

# 基於 PageRank 與聲譽模型之社會性書籤雲端平台系統

施柏宏

國立交通大學資訊管理研究所

a770401362@gmail.com

楊千

國立交通大學資訊管理研究所

professor.yang@gmail.com

摘要

隨著網際網路的蓬勃發展，資料量愈趨龐大，但目前主流的搜尋引擎卻是單使用者的設計模式，導致無法提供人們進行協同合作過濾龐大的資訊。本研究之主要貢獻是基於雲端運算所開發的社會性書籤網站—MyBook 及 IE Toolbar，協助人們進行協同合作與網站篩選，進而分享、學習知識。本研究主要利用使用者的書籤資訊，計算使用者彼此間的興趣相似程度，並結合 PageRank 與聲譽模式，作為網站推薦機制及交友機制，透過一星期的系統實測，蒐集 52 位使用者的行為資訊，並從中探討相似度、聲譽值、PageRank 與引用次數彼此間關係、使用者角色及社會性書籤系統設計等相關議題。

關鍵詞：社會性書籤網站、社會性標記、協同式網頁搜尋、聲譽模型、雲端運算

## 壹、緒論

### 一、 研究背景與動機

隨著電腦的進步與普及，使得資訊快速的成長與爆炸，並且各大主流的搜尋引擎也紛紛崛起(如：Google、Yahoo)，使得使用者可以輕易搜尋到大量的相關網站，但是並非每一個網站的資訊都是使用者所需要的，仍需要透過相當多時間過濾龐大的網站，才能得到所需之資訊。若能透過使用者間的協同合作，來分享彼此間的搜尋經驗，可以大大降低過濾的時間，但是目前各大主流的搜尋引擎主要是以單使用者的設計模式，使得無法有效達到使用者間的分工合作(McNally et al. 2010; Morris et al. 2007)，雖然使用者現階段會使用即時通訊、電子信箱及電話等來進行協同合作的方式，但是其效果皆有其限度(Morris et al. 2007)，因此本研究希望基於雲端運算所開發的社會性書籤網站—MyBook 及 IE Toolbar，能夠達到協同合作及知識分享等互助行為。

雖然現在已經有許多相關的研究及提供協同合作的社會性書籤網站(如：del.ici.ous、funp 推推王、HEMiDEMi 黑米書籤)，但是目前的社會性書籤網站，鮮少依照個體使用者量身打造，給予符合使用者需求之推薦，而且也鮮少有與現有的搜尋引擎作結合，故較少以「個人」為出發點及「強調使用者彼此之間關係連結」的研究與實作，因此本研究將社群網站與書籤作結合，將每個資料夾視為一本書、每個書籤視為一個頁面，而每個使用者都視為是一個作者，期望能夠達到以強調個人為出發點及交友的社會性書籤網站，並透過結合聲譽模式與 Google PageRank 主動推薦適合每位使用者之書籤與朋友。

### 二、 研究目的

本研究基於雲端運算平台所開發的 IE Toolbar 與社會性書籤網站，希望能夠提供使用者一個協同合作的網路搜尋方法，並且在分享經驗的同時也能夠結交到興趣相投的好朋友，因此本研究目的如下：

- (1) 藉由新的社群書籤網站介面，透過群體智慧並搭配聲譽模式及 PageRank 之下，提供使用者更多的參考數值，以改善目前使用者搜尋網頁之速度。
- (2) 透過使用者提供書籤資訊(如：標籤)，計算出使用者彼此間興趣之相似度，注重每個使用者個體，依照不同使用者，給予適合的好友/書籤推薦。
- (3) 藉由 IE Toolbar 整合目前主流的搜尋引擎，提供使用者能立即對網頁標記、收藏等分享方式，讓使用者能更便利的進行協同合作與知識分享。

## 貳、文獻探討

### 一、 社會性書籤網站(Social Bookmarking Website)

#### 1. 書籤

書籤功能的出現，主要便是讓使用者將個人有興趣或瀏覽過的網站之 URL 儲存在個人的電腦中，讓使用者依興趣與習慣進行整理、分類，以供日後方便再次瀏覽。由此

可見，雖然目前已有許多主流的搜尋引擎(如：Google、Yahoo)提供搜索，但是使用者面對龐大的搜尋結果仍需要有所記錄與標記，因此書籤的概念是不可獲缺的一環。不過即使現在所有的瀏覽器皆有提供書籤的功能，但卻只能將書籤儲存在個人的電腦之中，意即僅能由使用該電腦的使用者才能瀏覽與使用，致使使用者無法隨時隨地使用這些書籤，也無法公開地與其他使用者進行交流與分享其資訊。

## 2. 社會性書籤網站定義與品質

早期的網際網路(即 Web1.0)，強調所有網路資源與資訊皆由特定的對象提供(如：某公司的網頁)，一般的使用者僅能透過搜索，被動式地接收資訊；但隨著 Web2.0 的觀念出現，愈來愈強調以一般使用者為導向，更強調一般使用者的參與度(OReilly 2007)。而社會性書籤網站便是應用 Web2.0 的概念，由一般使用者依照自己的背景、習慣、興趣及經驗等，進行分類與整理專屬於個人的書籤，之後使用者透過網際網路的互動與分享，除了共同過濾搜尋引擎所提供的大量資訊之外，更能從中找到興趣相同的使用者。

