場、健康、校園、心情、親子、兩性、愛情、閱讀、寵物、日記、消費
財經	股票、理財、投資、金融、經濟、產業、國際
時尚	美妝、服飾、設計、保養、購物、髮型、美甲
新聞	國際、政治、社會、財經、評論、科技、運動、健康、娛樂
影片	KUSO、搞笑、廣告、娛樂、電影、MV、音樂、節目、動畫

四、資料庫設計

經 EXCEL 前置處理後，本研究將整理好的書籤資料、使用者資料、網路資源資料、標籤資料儲存至 MySQL 資料庫中以便進行後續的各項分析。本研究共設計了四個資料表，分別為書籤資料表、使用者資料表、標籤資料表、網路資源資料表，四個資料表的實體關係圖(E-R Diagram)，如圖 1 所

示。書籤資料表是此資料庫設計的核心：一個書籤可標記多個標籤，且一個標籤可被多個書籤使用，呈現多對多的關係；一個書籤僅由一位使用者所建立，並且只會針對一個網路資源作為收藏，各呈現一對一的關係。使用者資料表的部份，一位使用者可以標記多個書籤，一個書籤也可以被多位使用者命名使用，各呈現多對多的關係；一位使用者可以

擁有多個網路資源，一個網路資源也可以被多位使用者擁有，呈現多對多的關係。網路資源資料表與標籤資料表的關係，一個網路資源可以不被標記標

籤亦可以被多個標籤所標記，一個標籤可以標記多個網路資源，呈現多對多的關係。

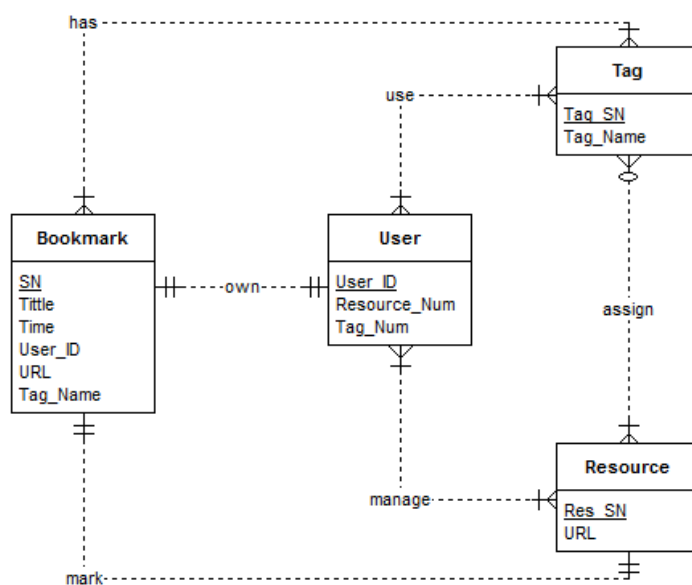


圖 1 資料庫實體關係圖

五、分析項目

完成書籤資料的蒐集後，本研究首先對 2010 年 7 月 27 日至 8 月 27 日為期一個月的資料進行書籤基本統計，項目包括：

1. 書籤之主題分類統計；
2. 書籤之標籤數統計；
3. 書籤之標記時序統計。

接著，再從一個月的書籤資料中擷取 2010 年 7 月 27 日至 8 月 2 日一週的書籤資料進行使用者、標籤、網路資源等三大元素的分析，項目包括：

1. 使用者、標籤、網路資源使用頻率分析；
2. 使用者、標籤、網路資源關聯分析；
3. 標籤與網路資源之連結探索。

肆、分析結果

一、書籤基本統計

本研究針對 2010 年 7 月 27 日至 8 月 27 日一個月的書籤資料，進行書籤之主題分類統計、標籤個數統計、標記時序統計，初步了解 funP 推推王的使用概況，表 2 為此部份樣本之基本資料。

表 2
書籤樣本之個數統計

	個數
書籤 $ B _{\text{month}}$	24,747
使用者 $ U _{\text{month}}$	2,849
標籤 $ T _{\text{month}}$	24,349
網路資源 $ R _{\text{month}}$	8,332

