

# 社會搜尋：應用社會網絡分析方法於網頁搜尋技術之研究

## Social Search: Applying Social Networks Analysis for Web Search Techniques

呂筱萱

國立高雄大學資訊管理學系

m0993309@gmail.com

丁一賢

國立高雄大學資訊管理學系

iting@nuk.edu.tw

王學亮

國立高雄大學資訊管理學系

slwang@nuk.edu.tw

### 摘要

近年來社群網站在人與人間的關係中扮演著重要的角色，隨著時間也讓人與人形成強連結的關係。對使用者來說，擁有強連結關係的朋友所提供的資訊，亦有高度興趣。目前網路上大多數搜尋平台是依據關鍵字和文章之相關程度，尚未加入文章擁有者與搜尋者間的關係，因此本研究將傳統搜尋引擎加入社會關係，預期可改善搜尋品質並提升搜尋者之滿意度。

本研究將透過 Facebook 之塗鴉牆資料作為社會搜尋之依據，接著進行 CKIP 詞庫小組處理和 TF-IDF 計算，最後結合字頻和社會關係並進行結果排名。因此，本研究會以真實狀況來驗證所提出之方法之成效。

關鍵字：搜尋引擎、社會網絡分析(SNA)、社群網站、社會搜尋(Social Search)。

## 壹、緒論(Introduction)

### 一、 研究背景與動機

目前在全球資訊網上的大量資訊中，大多都是非結構化的資料，使用者可利用搜尋引擎來進一步從大量資訊中發掘欲知的資訊。在使用搜尋引擎的過程中，首先使用者會先定義關鍵字並透過搜尋引擎得到相關資訊。在傳統搜尋引擎技術中，這些資訊都是透過關鍵字與內容的相關(相似)程度或透過網頁排名作分析，再利用分析所得的相關程度或網頁排名進行排序。若關鍵字與網頁內容之相關程度或網頁排名越高，則表示該內容越有可能是使用者所需之答案，這些內容則會在搜尋引擎結果中排名位居前面之位置。

由於資訊量過大，目前的搜尋引擎技術只與使用者所定義之關鍵字的相關文章進行排序並回傳結果給使用者，然而大多時候搜尋引擎所回傳之資訊需要耗費使用者時間來瀏覽每個回傳的結果才能找到相關的答案。這種狀況的主因，是因為目前搜尋引擎只針對內容相關程度或是網頁排名進行分析，忽略了文章作者與搜尋者之間關係的考量。

在過去研究中發現，有些學者已針對這個問題進行探討，在搜尋引擎中加入與使用者本身相關的資訊，這類型的搜尋技術稱之為個人化搜尋(Personalized Searching Technique)。關於使用者本身之相關資訊大多是指透過分析使用者在線上瀏覽的行為資料，包括從瀏覽網頁的歷史紀錄，和點擊相關連結的紀錄檔兩方面來研究，並藉以改善搜尋引擎排名的效率。這種方法所面臨之最大考驗 是使用者需要有長時間使用網路之經驗，才能透過使用紀錄來分析並提供搜尋結果。

近年來由於社群網站的掘起使得人與人之間的連結更加緊密且多樣。以社會網絡(Social Networks)為例，人與人之間的連結度有從聚程度、緊密程度等等關係(Hanneman, 2005)。透過這些關係可瞭解使用者與其他人之間的連結程度，從中可得知使用者對於關係程度較好的人所推薦之相關人事物也會有高度信任或喜愛(Yang et al., 2011)。

Han 在 2011 年提出在社群網站中“讚(Like)”扮演著使用者之間的一個重要關係，使用者分享文章或是表示自己對於這篇文章有高度的興趣，可以透過“讚”來表達他們的支持(Han et al., 2011)。Han 利用社群網站中都有“讚”的機制之特性來改善搜尋品質，但目前研究中尚未針對“讚”的定義來做分析，只知道是使用者對於人事物推薦之結果，倘若能進一步利用社群網站特性來衡量使用者與使用者之間的關係，就可得知使用者是否對於某個訊息具有潛在的興趣或是喜好的關係存在。

綜合以上所描述的研究背景來看，若能將傳統的搜尋引擎技術加入分析文件作者與搜尋者之間的關係，並且在搜尋結果中顯示文件與使用者之間的關係，如此可提高使用者對於該文件之喜好進而點擊瀏覽，進而改善搜尋引擎並提高搜尋者之滿意度的目標。所謂文件關係，是指衡量使用者之間的關係來提高使用者對於該文件內容之興趣，以降低使用者在瀏覽網頁時間之外，亦可透過使用者的朋友在社群網站中之喜好或興趣來推測出使用者之喜愛為何。這就是本研究加入文件關係之主要原因。

為了分析文章與使用者之間的關係，本研究計畫從人與人之間的關係切入。透過 Web 2.0 的應用，使用者不再只是單向的接收資料，而是以提供資料的新角色在網路上

成長，這種角色的成長快速，讓線上使用者面對的不再只是機器頁面，進而開始與其他使用者有直接或間接接觸。當人與人接觸之後，就會開始有了互動，倘若這種互動關係延伸出更多的關係時，這種現象我們便稱之為是線上的社會網絡現象。

