

# 國立交通大學

## 管理科學研究所 碩士論文

分析臉書粉絲團資訊以發展使用者特徵檔為基  
礎之推薦



Recommendations based on user profiles discovered from  
Facebook Like List

研究生：李榮維

指導教授：劉敦仁、林君信 教授

中華民國一百零一年二月

分析臉書粉絲團資訊以發展使用者特徵檔為基礎  
之推薦

Recommendations based on user profiles discovered from  
Facebook Like List

研究生：李榮維

Student: Rong-Wei Lee

指導教授：劉敦仁、林君信

Advisor: Dr. Duen-Ren Liu, Dr. Chiun-Sin Lin



A Thesis

Submitted to Institute of Management Science

College of Management

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science in Information Management

February 2012

Hsinchu, Taiwan, the Republic of China

中華民國 一百零一 年 二 月

# 分析臉書粉絲團資訊以發展使用者特徵檔為基礎之推薦

研究生：李榮維

指導教授：劉敦仁、林君信 博士

國立交通大學管理科學研究所

## 摘要

近年來社群網路的蓬勃發展，除了傳統的部落格網站，許多新型態的社群網站因應而生，如微型網誌 Twitter、Plurk、社交網站 Facebook、MySpace 等，隨著這些社群網站的興起，大大改變了人們原本的使用網路的習慣。

社群網站-Facebook 在使用的過程中，如張貼訊息、回應訊息、對訊息點選「讚」、訂閱粉絲團等資訊，都間接透露出使用者的需求與喜好等資訊。

本研究欲利用使用者訂閱的粉絲團清單，找出使用者的興趣關鍵字特徵檔，並進行相關應用：推薦粉絲團、推薦朋友、推薦廣告。本研究首先利用「正規化概念分析法」分析使用者訂閱的粉絲團之間彼此的關係，進而發掘代表使用者興趣的概念，產生使用者興趣關鍵字特徵檔，最後利用產生的興趣關鍵字特徵檔進行推薦。

本研究實地蒐集了 Facebook 的使用者資料以及 Yahoo 奇摩拍賣的內容式廣告建立離形系統，以離形系統展示研究成果及模擬使用者操作推薦粉絲團、推薦朋友之情況和推薦廣告的成果。

**關鍵字：**社交網路、推薦系統、正規化概念分析

# Recommendations based on user profiles discovered from Facebook Like List

Student: Rong-Wei Lee

Advisor: Dr. Duen-Ren Liu, Dr. Chiun-Sin Lin

Institute of Management Science

National Chiao Tung University

## Abstract

After the fever of web 2.0 when blogging embarked on a new era of internet community, social networks have flourished in recent years and become another Internet golden era. Besides the traditional blog website, many new types of social network websites such as Facebook, MySpace or Microblog Twitter, Plurk came up. Social network websites have changed user behaviors on Internet nowadays dramatically.



In facebook, personal interests or needs are disclosed when users post messages on the wall, reply friends' messages, push "like" messages or join "fan pages".

This research proposes to recommend fan pages, friends and advertisements through discovering user interests from Facebook users' "like list". In this thesis, Formal Concept Analysis (FCA) is adopted to analyze the relation between fan pages in the like list. Afterwards, concepts that are appropriate to represent users' interests are extracted to generate user profiles. Finally, a prototype System is developed to demonstrate the research result of recommending friends, fan pages and advertisements based on the discovered user profiles.

**Keyword:** Social Network, Recommendation System, Formal Concept Analysis

## 誌謝

經歷了許多的辛苦終於完成了自己的研究。雖然選擇了做資訊方面的研究，而不是管理的研究，但是覺得在這一路上學到了很多東西，因此很高興自己選擇了這條道路。而在這漫長的日子裡，接受了很多人的幫助，才讓我能順利的從事自己喜歡的研究。

首先最感謝的就是我的指導教授劉敦仁教授，不管在研究過程中我遇到甚麼困難，或是有一些疑問，甚至在我沒有方向的時候，都很有耐心的給我很多幫忙，真的非常感謝能遇到這麼好的教授。還要感謝我的另外一位指導教授林君信教授，在我的論文研究給了我很多的鼓勵，點出研究的方向。另外非常謝謝口試委員賴錦慧與羅濟群老師細心研讀了我的研究，給予我很多建議，讓我的論文能更加的完整。

其他要感謝的人實在太多了，包含管科所和我一起輔所的同伴王馨卉，每次在我遇到挫折的時候，總是會跳出來給我支持和力量；還有宇軒學長和 Hani 學長在學術上或是生活上也給予我很多的幫助；其他 DB LAB 的學弟妹們也使我的研究生活雖然辛苦但是一點都不無聊，真的非常謝謝大家。

最後要感謝的是給予我一切精神支柱的父母，在這兩年總是無私的為我付出，讓我能順利完成碩士學位。

李榮維 謹於  
交通大學  
中華民國一百零一年二月

# 目錄

中文摘要.....	I
英文摘要.....	II
誌謝.....	III
目錄.....	IV
圖目錄.....	VI
表目錄.....	IX
緒論.....	1
1.1 研究背景.....	1
1.2 研究動機.....	2
1.3 研究目的.....	4
2. 文獻探討.....	5
2.1 社交網路.....	5
2.2 推薦系統.....	5
2.2.1 內容過濾式演算法(Content-based Filtering).....	6
2.2.2 協同過濾演算法(Collaborative Filtering).....	6
2.3 正規化概念分析.....	7
2.3.1 正規文本(Formal Context).....	7
2.3.2 形式概念(Formal Concept).....	8
2.3.3 概念點陣(Concept Lattice).....	8
2.3.4 正規化概念分析案例.....	9
2.3.5 正規化概念分析的應用.....	16
3. 系統架構.....	17
4. 模型介紹.....	19
4.1 興趣關鍵字特徵檔模組.....	19

4.1.1	正規化概念分析.....	19
4.1.2	產生興趣關鍵字特徵檔.....	34
4.2	推薦粉絲團模組.....	37
4.2.1	推薦粉絲團模組架構.....	37
4.2.2	建立使用者與粉絲團清單之特徵矩陣.....	38
4.2.3	計算使用者之間的相似度.....	39
4.2.4	計算粉絲團分數並產生推薦結果.....	40
4.3	推薦朋友模組.....	41
4.4	推薦廣告模組.....	42
4.4.1	推薦廣告模組架構.....	42
4.4.2	廣告詞彙擴張.....	43
4.4.3	產生廣告推薦.....	44
5.	雛形系統建置與實作.....	45
5.1	開發工具與平台.....	45
5.2	建置雛形系統資料集.....	45
5.3	Facebook 推薦系統實作內容.....	50
5.4	推薦粉絲團系統實作內容.....	52
5.5	推薦朋友系統實作內容.....	55
5.6	廣告推薦實作說明.....	58
5.6.1	資料集.....	58
5.6.2	廣告推薦實作.....	60
6.	結論與未來研究方向.....	62
6.1	結論.....	62
6.2	未來研究方向.....	62
	參考文獻.....	65

## 圖目錄

圖 1.1 台灣人平均每月使用 Facebook 的時間(統計資訊截至 2010.1) ...	2
圖 1.2 台灣 Facebook 使用現況(統計資訊截至 2012.1) .....	2
圖 1.3 台灣排名前 10 名的粉絲團(統計資訊截至 2011.8) .....	3
圖 2.7 概念點陣(Concept Lattice) .....	15
圖 3.1 Facebook 推薦系統架構 .....	17
圖 3.2 文字廣告範例 .....	18
圖 4.1 正規化分析的 Object 及 Attribute .....	19
圖 4.2 建置概念矩陣之步驟 .....	20
圖 4.3 受試者所訂閱的粉絲頁面清單 .....	20
圖 4.4 user_likes 物件 .....	21
圖 4.5 粉絲團清單 .....	22
圖 4.6 粉絲團的基本資訊 .....	23
圖 4.7 粉絲團塗鴉牆之內容 .....	24
圖 4.8 swf 類型的 Facebook 訊息 .....	25
圖 4.9 取得所需之欄位值(swf) .....	26
圖 4.10 video 類型的 Facebook 訊息 .....	26
圖 4.11 取得所需之欄位值(video) .....	27
圖 4.12 link 類型的 Facebook 訊息 .....	27
圖 4.13 取得所需之欄位值(link) .....	28
圖 4.14 photo 類型的 Facebook 訊息 .....	28
圖 4.15 取得所需之欄位值(photo) .....	29
圖 4.16 music 類型的 Facebook 訊息 .....	29
圖 4.17 取得所需之欄位值(music) .....	30



圖 4.18 photo 類型的 Facebook 訊息 .....	30
圖 4.19 取得所需之欄位值(question) .....	31
圖 4.20 status 類型的 Facebook 訊息 .....	31
圖 4.21 取得所需之欄位值(status) .....	32
圖 4.22 回傳的 XML 檔案 .....	33
圖 4.23 YAHOO API 提供的參數設定 .....	33
圖 4.24 產生興趣關鍵字特徵檔之演算法 .....	37
圖 4.25 使用者與粉絲團清單之特徵矩陣 .....	38
圖 4.26 Jaccard 相似度的計算 .....	39
圖 4.27 排序後的使用者與粉絲專頁特徵矩陣 .....	40
圖 4.28 依照分數排序的使用者與粉絲專頁特徵矩陣 .....	41
圖 4.29 推薦朋友模組架構圖 .....	42
圖 4.30 廣告推薦模組架構 .....	42
圖 4.31 利用 Wikipedia 從事詞彙擴張 .....	43
圖 5.1 匿名使用者 Nancy 之概念點陣圖 .....	48
圖 5.2 節點 A 之子孫 1 .....	49
圖 5.3 節點 A 之子孫 2 .....	49
圖 5.4 Facebook 推薦系統首頁 .....	51
圖 5.5 Facebook 推薦系統操作流程圖 .....	51
圖 5.6 使用者登入後頁面 .....	52
圖 5.7 選擇推薦粉絲團系統 .....	53
圖 5.8 推薦給 Nancy 之粉絲專頁 .....	53
圖 5.9 選擇推薦朋友系統 .....	55
圖 5.10 推薦朋友系統頁面 .....	55
圖 5.11 系統推薦男性朋友清單 .....	56

圖 5.12 系統推薦女性朋友清單 .....	57
圖 5.13 系統推薦不分性別朋友清單 .....	57
圖 5.14 奇摩拍賣女性長 T 頁面 .....	59
圖 5.15 奇摩拍賣女性長 T 之分類廣告 .....	60
圖 5.16 推薦給 Nancy 之廣告 .....	61



## 表目錄

表 2.2 正規化概念分析案例 .....	9
表 2.3 產生 Formal Context .....	12
表 2.4 所有的 Concept .....	12
表 4.1 解析後的結果 .....	34
表 5.1 系統開發平台與工具 .....	45
表 5.2 匿名使用者 Nancy 訂閱之粉絲團 .....	46
表 5.3 使用者 Nancy 之興趣關鍵字 .....	49
表 5.4 推薦粉絲團結果 .....	54
表 5.5 Nancy 和其他使用者之間的相似度 .....	58
表 5.6 詞彙擴充之推薦廣告結果 .....	61



# 1. 緒論

## 1.1 研究背景

隨著資訊科技的迅速發展，如今早已邁入 Web2.0 的時代，直到現在，Web 已經和我們生活產生密切的結合。而新型社群網站的興起，如微型網誌 (microblog) Twitter、Plurk 及社交網路 (social network) Facebook、MySpace 等的盛行，更是改變了網路世界及人們生活習慣。龐大的用戶人數代表無限的商機，商家可以直接在 Facebook 刊登廣告、成立自己的粉絲團或是利用開發 Facebook 的 App 提高商品及品牌曝光率。

從前人們上網也許會先從 Google、Yahoo 這些入口網站 (Portal) 進入網路世界。但現在許多人上網的第一件事就是登入社群網站，特別是現在最熱門的社群網站—Facebook。依據 ARO 資料顯示[12](圖 1.1)，台灣人每個月使用 Facebook 的時間約 435 分鐘以上，已經逼近傳統的入口網站 Yahoo 奇摩，且使用的時間還在持續成長中。觀察這些現象可以發現這些新型社群網站儼然已經成為另一種型態的入口網站。

特別是時下最最熱門的社群網站 Facebook，截至 2012 年 1 月，依據市調網站 CheckFacebook.com 的資料顯示 Facebook 全球用戶數已達近 80 億[13]，成為全球最大的國際化社區。

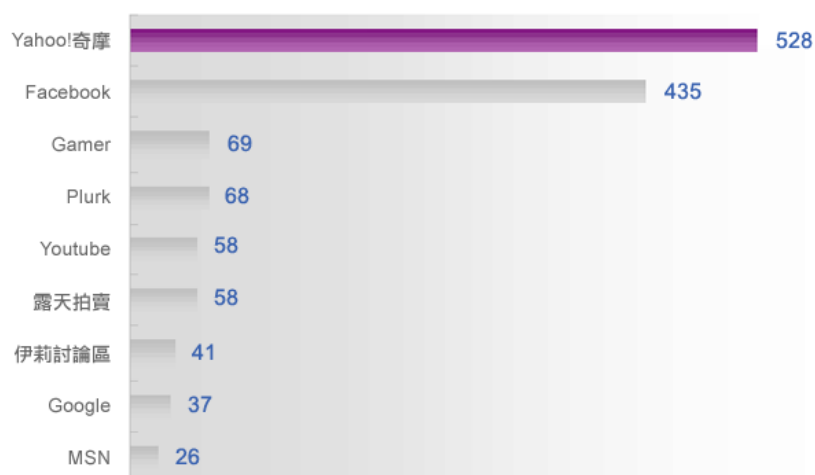


圖 1.1 台灣人平均每月使用 Facebook 的時間(統計資訊截至 2010.1)

## 1.2 研究動機

許多人擁有臉書(Facebook)的帳號。最新 CheckFacebook.com 的臉書使用統計資料顯示台灣使用 Facebook 人數已超過 1000 萬人(圖 1.2)[13]。統計截至 2012 年 1 月止,Facebook 帳號在台灣共有 1160 萬個,且其用戶分布不分男女,甚至不分年紀。

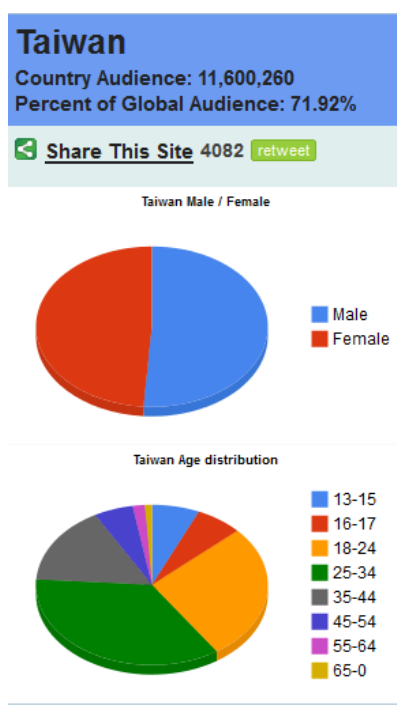


圖 1.2 台灣 Facebook 使用現況(統計資訊截至 2012.1)

在擁有那麼多會員的情況下，Facebook 產生出了無限的商機。因為龐大用戶人數代表著廣大的市場，而社群網站分享的特點代表著豐富的市場資訊，社群網站提供了取得使用者特徵的來源，例如使用者的喜好、需求與習慣等資訊，且現在已經有許多使用者表示願意對外公開自己的訊息。這些資訊可概分為顯性使用者資訊及隱性使用者資訊：