另外在社會性書籤網站中，最重要的關鍵便是社會性標記。其社會性標記，主要是指在網際網路的環境中，由使用者共同標記與分類資源(Trant 2009)，亦即由使用者依照自己的經驗與習慣，給予每個書籤相關的關鍵字詞，共同將書籤進行整理分類的機制。雖然一般大眾對於網站的標記無法像專家般的精確，也可能會出現標記內容與網站標題或網域名稱相同的情況，但是社會性標記儼然已成為一種趨勢(Gupta et al. 2010)，而且共同標記非常能夠顯示出使用者的需求與認知，只是應要對不同的使用者，給予不同的信賴程度，以鑑定使用者的標記品質，解決其使用者知識水平不同的問題，所以本研究系統主要針對使用者間進行相似度分析與計算其使用者聲譽值，並同時利用 Google PageRank 計算網站之重要程度，以精確排序推薦之結果予以使用者。

## 二、 網頁搜尋(Web Search)相關研究

### 1. 相似度排序法

在過去已經有許多研究，提出關於相似度排序法之模型，而相似度排序，顧名思義就是計算關鍵字與網頁之間的相似程度，若相似度愈高則排名愈高；反之則愈低。但隨著技術的演進，已有許多研究，不僅僅只透過關鍵字來計算相似程度，而是開始藉由額外的網頁元數據(Metadata)來強化相似度計算之精確，例如：網頁標題、超鏈結以及使用者的檢索記錄。幸運的是，隨著 Web2.0 的觀念出現，以及近年來的社群網路概念，愈來愈強調由一般使用者進行協同合作，共同標記分類大量的網頁資源，因此社會性標記成為了新的元數據資源。目前有許多學者提出相關之研究，欲透過使用者所提供之標籤資源，更加強化相似度計算之精確度。例如：SocialSimRank(Bao et al. 2007)。

本研究採用 Haesung et al. (2011)所提出的文件相似度計算方法，此方法基於社群網路、圖形理論以及情境感知運算(Context-aware)之架構，故本研究想透過此方法欲計算出使用者與社交圈朋友之間標籤的相似程度，即計算使用者之間是否具有相同之興趣，進而達到好友推薦及書籤推薦，其詳細方法將於系統架構中介紹。

## 2. 靜態排序法

自從 Brin 和 Page 在 1998 年提出 PageRank 之概念後(Page et al. 1998)，許多研究便開始朝向網頁的靜態排序法邁進。PageRank 主要是根據網頁之間的超連結架構，來計算網頁的重要程度，並加以排序。

雖然從過去到現在有許多學者提出不同的運用(Richardson et al. 2006; Bao et al. 2007; Chia-Hao et al. 2008; Haesung et al. 2011)，如表 1。但是目前最知名且品質高的演算法仍是 Google PageRank。雖然我們無法得知 Google 如何區分正常的連結和不正常的連結交換，但是幸運的是我們仍然可以透過 Google 取得所有網頁的 PageRank 值，因此本研究採用 Google PageRank 演算法作為排序網頁推薦之結果。

表 1 靜態排序法之文獻整理

| 年份   | 作者                            | 方法   |
|------|-------------------------------|--|
| 2006 | Richardson,Prakash, and Brill | fRank：利用網頁的內容、使用者點擊率面向，計算網頁之重要性。   |
| 2007 | Bao, Wu, Fei, Xue, Su and Yu  | SocialPageRank：類似 PageRank 之作法，不同在於加入社會性標記元素，認為受歡迎之網頁，通常會被活躍的使用者標記許多熱門的標籤，因此具有較高之品質。 |
| 2008 | Lo,Peng,and Chiang            | RecommendationPageRank：採用的是 PageRank 之方法，但是加入社會性書籤概念及標記元素，利用使用者與標籤之間的關係，進行推薦網頁。      |
| 2011 | Lee and Kwon                  | GPR(V,E)：採用 PageRank 方法，但加入大眾分類法之元素，採用無方向性的加權重圖形理論計算網頁之重要性。                          |

## 三、 雲端運算(Cloud Computing)

簡單的說，雲端運算是將所有的運算資源與儲存資源都架設於網路之中，所有的運算及儲存的功能皆在網路上進行處理，因此在這樣的架構之下，人們不需要一台強大的電腦，僅需要一台能夠連接雲端運算服務的裝置，就能進行許多工作。而人們一貫以雲朵的方式來代表網際網路，因此雲代表著網際網路；而端代表的是任何一種可以連接至雲端運算服務之裝置，因此端代表著連結之裝置，故雲端運算因此而命名(Maggiani 2009)。

雲端運算根據使用者的需求，提供三種雲端的拓模供使用，分別為公有雲(Public Cloud)、私有雲(Private Cloud)以及混合雲(Hybrid Cloud)。公有雲主要是第三方提供服務給其他使用者作使用，因此非常適合於服務提供商使用；私有雲則是所有運算及儲存的軟硬體設備都在公司內部，並只提供給企業內部專屬使用(Armbrust et al. 2010);最後混合雲則是公有雲與私有雲的混合運用。又以服務類型作為分類時，可分為三個層級的服務，分別為 IaaS、PaaS 以及 SaaS(Vaquero et al. 2008)。

- (1) IaaS：雲端基礎設備服務的供應商，主要提供一個虛擬化的運算基礎設施。使用者只需要藉由網路的連結，就能享有與實體相同的伺服器與儲存裝置。使用

者不需要購買這些基礎設施，更不需要去維護這些設施，因此為使用者降低了許多的建置成本。

- (2) PaaS：雲端平台服務的供應商，主要提供一個程式開發平台、作業系統以及伺服器等資源，讓使用者同樣地只需要藉由網路的連結，就能享有這些平台環境，使用者不必再自行建置系統的平台或是作業系統。例如：本研究所採用的 Windows Azure。
- (3) SaaS：雲端應用服務的供應商，主要提供一個已開發的雲端運算軟體，而且這個軟體不需要使用者安裝在每台電腦上，使用者僅需要有可以連接至雲端運算服務之裝置(如：智慧型手機)，輸入一組可以辨識身份之認證碼(如：帳號、密碼)，便可能使用雲端運算軟體的所有相關服務。例如：本研究開發之系統 MyBook、Google Earth、Youtube、FaceBook。