因此，本研究將在傳統搜尋引擎分析網頁內容基礎上結合社會網絡分析技術之期望可以提高搜尋引擎之效率與使用者使用經驗之提升，我們將這種新的搜尋技術稱之為社會搜尋(Social Search)。因此，本研究計畫之主題訂為“應用社會網絡分析方法於網頁搜尋技術之研究”。

## 貳、文獻探討

### 一、搜尋(Search)

由於線上使用者欲搜尋之資料不同和所使用搜尋方法也不盡相同，因此，本節將搜尋分為資料搜尋目的和資料搜尋步驟兩部份來看。

#### (一) 資料搜尋步驟

不同專家學者對於資料搜尋過程之步驟提出不同的看法，分為三種不同說法：(1) Kuhlthau 在 1991 年提出資料搜尋任務步驟為：開始搜尋(Initiation) → 確認主題(Selection) → 勘查探索(Exploration) → 規劃構想(Formulation) → 收集資料(Collection) → 呈現結果(Presentation)；(2) Marchionini 在 1995 年提出資料搜尋任務步驟為：確認問題(Recognize Accept) → 定義問題(Define Problem) → 選擇搜尋系統(Select Source) → 闡述疑問(Formulate Query) → 執行搜尋(Execute Query) → 審查結果(Examine Results) → 摘錄資料(Extract Info)；(3) Iwan 等學者則在 2005 年提出與上述不同的資料搜尋任務步驟：定義資料問題(Define the information problem) → 選擇資料來源(Select sources of information) → 搜尋資料(Search and find information) → 處理資料(Process information) → 組織並呈現資料內容(Organize and present information)。

從上述學者對於資料搜尋過程之步驟看來，可得知資料搜尋的過程並非直線式進行，都是需要透過部分步驟需反覆進行，一再確認直到獲得搜尋結果(Marchionini, 1995)。Marchionini (1995)所提到的資料搜尋步驟中，可得知搜尋前需先對問題或搜尋目的的方向清楚，接著進一步嘗試定義關鍵字、闡述詢問，再進行搜尋。本研究使用的是 Iwan 等學者(2005)所提出對於資料搜尋的定義，進行搜尋步驟。

#### (二) 目前搜尋遭遇之問題

目前搜尋平台種類繁多，但大多數的搜尋平台是使用文章中與使用者輸入之關鍵字作相似度計算，由於目前搜尋平台只單方面針對文章的相似度排序(Salton & McGill, 1983)，因此使用者必須點擊每一筆結果之連結來得知網頁內容是否為自己所需。近年來社群網站連結起線上人與人之間的關係(Tong, 2008)，Yang 在 2011 年提出興趣相同的兩個人會因共同的興趣而互為朋友。因此，本研究利用認識的人所分享之人事物對於搜尋者會有高度興趣之特性(Mislove et al., 2007)，並且傳回的結果讓使用者能得知朋友之喜好程度並提供使用者相關資訊 (Li & Lin, 2009; Brzozowski, 2008)。

## 二、社會搜尋 (Social Search)

過去針對搜尋加入社會關係之相關研究是很少的，目前現有之研究中發現以社會搜尋為基礎之應用方法有針對標籤(Zanardi & Capra, 2008)、“讚”(Han, 2011)等等。大量使用者投入社群網站，除了讓人與人之間關係更加緊密之外，線上人與人互動(Jiang, 2010)逐漸頻繁，便會產生大量互動資料。當其他使用者可透過搜尋來快速得到這些互動資料(Zinoviev, 2009)。因此，社群網站的關係(Baer, 2010)對於進行社會搜尋是重要的。因此，本研究將社會搜尋分別從社會關係和社會關係排名兩部份來看。

### (一) 社會關係 (Social Relationship)

社會關係早在社群網站出現就一直受到學者的關注，，例如：Facebook、Twitter、LinkedIn 等等，透過這些社群網站中每位使用者之朋友結構或其互動狀況來確定雙方之社會關係。社會關係不只是人與人之間的關係，其中還包括人與事物之間的關係和事物與事物之間的關係，以下針對過去學者研究中，針對社會關係所使用之方法，如下表 1 所示。

表 1 社會關係之相關研究

學者	方法
Han et al. (2011)	Facebook 中“讚”的機制
Yang et al. (2011)	興趣相同者
Gou et al. (2010)	文件或影音檔
Zanardi & Capra. (2008)	標籤(tag)
Velardi et al. (2008)	互動內容
Vieira et al. (2007)	搜尋 log 檔
Mislove et al. (2006)	超連結(hyperlink)

(資料來源：本研究)