- 顯性使用者資訊：指使用者在註冊時所主動填寫的基本資訊。包含：學歷、工作、興趣、性別、生日……
- 隱性使用者資訊：指使用者在使用過程中所被動顯漏之資訊。如：使用者推文內容、使用者按「讚」的文章、使用者塗鴉牆內容、使用者所訂閱的粉絲團清單等……

現今許多 Facebook 使用者都有訂閱粉絲團的習慣，因為加入粉絲團不需填寫額外的資訊或是做申請的動作，只需在頁面點選「讚」即加入該粉絲團。CRM2. tw[14]就做了一個統計，台灣擁有超過 20 萬以上粉絲的粉絲團就有 42 個(圖 1.3)。



排名	粉絲團名稱	粉絲數
1	周杰倫 Jay Chou	1562162
2	7-ELEVEN	1412735
3	羅志祥 SHOW	940509
4	不正常人類研究中心	938963
5	馬英九總統	848633
6	統一星巴克咖啡同好會	793333
7	楊丞琳 Rainie Yang	639896
8	隋棠	638626
9	Hebe Tian	602835
10	羅志祥	548632

圖 1.3 台灣排名前 10 名的粉絲團(統計資訊截至 2011.8)

由於訂閱粉絲團的普及，因此使用者訂閱了哪些粉絲團可以成為我們了解使用者喜好及需求的重要資訊。使用者訂閱的粉絲團清單(Like List)會被

Facebook 彙總在 user\_likes 物件，但由於使用者訂閱的粉絲團列表(user\_likes 物件)資料雜亂無章，很難輕易的從粉絲團列表就看出其中隱含的使用者資訊。因此本研究欲利用粉絲團列表所隱含之資訊，找出使用者興趣特徵檔，並以此特徵檔進行相關應用。

### 1.3 研究目的

本研究希望達到的目的是從「使用者訂閱的粉絲團清單(Like List)」去發掘使用者的興趣，並以此建立「使用者興趣關鍵字特徵檔」。為了達成此目標，我們會利用正規化概念分析(Formal Concept Analysis, FCA)去找尋使用者訂閱的粉絲團之間相互關係，之後利用一個興趣關鍵字演算法找出哪些粉絲團較能用來表示使用者的興趣，並用這些粉絲團的關鍵字建立使用者興趣關鍵字特徵檔。

在建立使用者興趣關鍵字特徵檔後，本研究會進行相關的應用，包含推薦粉絲團、推薦朋友以及推薦廣告。

最後本研究會從實際蒐集到的 Facebook 使用者資料中隨機選出一個 Facebook 使用者做為目標使用者，並以此使用者的粉絲團清單(Like List)進行挖掘興趣，進而做推薦粉絲團、推薦朋友以及推薦廣告等動作，最後以一個離型系統的方式呈現研究成果。

## 2. 文獻探討

本研究欲利用現今最熱門的社交網路網站「Facebook」所提供的加入粉絲團資訊，進行使用者興趣關鍵字特徵檔的挖掘。並應用所得到的使用者興趣關鍵字特徵檔建立一個 Facebook 的推薦系統，系統包含三個模組：推薦粉絲團模組、推薦朋友模組、推薦廣告模組。

使用者興趣關鍵字特徵檔的挖掘和 Facebook 推薦系統各使用不同技術：

- 挖掘使用者興趣關鍵字特徵檔：正規化概念分析(Formal Concept Analysis)
- Facebook 推薦系統：協同過濾演算法(Collaborative Filtering Algorithm)

本章將對相關主題進行探討。

### 2.1 社交網路



在現代社會人們互相建立、維持關係的平台逐漸由現實社會移轉至虛擬線上世界，人們利用社交網路(Social Network, SN)交朋友。因此社交網路服務(social network service)逐漸吸引越來越多學者的注意，社交網路也成為現今熱門的研究主題。而在本研究中，我們把焦點聚集在一種特殊的 Social Network: Facebook。

INSNA [15]定義社交網路包含了三種元素：邊(edges)、節點(node)、資訊的傳遞，社交網路利用節點之間互相連接的邊，進行資訊的溝通及傳遞。

### 2.2 推薦系統

現今推薦系統大致可以分成兩種類型：

1. 內容過濾式演算法(內容過濾式演算法(Content-based Filtering))
2. 協同過濾演算法(Collaborative Filtering)

以下兩小節茲對此兩種演算法做文獻回顧。



### 2.2.1 內容過濾式演算法(Content-based Filtering)

內容過濾式推薦系統利用項目內容(Content)間的相似度來從事推薦，因此設計內容過濾式推薦系統有兩個重要的議題：

- (1) 表現項目的方式
- (2) 建立項目的特徵檔

在很多內容過濾式推薦系統是以文章中重要的文字來表示一篇文件，例如 Balabanovic 和 Shoham [1]利用文章中 TF-IDF 值排名前 100 名的文字來表示一篇文件。

### 2.2.2 協同過濾演算法(Collaborative Filtering)

協同過濾演算法(Collaborative Filtering, CF)為現今應用在線上推薦系統最成功的技術之一[5, 6]，協同過濾演算法可以分成兩種類型：

- (1) 以使用者為基礎的協同過濾(user-based CF)

用相似統計的方法得到和目標使用者具有相似愛好或興趣的相鄰使用者，彙總這些評分資訊去預測目標使用者尚未使用的產品[2]。Herlocker、Konstan 和 Ried [5, 17]整理了方法步驟：

#### 1. 收集使用者資訊

收集可以代表使用者興趣的資訊。一般的網站系統使用評分或是給予評價的方式，這種方式被稱為「主動評分」。另外一種是「被動評分」，是根據使用者的行為模式由系統代替使用者完成評價，不需要使用者直接打分或輸入評價資料。

#### 2. 最近鄰搜索(Nearest neighbor search, NNS)

計算全部使用者和目標使用者之間的相似度，選擇最相近使用者的子集作為最鄰近集合(Nearest neighbor)，用來預測待預測項目的分數。

#### 3. 產生推薦結果

利用最鄰近集合(Nearest neighbor)的評分分數計算目標使用者對於待

預測項目的分數。

## (2) 以項目為基礎的協同過濾(item-based CF)

為了提升系統的可擴充性(scalability),Linden、Jacobi 和 Benson [6, 8] 提出了以項目為基礎的協同過濾法。以項目為基礎的協同過濾法有一個基本的假設「能夠引起使用者興趣的項目,必定與其之前評分高的項目相似」,透過計算項目之間的相似性來代替使用者之間的相似性。方法步驟:

### 1. 收集使用者資訊

方法同以使用者為基礎的協同過濾。

### 2. 針對項目的最近鄰搜索

先計算已評價項目和待預測項目的相似度,並以相似度作為權重,加權各已評價項目的分數,得到待預測項目的預測值。

### 3. 產生推薦結果

因為項目之間的關係比使用者之間的關係較為穩定,所以以項目為基礎的協同過濾在某些情況能節省線上運算的成本而不降低預測品質。

## 2.3 正規化概念分析

正規化概念分析(Formal Concept , FCA)是由 Rudolf Wille[10]於1982年提出來的。Ganter&Wille[4]解釋 FCA 的核心構想是一種稱之為「Galois connection」的對偶性(duality),這種對偶性可以用來觀察兩種類型的項目(items)之間的關係,例如物件(Objects)和屬性(Attributes)、文件(Documents)和詞彙(terms)。

### 2.3.1 正規文本(Formal Context)

正規化概念分析可由正規文本(Formal Context)開始定義[10]:

正規文本(Formal Context)是由三個部分組成: $(G, M, I)$ ,  $G$  和  $M$  皆為一個集合, $I$  則代表  $G$  和  $M$  的二元關係。

■  $G$  為 Formal Context 的物件集合。

- $M$  為 Formal Context 的屬性集合。
- $I$  為  $G$  和  $M$  之二元關係。如果物件  $g$  和屬性  $m$  具有關係，可以用  $gIm$  或是  $(g, m) \in I$  來表示。

### 2.3.2 形式概念(Formal Concept)

定義完 Formal Context，接著我們定義何謂形式概念(Formal Concept)[10]

在  $A \subseteq G$  且另一方面  $B \subseteq M$  的情況下，定義：

- $A' = \{m \in M \mid (g, m) \in I \text{ for all } g \in A\}$
- $B' = \{g \in G \mid (g, m) \in I \text{ for all } m \in B\}$

如果  $A$ 、 $B$  滿足以下三個條件

1.  $A \subseteq G$
2.  $B \subseteq M$
3.  $A' = B \wedge B' = A$

則  $(A, B)$  可稱為 Formal Concept (圖 2.4)。也就是當物件集合  $A$  中的所有物件所共有的屬性為  $B$  集合時，另一方面屬性集合  $B$  之所有屬性被所有  $A$  物件所共享，則  $(A, B)$  或  $(\{g_2, g_3\}, \{m_3, m_4\})$  即為一個 Formal Concept，簡稱 Concept (概念)。此時  $A = \{g_2, g_3\}$  可稱做此概念  $(A, B)$  之範圍(extent)， $B = \{m_3, m_4\}$  可稱作此概念  $(A, B)$  之內涵(intent)。

### 2.3.3 概念點陣(Concept Lattice)

Formal Context 的 Concept 之間存在有階層關係，也就是子概念-父概念的關係(subconcept-superconcept relation)，經由這些關係，可以描繪出整個 Formal Context 概念間的關係，即概念點陣(Concept Lattice)。概念間的 subconcept-superconcept relation( $\leq$ )，可以由以下關係式定義：

如果概念  $C_1 (A_1, B_1)$  為概念  $C_2 (A_2, B_2)$  的子概念：

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \quad (2.1)$$

或

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow B_1 \supseteq B_2 \quad (2.2)$$

式 2.1 與式 2.2 須滿足其中一個條件，即代表 Concept  $C_1 (A_1, B_1)$  為 Concept  $C_2 (A_2, B_2)$  的子概念 ( $C_1 \leq C_2$ )，換句話說當：

1.  $C_1$  範圍(extent)為  $C_2$  範圍的子集合

或

2.  $C_2$  內涵(intent)為  $C_1$  內涵的子集合

則  $C_1$  為  $C_2$  之子概念。

#### 2.3.4 正規化概念分析案例

在 2.3 最前面提到正規化概念分析可以觀察兩種項目的關係，例如物件及屬性、文件及詞彙。在此我們利用 delicious.com 交流書籤的網站，作為正規化概念分析說明的範例。

假設一共有五個書籤 [Facebook](#)、[Twitter: What are you doing?](#)、[Welcome to Flickr - Photo Sharing](#)、[Wordle - Beautiful Word Clouds](#)、[YouTube - Broadcast Yourself](#)，而每個書籤都有做標籤的動作(表 2.2)：

表 2.1 正規化概念分析案例

書簽名稱	標籤
<a href="#">Facebook</a>	social facebook networking community friends web2.0 socialnetworking entertainment

	network bookmarkbar
<a href="#">Twitter: What are you doing?</a>	twitter social web2.0 community blog socialnetworking networking mobile tools technology
<a href="#">Welcome to Flickr - Photo Sharing</a>	photos flickr photography photo sharing images web2.0 community social tools
<a href="#">Wordle - Beautiful Word Clouds</a>	visualization design cloud

	words tools web2.0 generator typography art wordle
<a href="#">YouTube - Broadcast Yourself.</a>	video youtube videos entertainment media web2.0 social fun community music

我們把書簽名稱(bookmark)做為 Object，標籤(tag)做為 Attribute，來進行正規化概念分析。第一步驟我們要先產生 Formal Context，因為 Attribute(標籤)較多，所以在此只列出部分的 Attribute(表 2.3)

表 2.2 產生 Formal Context

Attribute											
Object Name	art	blog	Bookmarkbar	cloud			community	web2.0	wordle	words	youtube
Facebook			X				x	x			
Twitter: What are you doing?		x					x	x			
Welcome to Flickr - Photo Sharing					.....		x	x			
Wordle - Beautiful Word Clouds	x			x				x	x	x	
YouTube - Broadcast Yourself.							x	x			x

第二步驟為找出 Context 中的所有 Concept(表 2.4)，利用 Formal concept analyse 工具產生 Formal Context 的所有 Concept[16]，之後我們都會利用 van der Merwe、Obiedkov 和 Kourie[9]所提出的 AddIntent 演算法來建置出 Concept Lattice(圖 2.7)。

表 2.3 所有的 Concept

All concepts of conceptual lattice		
ConceptID	Extent	Intent
c(0)	{Facebook; Twitter: What are you doing?; Welcome to Flickr - Photo Sharing; Wordle - Beautiful Word Clouds; YouTube - Broadcast Yourself. }	{web2.0}

c(1)	{Twitter: What are you doing?; Welcome to Flickr - Photo Sharing; Wordle - Beautiful Word Clouds}	{tools ; web2.0}
c(2)	{Facebook; Twitter: What are you doing?; Welcome to Flickr - Photo Sharing; YouTube - Broadcast Yourself.}	{community ; social ; web2.0}
c(3)	{Twitter: What are you doing?; Welcome to Flickr - Photo Sharing}	{community ; social ; tools ; web2.0}
c(4)	{Facebook; Twitter: What are you doing?}	{community ; networking ; social ; socialnetworking ; web2.0}
c(5)	{Welcome to Flickr - Photo Sharing}	{community ; flickr ; images ; photo ; photography ; photos ; sharing ; social ; tools ; web2.0}
c(6)	{Facebook; YouTube - Broadcast Yourself.}	{community ; entertainment ; social ; web2.0}
c(7)	{YouTube - Broadcast Yourself.}	{community ; entertainment ; fun ; media ; music ; social ; video ; videos ; web2.0 ; youtube}
c(8)	{Facebook}	{bookmarkbar ; community ; entertainment ; facebook ; friends ; network ; networking ; social ; socialnetworking ; web2.0}
c(9)	{Twitter: What are you doing?}	{blog ; community ; mobile ; networking ; social ;



		socialnetworking ; technology ; tools ; twitter ; web2.0}
c(10)	{Wordle - Beautiful Word Clouds}	{art ; cloud ; design ; generator ; tools ; typography ; visualization ; web2.0 ; wordle ; words}
c(11)	{}	{art ; blog ; bookmarkbar ; cloud ; community ; design ; entertainment ; facebook ; flickr ; friends ; fun ; generator ; images ; media ; mobile ; music ; network ; networking ; photo ; photography ; photos ; sharing ; social ; socialnetworking ; technology ; tools ; twitter ; typography ; video ; videos ; visualization ; web2.0 ; wordle ; words ; youtube}