## 參、系統架構

### 一、系統架構

由於本研究之目的希望能強化目前搜尋引擎的不足，提供使用者一個協同合作的網路搜尋平台來分享個人蒐集到的書籤。本研究的系統架構共分為三個階層，如圖 1 所示：

- (1) Level 1：為外部系統，使用者透過使用 IE 瀏覽各大搜尋引擎，並在各大搜尋引擎中進行搜尋網頁。
- (2) Level 2：為前端系統，分別為 IE Toolbar 與 MyBook。主要的特色有四點。第一，使用者可以自行在 MyBook 中，自行創建資料夾，來分類他們的各種書籤；第二，可以透過 IE Toolbar 即時將網站加入到 MyBook 中；第三，MyBook 除了提供好友機制外，還會主動比對使用者間相似度，進行好友推薦的動作；第四，MyBook 使用標籤結合 Google PageRank 的方式，來排序網站推薦之結果。
- (3) Level 3：為後端系統，本研究主要架設於 Windows Azure 的環境之中，後端系統分別為 Google PageRank 與 SQL Azure，前者主要用來取得網頁之重要程度之分數；後者則是資料庫，負責儲存所有相關資料。

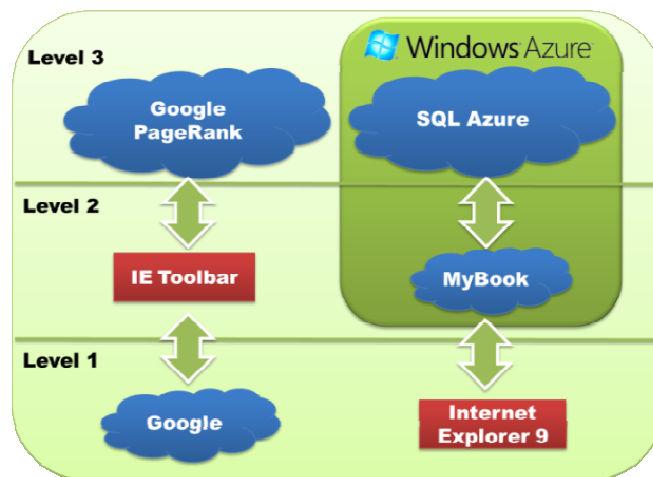


圖 1 系統架構圖

## 二、 系統流程

本研究系統流程共分為五個步驟：

- (1) 步驟一：計算使用者間的相似程度。透過計算使用者之間標籤的相似程度，以辨別是否與使用者為興趣相同的人，而成為好友候選人，其詳細方法將於下一小節介紹。
- (2) 步驟二：計算使用者的聲譽值。因為即使與該使用者興趣彼此相似，也未必是個有經驗或優良的使用者(有可能是惡意的使用者)，因此透過聲譽值的過濾，以找出更加適合的推薦對象，其詳細方法將於下一小節介紹。
- (3) 步驟三：排序好友推薦之結果。主要利用步驟一、二之方法，首先將好友候選人依相似度遞減排列；次之再由聲譽值遞減排列，再將其好友推薦結果呈現給使用者。
- (4) 步驟四：計算網頁之 PageRank 值。因為本研究認為若一個網頁若被許多網頁所連結，則此網站相對來說會比較容易受到歡迎與注意，也有較大的機會適合於使用者，因此透過 PageRank 作為書籤品質判斷之依據，其詳細方法將於下一小節介紹。
- (5) 步驟五：排序書籤推薦之結果。由於網頁之推薦主要是藉由使用者社交圈朋友中的書籤。當使用者與好友間具有標記相同之標籤時，則將具有相同標籤的書籤(且使用者目前所沒有之書籤)，依據 PageRank 值遞減排列，進行書籤推薦，並在此結果之後附加該好友前十大 PageRank 值之書籤；若當使用者與好友間不具有任何相同之標籤時，則直接依據 PageRank 值遞減排列，進行前十大 PageRank 值之書籤推薦。

## 三、 推薦機制

### 1. 使用者相似度之計算

在直覺上的想法，本研究認為當使用者彼此的標籤相似度愈高，表示使用者彼此的興趣愈相近，則成為好朋友的機會就大幅提升。因此本研究採用 Lee et al. (2011)所提出的文件相似度計算方法，這個方法是採用加權重圖形理論，可表達為  $Gsim(V,E)$ ，主要用來顯示某使用者與社交圈中所有朋友的相似程度，如圖 2 所示。而為了要計算使用者間的相似程度，此方法首先利用皮爾森相關係數(Pearson correlation coefficient)，透過彼此擁有相同的標籤，計算出使用者與社交圈中的朋友之相似程度，如公式(1)所示。

$$\gamma_{ij} = \frac{\sum_k (s_{ik} - \bar{s}_i)(s_{jk} - \bar{s}_j)}{\sqrt{\sum_k (s_{ik} - \bar{s}_i)^2 \sum_k (s_{jk} - \bar{s}_j)^2}} \quad (1)$$

在上述公式中， $\gamma_{ij}$ 為使用者  $i$  與使用者  $j$  相似度之值； $s_{ik}$ 則代表使用者  $i$  標記某網頁時，使用關鍵字  $k$  的次數；同理， $s_{jk}$ 代表使用者  $j$  標記某網頁時，使用關鍵字  $k$  的次數；而 $\bar{s}_i$ 為使用者  $i$  的平均標記次數；同理， $\bar{s}_j$ 亦表示使用者  $j$  的平均標記次數。