Han 在 2011 年研究以 Facebook 中“讚”的機制作為朋友關係的依據，透過朋友與朋友之間或是朋友對於粉絲頁所給予的“讚”，得知朋友對於某個人或是粉絲頁是具有高度興趣或欲快速分享資訊給朋友，也可底透過“讚”的機制來完成，但是 Han 只針對“讚”的機制來作為關係強度之依據，如此之關係資訊實屬不足。Yang 在 2011 年表示在社群網絡中任意兩個人會因為興趣相同進而成為好友，但在同一興趣領域中的兩人卻不一定會經常性的互動。當互動資訊少時，其中一位使用者所喜愛的人事物之相關資訊，就無法提供另一位使用者在瀏覽資料的依據。

Velardi 在 2008 年提出以互動內容來定義朋友之間關係的權重，先利用 k-Means 將擁有共同興趣或喜好之使用者分類，再分別從分類好的子群中，探討各子群中的網絡關係，與 Yang (2011)所提出的概念相似。另外，尚有透過文件或影音檔(Gou et al., 2010)、標籤(Zanardi & Capra, 2008)、搜尋 log 檔(Vieira et al., 2007)、超連結(Mislove et al., 2006)等等作為任意兩人之間之關係依據。

### (二) 社會關係排名 (Social Ranking)

除了找出社會關係之外，尚有許多學者針對關係排名計算之研究，以下幾種方法是針對朋友之間的關係排名之計算，如表 2 所示：

表 2 社會關係排名之相關研究

方法	學者
MAS algorithm	Gou et al.(2010)
Cosine similarity	Zanardi & Capra.(2008)
Shortest paths	Vieira et al.(2007)
LHM algorithm	Leicht et al.(2006)

(資料來源：本研究)

在過去研究中，學者針對已建構好之社會網絡架構來計算彼此間相關程度並進行排名(Zanardi, 2008; Leicht, 2006)，Zanardi 等學者在 2008 年利用 Cosine similarity 來計算標籤與使用者的相關程度，但 Cosine similarity 的計算前提是在社會網絡中任意兩點必須是有連線的狀況下，才能進行運算；Leicht 等學者在 2006 年提出 LHM algorithm 來計算在社會網絡中的任意兩點只要有共同事物，便可計算使用者之間之相關程度，但 LHM algorithm 方法若要處理大規模的社會網絡結構時，其耗時相當高。Gou 等學者在 2010 年提出 MAS algorithm 的方法，結合 Cosine similarity 和 LHM algorithm 的方法，計算出在社群網站中已建立好的朋友架構，透過這個朋友架構可以計算在任意沒有連線的兩人，可以延伸 LHM algorithm，針對 MAS algorithm 的方法，下列作個詳細之說明，要得知在社會網絡中任意未連線之兩人之相似度，必須進行四步驟之計算：首先計算兩人各自在群體中之相似程度，接著計算群體與群體間之相似程度後，利用階層化方式顯示兩人在各群之位置，確定兩人之位置後，便可計算出兩人之相似程度，如圖 1 所示。

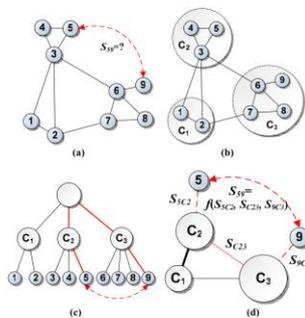


圖 1 MAS algorithm 計算步驟

(資料來源：Gou et al., 2010)

### (三) 目前搜尋遭遇之問題

綜合以上過去學者所作之相關研究，可發現大多數都以共同事物作為對於朋友關係的確定，也表示是在已知的社會網絡架構下所發展之研究與方法，但這些方法都未考慮雙方之間的互動狀況。因此，本研究所加入的社會關係中，除了利用按“讚”的內容之外，還加入使用者與朋友之間在塗鴉牆上互動的狀況，互動狀況會作為使用者與朋友之間互動權重，將互動權重與關係權重結合，預期提升搜尋結果之滿意度。