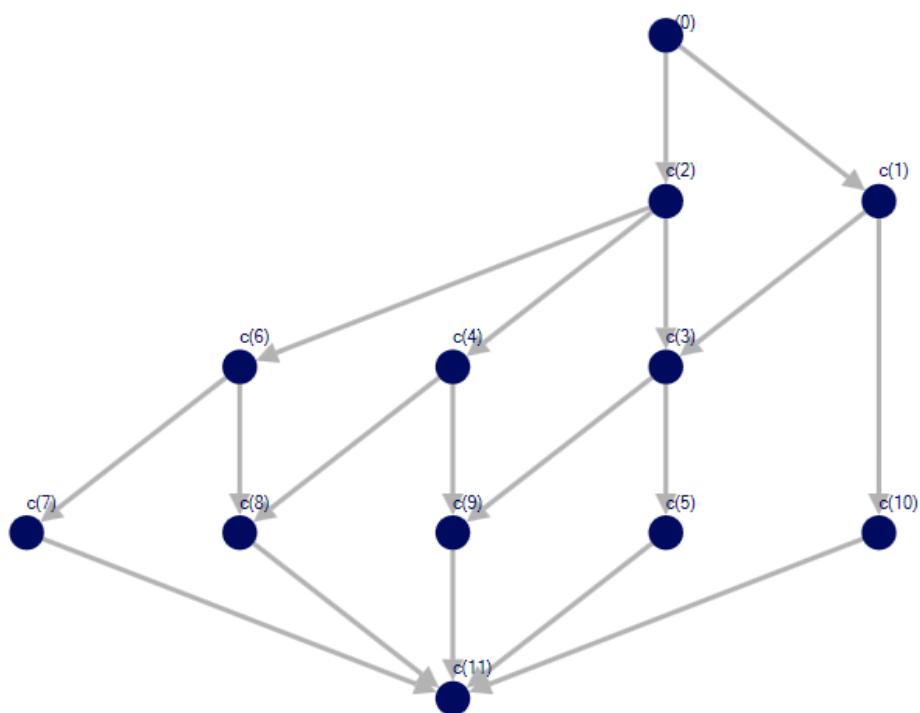


圖 2.1 概念點陣(Concept Lattice)

觀察表 2.4 的 Concept 可以發現當 intent 包含的標籤(tag)的數量增加，extent 的書籤(bookmark)數量就會減少，最極端的例子為 Concept c(11)，intent 包含了所有的標籤，但 extent 所包含的書籤則為空集合；反之當 extent 中書籤數量減少，intent 所能包含的標籤就會增加，最極端的例子為 c(0)，extent 包含了所有的書籤，intent 中的所有書籤共享了一個標籤「web 2.0」。

再觀察圖 2.7 的概念點陣，從最高的節點 c(0)往下走訪最低的節點 c(11)，Concept 的 intent 所包含的屬性會越來越多；另一方面從最低的節點 c(11)往上走訪最高的節點 c(0)，Concept 的 extent 所包含的物件同樣也會越來越多。

因此可以得到結論，當 Concept 在概念點陣的越上方時，Concept 的範圍(extent)所包含的物件愈多，代表 Concept 較為一般化(general)；當 Concept 在概念點陣的越下方時，Concept 的範圍(extent)所包含的物件越少，代表 Concept 較為較特定化(specific)。

### 2.3.5 正規化概念分析的應用

在 Wille 提出正規化概念分析(FCA)後的起初十年，並沒有特別受到學術界的重視，只有 Wille 及他的學生在德國鑽研這塊領域。雖然 FCA 的研究有擴散到歐洲的其他地區，如東歐、法國等，但是在美國，FCA 在資訊科技的研究領域並沒有知名度。根據 Priss [7]指出，只有 Caprineto & Romano [3]提出的一篇 paper 被刊登在美國的 JASIS(Journal of the American Society for Information Science)期刊上。

在近十年來，FCA 因為出現在一些著名的 paper 及研討會，開始受到學術界的重視，被應用到各種領域，包含語言學(linguistic)、軟體工程、心理學、人工智慧、資訊檢索(information retrieval)。



### 3. 系統架構

本研究欲利用使用者所訂閱的粉絲團清單(Like List)，來挖掘出目標使用者所真正感興趣的議題，產生目標使用者的興趣關鍵字特徵檔，並以此進一步做應用。

而要如何從使用者訂閱數量繁多的粉絲團找出使用者真正感到興趣的關鍵字，是一項困難的課題。在本研究利用正規化概念分析(Formal Concept Analysis)技術，以粉絲團做為物件(Object)，而粉絲團的關鍵字作為屬性(Attribute)，進行分析產生出 Formal Context。在產生出 Formal Context 之後，進一步的描畫出概念點陣圖(Concept Lattice)，經由圖型化的方式可以觀察概念(Concept)之間的關係，利用興趣關鍵字演算法，挑選數個最能代表目標使用者的概念，並取出概念的屬性，即粉絲團的關鍵字，用來作為目標使用者的興趣關鍵字特徵檔。

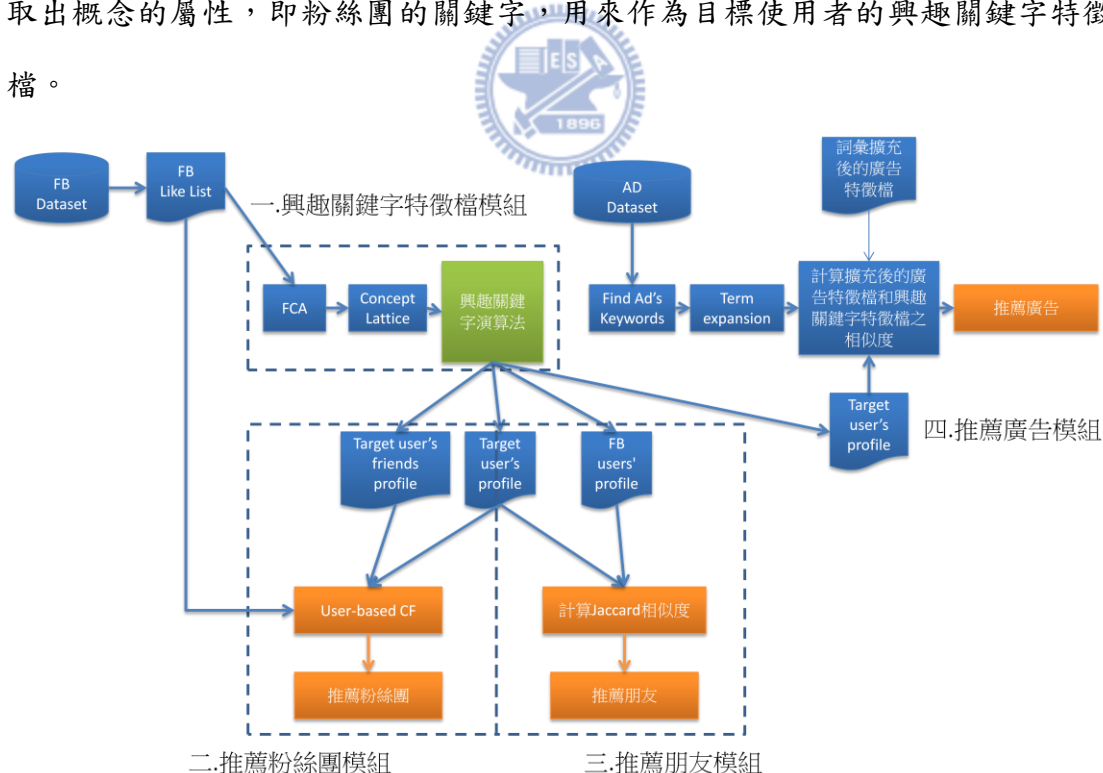


圖 3.1 Facebook 推薦系統架構

如圖 3.1 系統共分為四個模組：興趣關鍵字特徵檔模組、推薦粉絲團模組、推薦朋友模組、推薦廣告模組：

1. 興趣關鍵字特徵檔模組：使用正規化概念分析、概念點陣以及本研究提出的興趣關鍵字演算法找出使用者興趣關鍵字特徵檔。
2. 推薦粉絲團模組：使用以使用者為基礎的協同過濾演算法 (User-based Collaborative Filtering)，Facebook 目標使用者的朋友做為 user，粉絲團作為 item；利用使用者興趣關鍵字特徵檔來計算目標使用者及其朋友之間的相似度，最後依照計算出的粉絲團分數進行推薦。
3. 推薦朋友模組：利用從粉絲團清單擷取出的興趣關鍵字特徵檔，計算使用者之間興趣關鍵字特徵檔的相似度，推薦擁有相似興趣的人給目標使用者。
4. 推薦廣告模組：目的為推薦「文字廣告」給目標使用者，其運作原理和推薦朋友模組相似。相異之處為：由於粉絲團有塗鴉牆訊息、粉絲團資訊等，能擷取出的關鍵字較多；相較下廣告通常只有包含三個部分：標題、摘要、網址(圖 3.1)。因此在本研究廣告關鍵字利用 Wikipedia 資源進行廣告特徵檔的擴張，最後再和使用者興趣關鍵字特徵檔進行相似度比對，推薦使用者較可能有興趣，即和使用者興趣關鍵字特徵檔相似度較高之廣告給目標使用者。



Google首屆校園大使選拔  
www.google.com.tw/ambassador +1  
你是同學心中神人嗎? 快加入G+秀自己  
呼朋引伴來相挺 Google實習等著你!

圖 3.2 文字廣告範例

## 4. 模型介紹

以下，我們先對於本研究如何產生使用者興趣關鍵字特徵檔作說明，其後將針對三個模組：推薦粉絲團模組、推薦朋友模組、推薦廣告模組，做詳細的介紹。

### 4.1 興趣關鍵字特徵檔模組

在本研究提出了一個「興趣關鍵字演算法」產生使用者興趣關鍵字特徵檔，此演算法分成兩個步驟：

1. 正規化概念分析(Formal Concept Analysis)：分析使用者訂閱粉絲團清單(Like List)中粉絲團之間的關係。
2. 產生興趣關鍵字特徵檔：找出數個足以代表使用者興趣的粉絲團，並以這些粉絲團的關鍵字作為使用者興趣關鍵字特徵檔。

#### 4.1.1 正規化概念分析

利用正規化概念分析，把使用者訂閱的每一個粉絲團當作 Object，而每一個粉絲團我們都會擷取相對應的關鍵字，這些關鍵字我們把它作為 Attribute，如圖 4.1。即 Object=使用者訂閱的粉絲團，Attribute=此粉絲團所對應的關鍵字，以此方式進行正規化概念分析。



圖 4.1 正規化分析的 Object 及 Attribute

圖 4.2 歸納如何建置 Formal Context 可分為以下步驟：

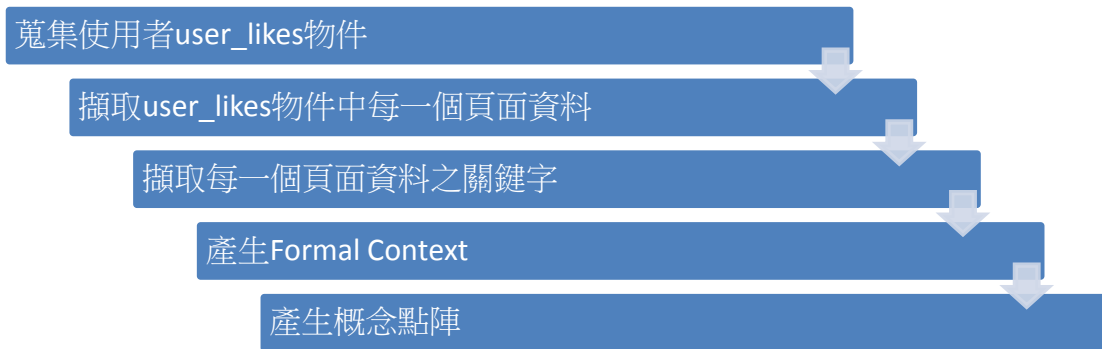


圖 4.2 建置概念矩陣之步驟

#### (1) 蒐集使用者 user\_likes 物件

首先我們必須蒐集使用者所訂閱的粉絲團清單(圖 4.3)。利用 Facebook 官方為 Facebook 應用程式開發者所提供的工具 Graph API Explorer (<https://developers.facebook.com/tools/explorer/>)，先取得使用者的 user\_likes 權限，接著 Graph API Explore 會回傳我們該使用者的 user\_likes 物件 (圖 4,4)。user\_likes 物件為一個 JSON 的物件檔案，裡面包含使用者按過「讚」的頁面記錄。

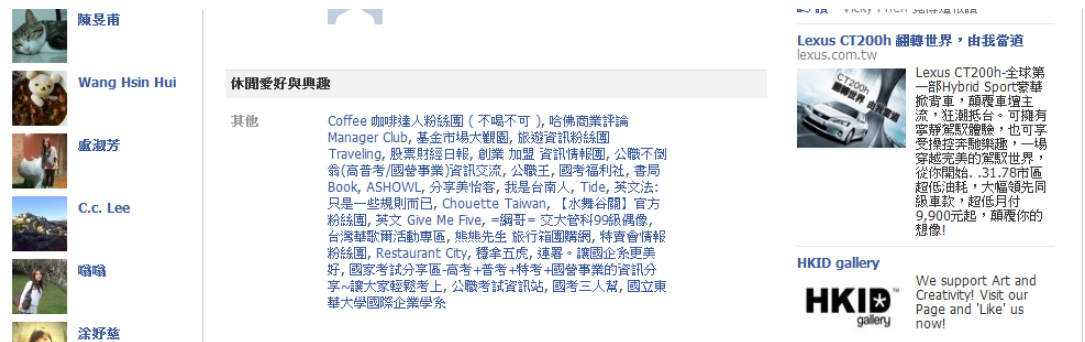


圖 4.3 受試者所訂閱的粉絲頁面清單



```

{
  "data": [
    {
      "name": "Garena 《英雄聯盟 LOL》 ",
      "category": "Games/toys",
      "id": "179663418728712",
      "created_time": "2011-11-22T15:21:59+0000"
    },
    {
      "name": "寵物資訊粉絲團 Pet info",
      "category": "Media/news/publishing",
      "id": "246032849856",
      "created_time": "2011-11-09T15:04:45+0000"
    },
    {
      "name": "好好笑",
      "category": "Application",
      "id": "108254609256337",
      "created_time": "2011-11-09T04:25:30+0000"
    },
    {
      "name": "Spry Fox",
      "category": "Media/news/publishing",
      "id": "164186300287778",
      "created_time": "2011-10-24T05:16:17+0000"
    },
    {
      "name": "Unlight (中文版)",
      "category": "Application",
      "id": "148740785176612",
      "created_time": "2011-10-20T05:06:11+0000"
    },
    {
      "name": "蠟筆小新驚點語錄",
      "category": "Community",

```

圖 4.4 user\_likes 物件

## (2) 擷取 user\_likes 物件中每一個頁面資料

user\_likes 物件裡 data 欄位對應著一個 JSON Array，而此 JSON Array 又包含數個 JSON Object，每個 JSON Object 皆為使用者訂閱的粉絲團，所以整個 JSON Array 就為使用者訂閱的粉絲團清單(Like List)。我們把 data 欄位對應的 JSON Array 中的每一個 JSON Object(圖 4.5)做為正規化概念分析之物件(Object)，每個物件 id 欄位之值皆有