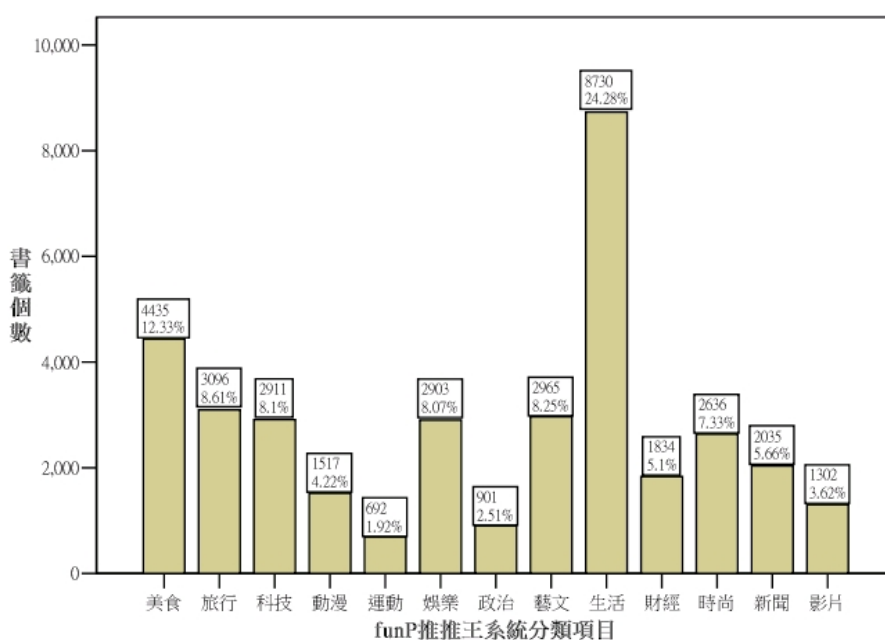


圖 2 書籤之主題分類統計

(一)書籤之主題分類統計

本研究以 funP 推推王所訂定的內建主分類標籤，對書籤進行主題分類。舉例來說，若有一筆書籤所標記的標籤包含有內建主分類標籤美食與生活，以及非內建主分類標籤小吃，則此書籤便歸類在美食與生活兩個分類項目中。圖 2 為書籤分類的統計結果：主分類為生活的書籤有 8,730 筆，比例

佔 24.28%，是 funP 推推王系統比例最高的分類；主分類為運動的書籤有 692 筆，比例佔 1.92%，是 funP 推推王系統比例最低的分類。

(二)書籤之標籤數統計

funP 推推王系統限制每個書籤最多可標記 21 個標籤，並且至少要標記 1 個標籤。舉例來說，若一筆書籤所擁有的標籤包含有科技、iPhone、

APPLE，則此書籤擁有的標籤個數為 3 個。將統計數據繪製成圖 3，可以發現擁有標籤為 4 至 6

個的書籤個數居多；擁有 6 個標籤以上的書籤個數則漸次減少；擁有 21 個標籤的書籤個數最少。

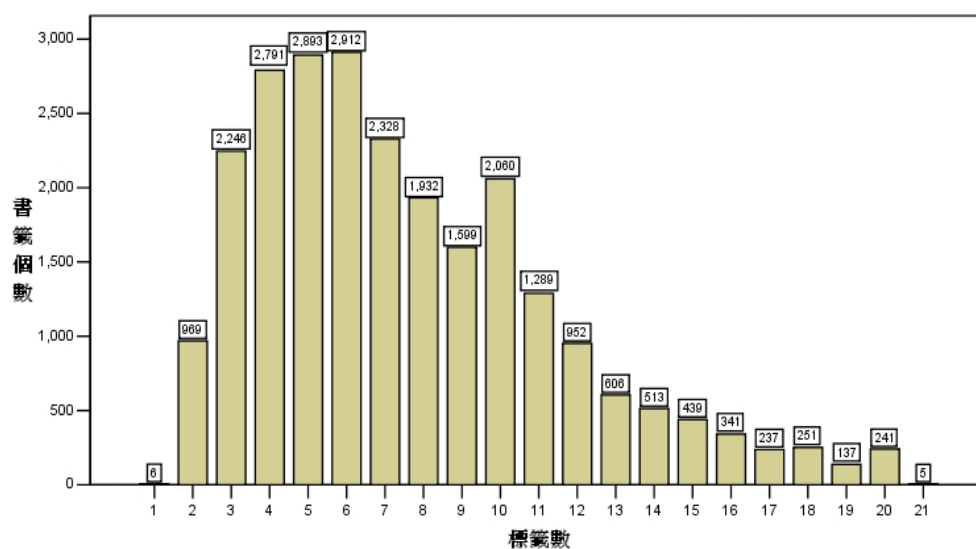


圖 3 書籤之標籤數統計

(三)書籤之標記時序統計

圖 4 則為每日新增書籤個數統計。以一週七天為週期觀察，可以發現星期一與星期二的書籤新增個數較高；星期六、星期日的書籤新增個數較低；愈接近週末，書籤的新增個數愈低。以一天二十四小時為週期觀察，可以發現 0 時至 4 時的新增書籤平均個數為下降的趨勢；5 時至 11 時的新增書籤

平均個數為上升的趨勢；11 時是當日新增書籤平均個數最高的時間點；12 時至 15 時的新增書籤平均個數皆為上升的趨勢，唯獨星期日是呈現下降的趨勢，15 時是當日新增書籤平均個數次高的時間點；15 時至 18 時的新增書籤平均個數皆為下降的趨勢，唯獨星期日是呈現上升的趨勢；自 20 時開始，愈接近 23 時，新增書籤平均個數則再增高。