## 參、研究方法

在本章節會探討本研究之主軸—社會搜尋，主要所使用之相關技術與概念包括：TF-IDF、社會關係計算和社會關係排名之計算。

### 一、系統架構

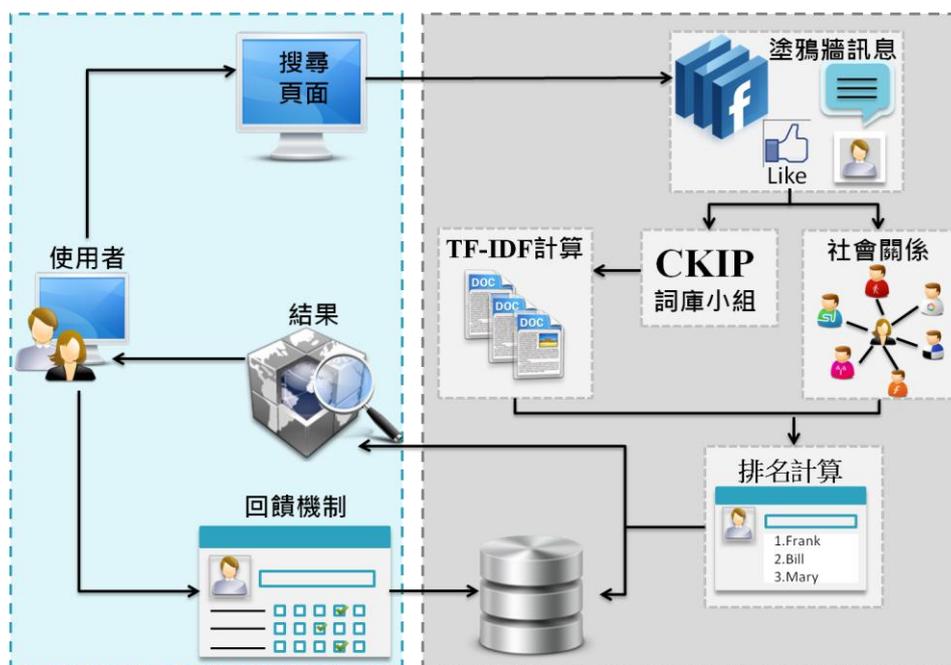


圖 2 本研究之系統架構

使用者可以透過本研究所提供之社會搜尋進行問題之查詢，透過一系列社會關係計算和 TF-IDF 計算處理後，使用者可以透過系統傳回之結果進行相關任務問題，針對結果是否符合使用者問題之答案之相關問題，來進行評分。當分數越高則表示本研究所提出之計算方法準確度越高；反之亦然。最後使用者評分結果會再傳回系統。

### 三、前置資料處理與分析

本研究主要探討在社群網站中使用者與朋友之間的互動關係，目前線上著名的社群網站有：Facebook、Twitter、LinkedIn 等等。根據 Alexa 統計調查網站 (<http://www.alexa.com/topsites/global>) 在 2011 年 11 月的調查報告中，Facebook 是在社群網站分類中是位居第一。

由於 Facebook 網站註冊的會員數眾多，因此 Facebook 對於每位使用者會希望以實名制來註冊，所以可確定每個帳號都是真人註冊的，也可得知人與人之間的關係連線是真實的。Viswanat 等學者在 2009 針對 Facebook 中之塗鴉牆內容作研究，並加入時間之因素來探討使用者在 Facebook 中的活躍狀況，其中作者發現在每位使用者之塗鴉牆上，只有在生日當天塗鴉牆之訊息量會突然爆增，但 Viswanat 等學者只針對塗鴉牆之訊息量而未加入訊息內容之考量。

礙於目前線上資料過於繁多，格式也不盡相同，線上亦無一套既定的方法，使得資料明顯呈現非結構化。因此本研究將以 Facebook 以 XML 格式之特性，作為社群網站資

料來源之基礎，探討朋友之間互動狀況，並將這種互動狀況稱之為互動權重，進行一系列的計算。

### (一) 資料擷取方式

主要利用 Facebook 所提供 API 為資料擷取方法—Graph API。透過 Facebook 的應用程式會員可允許擷取該會員的相關資料之機制，再進一步的擷取該會員之相關資料。因此，在 Facebook 的會員個資部份是不會隨意的擷取，並且保證個資擷取後並不公開，只單純作為學術研究之用。因此，在 Facebook 中的會員無須擔心自己的個資是否會被使用去進行違法之相關事務。

### (二) CKIP 中文斷詞技術

本研究關於詞彙擷取的部份，以中研院資訊科學研究所詞彙小組(CKIP)所開發的中文斷詞系統為工具，該系統結合詞庫式斷詞法及統計式斷詞法之優點的混合式斷詞法，將使用者所輸入之文章或文句自動斷詞後，再標示出每個詞的詞類標記。此系統分詞依據為此一詞彙庫及定量詞，重疊詞等構詞規律及線上辨識的新詞，並解決分詞歧義問題。舉例來說，「我下禮拜看福爾摩斯」，CKIP 將回傳：「我 (N) 下 (DET) 禮拜 (N) 看 (Vt) 福爾摩斯 (N)」，括號中的 N 代表名詞、DET 代表特指定詞、Vt 代表動詞。利用這些資訊，即可進行後續的詞彙分析。

### (三) TF-IDF 關鍵字計算

在 TF-IDF 關鍵字計算(Salton & McGill, 1983)中，透過上述 Facebook 所開發之應用程式之後，本研究主要擷取 Facebook 會員資料可分為二部份：一是會員之好友列表；二是使用者之間互動內容，即為 Facebook 之塗鴉牆，並且在塗鴉牆內容的部份進行 CKIP 和 TF-IDF 之計算。

透過 CKIP 斷詞斷字後，將進行詞頻(Term Frequency, TF)作計算，所謂的詞頻是指某一個給定的詞語在該文件中所出現之次數。 $n_i$  表示某個關鍵字; $n_j$  表示某一篇文章; $n_k$  則表示在一篇文章中所有關鍵字。