對應到一個 Facebook 頁面(圖 4.5)，之後實際頁面內容之關鍵字做為屬性(Attribute)，用以進行正規化概念分析。



圖 4.5 粉絲團清單

在進行正規化概念分析之前，首先我們要先取得粉絲團的內容，以擷取該粉絲團的關鍵字作為屬性。粉絲團的內容包含了：

- 粉絲團基本資訊
- 粉絲團塗鴉牆訊息

## ■ 粉絲團基本資訊

每一個 Facebook 的粉絲團皆有一般基本公開資訊可以提供我們存取，每個粉絲團提供的公開資訊不一定一樣，例如粉絲團的 id、粉絲團的名稱、類別、like 數等(圖 4.6)……這些都是非常重要的資訊。利用圖 4.5 粉絲團清單可以取得使用者訂閱所有粉絲團的 id，利用粉絲團 id 可以透過連接到網址 [https://graph.facebook.com/like\\_id/](https://graph.facebook.com/like_id/) 取得粉絲團基本資訊，譬如如果想要取得寵物資訊粉絲團的基本資訊，連接到 <https://graph.facebook.com/246032849856/>，即可得到此粉絲團的基本資訊(JSON 格式)(圖 4.7)。基本資訊包含了很多欄位，最後我們一共擷取了以下欄位："affiliation", "artists\_we\_like", "awards", "band\_interests", "band\_members", "bio", "company", "company\_overview", "description", "directed\_by", "features", "general\_info", "general\_manager", "genre", "influences", "mission", "name", "network", "personal\_info", "personal\_interests", "plot\_outline", "produced\_by", "products", "public\_transit", "record\_label", "schedule", "screenplay\_by", "starring", "studio", "written\_by"。

```
{
  "id": "40796308305",
  "name": "Coca-Cola",
  "picture": "https://fbcdn-profile-a.akamaihd.net/hprofile-ak-snc4/276879_40796308305_15784201",
  "link": "https://www.facebook.com/coca-cola",
  "likes": 37006328,
  "category": "Food/beverages",
  "website": "http://www.coca-cola.com",
  "username": "coca-cola",
  "founded": "1886",
  "description": "On May 8, 2011, Coca-Cola marked its 125th anniversary. Created in 1886 in At",
  "can_post": true,
  "checkins": 77,
  "talking_about_count": 382939,
  "type": "page"
}
```

圖 4.6 粉絲團的基本資訊

## ■ 取得粉絲團塗鴉牆資料

同樣的利用圖 4.5 粉絲團清單可以取得使用者訂閱所有粉絲團之 ID，利用 ID 可以連接到網址 [https://graph.facebook.com/like\\_id/feed](https://graph.facebook.com/like_id/feed) 取得塗鴉牆的內容，譬如如果想要取得寵物資訊粉絲團的塗鴉牆內容，連接到 <https://graph.facebook.com/246032849856/feed>，就可以取得此粉絲團的塗鴉牆內容(JSON 格式)(圖 4.7)。



圖 4.7 粉絲團塗鴉牆之內容

Facebook 粉絲團塗鴉牆之訊息，在 [https://graph.facebook.com/like\\_id/feed](https://graph.facebook.com/like_id/feed) 回傳的 JSON 檔是以 JSON Array 型態表示，JSON Array 包含數個物件，每個物件就是一則訊息。在這裡我們只截取粉絲團發佈的資料，而過濾掉其他粉絲在粉絲團的貼文資料。

在我們所取得的 Facebook 粉絲團訊息中，可以發現 Facebook 的粉絲團訊息共可分為 7 大類(也就是在 JSON Object 之 message-type 對應之值):swf、video、link、photo、music、question、status，對於每一種不同的粉絲團訊息，我們會擷取不同欄位之值，最後彙總粉絲團的所有訊息。

## 1. swf

第一種訊息類型為 swf 格式之影片。



圖 4.8 swf 類型的 Facebook 訊息

在這一類型的 Facebook 訊息中，我們取得 JSON Object 以下屬性之值(圖 4.8、圖 4.9)：

- description：鏈結的說明文字（出現在鏈結名稱之下）
- message：粉絲團張貼的訊息
- name：鏈結的名稱

"message": "Check out K-Pop sensation Super Junior's new video, 'Mr. Simple'. Spot who's wearing PUMA Easy Wing Lows and Vesta Runners. You know you love them those guys and their kicks.",

"picture": "https://www.facebook.com/app\_full\_proxy.php?app=121559351219258&v=1&size=p&checksum=0a6c4a606a9cc8ac6e5bf1a15af2dc12&src=http%3A%2F%2Fyimg.com%2Fvi%2F6TwzSGYycM%2Fdefault.jpg",

"source": "http://www.youtube.com/watch?v=6TwzSGYycM?version=3&f=videos&app=youtube\_gdata",

"description": "Super Junior 슈퍼주니어\_Mr.Simple\_MUSICVIDEO",

"name": "Super Junior 슈퍼주니어\_Mr.Simple\_MUSICVIDEO",

"likes":

圖 4.9 取得所需之欄位值(swf)

## 2. video

第二種訊息類型為影片。



The screenshot shows a Facebook post from the page 'PAPAGO'. The post content includes a video player for 'PAPAGO! 行車記錄器-P2產品介紹' and a description in Chinese: 'PAPAGO! 行車記錄器-P2- 120度超廣角、1080P高畫質、內建GPS、重力感測、固定式測速照相器語音提醒。 PAPAGO! 行車安全，第一品牌'.

Annotations on the image:

- message**: Points to the URL `http://www.youtube.com/watch?feature=player_embedded&v=ngxSac1ahgQ`.
- name**: Points to the video title `PAPAGO! 行車記錄器-P2產品介紹` and the domain `www.youtube.com`.
- description**: Points to the detailed Chinese text describing the product features.

Below the video, there are interaction buttons: '讚' (Like), '留言' (Comment), '分享' (Share), and a timestamp '12月1日 10:22'. It also shows '26 個人都說讚' (26 people said like), '查看全部 5 則留言' (View all 5 comments), and '被轉分享 2 次' (Shared 2 times).

圖 4.10 video 類型的 Facebook 訊息

在此類型要擷取的資料欄位一樣為：

- description、message、name (圖 4.10、圖 4.11)

```

"id": "174178510914_270636532988471",
"message": "http://www.youtube.com/watch?feature=player_embedded&v=ngxSaclahgQ",

"picture": "https://s-external.ak.fbcdn.net/safe_image.php?id=AQDaP0xoEy3NjmJB&w=130&h=130&url=http%3A%2F%2Fi3.yimg.com%2Fvi%2FngxSaclahgQ%2Fhqdefault.jpg",
"source": "http://www.youtube.com/v/ngxSaclahgQ?version=3&autohide=1&autoplay=1",
"description": "PAPAGO! 行車記錄器-P2- 120 度超廣角、1080P 高畫質、內建 GPS、重力感測、固定式測速照相器語音提醒。 PAPAGO! 行車安全，第一品牌",
"name": "PAPAGO! 行車記錄器-P2 產品介紹",

"likes": {
  "count": 24,
  "data": [
    {
      "id": "100000028158147",
      "name": "張家欣"
    }
  ]
},

```

圖 4.11 取得所需之欄位值(video)

### 3. link

第三種類型為轉貼連結類，可以從 Facebook 站外或是站內轉貼。



message

name

description

圖 4.12 link 類型的 Facebook 訊息



在此類型要擷取的資料欄位一樣為：

- description、message、name (圖 4.12、圖 4.13)

"message": "【墾丁大街拍拍照】網友墾丁遊記：\n 墾丁大街吃的喝的玩的~通通都有詳細介紹~有圖有真相！\n 墾丁必逛的景點：一到晚上，墾丁街將搖身成繁華熱鬧街景，街上 PUB 音樂炒熱氣氛，小攤位各式各樣的趣味遊戲、小吃，沿路都有具有紀念價值的個性商店，還有\n2026 還有撲鼻而來的異國/南洋料理香味，快來墾丁，享受墾丁獨一無二的夜生活吧!",

'picture': "https://s-external.ak.fbcdn.net/safe\_image.php?id=AQBOOE TaPDhPk13i&w=90&h=90&url=http%3A%2F%2Fblog.yimg.com%2F2%2F8n8cqTp7s5\_9dY75jZyZ6hgPg245KBd4duVTwQOWRR96OWY6bINYHQ--%2F83%2Fm%2FhDc6P0vfT9qR3GbLRB7Ca.jpg",

"description": "來到恆春半島，越夜越美麗的「墾丁大街」絕對是大家必逛的觀光聖地！綿延一公里多的「墾丁大街」泛指「墾丁國家公園牌樓」到「凱薩大飯店」之間的「墾丁路」，這裡的白天是台 26 線的主要道路，每 ...",

"likes": {  
 "count": 88,  
 "data": [{  
 "id": "100001167177513",  
 "name": "墾丁住宿"  
 }]  
 },

"name": "墾丁大街拍拍照 - Family 的 Yahoo!奇摩部落格",  
 "created\_time": "2011-11-17T09:41:18+0000",

圖 4.13 取得所需之欄位值(link)

#### 4. photo

photo 類型之訊息為粉絲團張貼自己或分享他人的相片、相簿。



健康減肥 快樂人生分享了 Carol Lin 的相簿：Business Mastery Classes。



**Business Mastery Classes** name

That was a exhausting but rewarding week! But well Worth it. Taking Internet Marketing Class from Frank Kern was a pleasure. He is such an easy going, relaxed and approachable person. I am glad that I am going again in Las Vegas after Tony Robbins Date With Destiny!

發佈者：Carol Lin

相片：4

讚 · 留言 · 分享 · 11月20日 17:25 ·

8 個人都說讚。

查看 1 則回應 被轉分享 1 次

留言.....

description

圖 4.14 photo 類型的 Facebook 訊息

在此類型要擷取的資料欄位一樣為：

- description、message、name(圖 4.12、圖 4.13)

```

    "story": "健康減肥 快樂人生 shared Carol Lin's album: Business Mastery Classes.",
    "description": "That was a exhausting but rewarding week! But well Worth it. Taking Internet Marketing Class from Frank Kern wa  

    a pleasure. He is such an easy going, relaxed and approachable person. I am glad that I am going again in Las Vegas after Tony Robbins  

    Date With Destiny! ",
    "likes": {
      "count": 8,
      "data": [
        {
          "id": "100000159222430",
          "name": "李寧寧"
        }
      ]
    },
    "name": "Business Mastery Classes",
    "created_time": "2011-11-20T09:25:45+0000",
    "actions": [
      {
        "name": "Comment",
        "link": "https://www.facebook.com/141568392552227/posts/245178362209231"
      }
    ]
  },
  {
    "name": "Business Mastery Classes",
    "created_time": "2011-11-20T09:25:45+0000",
    "actions": [
      {
        "name": "Comment",
        "link": "https://www.facebook.com/141568392552227/posts/245178362209231"
      }
    ]
  }
}

```

圖 4.15 取得所需之欄位值(photo)

## 5. music

此類型訊息為粉絲團分享音樂。



The image shows a Facebook post from the user 'deadmau5'. The post includes a profile picture, the name 'deadmau5', and a message: 'track of the day for me: http://www.beatport.com/track/god-is-green-phil-kieran-club-mix/1625431 Gotta hand it to Phil Kieran / Alloy Mental.. always loved all his stuff!'. Below the message is a video player for 'Skint Presents Legends - Volume 1 [Skint Records]' with a duration of 0:07. The video title is 'God Is Green (Phil Kieran Club Mix)' and the URL is 'www.beatport.com'. The description of the video is 'God Is Green (Phil Kieran Club Mix) is available for download on Beatport, the world's largest music store for DJs'. The post has 1,166 likes and 216 comments. The user's name 'deadmau5' is highlighted with a box and labeled 'name'. The message text is highlighted with a box and labeled 'message'. The video title 'God Is Green (Phil Kieran Club Mix)' is highlighted with a box and labeled 'name'. The video description 'God Is Green (Phil Kieran Club Mix) is available for download on Beatport, the world's largest music store for DJs' is highlighted with a box and labeled 'description'.

圖 4.16 music 類型的 Facebook 訊息



在此類型要擷取的資料欄位一樣為：

- description、message、name(圖 4.16、圖 4.17)

```
"id": "9980651805_115648778551423",  
"message": "track of the day for me: http://www.beatport.com/track/god-is-green-phil-kieran-club-mix/1625431 Gotta hand it to Phil Kieran / Alloy Mental, always loved all his stuff! ",  
"source": "http://geo-samples.beatport.com/items/volumes/volume3/items/1000000/500000/20000/5000/400/30/1625431.LOFL.mp3",  
"description": "God Is Green (Phil Kieran Club Mix) is available for download on Beatport, the world's largest music store for DJs",  
"likes": {  
  "count": 1138,  
  "data": [{  
    "id": "100001622637353",  
    "name": "Anthony Constant"  
  }]  
},  
"name": "God Is Green (Phil Kieran Club Mix)",  
"created_time": "2011-11-29T12:07:17+0000",  
"actions": [{  
  "name": "Comment",
```

圖 4.17 取得所需之欄位值(music)

## 6. question

此類型之訊息為粉絲團發起之問券調查。



**Cold Stone Creamery 酷聖石冰淇淋**問了「【歡樂投票時間】  
星期一天氣晴，粉絲們也要有開心的心情唷！這個週末好不容易雨停了，一定要趁著機會享受歡樂的冰淇淋時光！粉絲們喜歡哪一款『愛爾蘭奶酒冰淇淋』呢？請投下歡樂的一票喔！」問題

1. 我是派對王！最喜歡『櫻桃狂歡派對』！  
2. 我愛聖誕節！最喜歡『聖誕枯杳派對』！  
3. 我是狂歡王！最喜歡『奶酒淇淋』！  
4. ....太難選了！我全部都超愛的啊！

追蹤 · 問朋友 · 11月21日 14:38 ·

圖 4.18 photo 類型的 Facebook 訊息

此類型之 Facebook 訊息我們只擷取問題的部分，也就是 story 欄位的資料，而

不擷取選項的資料(圖 4.18、4.19)。

```
{  
  "id": "153845928748_10150411586553749",  
  "story": "Cold Stone Creamery 酷聖石冰淇淋 asked: 【歡樂投票時間】星期一天氣晴，粉絲們也要有開心的心情唷！這個週末好不容易雨停了，一定要趁著機會享受歡樂的冰淇淋時光！ 粉絲們喜歡哪一款『愛爾蘭奶酒冰淇淋』呢？請投下歡樂的一票喔！",  
  "icon": "https://s-static.ak.facebook.com/rsrc.php/v1/yy/h/pz5wRf7MB0H.png",  
  "application": {  
    "id": "10150110253435258",  
    "name": "Questions"  
  },  
  "object_id": "10150411586528749",  
  "from": {  
    "id": "153845928748",  
    "category": "Product/service",  
    "name": "Cold Stone Creamery 酷聖石冰淇淋"  
  }  
}
```

圖 4.19 取得所需之欄位值(question)

## 7. status

此類型為粉絲團發的一般文字訊息。



圖 4.20 status 類型的 Facebook 訊息

在此類型之 Facebook 訊息只有 message 欄位可以擷取(圖 4.20、圖 4.21)。

```

    "id": "135542383583_10150442812208584",
    "message": "大學生了沒錄影徵求!!!\n 錄影時間:12/5(一)\n 徵求主題:[大學生金錢觀]\n 徵求對象:\n1.買東西不手軟,常常刷爆卡的大學生\n2.月初有錢,月底窮到要借錢的大學生\n3.沒有記帳習慣,總是不知道錢花到哪裡去的大學生\n4.信用卡費每次都無法一次付清的大學生\n203b 若你符合以上任何一個條件,請踴躍報名參加\n 節目將教你解決債務的方法!\n\n 請來信說明你要報名的主題\n 並附上你的姓名/學校科系/聯絡電話/fb 或無名連結 mail 到\nfar@wdfine.com 小 FAR 收,或打(02)2706-8088 #116",
    "likes": {
      "count": 114,

```

圖 4.21 取得所需之欄位值(status)

### (3) 擷取每一個頁面資料之關鍵字

由於如何擷取關鍵字並不在本研究的範疇，因此擷取粉絲團關鍵字的部份，我們使用了 Yahoo 提供的關鍵字擷取 API。擷取關鍵字的流程為：

#### i. 利用 Yahoo 關鍵字 API 進行關鍵字擷取

如圖 4.22，經由 Yahoo API 請求，會回傳一個 XML 檔案，裡面包含了此文件檔(粉絲團)的關鍵字(token)，以及分數(score)。在 Yahoo 關鍵字 API 預設 score 的門檻值(threshold)是 30，可以透過參數傳遞進行設定(圖 4.23)，在此使用 Yahoo 的預設值。另外也可以設定每個頁面要取用的關鍵字數目，在此同樣使用 Yahoo 的預設值，一個粉絲團最多取 10 個關鍵字。