經過使用者間的相似度計算之後，採用 Dijkstra 之演算法，從中挑選出相似度(即權重)最高之社交路徑到達下一個使用者(節點)，再由之後的使用者重覆上述之計算，最後即可選出單一最大權重之社交路徑。

以使用者  $C$ (節點  $C$ )為例，我們可以選出的單一最大權重路徑為 $\{C \rightarrow B \rightarrow A \rightarrow R \rightarrow D \rightarrow T\}$ ，此  $B$ 、 $A$ 、 $R$  以及  $T$  所收藏的書籤，使用者  $C$  感興趣的可能性就非常高，故本研究會將使用者  $B$ 、 $A$ 、 $R$  以及  $T$  的書籤推薦給使用者  $C$ ；另外使用者  $D$  原先並不在使用

者  $C$  的社交圈當中，但是經過相似度計算之後，使用者  $D$  與使用者  $C$  的興趣可能相近，因此使用者  $D$  所收藏之書籤可能會適合於使用者  $C$ ，故會將使用者  $D$  推薦給使用者  $C$ 。

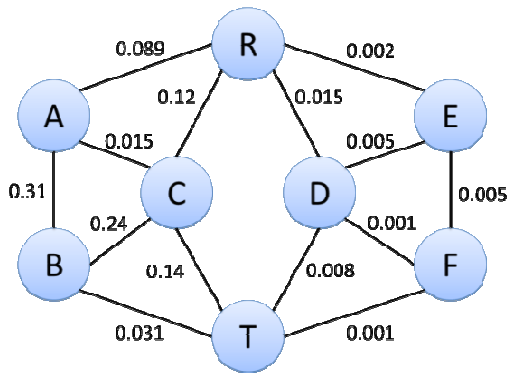


圖 2 加權社群網路

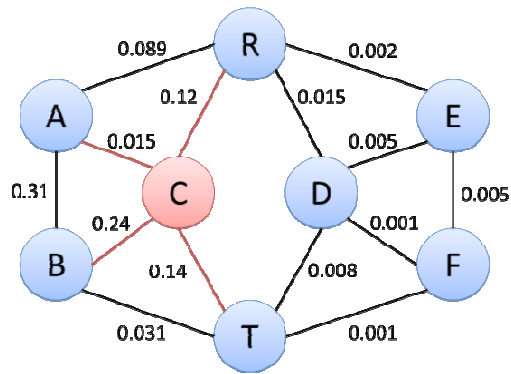


圖 3 以使用者  $C$  為例

## 2. 使用者聲譽值之計算

在過去的研究當中，McNally et al. (2010) 提出的使用者聲譽模型便認為有經驗的使用者其推薦之網站結果，應該要排列在無經驗的使用者之前，並且該研究指出聲譽模型除了能避免有惡意的使用者來破壞系統的品質之外，還能激勵使用者未來對系統有更多的貢獻。因此本研究基於 McNally et al. (2010) 所提出的聲譽模型，進行小幅度修改應用在本研究上。此聲譽模型，將使用者區分為生產者(Producer)與消費者(Consumer)，在本研究中，生產者意指將書籤推薦給系統中其他人的使用者；而消費者意指接受生產者推薦之書籤的使用者。當消費者接受生產者之書籤推薦時，則系統便會給予生產者聲譽值。

但是生產者擁有的同一個書籤可能會有許多的消費者，相對來說接受書籤推薦的這些消費者，某種程度上對於書籤篩選的貢獻也具有一定的程度，因此將其聲譽值平均分配給生產者以及所有的消費者。因此若當時間為  $t$  時，有  $k$  個消費者接受過此書籤之推薦，則生產者與  $(k-1)$  個消費者可獲得  $1/k$  聲譽值。如圖 4，最後在這個情境中，可以發現  $u_1$ 、 $u_2$ 、 $u_3$  以及  $u_4$ ，依序分別共獲得  $\frac{17}{6}$ 、 $\frac{5}{6}$ 、 $\frac{1}{3}$  以及 0。

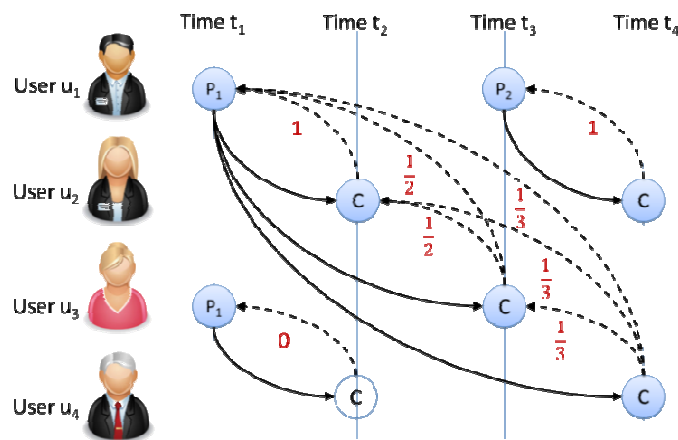


圖 4 聲譽值計算之情境應用示意圖

### 3. 網站推薦機制

本研究採用 Google PageRank 演算法，來作為網頁排序的基礎，透過 Google 計算出使用者書籤的 PageRank 值，進而排序其書籤之推薦結果。Google PageRank 將網頁彼此之間的超連結關係視為是一種投票機制，並透過此超連結關係來計算網頁之重要程度，其值介於 0 至 10 之間。例如：Google 將  $P_1$  網頁連結到  $P_2$  網頁的超連結，視為  $P_1$  網頁投票至  $P_2$  網頁。Google 會根據  $P_1$  網頁的上游網頁以及  $P_2$  網頁的等級，來決定新的 PageRank 值。簡單的說，Google PageRank 傾向採用 HITS(Hyperlink-Induced Topic Search) 方式，認為一個網頁若被許多網頁所連結，則其重要程度便會提高，並且一個 PageRank 值高的網頁，若連結至 PageRank 值較低的網頁，會有助於提高低等級網頁的 PageRank 值，而 PageRank 的計算方式為公式(2)所示，詳細內容請見 Page et al. (1988)。

$$PageRank(P_i) = \frac{1-d}{N} + d \cdot \sum_{P_j \in E(P_i)} \frac{PageRank(P_j)}{OutLink(P_j)} \quad (2)$$