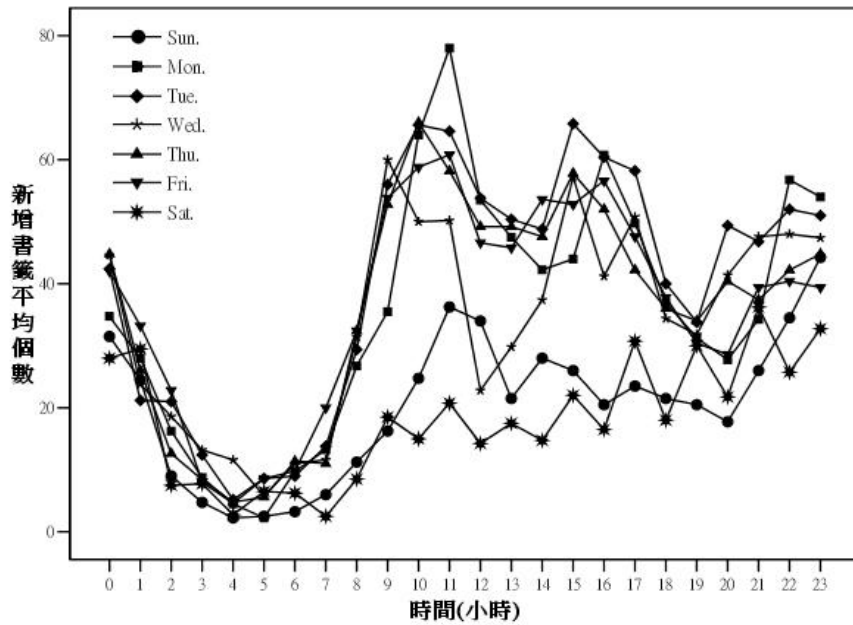


圖 4 書籤之標記時序統計

二、使用者、標籤、網路資源三大元素分析

由於 funP 推推王書籤資料龐雜，受限於分析系統的軟硬體限制，無法完整分析一個月的書籤資料的細部分析，因此本研究再挑出 2010 年 7 月 27 日至 8 月 2 日為期一週的書籤樣本，進行使用者、標籤、網路資源三大元素分析。此部份樣本的書籤數 $|B|_{\text{week}}=5,466$ ，其中使用者數 $|U|_{\text{week}}=1,295$ ，標籤數 $|T|_{\text{week}}=7,487$ ，網路資源數 $|R|_{\text{week}}=1,840$ 。

(一)三大元素使用頻率分析

圖 5 為使用者、網路資源、標籤三大元素在書籤中的使用頻率分佈，表 3 為分佈之敘述統計。本研究發現三大元素的使用頻率排名均符合齊夫定律，其趨勢線方程式與解釋度整理於表 4，使用者與網路資源的解釋度較高；標籤的解釋度則較低。