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

以上式子中  $n_{i,j}$  是某個詞彙在某文件  $d_j$  中的出現次數，而分母則是在某文件  $d_j$  中所有詞的出現次數之和。

接著進行逆向文件頻率 (Inverse Document Frequency, IDF) 計算，逆向文件頻率計算是指在總文件中，某一個給定的詞語所出現之次數。以上述例子來說，公式如(3,2)式。

$$idf_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (2)$$

其中  $|D|$  是指文件總數， $d_j$  是指第  $j$  份文章， $t_i$  是指第  $i$  個關鍵字。關於 IDF 的概念，首先將某個詞彙設為  $t$ ，如果包含  $t$  的訊息或文章數量越少，也就是指  $n$  越小，IDF 越大，則說明  $t$  具有很好的類別區分的能力。

最後將 TF 和 IDF 所求得之值相乘， $TF-IDF=0.05 \times 2.3025=0.1151$ ，公式如(3,3)式。

$$TF-IDF_{i,j} = tf_{i,j} \times idf_i \quad (3)$$

我們可以透過上述所求得的值得知，本研究所蒐集的訊息或文章數中所代表之關鍵字為何，接著每則訊息或文章會與下列社群關係排名計算結合。

#### 四、社會搜尋計算(Social Search Algorithm)

在社群關係排名計算中會分成三部份，分別為(1)社群關係計算；(2)排名引擎；(3)使用者回饋機制訂定。

##### (一) 社群關係 (Social Relationship) 計算

透過 Facebook 所蒐集而來之塗鴉牆內容列表，進一步作互動權重分析，以圖 3、圖 4 和圖 5 來說明使用者之間的權重分析之給定，並且以  $W(\text{User}_A, \text{User}_B, \text{Value})$  來表達使用者 A 對於朋友 B 的回應值。



圖 3 使用者 A 張貼訊息給使用者 B

以圖 3 來看，當使用者 A 在公開權限狀況下發布一則訊息，權重值則表示使用者 A 對於他所有的好朋友的權重值皆增加為一筆，其表示法為  $W(A, *, 1)$ 。假設使用者 A 只允許使用者 B 觀看自己所發布之訊息時，此時表示為  $W(A, B, 1)$ 。(其中\*表示在公開權限下張貼訊息或文章，因此所有人都可以看到。)



圖 4 使用者 B 回應使用者 A 訊息

以圖 4 來看，當使用者 A 在公開權限的狀況下，發布一則訊息，使用者 A 的所有好友看到此篇訊息而進行回應的動作，其權重值則表示使用者 A 的好友群對於使用者 A 的權重值皆增加為一筆，其表示法為  $W(*, A, 1)$ 。倘若今日為使用者 A 在限定使用者 B 觀看使用者 A 所發布之訊息時，使用者 B 回應了使用者 A 的訊息，此時表示法則為  $W(B, A, 1)$ 。



圖 5 使用者 A 與使用者 B 之關係

綜合圖 3 與圖 4 來看，可得知狀況為圖 5 表示使用者 A 和使用者 B 互為朋友之外，也可得知雙方是有互動的。

## (二) 社群排名引擎 (Ranking Engine)

結合以上二種計算權重的方法後，我們可得知在某篇訊息或文章中，互動權重值最高與所提及之關鍵字最多的是什麼。在內容互動權重值中，本研究分為兩部份來計算，一是針對“讚”的回應；二是針對塗鴉牆的內容回應。最後將關係權重與內容互動權重結合。由於目前尚未有研究針對“讚”與塗鴉牆內容的重要性比較，因此本研究將“讚”的機制與塗鴉牆內容之權重定義等同為各為 0.5，並且利用上述例子作說明：

本研究結合“讚”與塗鴉牆內容之回應狀況來作為社會關係排名的計算。本研究將此權重表示為下述公式(4)所示。

$$\begin{aligned} &W_l(\text{User}_A, \text{User}_B, \text{Value}) + W_m(\text{User}_A, \text{User}_B, \text{Value}) \\ &= W_{l+m}(\text{User}_A, \text{User}_B, \text{Value}) \end{aligned} \quad (4)$$

透過以上計算結果後，將關係權重和 TF-IDF 結合後，該數值表示在某篇訊息或文章某個人所提到的關鍵字。其公式如(5)式。

$$\text{Rel}_{A,B} = \alpha \cdot W_{l+m} \times \text{TFIDF}_j + \beta \cdot W_r \times \text{TFIDF}_j \quad (5)$$

其中  $W_r$  表示使用者 A 與使用者 B 之間的關係，目前將關係指訂定在是否為好友，因此  $W_r$  值為 0、1 的二元值， $\alpha$  值表示互動權重常數， $\beta$  值表示雙方是否互為好友權重常數。

$$\text{Rel}_{A,B} = 0.5W_{l+m} \times \text{TFIDF}_j + 0.5W_r \times \text{TFIDF}_j \quad (6)$$