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <KeywordExtractionResult xmlns="http://wl.openapi.ssearch.tpc"
3   <Keyword>
4     <token>youtube</token>
5     <score>109.66935109959</score>
6   </Keyword>
7   <Keyword>
8     <token>Apple store</token>
9     <score>102.39414053559</score>
10  </Keyword>
11  <Keyword>
12    <token>wood</token>
13    <score>100.04106202792</score>
14  </Keyword>
15  <Keyword>
16    <token>billboard</token>
17    <score>92.267387766868</score>
18  </Keyword>
19  <Keyword>
20    <token>promos</token>
21    <score>78.621264300278</score>
22  </Keyword>
23  <Keyword>
24    <token>The Oxford Murders</token>
25    <score>73.783474479286</score>
26  </Keyword>
27  <Keyword>
28    <token>video</token>
29    <score>72.487549115774</score>
30  </Keyword>
31  <Keyword>
32    <token>Maggie Q</token>

```

圖 4.22 回傳的 XML 檔案



欄位名稱	是否必填	預設值	說明
appid	是		應用程式授權碼
content	是		UTF-8編碼的純文字內容
threshold	否	30	只輸出分數高於此門檻值的關鍵字
maxnum	否	10	最多只輸出此個數的關鍵字

圖 4.23 YAHOO API 提供的參數設定

ii. 把回傳的 XML 檔案進行解析取得關鍵字

回傳的檔案經由 Parser 取得每個粉絲頁面的關鍵字(表 4.1)

表 4.1 解析後的結果

id	keyword	score
10006554849	mtv	66.015755446966
10006554849	Apple store	102.39414053559
10006554849	wood	100.04106202792
10006554849	never stop	64.7119859617
10006554849	billboard	92.267387766868
10006554849	youtube	109.66935109959
10006554849	promos	78.621264300278
10006554849	Maggie Q	71.607537956817
10006554849	video	72.487549115774

#### (4) 產生 Formal Context



在擷取完每個專頁的關鍵字後，即可以產生出一個 Formal Context。

每個粉絲團為 Formal Context 的物件(Object)，而粉絲團的關鍵字則為屬性(Attribute)。

#### (5) 產生概念點陣圖

再產生 Formal Context 之後，依照所有概念(Concept)的關係，利用 Formal concept analyse [16]，可以描繪出概念點陣圖。

#### 4.1.2 產生興趣關鍵字特徵檔

「興趣關鍵字演算法」第二步驟為依據演算法的第一步驟正規化概念分析所產生之 Concept Lattice，觀察 Concept Lattice，找出數個重要的概念(Concept)，並以這些概念之內涵(intent)之聯集產生使用者興趣關鍵字特徵檔，為一種 Heuristic Algorithms(啟發式演算法)。

因為 Concept Lattice 越往上方，所得到的概念會愈一般化，而越往下方，

所得到的概念會越特定化，因此我們的演算法必須避免我們擷取到過於特定化或是一般化的概念。

在此演算法，我們會使用到 Zhang 和 Feng[11]所定義的兩個概念特徵 (Concept Feature)：

(1) 概念重要性(Concept Importance, CI)

概念(Concept)之範圍(extent)越廣，代表涵蓋越多物件，在此也可以說此概念包含越多粉絲團。因為包含較多粉絲團，所以此概念的內涵也就越可以用來描述使用者的興趣。由於本想法類似詞頻(term frequency)之概念，所以在此用 TF 符號表示。

$$TF(C_i) = \text{concept } C_i \text{ 之 extent 所包含的 Object 數量} \quad (4.3)$$

但如果 Concept 在越高層，代表此 Concept 越一般化，所包含的共通關鍵字越少，可在本研究欲找到粉絲團的共通關鍵字當作使用者的興趣關鍵字。因此為了避免概念過度一般化，本研究把概念的子孫數作為除項，因類似 DF 概念，在這裡以 DF 符號表達。

$$DF(C_i) = \text{concept } C_i \text{ 之子孫數} \quad (4.4)$$

由此可推 IDF 的公式為：

$$IDF(C_i) = 1 + \frac{N}{\log DF(C_i)} \quad (4.5)$$

可得 CI 為：

$$CI(C_i) = TF(C_i) \times IDF(C_i) \quad (4.6)$$

(2) 概念相似性(Concept Similarity, CS)

概念相似性顧名思義即為 2 個 Concept 的 extent 所包含之粉絲團集合相似度，因避免加入過度相似的 Concept，所以在此定義這個公式：

$$CS(C_1, C_2) = \left| \frac{\text{extent}(C_1) \cap \text{extent}(C_2)}{\text{extent}(C_1) \cup \text{extent}(C_2)} \right| \quad (4.7)$$

分母為兩個概念  $C_1$ 、 $C_2$  粉絲團集合聯集的數量，分子為概念  $C_1$ 、 $C_2$  粉絲團集合交集的數量。

#### ■ 產生興趣關鍵字特徵檔

圖 4.24 為產生興趣關鍵字特徵檔之演算法，其目的是為了選出適當的、不會過於一般化或是特定化的概念(Concept)，並以這些概念之內涵(intent)做為使用者興趣關鍵字特徵檔的來源。以下簡述此演算法之步驟：

- 第一步驟：計算使用者所有概念的 CI 值，並由大到小排序。
- 第二步驟：先把 CI 值由大到小逐一加入 ISet(interest Set)中，每欲加入一個 Concept，就必須把此 Concept 和 ISet 中所有已加入 ISet 之 Concept 計算兩兩的 CS 值，如果其中最大的 CS 值沒超過門檻值(threshold)，就把此 Concept 加入 ISet，反之則不予加入 ISet。一直重複此步驟，直到取得足夠的 Concept 或是已經讀完所有的 Concept。
- 第三步驟：把所有 ISet 的內涵(Intent)做 Union 集合運算，最後即獲得 FSet(Feature Set)，即使用者興趣關鍵字特徵檔。

CSet (Concept Set) : 在 concept lattice 之所有 concept 所構成的集合。

ISet (Interest Set) : 目標使用者感興趣的 concept 集合。

FSet (Feature Set) : 目標使用者所感興趣的關鍵字集合。

K: 控制 ISet Concept 之數量。

N: 在 concept lattice 之所有 concept 數量。

Step1: 計算 CSet 所有 Concept 的 CI 值，並由大至小排序。

Step2: 尋找目標使用者感興趣之 Concept。