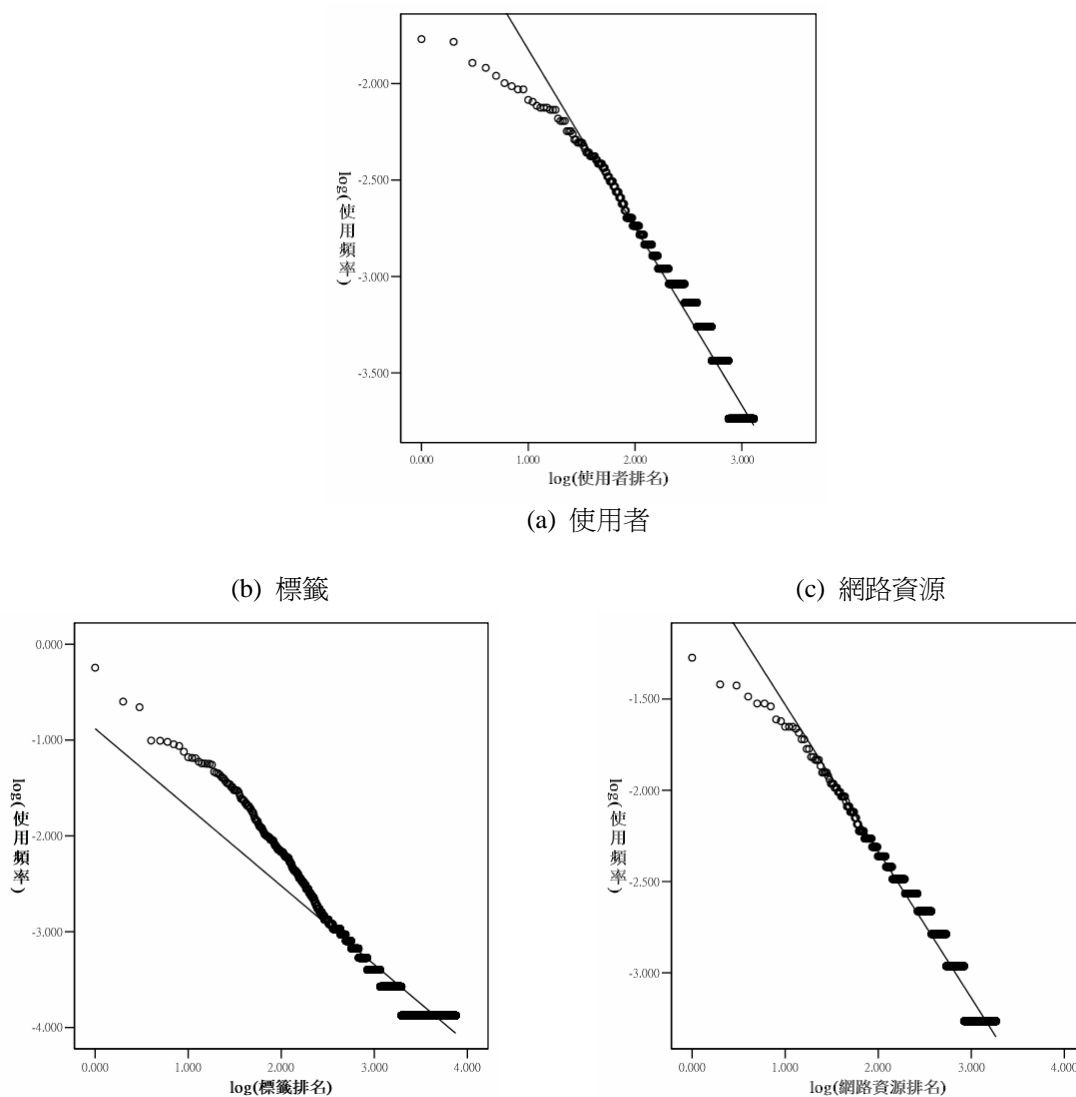


圖 5 三大元素使用頻率分佈

表 3

三大元素使用頻率分佈之敘述統計

統計項目	個數	平均數	標準差	眾數	中位數	最大值	最小值
書籤元素							
使用者	1,295	4.221	7.747	1	2	93	1
標籤	7,487	5.698	64.362	1	1	4,265	1
網路資源	1,840	2.971	5.832	1	1	97	1

表 4
三大元素使用頻率分佈之趨勢線方程式

	趨勢線方程式	R ²
使用者	$y = -0.9229x - 0.9004$	0.9493
標籤	$y = -0.8206x + 0.8807$	0.8718
網路資源	$y = -0.8035x - 0.7277$	0.9503

(二)三大元素關聯分析

爲了探討 funP 推推王的使用者、標籤、網路資源之間的關聯關係，本研究以 Schenkel et al. (2008) 分析使用者、標籤、網路資源之間的關聯分析法爲基礎，除正向主動關聯分析外(使用者標記標籤、標籤隸屬網路資源、使用者收藏網路資源)，並加入反向被動關聯分析(標籤被使用者標記、網路資源被使用者收藏、網路資源被標籤標記)。因此，本研究定義了六種關聯關係如下：

- 1.使用者-標籤：使用者與標記標籤個數的關聯關係。
- 2.使用者-網路資源：使用者與擁有網路資源個數的關聯關係。
- 3.標籤-使用者：標籤被多少位使用者標記次

數的關聯關係。

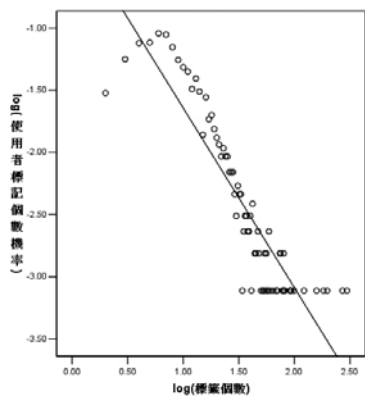
- 4.標籤-網路資源：標籤被多少網路資源使用次數的關聯關係。
- 5.網路資源-使用者：網路資源被多少位使用者收藏次數的關聯關係。
- 6.網路資源-標籤：網路資源擁有標籤個數的關聯關係。

根據六種關聯關係定義，表 5 爲關聯關係之敘述性統計；圖 6 則爲關聯關係之雙對數機率分佈圖；表 6 爲相關趨勢線方程式之整理。本研究發現網路資源-使用者是六種連結關係中最符合冪次定律；使用者-標籤、使用者-網路資源、網路資源-標籤的分佈相似；標籤-使用者、標籤-網路資源的解釋度則較低。

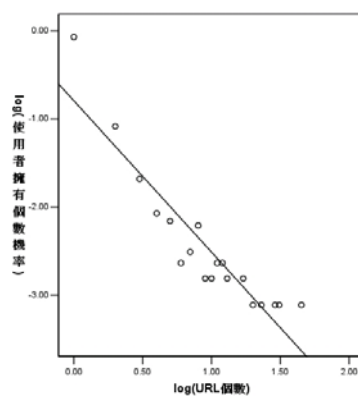
表 5
關聯關係之敘述統計

統計項目 三方關係	個數	平均數	標準差	眾數	中位數	最大值	最小值
使用者-標籤	1,295	13.5027	18.4146	6	9	295	2
使用者-網路資源	1,295	1.4795	2.3241	1	1	44	1
標籤-使用者	7,487	2.3383	17.9388	1	1	1,053	1
標籤-網路資源	7,487	2.9338	23.2452	1	1	1,403	1
網路資源-使用者	1,840	1.0413	0.2807	1	1	6	1
網路資源-標籤	1,840	11.3413	12.9602	7	8	195	2