或

$$\text{Rel}_{A,B} = (0.5W_{l+m} + 0.5W_r) \times \text{TFIDF}_j \quad (7)$$

最後依分數高低進行排序，其表示法如(8)式。

$$\text{Social Ranking} = \text{Rel}_{A,B} \quad (8)$$

## 肆、系統實作

本研究資料擷取範圍可分為以下二種：一是從使用者與朋友之間之回應內容，主要以塗鴉牆內容為主；二是從使用者的朋友列表。在實驗進行時，使用者會透過 SocialSearch 之實驗界面進行搜尋，如圖 6 所示。

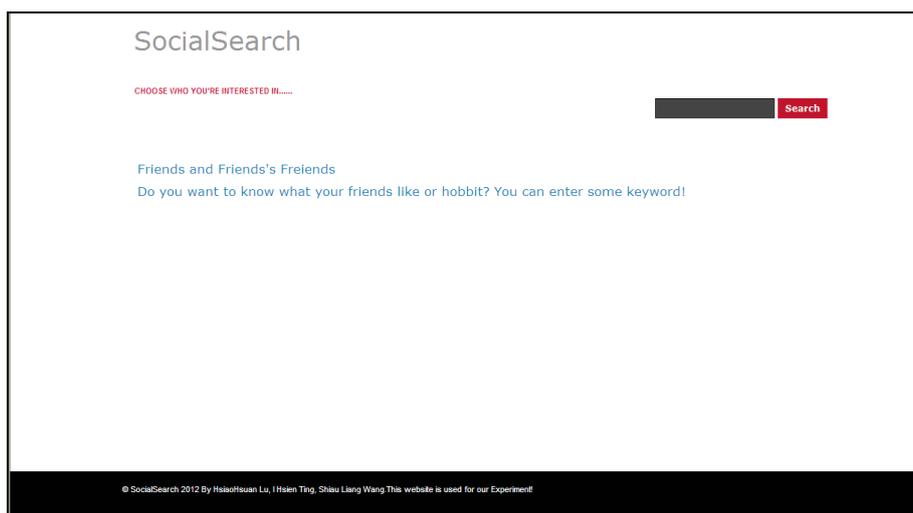


圖 6 本系統之 SocialSearch 界面

使用者可透過本研究所設計之問題進行查詢動作，並且透過傳回之結果進行分數給定，使用者之分數評判標準主要以結果是否為自己所需之資訊為主。

本系統所傳回之結果主要是透過 Facebook 中使用者所提供之塗鴉牆資料。在 Facebook 中使用者之間的回應狀況，以下圖 7 所示：

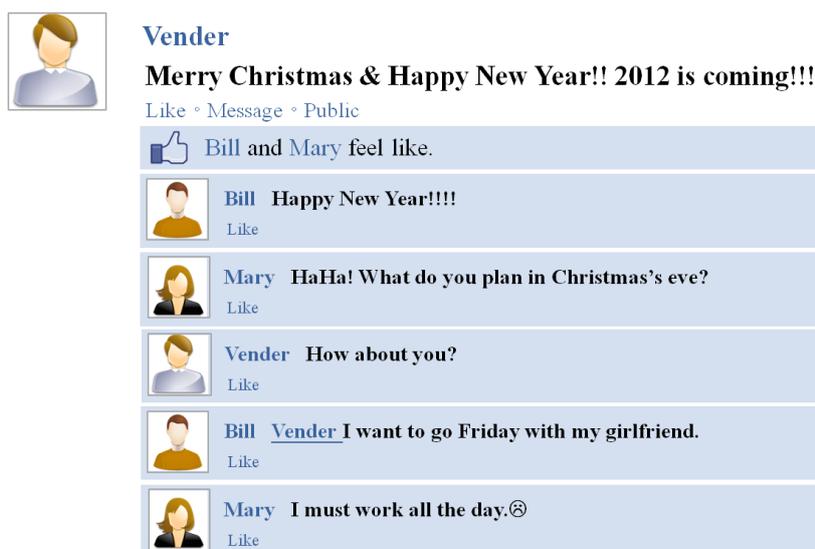


圖 7 Facebook 中使用者與朋友回應文章狀況

圖 7 中，我們可以看到有主要張貼訊息的人(Vender)、回應訊息的人(Bill 和 Mary)、利用“@”針對某人(Bill 針對 Vender)進行回應以及按“讚”的人(Bill 和 Mary)，這四部份，以下我們將圖的結果分為二部份來看：一是針對“讚”的回應，如表 5；二是針對訊息的回應，如表 6：