在上式公式中， $P_i$ 代表所要計算之網頁； $N$ 代表網頁頁面的總數量； $d$ 為阻尼係數(damping factor)，目的用來估算網頁可能被使用者放入書籤的機率，值為  $d = 0.85$ ； $E(P_i)$ ，則代表有超連結至  $P_i$ 之網頁；而  $OutLink(P_j)$ 則代表  $P_j$ 連結至外部網頁的超連結數量。

## 肆、系統實作與說明

### 一、 BLOB

BLOB 是 Windows Azure 提供的儲存服務。其中一種為 BLOB Container，主要是用來收納 BLOB 物件之用，簡單的說，可以將 BLOB Container 視為我們一般在使用 Windows 時的資料夾。本研究使用 BLOB Container 主要是用來儲存使用者的大頭照片以及資料夾圖片，當使用者上傳照片時，便會新增屬於該使用者的 BLOB Container，儲存使用者的所有照片與圖片，其語法如圖 5。

```
private void SaveToBlob(string fileName, string contentType, byte[] data)
{
    var account = CloudStorageAccount.FromConfigurationSetting("連結字串");
    var client = account.CreateCloudBlobClient();
    var container = client.GetContainerReference("使用者帳號");
    container.CreateIfNotExist();
    var permissions = container.GetPermissions();
    permissions.PublicAccess = BlobContainerPublicAccessType.Container;
    container.SetPermissions(permissions);
    var blob = container.GetBlobReference(fileName);
    blob.Properties.ContentType = contentType;
    blob.UploadByteArray(data);
}
```

圖 5 BLOB Container 語法



## 二、 系統實作畫面

### 1. IE Toolbar

本研究為了提供使用者一個協同合作的網路搜尋方法，因此藉由開發 IE Toolbar，與目前主流的搜尋引擎作整合，使用者僅需輸入帳號與密碼即可對網頁進行收藏，讓使用者能夠更便利進行協同合作與知識分享，如圖 7(a)。

### 2. MyBook

在使用者個人資訊的部份，如圖 7(b)，若使用者想讓其他使用者更瞭解自己，可於「我的布馬克」中的「個人資訊」選項，進行編輯個人資訊的動作，但是在使用者興趣的部份，則是由系統根據使用者所使用的標籤來自動產生，故使用者不需要進行編輯。另外在「個人資訊」中，也提供顯示使用者個人的聲譽值。

在資料夾的部份，如圖 7(c)，使用者可依照自身的需求或習慣，創建不同名稱的資料夾，來分類所加入的新書籤。更值得一提的是，為了要使每個使用者都成為作者，使其書籤資訊更具有個別使用者的獨特性以及創造性，MyBook 提供使用者變更資料夾之圖像顯示，使其使用者彷彿為著書之作者

在個人書籤資訊的部份，在創建完資料夾之後，點擊資料夾圖片，便可進入資料夾。MyBook 提供使用者六種方式新增書籤資料至系統中：(1)透過 IE Toolbar 直接加入至系統中、(2)使用上傳個人電腦內「我的最愛」之書籤(副檔名為.url)、(3)藉由手動方式，輸入書籤之相關資訊、(4)運用搜尋方式、(5)從好友的書籤加入以及(6)系統推薦方式，如圖 7(e)等方式。無論採用上述何種方式，都能順利將書籤新增至 MyBook 系統當中，新增成功之後，會如圖 7(d)，系統會顯示其書籤標題、書籤網址、引用人數、標籤、更新時間以及 PageRank 值，以供使用者參考。

另外 MyBook 與一般的書籤網站最大的不同，便在於融合了現下社交網站的運作模式，並提供三種交友模式：(1)透過使用者自行加入成為好友、(2)確任其他使用者的申請加入以及(3)透過 MyBook 推薦進而成為好友，如圖 7(f)。當建立起自我的社交圈之後，可以在標題列「我的布馬克」中的「我的朋友」，查看自己的社交圈；另外亦可透過標題列「首頁」中的「好友近況」，快速獲得好友近期書籤收集之近況。

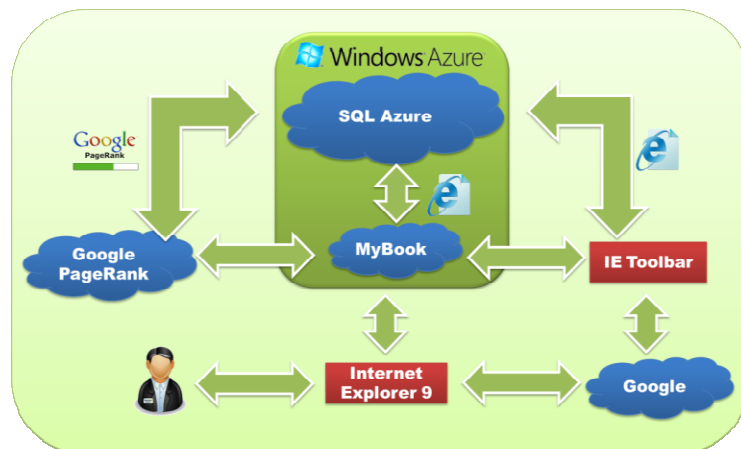
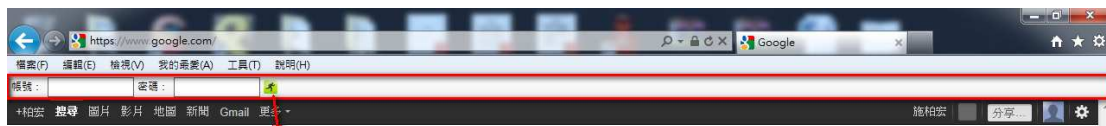