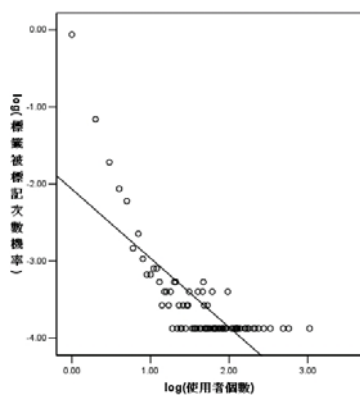
(a) 使用者-標籤



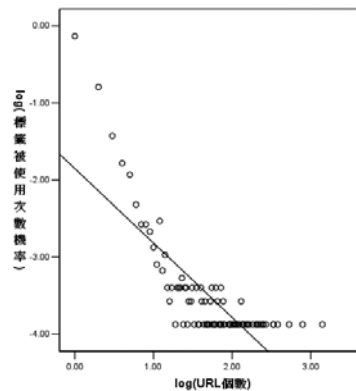
(b) 使用者-網路資源



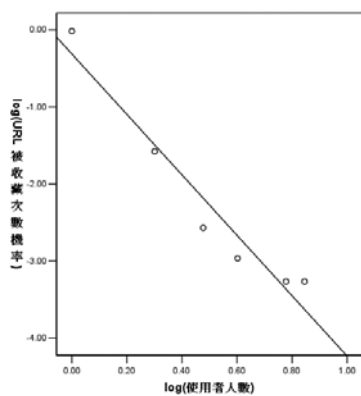
(c) 標籤-使用者



(d) 標籤-網路資源



(e) 網路資源-使用者



(f) 網路資源-標籤

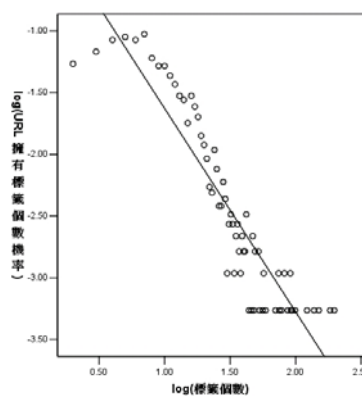


圖 6 三大元素關聯機率分佈

表 6
三大元素關聯機率分佈之趨勢線方程式

	趨勢線方程式	R^2
使用者-標籤	$y = - 1.4461x - 0.1958$	0.8194
使用者-網路資源	$y = - 1.7193x + 0.7914$	0.8538
標籤-使用者	$y = - 0.8971x - 2.0657$	0.5933
標籤-網路資源	$y = - 0.9666x - 1.8505$	0.6356
網路資源-使用者	$y = - 3.9245x - 0.3101$	0.9425
網路資源-標籤	$y = - 1.6507x + 0.0234$	0.8535

(三)標籤與網路資源之連結探索

在發現使用者、標籤、網路資源等三大元素的使用頻率均符合齊夫定律，以及六種關聯關係均符合冪次定律之後，本研究確認使用頻率較高的少數使用者、少數標籤、以及少數網路資源主宰了多數的書籤：在一週共 5,466 個書籤下，Top 20 使用者涵蓋 18.9%的書籤(1,034/5,466)；Top 20 標籤涵蓋 96.9%的書籤(5,296/5,466)；Top 20 網路資源則涵蓋 17.2%的書籤(938/5,466)。因此，本研究可以藉檢視使用頻率較高的少數元素來了解整體社會性書籤系統的大概樣貌。不過受限於 funP 推推王使用者資訊的隱私，本研究僅列出 Top 20 標籤與 Top 20 網路資源，探索熱門標籤與網路資源的代表性。如表 7 左大欄所示，部落格是標記次數最多的標籤，且次數遠多於其他 19 名的標籤。然而 Top 20 標籤除了部落格、Pixnet、無名、台北之外，其餘 16 個標籤皆為 funP 推推王之系統內建分類標籤。因此，本研究再將系統內建分類標籤排除，重新統計 Top 20 標籤於表 7 右大欄。結果發現部落格仍為標記次數最多的標籤，但相較於包含系統內建分類標籤之統計數據，標記次數和其他標籤的標記次

數差距則更加擴大。

由於部落格的標記次數最多，表示與其他標籤共同標記於書籤的可能性很高。因此，本研究針對部落格標籤與其他標籤進行共同標記分析，並區分為包含系統內建分類標籤與排除系統內建分類標籤兩種情況。在一週的書籤中，部落格被 4,265 個書籤標記過，共有 29,527 個標籤與部落格共同被標記在書籤中，若排除重複出現的共同標記標籤則有 618 個，並標記於 1,309 個不同的網路資源。接著，將所有與部落格共同標記過的標籤計算個別的共用機率，繪製於圖 7：其中圖 7(a)包含系統內建之分類標籤；圖 7(b)則排除系統內建之分類標籤。圖中並分別註明與部落格共用機率最高的前 5 名標籤：圖 7(a)中的前 5 名標籤分別為 Pixnet、生活、美食、旅行、無名；圖 7(b)中的前 5 名標籤分別為 Pixnet、無名、Xuite、im.tv、台北。這些共用標籤均為部落格相關詞彙。另外，共用標籤分佈也均符合齊夫定律(包含系統內建分類標籤之 $R^2=0.9406$ ；排除系統內建分類標籤之 $R^2=0.9401$)。