表 5 使用者與朋友之間互動訊息中對於“讚”的回應

	Vender	Bill	Mary
Vender	0	1	1
Bill	0	0	0
Mary	0	0	0

表 6 使用者與朋友之間互動訊息中對於訊息的回應

	Vender	Bill	Mary
Vender	0	2	2
Bill	1	0	2
Mary	2	0	0

透過上述二表的表示法，首先以“讚”的回應為例，可以看到 Bill 和 Mary 對於 Vender 的回應按了“讚”，因此在權重表示法則  $W_l(\text{Bill}, \text{Vender}, 1)$  和  $W_l(\text{Mary}, \text{Vender}, 1)$ ，其它狀況皆為 0，則表示為  $W_l(\text{Vender}, \text{Bill}, 0)$ 、 $W_l(\text{Vender}, \text{Mary}, 0)$ 、 $W_l(\text{Mary}, \text{Bill}, 0)$ 、 $W_l(\text{Bill}, \text{Mary}, 0)$ 、 $W_l(\text{Vender}, \text{Vender}, 0)$ 、 $W_l(\text{Bill}, \text{Bill}, 0)$ 、 $W_l(\text{Mary}, \text{Mary}, 0)$ 。接著以訊息的回應為例，其權重值表示法依高低順序排列  $W_m(\text{Vender}, \text{Mary}, 2)$ 、 $W_m(\text{Bill}, \text{Vender}, 2)$ 、 $W_m(\text{Mary}, \text{Bill}, 2)$ 、 $W_m(\text{Mary}, \text{Vender}, 2)$ 、 $W_m(\text{Vender}, \text{Bill}, 1)$ 、 $W_m(\text{Bill}, \text{Mary}, 0)$ 、 $W_m(\text{Vender}, \text{Vender}, 0)$ 、 $W_m(\text{Bill}, \text{Bill}, 0)$ 、 $W_m(\text{Mary}, \text{Mary}, 0)$ 。

結合以上二種計算權重的方法後，我們可以得知在某篇訊息或文章中，互動權重值最高與所提及之關鍵字最多的是什麼？在互動權重值中，本研究分為兩部份來計算，一是針對“讚”的回應；二是針對塗鴉牆的內容回應。最後將關係權重與互動權重結合，如圖 7 所示。由於目前尚未有研究針對“讚”與塗鴉牆內容的重要性比較，因此本研究將“讚”的機制與塗鴉牆內容之權重定義等同為各為 0.5，並且利用上述例子作說明：



圖 8 使用者關係與訊息或文章之 TF-IDF 結合

首先處理 Vender、Bill 和 Mary 三人在 Facebook 中的狀況。透過上述例子我們重新整理這三人之權重表示： $W_{l+m}(\text{Bill}, \text{Vender}, 1)$ 、 $W_{l+m}(\text{Mary}, \text{Vender}, 1)$ 、 $W_{l+m}(\text{Vender}, \text{Mary}, 2)$ 、 $W_{l+m}(\text{Mary}, \text{Bill}, 2)$ 、 $W_{l+m}(\text{Vender}, \text{Bill}, 1)$ 、 $W_{l+m}(\text{Bill}, \text{Mary}, 0)$ 、 $W_{l+m}(\text{Vender}, \text{Vender}, 0)$ 、 $W_{l+m}(\text{Bill}, \text{Bill}, 0)$ 、 $W_{l+m}(\text{Mary}, \text{Mary}, 0)$ 。透過以上計算方法後求得式子(8)之結果為本研究之社會關係排名。因此，透過本系統輸入“Christmas”之關鍵字，加入圖 8 所示之權重值計算之後，我們會得到 Vender 和 Mary 曾經提到與

Christmas 相關之訊息。因此，使用者若想要知道朋友在“Christmas”做了什麼？本系統便可得到與朋友互動的結果，如圖 9 所示。

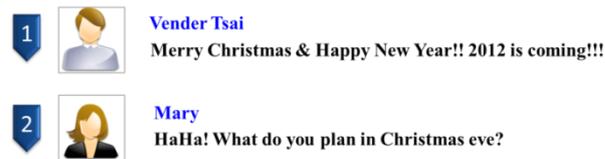


圖 9 符合 Christmas 之訊息

## 伍、結論

隨著時間成長，社群網站在人與人間的關係中已扮演著重要的角色，讓人與人之間形成了強連結。本研究將傳統搜尋引擎加入社會關係，改善了以往搜尋引擎未考慮文章擁有者與搜尋者之間的關係，加強了搜尋之品質。首先，透過 Facebook 之塗鴉牆資料作為搜尋之依據，接著進行 CKIP 詞庫小組處理和 TF-IDF 計算，最後結合字頻和社會關係並進行結果排名。本研究會以真實狀況來驗證所提出之方法之成效。

礙於 Facebook 隱私權的問題，本研究透過 Facebook 所搜尋回來的資料，只有第一層（使用者）資料，而無法進一步擷取第二層（使用者的朋友）資料，因此實驗進行時只針對使用者的塗鴉牆資料作分析，無法進一步針對與使用者朋友的塗鴉牆資料作分析。因此，提供使用者進行 Social Search 時，導致主題性之資料量缺乏。未來 Facebook 若開放個資的權限後，Social Search 所需之資料亦會增加，也會提升使用者進行搜尋的使用。再者，我們更能夠透過 Social Search 中所提供的社會關係來幫助我們搜尋時的依據，可讓我們得到更多推薦之相關資訊並進行判斷。