圖 6 系統運作流程圖



(a)

MyBook 歡迎 9934501! [登出]

首頁 我的布馬克 好友/書籤推薦 搜尋 搜尋

施柏宏 8.5 ← 使用者個人聲譽值 編輯個人資料

興趣

[fun] [Google] [NCTU] [new] [交通大學] [台鐵] [圖書館] [搜尋引擎] [時刻] [查詢] [火車] [網誌]

(b)

MyBook 歡迎 9934501! [登出]

首頁 我的布馬克 好友/書籤推薦 搜尋 搜尋

ASP.NET 4.0 技術開始

C:\Users\userID\瀏覽...

更新 刪除 取消

藏書-NCTU

浩然圖書館

網址: <http://www.lib.nctu.edu.tw/>

引用: 1次

標籤: 交通大學, 圖書館

更新: 3/10/2012 10:05:40 AM

PageRank: ★★★★★☆☆☆☆

自訂標題  
書籤網址  
載自目前加入的人數  
標籤  
更新時間  
書籤之重要程度(分數)

編輯 刪除

(c)

(d)

MyBook 歡迎 9934501! [登出]

首頁 我的布馬克 好友/書籤推薦 搜尋 搜尋

相似度: 100%  
聲譽值: 1

書籤推薦

好友推薦

amilyn

相似度: 0%  
聲譽值: 5.5

加入 黑名單

Yuling

相似度: 0%  
聲譽值: 5

加入 黑名單

(e)

(f)

圖 7 系統實作畫面(a) IE Toolbar、(b)~(f)為社群書籤雲端平台系統

## 伍、實驗結果

### 三、 資料來源

本研究針對國立交通大學 52 名學生進行為期一個星期實驗，期間可自由使用本研究之系統，其中 26 名為男生；剩餘 26 名為女生。男生平均年齡約為 20 歲；而女生平均年齡約在 19 歲左右。實驗之所有對象均為資訊相關之科系，因此相當熟識於電腦之使用，並且均接受過本研究系統操作之訓練，最後以問卷從受測者方得到實驗之回饋。

### 四、 實驗結果

整體而言，為期一星期之系統測試，受測者共增加 134 筆書籤、228 筆標籤，若去除使用者間重覆之書籤則共有 89 種網址；而去除重覆之標籤則共有 131 種標籤。

#### 1. 聲譽值、PageRank 與引用次數

首先關於使用者聲譽值超過 10 者，僅有 11 位使用者，如圖 8。另外在問卷的部份顯示，總體而言，60% 的使用者對於較不熟悉之主題，會由於 PageRank 之值較高，而加入此書籤。若以性別的面向來看，男性使用者會考慮 PageRank 值的有 50%，而不考慮者亦有 50%，故男生較無顯著之趨勢；但以女性使用者來說，會考慮 PageRank 值的有 69.23%，而不考慮者僅有 30.76%。

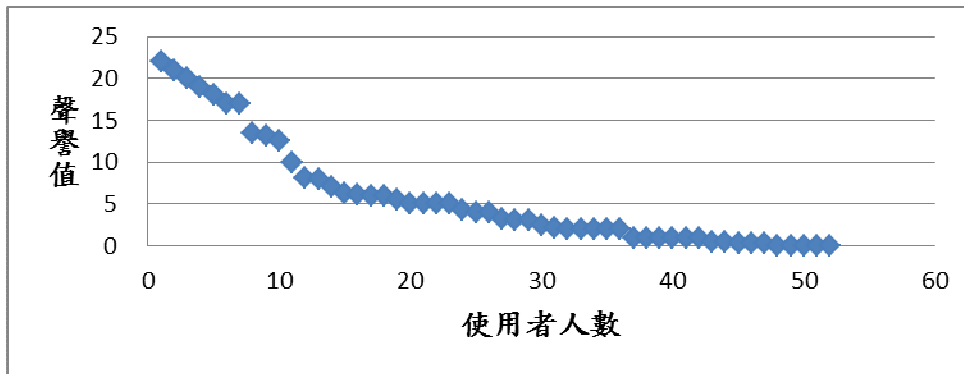


圖 8 使用者聲譽值之情況

#### 2. 生產者與消費者

本研究在此發現，大多數的使用者較傾向於消費者之角色，即大多數的使用者較偏向觀看與加入其他使用者之書籤資料，來增加自身的書籤，如圖 9。簡單地說，就如同目前網際網路中之論壇，往往回覆者(消費者)總比發文者(生產者)多，甚至回覆者之人數是微乎其微。更值得一提的是，女性使用者傾向於生產者的比例相當低落，甚至出現 0% 的完全生產者，因此數據中可發現，大部份擔任生產者角色的還是以男性使居多。

#### 3. 相似度與聲譽值

整體使用者而言，如圖 10，有 59.61% 的使用者會選擇依照相似度的高低來選擇好友；而有 40.38% 的使用者則會選擇依照聲譽值的高低來選擇好友。雖然大部份的使用者會依照與其他使用者興趣之相似度來選擇好友，但是聲譽值之部份，也將近有一半的使用者作為其考慮之依據，此外聲譽值與 PageRank、引用次數等為正向關係，而根據上節所述，60% 的使用者會去加入 PageRank 值較高之網頁，因此本研究認為聲譽值與相似度是屬相輔相成之面向，二者皆有助於進行社交，促進使用者之好友數量。