```
ConceptNo=1;//ISet 的 Concept 編號。
i=1;
similarityValue =0;//相似度。

CS_Threshold: 兩 concept 相似度的上界。

While (conceptNo<=K&& i<=N)//讀取 CSet 所有的 Concept。
{
    Ci=從 CSet 中取出 CI 值為第 i 高的 Concept;
    i++;
    similarityValue=max{CS(Ci,Cj)|Cj∈ISet};//計算目前的 concept Ci 和 ISet 中所有
    Concept 的相似度並取最大值。
    if (similarityValue<CS_Threshold){
        ISet=ISet ∪ Ci;//如果 CS<Threshold，就把此 Concept Ci 加入 ISet 集合中。
        conceptNo++;
    }
}
```

Step3: 產生興趣關鍵字特徵檔: ISet 的屬性聯集即為目標使用者的興趣關鍵字集合 FSet。

圖 4.24 產生興趣關鍵字特徵檔之演算法

## 4.2 推薦粉絲團模組

### 4.2.1 推薦粉絲團模組架構

此模組是利用以使用者為基礎的協同過濾演算法(user-based CF)。步驟如下：

1. 取得目標使用者以及其朋友之訂閱粉絲團清單(Like List)，建立使用



者與粉絲團清單之特徵矩陣。


2. 相似度計算使用 Jaccard 相似度公式，計算每位好友和目標使用者之間「興趣關鍵字特徵檔」的相似度。
3. 依照相似度將使用者的朋友排序，選出  $N$  個的好友作為目標使用者 (target user) 之最鄰近集合 (Nearest neighbor)
4. 計算粉絲團之分數，並產生推薦結果。

#### 4.2.2 建立使用者與粉絲團清單之特徵矩陣

圖 4.25 為利用圖 4.3 蒐集到的粉絲團清單建立一個使用者與粉絲團清單的特徵矩陣。

首先解釋圖 4.25 符號所代表之含義：

- $u_i$ ： $u_0$  代表目標使用者， $u_1 \sim u_m$  代表目標使用者在 Facebook 的朋友。
- $p_j$ ： $p_1 \sim p_n$  代表 Facebook 的粉絲團
- $X_{ij}$ ：為二元值  $\{0, 1\}$ 。如果  $X_{ij}=0$ ，代表使用者  $u_i$  並沒有訂閱  $p_j$  粉絲團；如果  $X_{ij}=1$ ，代表使用者  $u_i$  有訂閱  $p_j$  團。



	$p_1$	$p_2$	$p_3$	$p_4$	$p_j$	...	...	...	...	...	$p_n$
$u_0$											
$u_1$											
$u_2$											
$\vdots$											
$u_i$					$X_{ij}$						
$\vdots$											
$u_m$											

圖 4.25 使用者與粉絲團清單之特徵矩陣

#### 4.2.3 計算使用者之間的相似度

而本研究是使用 Jaccard 相似度作為計算使用者之間「興趣關鍵字特徵檔」之相似度，Jaccard 相似度(Jaccard's similarity coefficient)的計算公式為：

$$Jaccard(X, Y) = \frac{X \cap Y}{X \cup Y} \quad (4.1)$$

興趣關鍵字	k1	k2	k3	k4	k5	k6	k7	k8	k9
目標使用者	1			1	1			1	1
朋友 A		1	1	1			1		
朋友 B	1	1		1				1	

圖 4.26 Jaccard 相似度的計算

以圖 4.26 為例，朋友 A 和目標使用者興趣關鍵字聯集為 {k1, k2, k3, k4, k5, k7, k8, k9}，而其中兩人有共同的興趣關鍵字(交集){k4}，因此兩人的相似度為 1/8；朋友 B 和使用者興趣關鍵字聯集為 {k1, k2, k4, k5, k8, k9}，而兩人都有訂閱 {k1, k4, k8}，因此兩人的 Jaccard 相似度為 3/6。由此可見 B 和目標使用者的相似度(0.5) 比 A 和目標使用者的相似度(0.125) 為高。

在計算完每位使用者和目標使用者之間的興趣關鍵字特徵檔相似度之後，把目標使用者的朋友依相似度由高至低將原本的使用者與粉絲團清單矩陣進行重新排序，並取 N 個使用者作為鄰居(Neighborhood)，對目標使用者進行粉絲團的推薦。如圖 4.27，顏色越深的列，代表和目標使用者相似度越高的朋友。

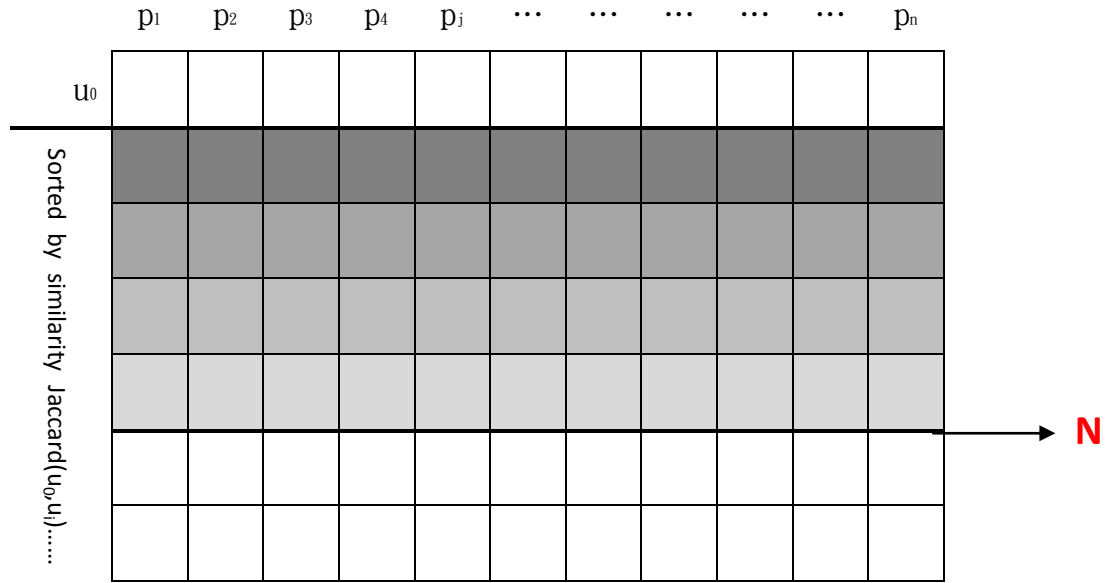


圖 4.27 排序後的使用者與粉絲專頁特徵檔矩陣

#### 4.2.4 計算粉絲團分數並產生推薦結果

推薦粉絲團模組的最後一個動作為計算每一個目標使用者尚未訂閱的粉絲團的分數，並依分數為粉絲團排序，選出 TOP N 的專頁推薦給使用者。計算粉絲團分數的公式為：

$$\text{Score}(p_j) = \frac{\sum_{u_i \in \text{neighbobohod}} \text{Jaccard}(u_i, u_0) X_{ij}}{\text{MAX}(\text{Score})} \quad \begin{matrix} 0 \leq \text{Score}(P_i) \leq 1 \\ X_{ij} = \{0,1\} \end{matrix} \quad (4.2)$$

式 4.2， $u_0$  為目標使用者，所以分子為以 Jaccard 相似度  $\text{Jaccard}(u_i, u_0)$  做為權重，乘上朋友  $u_i$  是否訂閱粉絲專頁  $P_j$  之加總，而分母  $\text{MAX}(\text{Score})$  則是除以所有計算出粉絲團中分數最高的分數，使最後總體分數  $\text{Score}(P_j)$  對應到 0 至 1 之間。

同樣以圖 4.25 做為例子，如要推薦目標使用者沒訂閱的粉絲團  $P_2$ 、 $P_3$ 、 $P_6$ 、 $P_7$ ，可以計算分數為：

$$\text{score}(P_2) = 1 * 0.125 + 1 * 0.5 = 0.625$$

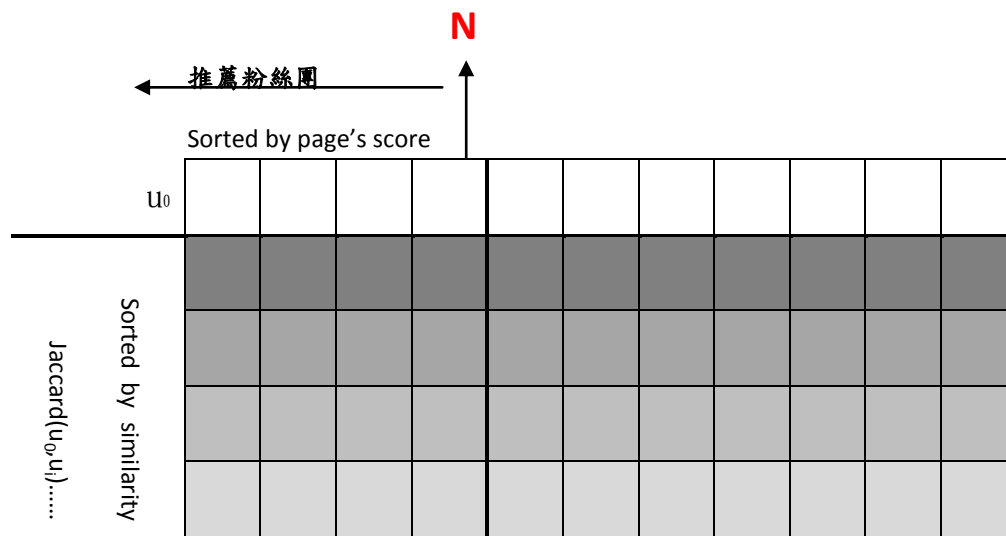
$$\text{score}(P_3) = 1 * 0.125 = 0.125$$

$$\text{score}(P6)=0$$

$$\text{score}(P7)=1*0.125=0.125$$

因此如要推薦粉絲團，會優先選擇推薦粉絲團 P2。

最後使用者與粉絲團清單特徵矩陣依照推薦分數做為排序，推薦 TOP N 的粉絲團給目標使用者(如圖 4.28)



### 4.3 推薦朋友模組

推薦朋友模組使用之方法為利用 4.1 所述方法，得到所有使用者的興趣關鍵字特徵檔，再利用興趣關鍵字特徵檔做目標使用者和其他使用者間的 Jaccard 相似度計算(式 4.1)，最後計算出分數(式 4.2)，選擇和目標使用者相似度最高的 TOP N 使用者，推薦給目標使用者加入好友。

因為每個人使用推薦朋友模組的目的不一，有些人可能為了擴展交友圈，而有些人可能是為了找尋伴侶，因此本模組一開始先讓使用者選擇系統推薦朋友之性別，可以選擇男性、女性、或是皆可。系統的運作流程可以參考圖 4.29。

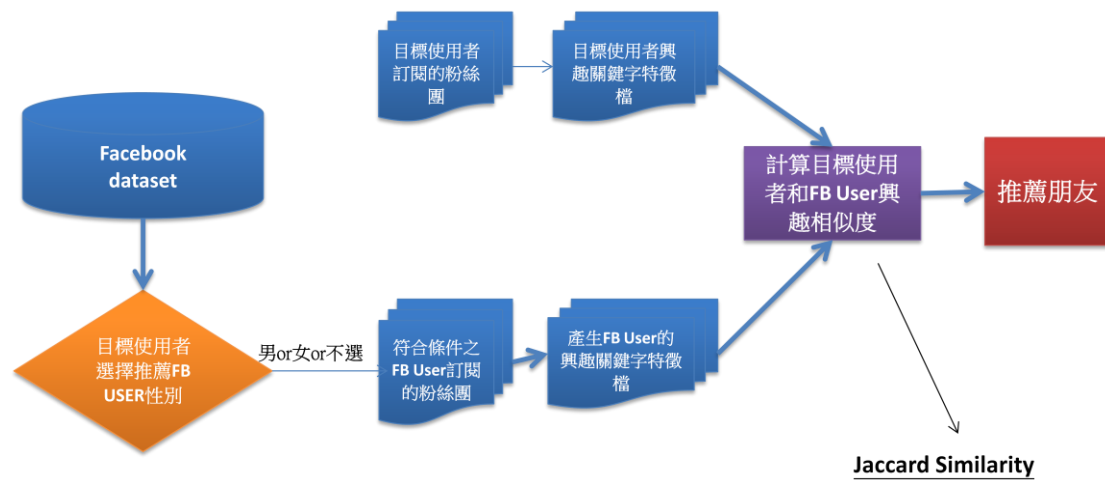


圖 4.29 推薦朋友模組架構圖

#### 4.4 推薦廣告模組

#### 4.4.1 推薦廣告模組架構

廣告推薦模組是利用先前產生的使用者興趣關鍵字特徵檔，再擷取廣告關鍵字，最後進行廣告和使用者興趣關鍵字特徵檔相似度比對，取得和目標使用者興趣相似的廣告做為推薦（圖 4.30）。

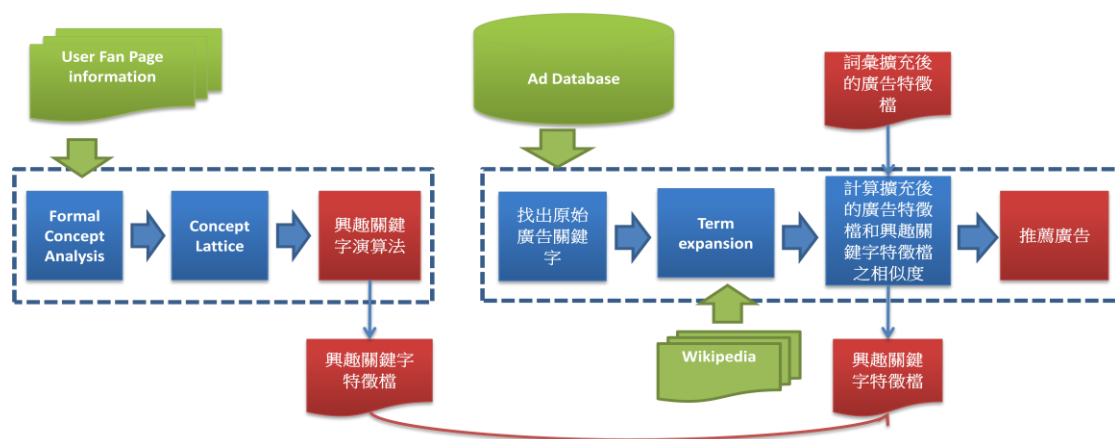


圖 4.30 廣告推薦模組架構

因為本模組是附加在其他模組之下，換言之使用者在使用朋友推薦或是粉絲團推薦服務時，廣告會出現在畫面之上，為了避免使用者對廣告產生反感，因而

在此只推薦三則廣告給目標使用者。

而在擷取廣告關鍵字方面，由於一般文字廣告只會呈現標題、摘要以及網址，這樣會造成只能擷取到少量的關鍵字，而無法和使用者興趣關鍵字特徵檔做配對的情況。因此在取得廣告關鍵字之後，我們會先經過詞彙擴充的程序，擴充廣告的關鍵字之後，再和使用者興趣關鍵字特徵檔做相似度計算。相似度計算在此模組同樣也用 Jaccard 相似度，最後取得相似度最高的前三則廣告做為推薦。

#### 4.4.2 廣告詞彙擴張

擴充廣告詞彙的方法，本研究利用了 Wikipedia 來做為詞彙擴張的來源。因為 Wikipedia 擁有豐富的資源，『是一個全球性多語言百科全書協作計劃，同時也是一部在網際網路上呈現的網路百科全書網站』。

詞彙擴張的步驟如圖 4.31 所示：

首先先將由 Yahoo 關鍵字 API 所得到的廣告關鍵字做為 seed，而後把每個 seed 傳送到 Wikipedia，進入 Wikipedia 的 entry page，利用 Yahoo 的關鍵字 API 取得此頁面的關鍵字，在此設定傳入參數 maxnum 等於 5，選擇此頁面在 Yahoo 關鍵字 API 計算中分數前五高的詞做為擴張的詞彙。