## 陸、參考文獻

1. Ave, N. G., Wang, C., & Luo, J. (2011). LikeMiner : A System for Mining the Power of “ Like ” in Social Media Networks. *In Proceedings of Conference on Knowledge Discovery and Data Mining*, August 21-24, 2011, San Diego, California, USA, pp. 1-4
2. Baer, M. (2010). The Strength-of-weak-ties Perspective on Creativity: a Comprehensive Examination and Extension. *Journal of applied psychology*, 95(3), pp. 592-601.
3. Brzozowski, M. (2011). Who Should I Follow? Recommending People in Directed Social Networks. *In Proceedings of Conference on Computer Supported Cooperative Work*, March 19-23, 2011, Asia.
4. Gou, L., Zhang, X. L., Chen, H. H., Kim, J. H., & Giles, C. L. (2010). Social network document ranking. *In Proceedings of Joint Conference on Digital Libraries*, June 10-14, 2010, Washington, DC, pp. 313-322.
5. Hanneman, R. A., Riddle, M., & Robert, A. (2005). Table of Contents and Methods Preface. *Introduction to Social Network Methods*, pp. 1-3.

6. Jiang, J., Wilson, C., Wang, X., Huang, P., & Sha, W. (2010). Understanding Latent Interactions in Online Social Networks. *In Proceedings of Internet Measurement Conference*, November 1-3, 2010, Melbourne, Australia.
7. Kuhlthau, C. C. (1991). Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science*, 42(5), Baton Rouge, pp. 361-371
8. Li, C.-T., & Lin, S.-D. (2009). Egocentric Information Abstraction for Heterogeneous Social Networks. *In Proceedings of Conference on Advances in Social Networks Analysis and Mining*, July 20-22, 2009, Athens, Greece, pp. 255-260
9. Mislove, A., & Gummadi, K. P. (2006). Exploiting Social Networks for Internet Search. *The Workshop on Hot Topics in Networks*, November 29-30, 2006, Irvine, California, USA, pp.79-84
10. Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and Analysis of Online Social Networks. *In Proceedings of Internet Measurement Conference*, June 7-10, 2007, Barge, France, pp. 29-42
11. Salton, G., & McGill, M. J. (1983). *Introduce to Modern Information Retrieval*, McGraw-Hill, Inc. New York, NY, USA, pp. 201-215
12. Tong, S. T., Van Der Heide, B., Langwell, L., & Walther, J. B. (2008). Too Much of a Good Thing? The Relationship between Number of Friends and Interpersonal Impressions on Facebook. *Journal of Computer-Mediated Communication*, 13(3), 531-549.
13. Vieira, M. V., Bruno M, F., Damaxio, R., Golgher, P. B., Reis, D. de C., & Ribeiro-Neto, B. (2007). Efficient Search Ranking in Social Networks Categories and Subject Descriptors. *In Proceedings of Conference on Information and Knowledge Management*, November 6-9, 2007, Lisboa, Portugal, pp. 563-572
14. Yang, S. H., Long, B., Smola, A., Sadagopan, N., Zheng, Z., & Zha, H. (2011). Like like alike: Joint Friendship and Interest Propagation in Social Networks. *In Proceedings of International World Wide Web Conference*, March 28-April 1, 2011, Hyderabad, India, pp. 537-546
15. Zanardi, V., & Capra, L. (2008). Social Ranking : Uncovering Relevant Content Using Tag-based Recommender Systems. *In Proceedings of International Conference on Recommender Systems*, July 7-11, 2008, Fort Lauderdale, Florida, USA.
16. Zinoviev, D. (2009). Toward Understanding Friendship in Online Social Networks. *Journal of Technology, Knowledge, and Society*.

# Social Search: Applying Social Networks Analysis for Web Search

## Techniques

Hsiao-Hsuan Lu

National University of Kaohsiung Information Management  
m0993309@gmail.com

I-Hsien Ting

National University of Kaohsiung Information Management  
iting@nuk.edu.tw

Shiau Liang Wang

National University of Kaohsiung Information Management  
slwang@nuk.edu.tw

### Abstract

In recent years, the social networking sites have become an important platform for users to establish the relationship between each other. As time goes by, the links between people will form the so-called “Strong Links”. For those users, information provided by the friends with strong link is considered as more interesting and useful. Currently, most of the search engines are designed based on only measuring the similarity between the keywords and articles. However, the social relations between the authors of articles and searcher have not been taken into account. Therefore, in order to improve the performance of search engines, we include the measurement of social relation to traditional search engine. We expect to improve search quality and to enhance the satisfaction of search.

In this study, we will train the data of from Facebook to calculate the social relationship and content. About the content, the data will be process by using CKIP and TFIDF. Finally, we combine the keyword frequency in the calculated values and social relations, and the value will be used as the key to rank the search results. In this paper, we will also demonstrate a real example to explain the proposed methodology as well as the system interface.

Keywords: search engine, social networks analysis (SNA), social networking sites, social search.