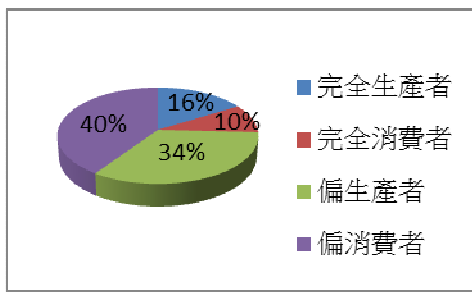


圖 9 整體使用者類型比例

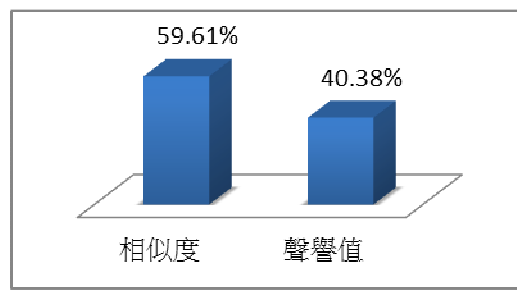


圖 10 選擇好友的依據比例

#### 4. 系統優點與建議

本研究將其整理，主要可分成七個部份，分別為：系統操作的易用性、個人書籤介面(如：資料夾功能)、社群方式呈現書籤(如：書籤上傳至雲端、可與好友進行互動、得知好友訊息)、系統推薦好友之功能(依照個人興趣進行推薦相符合之好友)、系統推薦書籤之功能(依照個人興趣進行推薦相符合之書籤)、系統社交功能(如：好友機制、聲譽機制)以及 IE Toolbar。而這七項被認為是優點的使用者人數百分比分別為：51.92%、59.61%、61.53%、30.76%、42.30%、44.23%以及 80.76%；而被提出改善意見的使用者人數百分比分別為：48.07%、40.38%、42.30%、32.69%、28.84%、50%以及 19.23%。

最值得一提的是，有八成(80.76%)的使用者認為本研究開發之 IE Toolbar 是本系統最大的優點，其中 78.84%的使用者認為能夠使用 IE Toolbar 結合現有的搜尋引擎及瀏覽器，能夠有助於他們快速且便利的將書籤上傳至雲端當中；而有 46.15%的使用者認為 IE Toolbar 之功能，能夠增加他們使用社會性書籤平台之意願。

另外系統社交功能的部份也相當值得注意，有 50%的人認為最需要作為改進的部份，主要的原因在於隱私權。因此建議系統可以提供書籤權限之設定、資料夾隱藏等功能。本研究亦認為這也是未來研究需要發展之方向，因為任何個人資料一旦放置於雲端，等同於將資料曝露於網際網路，若無適當之做法，容易使得個人的隱私外流。

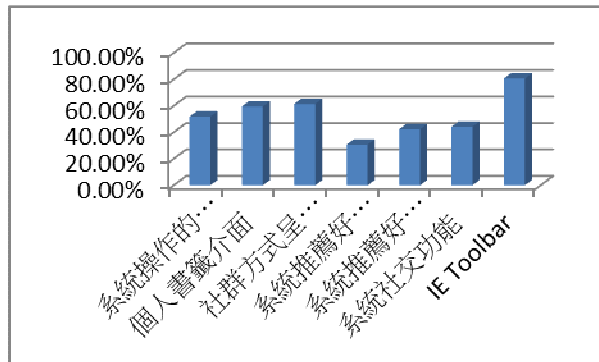


圖 11 系統功能優點之整體使用者分佈

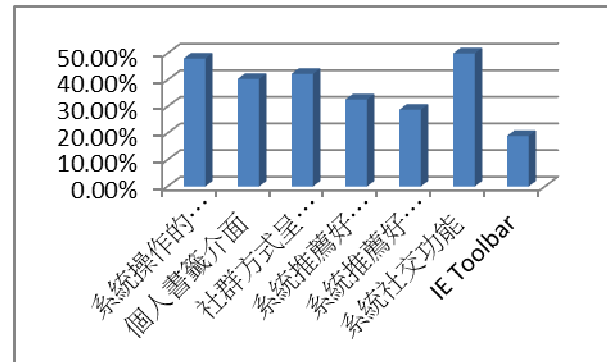


圖 12 系統功能建議之整體使用者分佈

## 陸、結論與未來展望

### 五、 結論

本研究透過 52 受測者作為社會性書籤平台系統實驗，本研究發現了以下幾點結論：

- (1) 聲譽值、PageRank 與引用次數為正向關係：使用者對於較不熟悉之主題，會由於 PageRank 之值較高，而加入此書籤。若使用者多加篩選網站之品質，並以 PageRank 作為網頁排名之基準，能夠有效的提高網站之引用次數，並能提高個人之聲譽值。
- (2) 多為消費者的角色：大多數的使用者，較傾向於消費者之角色，即大多數的使用者較偏向觀看與加入其他使用者之書籤資料，來增加自身的書籤資訊居多。
- (3) 相似度與聲譽值相輔相成：大部份的使用者會依照與其他使用者興趣之相似度來選擇好友。但是聲譽值與 PageRank、引用次數等為正向關係，因此聲譽值與相似度是屬相輔相成之面向，二者皆有助於進行社交，促進使用者之好友數量。
- (4) 個人化、互動多樣化、便利性：在社會性書籤系統設計上，使用者認為要能夠打造屬於個人之風格(例如：資料夾分類、資料夾圖示編輯)，並且能結合社群網路之概念，讓其使用者間能夠有許多的互動性(例如：留言、按讚)，此外更重要的是要能夠讓使用者便利的上傳書籤至雲端(例如：IE Toolbar 結合搜尋引擎)。