表 7
Top 20 標籤統計

包含內建分類標籤			排除內建分類標籤		
排名	標籤	標記次數	排名	標籤	標記次數
1	部落格	1,054	1	部落格	1,054
2	生活	572	2	Pixnet	486
3	Pixnet	486	3	無名	163
4	美食	334	4	台北	123
5	旅行	276	5	Xuite	108
6	藝文	236	6	優惠	50
7	科技	214	7	台中	48
8	娛樂	203	8	@100	47
9	時尚	175	9	@138	40
10	食記	174	10	高雄	34
11	無名	163	11	宜蘭	31
12	心情	159	12	台南	29
13	日記	137	13	相簿、英文網頁	21
14	台灣	129	14	桃園	20
15	消費	126	15	網路城邦	18
16	親子	125	16	FaceBook	17
17	健康	124	17	美女	16
18	台北	123	18	餐廳、花蓮、正妹	15
19	動漫	120	19	iPhone	13
20	新聞	114	20	@566、環保、新竹、ipad	12

從 Top 20 標籤的統計以及部落格標籤的共同標記分析得知，使用者所標記的網路資源大多以部落格相關詞彙為標籤，顯示使用者所收藏且有興趣的網路資源也應以部落格網站為主。因此，本研究

再統計 Top 20 網路資源於表 8：在包含重複排名的 Top 20 網路資源共有 24 個，其中有 21 個為部落格網站，確認 funP 推推王與部落格之間的確有著密切的關聯。

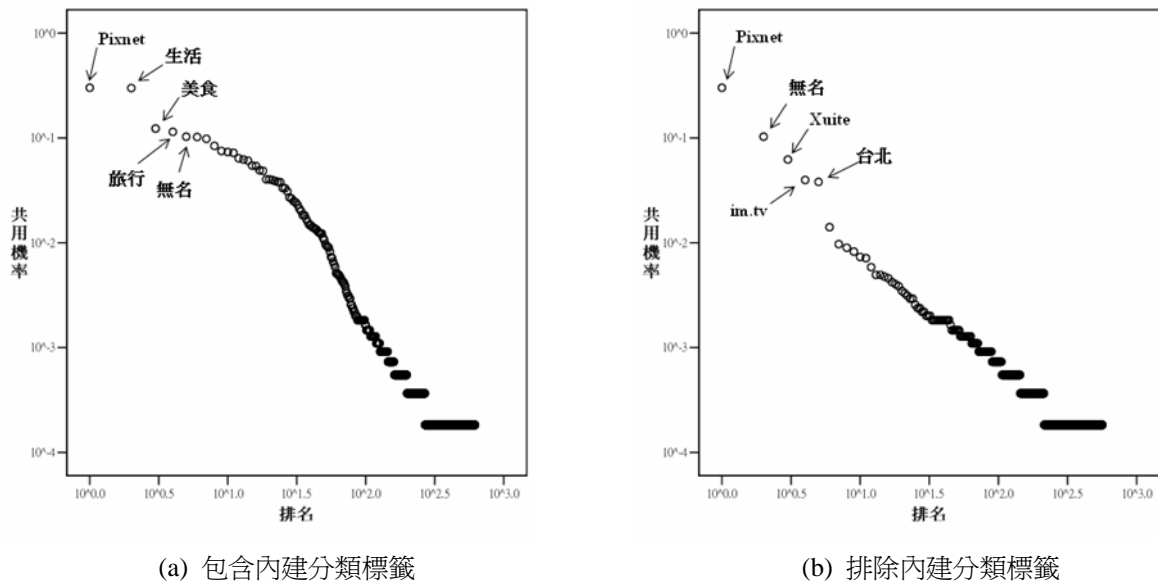


圖 7 部落格共同標記之標籤分佈

表 8

Top 20 網路資源

排名	URL	排名	URL
1	*http://www.wretch.cc/blog/jansen001	11	http://www.sitebro.tw/zhca49856
2	http://www.techorz.com	12	*http://blog.yam.com/michaelwu30 http://www.taiwantt.org.tw/tw/index.php
3	*http://blog.yam.com/meson	13	*http://nw0912.pixnet.net/blog
4	*http://neoformosamagz.blogspot.com	14	*http://roilyoko.pixnet.net/blog
5	*http://funiphone.pixnet.net/blog	15	*http://kurtz120.pixnet.net/blog *http://sagar.pixnet.net/blog
6	*http://pramep.com	16	*http://www.im.tv/blog/3370657
7	*http://tzoziing.pixnet.net/blog	17	*http://www.eslife.ws
8	*http://moonpoet.com	18	*http://www.im.tv/blog/120180703 *http://blog.yam.com/jasonjc
9	http://tw.news.yahoo.com *http://kuroro79.pixnet.net/blog	19	http://www.shinybeauty.com.tw/index.php
10	http://www.sitebro.tw/zhca49856	20	*http://tintinlee15.pixnet.net/blog