圖 4.31 利用 Wikipedia 從事詞彙擴張

#### 4.4.3 產生廣告推薦

本研究利用正規化概念分析以及興趣演算法分析使用者的粉絲團清單，來取得使用者的興趣關鍵字特徵檔，目標使用者的興趣關鍵字特徵檔和擴張後得到的廣告特徵檔計算相似度(式 4.1)，以及分數後(式 4.2)，最後選取三個廣告推薦給目標使用者。



## 5. 雛形系統建置與實作

依據前述架構，我們設計了一個 Facebook 推薦系統網站，來說明整個系統之運作架構。

### 5.1 開發工具與平台

本雛型系統的開發工具與平台如下表 5.1 所示：

表 5.1 系統開發平台與工具

系統平台	Operating System	Windows 7
	Application Server	Apache Tomcat 7.0.22
	Database Server	My SQL v5.5.17
開發語言工具	Server 端	JAVA 2.0、JSP2.0
	Client 端	JavaScript
網頁開發工具	Adobe Dreamweaver CS5	
正規化概念分析工具	FCA extension module for MS Excel 2007 based on NodeXL	
關鍵字擷取工具	Yahoo 關鍵字 API	

### 5.2 建置雛形系統資料集

為了實作出推薦系統的雛型架構，在本研究一共搜集了 70 位自願受試者的粉絲團清單(Like List)，作為分析 Facebook 使用者興趣之資料集，這 70 位使用者平均每人訂閱 44.56 個粉絲團。由於蒐集粉絲團清單牽涉到隱私的問題，故實驗蒐集之資料都以匿名來保護。

在之後的雛型系統示範，都會以匿名使用者「Nancy」作為目標使用者，模擬 Nancy 使用本系統的各種情境，包括使用推薦粉絲團系統、使用推薦朋友系統之情況，而廣告推薦的結果則會依據 Nancy 的興趣關鍵字特徵檔進行推薦，附加



在兩項服務(推薦粉絲團系統、推薦朋友系統)的頁面上。

使用者 Nancy 一共訂閱了 102 個粉絲團(表 5.2)，經過正規化概念分析後，可得到 Nancy 的 Concept Lattice(圖 5.1)，一共包含了 266 個概念(Concept)。Concept Lattice 中的每個節點都代表一個概念，圖 5.1 之節點 A 範圍(Extent)就包含了兩個粉絲團(物件)，而內涵(Intent)則包含了一個關鍵字電子商務，也可以視為兩個粉絲團之共同關鍵字。

表 5.2 匿名使用者 Nancy 訂閱之粉絲團

威秀影城粉絲團	David Bai	我是台北人	2012 跨年活動、跨年 去哪裡	(轉)一個女人突然決絕的跟 相愛五年的男友分了手，閃電 般嫁了別人。所有人都有個疑 問，明年你還愛我嗎？
Grandmon Nitti's Kitchen 中西美食	Shopping Design 設計採買誌	早晨起床後女生和 男生的差別 (超生 動超好笑!!!) - 天空の城	女生對男生行為的解 讀! [準到爆!!!笑死 我了~~] - 天空の城	商業周刊
吳尊	我可能不會愛你	【超爆笑!】各星 座等公車的反應 - 天空の城	iPhone & iPad 粉絲 同好會	程又青俱樂部
國立交通大學梅 竹後援會	MOS Burger 摩斯 漢堡「癮迷」俱樂 部	安心亞	大台北美食同好會	3M 台灣
Richard Gale Films	Dazzling Cafe ☆ honey toast 蜜糖 土司專賣店☆	韓國整型美女真實 紀錄-變美的代價	李大仁	報天氣
會長是女僕大人	程式/設計宅民歷	Dunkin' Donuts 台	21 Century 風味館	女人心事姊妹淘

		灣官方粉絲團		
長庚大學	簡易食譜	iLOOK 愛電影大學 電影社	長庚大學資訊管理學 系所	MUJI 無印良品生活研究所
熱門文章分享	你為什麼要檢舉我	食尚玩家 368 城鄉 故事	辛卯梅竹	屈臣氏 Watsons Taiwan
南崁高中	女人幫	長庚大學	【嘆】氣粉絲團	GOMAJI 夠麻吉
Badminton	研究生 2.0	物化男人	Bigtom 美國冰淇淋 文化館	BabyHome 寶貝家庭親子網
反"ㄟ一、物羅" 粉絲 CLUB	百萬人民站出來 - 大明星陳昱年國際 粉絲團	射 手 座 11/23~12/21	鳳眼	Turn Off the Lights
行銷就該這樣玩	美妝情報資訊粉絲 分享團	SALE 特賣會情報	阿默蛋糕	Uniqlo Taiwan
植物大作戰	DHC	哈 佛 商 業 評 論 Manager Club	福勝亭 TONKATSU 幸 福幸運的日式豬排專 賣	日本藥妝失心瘋俱樂部
GOHAPPY 線上快 樂購	HAPPY GO 點數・免 費賺！	我是桃園人	原來是美男啊~~真的 是美男唷?!	遠雄 easyHome 線上購物
心理測驗粉絲團	英文法：只是一些 規則而已	開心水族箱	新竹美食同好會	Kvideos
反范植偉!!!	樂 天 市 場 台 灣 (Rakuten Ichiba Taiwan)	Orzz	有這樣一種單身總有 人問你，有對像沒？ 呵呵，沒有呢。不會 吧，不可能吧 ……	7-ELEVEN
三 商 巧 福	havaianas 哈瓦仕	長庚大學	排隊快樂送，百萬粉	守口如瓶的感動

Always tasty , always for you	Taiwan		絲天天搶，獎品送不完	
馬英九總統	海賊王	Citytalk 城市通	不正常人類研究中心	壽星優惠情報站
小時候，幸福是一件簡單的事。 長大了，簡單是一件幸福的事。	生田斗真 ( Ikuta Toma )	驚!!!!!!! 日本變態正妹人皮面具 只要戴上人人都是正妹(6P) - 天空の城	MIX 時光機	裕榮食品公司蝦味先
反周玉搥粉絲團!!	日本的野心，核電廠背後的秘密！	驚人的擁抱	Mister Donut	Mister Donut 統一多拿滋官方粉絲團
加州椰子	台灣好市多 COSTCO 必買物			

64 c(61)	{123307891060857; 131307770301236}	(美英)
65 c(62)	{131307770301236}	(大聲公; 全家; 全球; 免費索取; 希望; 屈臣氏)
66 c(63)	{132670053446111}	(麻吉; 團購)
67 c(64)	{135084993203016}	{babyhome; BabyHome買家庭親子網; 大家訪}
68 c(65)	{107481239209090; 139064836124632}	{contact; youtube}
69 c(66)	{125585724754; 139064836124632}	{facebook; google}
70 c(67)	{139064836124632}	{amazon; browser; cheers; cinema; coming soon; c}
71 c(68)	{140574649321224}	{facebook; mail; unqlo; 上班族; 工作牌; 太平洋}
72 c(69)	{112294068801589; 141877778469}	(身分; 粉絲)
73 c(70)	{130324940344962; 141877778469}	(muji)
74 c(71)	{141877778469}	(br; Forever 21; Hello Kitty; muji; pixnet; 日本郵)
75 c(72)	{142655532432672}	(台東縣政府; 伊莎貝爾; 百事特童裝; 花蓮海)
76 c(73)	{108185025901296; 145312278848600}	(facebook; 喜歡)
77 c(74)	{119210047758; 145312278848600; 197910}	(謝謝)
78 c(75)	{145312278848600}	(facebook; fail; http; 一個人; 比基尼女郎; 告訴)
79 c(76)	{101077716629839; 145649977189}	(大遠百; 粉絲)
80 c(77)	{113224335370443; 145649977189}	(電子商務)
81 c(78)	{127962460631722; 145649977189}	(高雄夢時代)
82 c(79)	{130324940344962; 145649977189; 184645}	(生活)
83 c(80)	{145649977189}	(email; ibon; 大遠百; 加加; 生活; 卓越; 便當; 5)
84 c(81)	{107712535928724; 108185025901296; 147}	(第一次)
85 c(82)	{14711136868649}	(小友; 小故事; 小新; 卡內基; 平德鞋; 名偵探)
86 c(83)	{147994801904506}	(六福客棧; 炭火燒肉; 特香齋西餐廳; 情報站)
87 c(84)	{14885214185545}	(公司; 食品; 裕榮; 蝦味先)
88 c(85)	{119500671457377; 149342411791821; 624}	(日本)

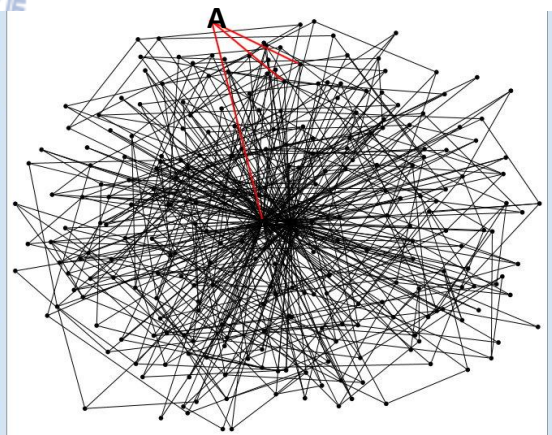