## 六、 未來展望

本研究主要是基於改善目前主流搜尋引擎的單使用模式，而開發社會性書籤系統協同合作平台與 IE Toolbar，與目前的搜尋引擎相互結合運用，並從中探討使用者行為與系統設計之優劣。建議未來之後續研究可朝下列方向發展：

- (1) 本研究所建構的社會性書籤系統，只考慮了使用者間的相似度與網頁的 PageRank 值，未來的研究可以考慮更多元數據(例如：引用次數、讚的次數、留言內容)。
- (2) 本研究僅考慮了使用者之間書籤互動性來增加聲譽值，未來的研究可加入其他不同互動方式，以增加使用者經營之意願，建立出具有高度互動性之社會性書籤系統。
- (3) 本研究尚未考慮到標籤的同義字詞之議題，未來的研究可以針對標籤的同義字詞進行研究，以精確的計算使用者彼此間興趣的相似程度。

## 柒、參考文獻

1. Agichtein, E., Brill, E., and Dumais, S. "Improving web search ranking by incorporating user behavior information," in: *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, Seattle, Washington, USA, 2006, pp. 19-26.
2. Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. "A view of cloud computing," *Commun. ACM* (53:4) 2010, pp 50-58.
3. Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., and Su, Z. "Optimizing web search using social annotations," in: *Proceedings of the 16th international conference on World Wide Web*, ACM, Banff, Alberta, Canada, 2007, pp. 501-510.
4. Chia-Hao, L., Wen-Chih, P., and Meng-Fen, C. "Ranking Web Pages from User Perspectives of Social Bookmarking Sites," *Web Intelligence and Intelligent Agent Technology*, 2008. WI-IAT '08. IEEE/WIC/ACM International Conference on, 2008.

5. Gupta, M., Li, R., Yin, Z., and Han, J. "Survey on social tagging techniques," *SIGKDD Explor. Newsl.* (12:1) 2010, pp 58-72.
6. Haesung, L., and Joonhee, K. "Improving Context Awareness Information Retrieval with Online Social Networks," *Computers, Networks, Systems and Industrial Engineering (CNSI), 2011 First ACIS/JNU International Conference on, 2011*, pp. 391-395.
7. Heymann, P., Koutrika, G., and Garcia-Molina, H. "Can social bookmarking improve web search?," in: *Proceedings of the international conference on Web search and web data mining*, ACM, Palo Alto, California, USA, 2008, pp. 195-206.
8. Maggiani, R. "Cloud computing is changing how we communicate," *Professional Communication Conference, 2009. IPCC 2009. IEEE International, 2009*, pp. 1-4.
9. McNally, K., O'Mahony, M. P., Smyth, B., Coyle, M., and Briggs, P. "Towards a reputation-based model of social web search," in: *Proceedings of the 15th international conference on Intelligent user interfaces*, ACM, Hong Kong, China, 2010, pp. 179-188.
10. Morris, M. R., and Horvitz, E. "SearchTogether: an interface for collaborative web search," in: *Proceedings of the 20th annual ACM symposium on User interface software and technology*, ACM, Newport, Rhode Island, USA, 2007, pp. 3-12.
11. O'Reilly, T. "What is Web 2.0: Design patterns and business models for the next generation of software,") 2007.
12. Page, L., Brin, S., Motwani, R., and Winograd, T. "The PageRank citation ranking: Bringing order to the web," *Stanford University, Computer Science Department Technical Report*,.
13. Peterson, E. "Beneath the metadata: Some philosophical problems with folksonomy," *D-Lib Magazine* (12:11) 2006.
14. Richardson, M., Prakash, A., and Brill, E. "Beyond PageRank: machine learning for static ranking," in: *Proceedings of the 15th international conference on World Wide Web*, ACM, Edinburgh, Scotland, 2006, pp. 707-715.
15. Schachter, J. "2006 Young Innovator. *Technology Review*.,") 2006.
16. Strohmaier, M., Körner, C., and Kern, R. "Why do users tag? Detecting users' motivation for tagging in social tagging systems," *International AAAI Conference on Weblogs and Social Media (ICWSM2010)*, Washington, DC, USA, 2010.
17. Trant, J. "Studying social tagging and folksonomy: A review and framework," *Journal of Digital Information*, 10(1) 2009.
18. Trant, J., and Wyman, B. "Investigating social tagging and folksonomy in art museums with steve. museum," *Proceedings of the WWW 2006 Collaborative Web Tagging Workshop*, 2006.
19. Vander Wal, T. "Explaining and showing broad and narrow folksonomies," 2005.
20. Vaquero, L. M., Roderer-Merino, L., Caceres, J., and Lindner, M. "A break in the clouds: towards a cloud definition," *SIGCOMM Comput. Commun. Rev.* (39:1) 2008, pp 50-55.

# The Cloud Computing of Social Bookmarking System Based on Google PageRank and Reputation Model

Po-Hong Shih

Department of Information Management  
National Chiao Tung University  
a770401362@gmail.com

Chyan Yang

Department of Information Management  
National Chiao Tung University  
professor.yang@gmail.com

## Abstract

While thriving development of Internet, amount of data is getting huge. Many search engines do not explicitly support collaboration during search and this let people can not collaborate to filter data. This paper develops a social bookmarking system based on cloud computing and IE toolbar to assist user collaboration from filtering websites. This paper supports users by harnessing the bookmark's metadata、users' similarity calculation、PageRank and Reputation Model as the base for websites and users recommendation. The experiment invites 52 users and they use this system about one week. The research findings show that reputation, PageRank and citation times are positive relationship. Personality and interactivity can enrich the content and enlarge the user interesting in this system.

Keywords: Social Bookmarking Website, Social Tagging, Collaborative Web Search, Reputation Model, Cloud Computing