備註：*為部落格網站

三、討論

本研究分析了 funP 推推王的使用概況，其結果與國外其他相關研究相較，在標記時序分析中，發現 funP 推推王愈接近週六、週日，使用者新增書籤的次數呈現減少的趨勢，與 Heymann, Koutrika, and Garcia-Molina (2008)分析 Delicious 的結果有顯著的不同。國外的使用者愈接近週六、週日，新增書籤次數呈現增加的趨勢，甚至是一週內新增次數最高的時段。如此使用型態不同之處可能來自生活文化上的差異，然而真正的原因則待進一步的研究探索。

本研究並發現使用者、標籤、網路資源等三大元素的使用頻率均符合齊夫定律，並且它們之間的六種關聯關係也皆符合冪次定律，此結果與 Delicious (Halpin et al., 2007; Li et al., 2008)、CiteULike、Connotea 以及 BibSonomy (Santos-Neto et al., 2007; Helic et al., 2010)等國外社會性書籤網站研究同樣遵循齊夫定律與冪次定律：少數使用者、少數標籤、以及少數網路資源主宰多數的書籤。

另外，在 Top 20 標籤與 Top 20 網路資源的統計中，本研究發現它們均與部落格有著密切的關聯：在 Top 20 標籤部份，使用者除了標記部落格的標籤之外，亦與其他相關詞彙搭配標記；在 Top 20 網路資源部份，使用者收藏的網路資源多以部落格網站為主，顯示出國內社會性書籤網站正扮演著傳播部落格資訊的媒介，而國外的相關研究也有類似的發現(Brooks & Montanez, 2006)。社會性書籤網站不僅可以利用標籤收藏部落格網站；部落格網站亦可透過標籤進行分類與管理。本研究的分析結果顯示：在 Web 2.0 群體智慧的环境下，社會性書籤網站與部落格網站正持續緊

密互動。

伍、結論

本研究藉由統計分析探索國內社會性書籤網站 funP 推推王的使用概況，並比較國內外書籤網站使用型態的差異。分析結果發現，在使用時序上，與國外社會性書籤網站相較，funP 推推王的週末使用率較低；使用者、標籤、網路資源等三大元素的使用頻率均符合齊夫定律；而使用者、標籤、網路資源三大元素關聯分析方面則均遵循冪次定律；從 Top 20 標籤與 Top 20 網路資源的統計中，則證實 funP 推推王與部落格網站有著緊密的標記連結。本研究之分析結果可為社會性書籤網站管理者參考：例如，內建分類標籤修訂時，可依使用頻率增刪標籤；可依共同標記分析重新規劃主分類與子分類的關係。

在研究限制方面，本研究僅針對 2010 年 7 月 27 日至 8 月 27 日一個月的書籤資料進行統計分析，無法完整代表 funP 推推王在其他時間的使用樣式；另外，礙於分析系統的軟硬體限制，亦僅能對 2010 年 7 月 27 日至 8 月 2 日一週的書籤資料進行使用者、標籤、網路資源等三大元素的細部分析。Golder and Huberman (2006)分析 Delicious 時，只採用四天的書籤資料，而後的社會性書籤相關研究的資料集則持續擴大(Korner et al., 2010)。資料集越大的分析當然越具有代表性，不過在資料集持續擴大的同時，也越來越需要大型資料探勘技術，甚至於計算社會科學(computational social science)的協助(Lazer et al., 2009)，才可能探索 Web 2.0 中大量快速產生的使用者產出內容(user-generated content)。

在未來研究方面，本研究認為社會性書籤網站與部落格網站的應用狀態的互動關係，以及部落格

網站內外的標籤標記模式分析,均是可行的研究方向(Brooks & Montanez, 2006);另外若能透過圖形理論與網絡分析的概念,將可具體地呈現出社會性書籤網站的社群架構,進一步探討 Web 2.0 群體智慧之下,使用者透過標籤與網路資源產生互動的社

群網絡、標籤之間的標籤語意呈現、以及網路資源之間的群集關係(Santos-Neto et al., 2007; Halpin et al., 2007)。

(收稿日期:2011年8月8日)

參考文獻

- funP 推推王官方網站 (2011)。funP 推推王。Retrieved August 4, 2011, from: <http://funp.com/push>
- 卜小蝶、張淇龍 (2009)。社會性書籤網站之使用者與標籤特性初探。《圖書資訊學研究》, 4(1), 1-26。
- 陳崇正 (2009)。應用網路書籤與 VSM 相似度演算法於強化實踐社群的形成。未出版之碩士論文, 國立中央大學資訊工程學系碩士在職專班, 中壢市。
- 曾姿婷 (2009)。運用時間序列分群於社會性標籤之研究。未出版之碩士論文, 國立交通大學資訊管理研究所, 新竹市。
- 鄧睿清 (2009)。以個人化標籤推薦系統探討網路標籤使用行為。未出版之碩士論文, 國立交通大學資訊學院碩士在職專班, 新竹市。
- 戴瑋 (2008)。應用社會化協同標籤於網路資源搜尋。未出版之碩士論文, 國立中央大學資訊工程研究所, 中壢市。
- Alexa (2010)。Alexa Internet - Website Information. Retrieved August 10, 2010, from: <http://www.alexa.com/siteinfo>.
- Brooks, C. H., & Montanez, N. (2006). Improved annotation of the blogosphere via autotagging and hierarchical clustering. *Proceedings of the 15th International Conference on World Wide Web*, 625-632.
- Cattuto, C., Schmitz, C., Baldassarri, A., Servedio, V., Loreto, V., Hotho, A., Stumme, G., & Grahl M. (2007). Network properties of folksonomies. *AI Communications*, 20(4), 245-262.
- Chi, E. H., & Mytkowicz, T. (2007). Understanding navigability of social tagging systems. *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia*, 81-88.
- Farooq, U., Song, Y., Carroll, J., & Giles, C. (2007). Social bookmarking for scholarly digital libraries. *IEEE Internet Computing*, 11(6), 29-35.
- Furnas, G. W., Fake, C., Ahn, L. V., Schachter, J., Golder, S. A., Fox, K., Davis, M., Marlow, C., & Naaman, M. (2006). Why do tagging systems work? *CHI Extended Abstracts on Human Factors in Computing Systems*, 36-39.
- Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198-208.
- Halpin, H., Robu, V., & Shepherd, H. (2007). The complex dynamics of collaborative tagging. *Proceedings of the 16th*

- International Conference on World Wide Web*, 211-220.
- Helic, D., Trattner, C., Strohmaier, M., & Andrews, K. (2010). On the navigability of social tagging systems. *IEEE International Conference on Social Computing/IEEE International Conference on Privacy, Security, Risk and Trust*, 161-168.
- Heymann, P., & Garcia-Molina, H. (2006). Collaborative creation of communal hierarchical taxonomies in social tagging systems. *InfoLab Technical Report 2006-10*, Computer Science Department, Stanford University.
- Heymann, P., Koutrika, G., & Garcia-Molina, H. (2008). Can social bookmarking improve web search? *Proceedings of the International Conference on Web search and Web Data Mining*, 195-206.
- Hoegg, R., Meckel, M., Stanoevska-Slabeva, K., & Martignoni, R. (2006). Overview of business models for web 2.0 communities. *Proceedings of GeNeMe 2006*, 23-37.
- Korner, C., Benz, D., Hotho, A., Strohmaier, M., & Stumme, G. (2010). Stop thinking, start tagging: Tag semantics emerge from collaborative verbosity. *Proceedings of the 19th International Conference on World Wide Web*, 521-530.
- Krause, B., Hotho, A., & Stumme, G. (2008). A comparison of social bookmarking with traditional search. *Proceedings of the IR Research, 30th European Conference on Advances in Information Retrieval*, 101-113.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, L.-A., Brewer, D., et al. (2009). Computational social science. *Science*, 323(5915), 721-723.
- Lerman, K. (2007). Social information processing in news aggregation. *IEEE Internet Computing*, 11(6), 16-28.
- Li, X., Guo, L., & Zhao, Y. (2008). Tag-based social interest discovery. *Proceeding of the 17th International Conference on World Wide Web*, 675-684.
- Liu, E., & Chang, Y. (2008). The learning opportunities of social bookmarking service: An example of funP. *WSEAS Transactions on Systems*, 7(10), 1196-1205.
- Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). Position paper, tagging, taxonomy, flickr, article, toread. In *Collaborative Web Tagging Workshop, 15th International World Wide Web Conference*.
- Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323-351.
- O'Reilly, T. (2007). What is web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*, 65, 17-37.
- Peters, I., & Stock, W. G. (2007). Folksonomy and information retrieval. *Proceedings of the American Society for Information Science and Technology*, 44(1), 1-28.
- Santos-Neto, E., Ripeanu, M., & Iamnitchi, A. (2007). Tracking usage in collaborative tagging communities. *Proceedings of Workshop on Contextualized Attention Metadata*.
- Schenkel, R., Crecelius, T., Kacimi, M., Michel, S., Neumann, T., Parreira, J., & Weikum, G. (2008). Efficient top-k querying over social-tagging networks. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 523-530.

- Yanbe, Y., Jatowt, A., Nakamura, S., & Tanaka, K. (2007). Can social bookmarking enhance search in the web? *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, 107-116.
- Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort*, Massachusetts: Addison-Wesley.
- Zollers, A. (2007). Emerging motivations for tagging: Expression, performance, and activism. *Proceedings of the 16th International Conference on World Wide Web*.