圖 5.1 匿名使用者 Nancy 之概念點陣圖

圖 5.2 c(23)和 5.3 c(80)兩節點皆為圖 5.1 A 節點的子孫。可以觀察出這兩個概念之範圍皆為圖 5.1 節點的子集合，而內涵則為圖 5.1 節點的超集合，也就是圖 5.2、5.3 這兩個概念皆有包含電子商務這個屬性(關鍵字)。

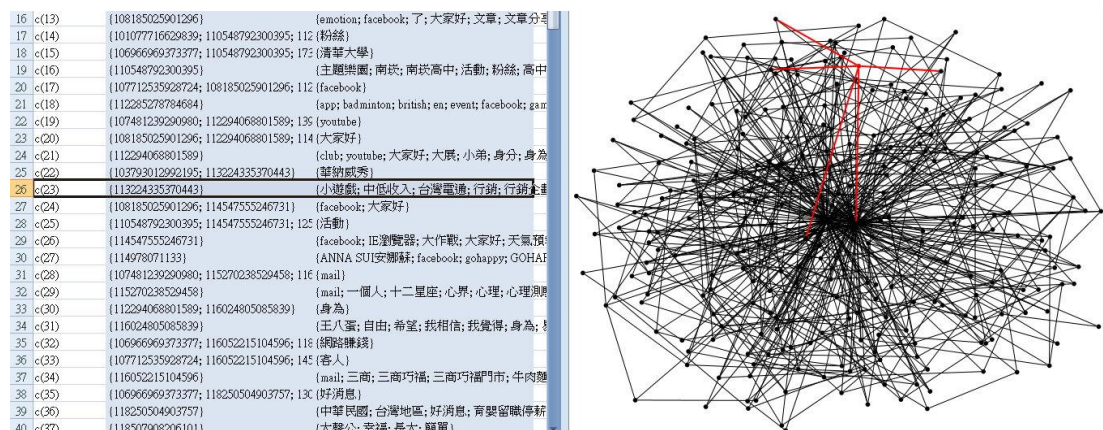


圖 5.2 節點 A 之子孫 1

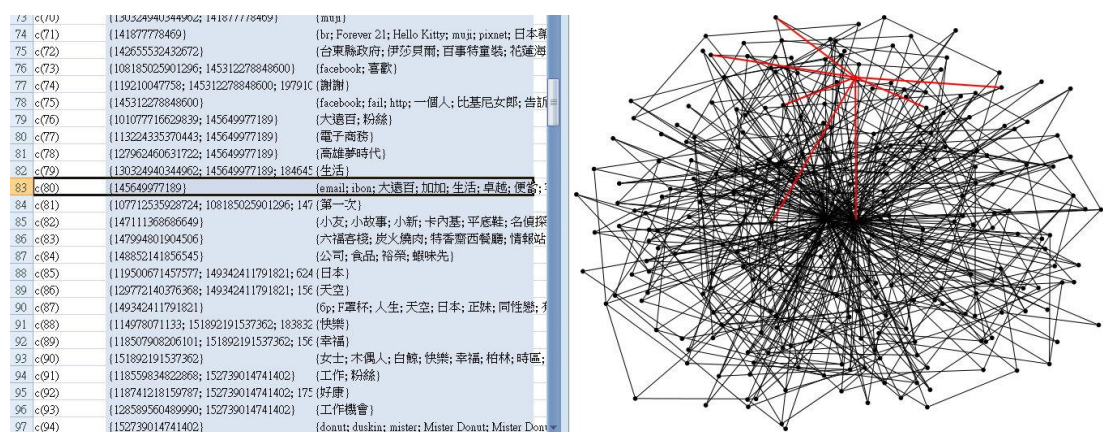


圖 5.3 節點 A 之子孫 2

得到 Nancy 之全部概念以及 Concept Lattice 後，經由興趣演算法，在本研究設定興趣演算法的參數為 CS\_THRESHOLD=0.5、K=10，最多只找 10 個概念作為使用者興趣關鍵字。最後可以得到使用者 Nancy 的興趣關鍵字特徵檔(表 5.2)，共 49 個詞彙，作為爾後系統個人化推薦之使用。

表 5.3 使用者 Nancy 之興趣關鍵字

1111 人力銀行	award	有趣圖片	天空	Adobe Flash Player
12 星座	babyhome	活動	女生	南崁
21 世紀	BabyHome 寶貝 家庭親子網	淡江	日本	高中

21 世紀烤雞	badminton	淡江大學	蘭陽平原	國立清華大學
3m	bai	清華大學	ANNA SUI 安娜蘇	Amazon
5566	BEVERLY HILLS	無名	圖片	南崁高中
7-11 ibon	bigtom	男生	主題樂園	101 大樓
7-11 禮券	blog	粉絲	app	Gmail
7-11 禮卷	british	長庚	暑期營隊	Animax
92 學年度	browser	長庚大學	大學	

### 5.3 Facebook 推薦系統實作內容

圖 5.4 為系統首頁。網頁左側列出系統功能，右側則有登入介面，使用本系統必須先登入，才能取得使用者的資訊。

圖 5.5 為整個操作流程圖，在使用者登入完網站後，系統會依據使用者的粉絲團清單來產生出使用者的興趣關鍵字，並顯示在首頁下方(圖 5.6)。登入完畢系統取得使用者關鍵字後就可以選擇使用推薦朋友系統或是推薦粉絲團系統，如果選擇推薦朋友系統，另外還可以選擇要推薦男性朋友、女性朋友或是都可。



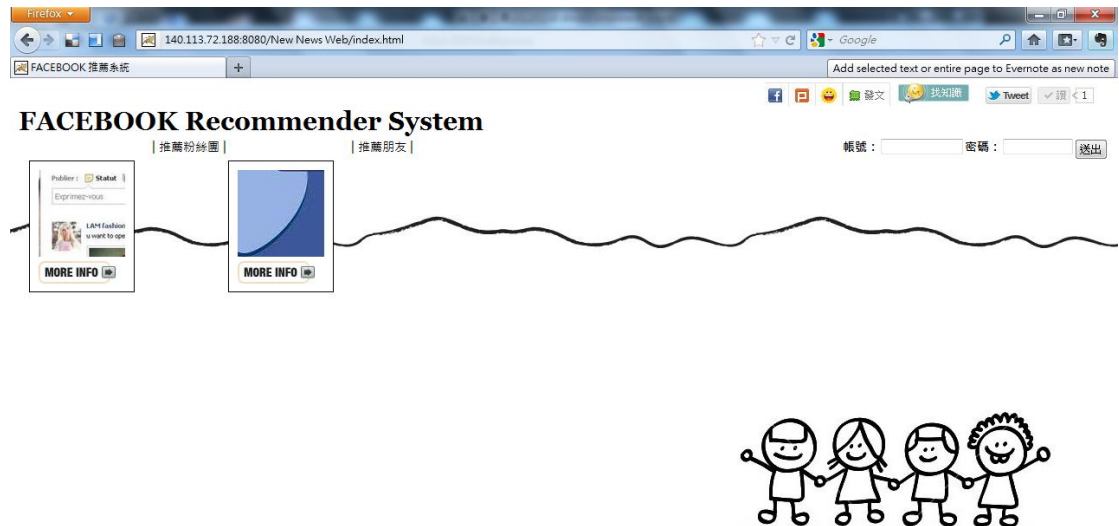


圖 5.4 Facebook 推薦系統首頁

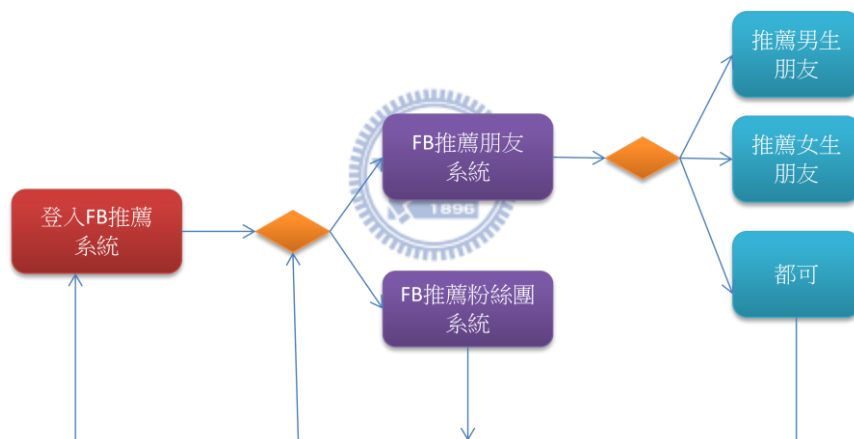


圖 5.5 Facebook 推薦系統操作流程圖

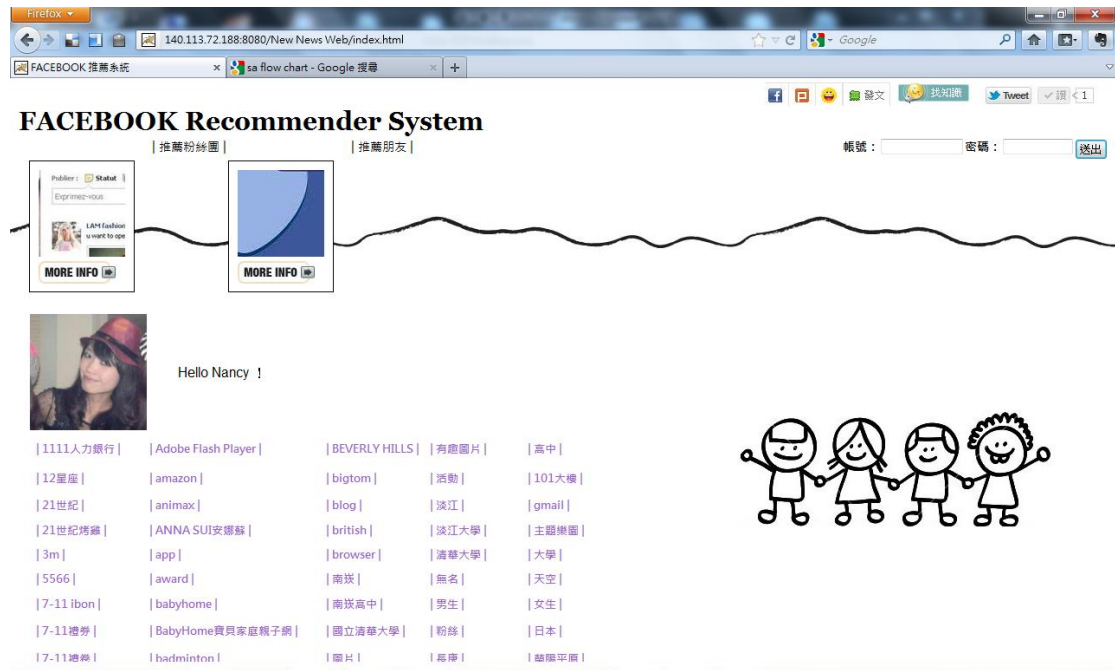


圖 5.6 使用者登入後頁面

#### 5.4 推薦粉絲團系統實作內容

選擇推薦粉絲團系統(圖 5.7)，接著會跳到另外一個頁面，依據先前產生的使用者的興趣關鍵字特徵檔以及粉絲團關鍵字進行相似度計算後，利用協同過濾法做粉絲團推薦，給予使用者推薦結果(圖 5.8)，而右方的廣告則是依照使用者興趣關鍵字和廣告做相似度計算後得到的結果，5.6 節會再詳加說明。



圖 5.7 選擇推薦粉絲團系統

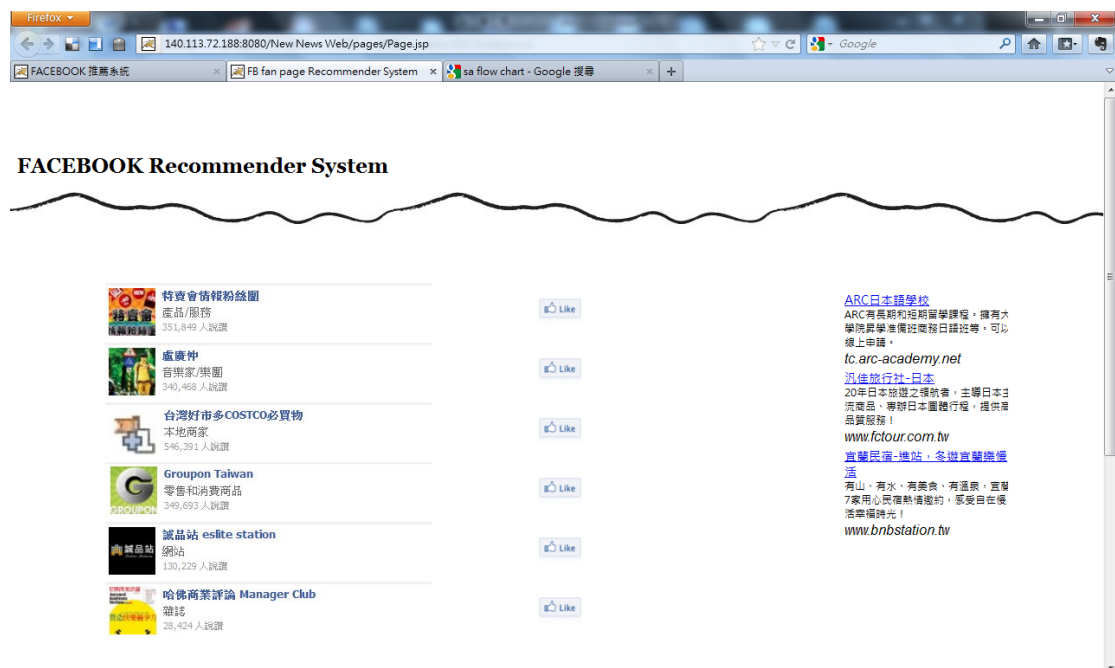


圖 5.8 推薦給 Nancy 之粉絲專頁

在推薦粉絲團系統，會推薦給目標使用者 Nancy 6 個粉絲團。如圖 5.8 所示推薦特賣會情報粉絲團、盧廣仲、台灣好市多 COSTCO 必買物、Groupon Taiwan、



誠品站 eslite station、哈佛商業評論 Manger Club 此六個網站。

而推薦此六個粉絲團給 Nancy 的原因，可以參考表 5.3 之數據。因為本雛型系統只使用了 70 人做建置，無法取大量人數做為鄰居，所以只選擇和 Nancy 之間相似度最大的前十名做為 Nancy 的最鄰近集合(Nearest neighbor)從事推薦。

最後使用式 4.1、4.2 計算出相似度及分數，取出 TOP 6 推薦給 Nancy。

表 5.4 推薦粉絲團結果

		1	2	3	4	5	6
	page id	140296556956	25444417410	119210047758	163125670396075	118569191022	208963391067
Jaccard Similarity	page name	特賣會情報粉絲團	盧廣仲	台灣好市多 COSTCO 必買物	Groupon Taiwan	誠品站 eslite station	哈佛商業評論 Manger Club
0.0645	Abigale						
0.0674	Aimee	1		1			
0.0842	Ashely		1		1		
0.0682	Hellen	1		1	1	1	
0.0745	Lichin		1		1	1	
0.0833	Maria						1
0.0755	Peter Yang	1		1		1	1
0.0714	Sophia	1	1				
0.1375	Terry Liu	1	1	1			1
0.0879	Vivian				1	1	
原始分數		0.4200	0.3676	0.3486	0.3148	0.3060	0.2963
score		1.0000	0.8753	0.8299	0.7495	0.7287	0.7055

## 5.5 推薦朋友系統實作內容

在首頁選擇推薦朋友系統之後(圖 5.9)，即會跳到推薦朋友系統頁面(圖 5.10)。此功能同樣是要登入後取得使用者特徵關鍵字後方可使用。



圖 5.9 選擇推薦朋友系統

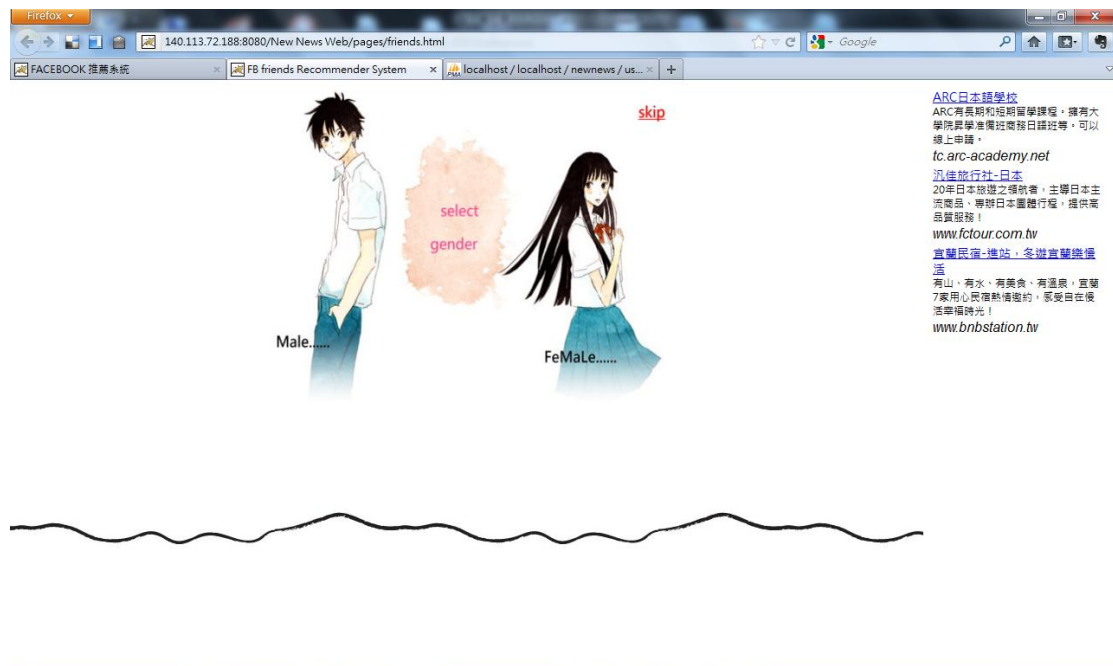


圖 5.10 推薦朋友系統頁面

在圖 5.10 中可以發現右邊同樣依據使用者興趣關鍵字顯示出了廣告。而在主頁面的圖片有三個連結可以點選，可以點選「Male、Female、Skip」，Skip 代表不挑選推薦朋友的性別。

點選 Male 圖像連結之後，可以發現圖像左方出現推薦朋友名單(圖 5.11)，系統幫 Nancy 推薦了男性的 Facebook 使用者做為加入朋友的參考；點選右方 Female 圖像連結，會在畫面右方出現女性的推薦朋友名單(圖 5.12)；點選 skip 圖像連結，則會在畫面下方出現不分性別的朋友推薦清單(圖 5.13)。

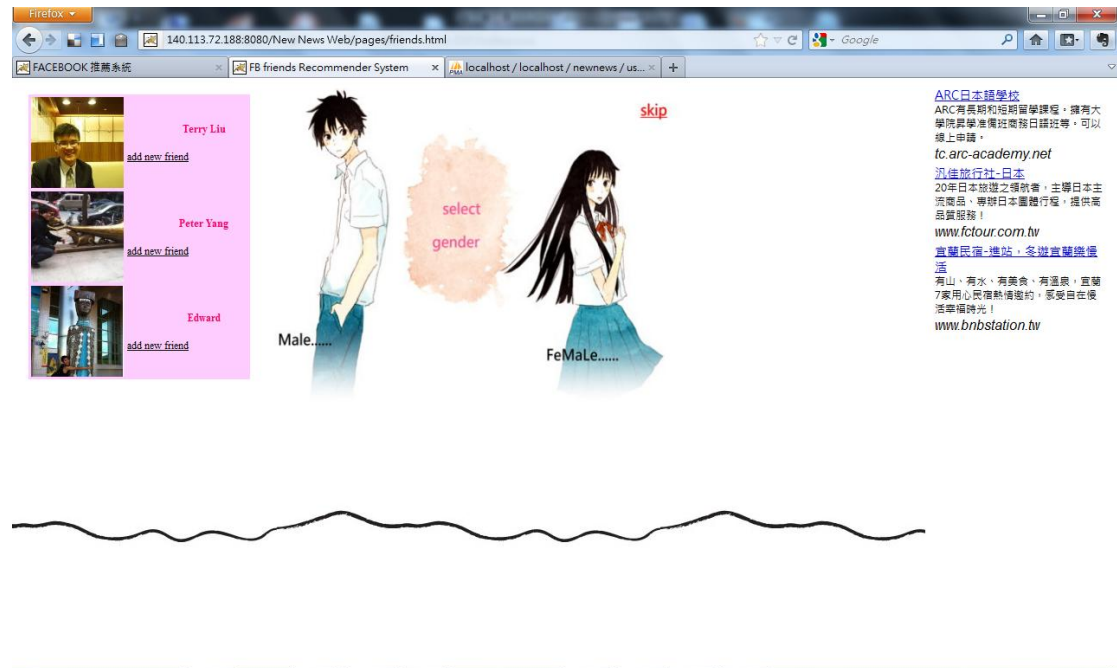


圖 5.11 系統推薦男性朋友清單

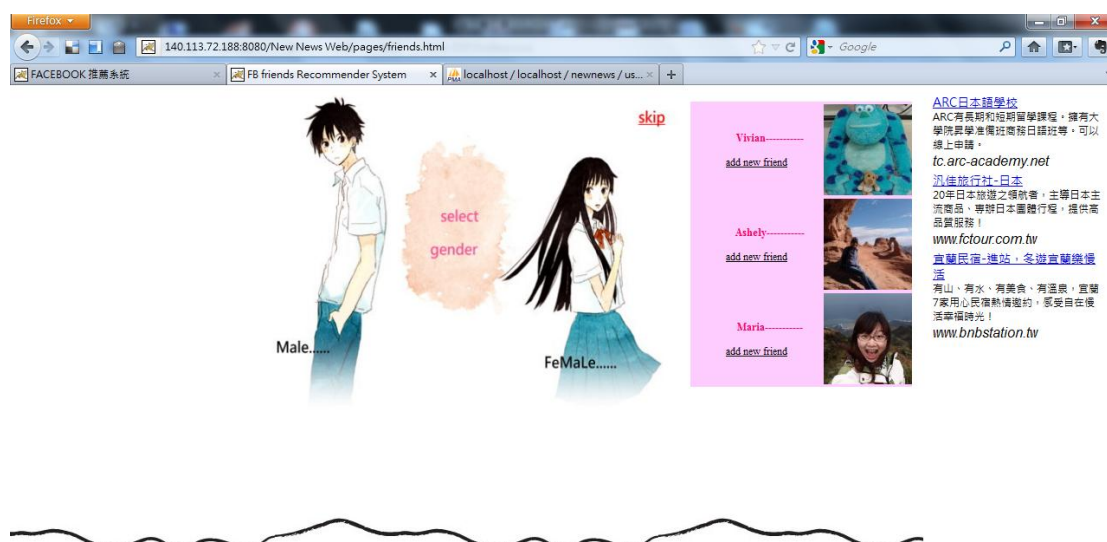


圖 5.12 系統推薦女性朋友清單

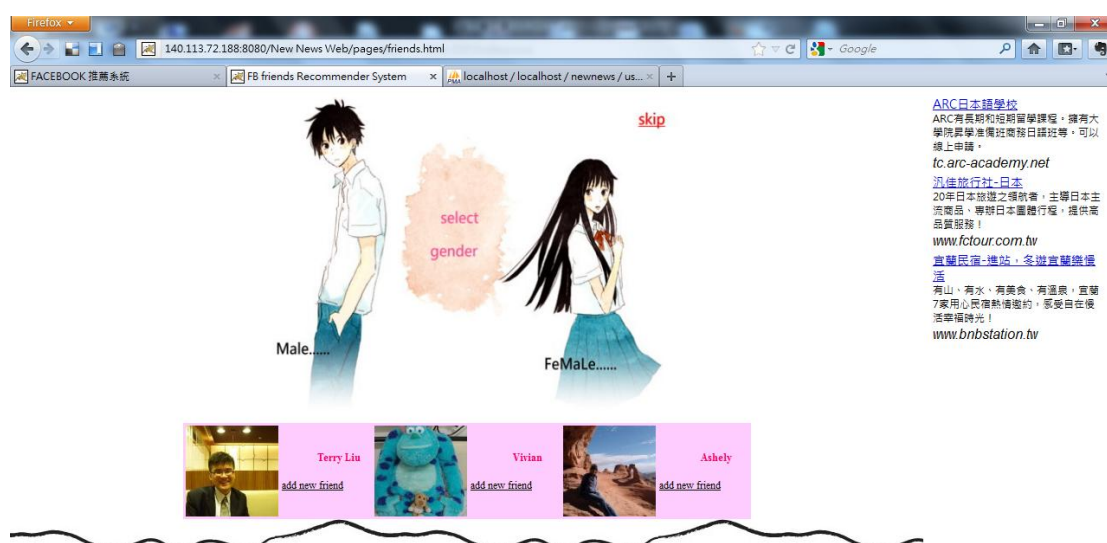


圖 5.13 系統推薦不分性別朋友清單

如的四章所敘，本推薦系統推薦朋友為依照使用者間之興趣關鍵字特徵檔相似度做 Jaccard 相似度計算來從事推薦，計算出結果如表 5.4。由表觀察，如果

推薦朋友，系統會推薦 Terry Liu、Vivian、Ashely；如果指定推薦男生朋友，系統會推薦：Terry Liu、Peter Yang、Edward；如果指定推薦女生朋友，系統則會推薦 Vivian、Ashely、Maria。使用者可以點選連結到推薦朋友之頁面，再決定是否要傳送邀請。

表 5.5 Nancy 和其他使用者之間的相似度

user	similarity	gender
Terry Liu	0.1375	M
Vivian	0.087912	F
Ashely	0.084211	F
Maria	0.083333	F
Peter Yang	0.075472	M
Lichin	0.074468	F
Sophia	0.071429	F
Hellen	0.068182	F
Aimee	0.067416	F
Abigale	0.064516	F
Ada	0.063291	F
Ailsa	0.0625	F
Edward	0.0625	M

## 5.6 廣告推薦實作說明

### 5.6.1 資料集

本雛型系統之廣告推薦模組的廣告來源是 Yahoo 奇摩拍賣的內容廣告。Yahoo 奇摩拍賣把商品一共分為 25 種類，分別為女裝與服飾配件、女包精品與

女鞋、美容保養與彩妝、男性精品與服飾、手錶與飾品配件、嬰幼兒與孕婦、電腦軟硬體與 PDA……而進入每一種大分類的子分類，分類下方都會顯示與該分類有關的廣告內容，如圖 5.14、15 為女裝服飾與配件之子分類(圓領長 T)該分類的下方所顯示的廣告。

為了能蒐集各種類型的廣告，本研究一共蒐集了 25 種類中較熱門的 16 類，每類蒐集 10 則，一共 160 則廣告。16 類廣告分別為：女裝與服飾配件、女包精品與女鞋、美容保養與彩妝、男性精品與服飾、手錶與飾品配件、嬰幼兒與孕婦、電腦軟硬體與 PDA、相機攝影與視訊、手機與通訊、電玩遊戲與主機、運動戶外與休閒、寵物用品與水族、玩具模型與公仔、圖書與雜誌、旅遊、生活服務等。

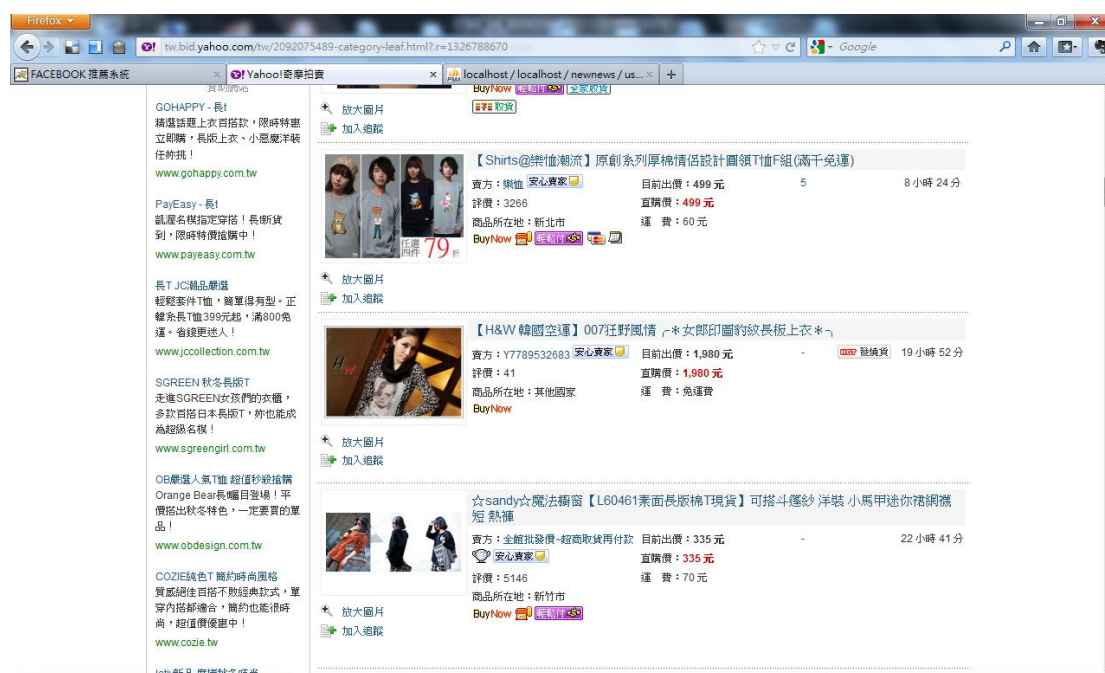


圖 5.14 奇摩拍賣女性長 T 頁面





圖 5.15 奇摩拍賣女性長 T 之分類廣告

#### 5.6.2 廣告推薦實作

在本研究廣告推薦的方法和朋友推薦的方法相似，差異點在於使用者的興趣特徵檔是由許多粉絲團之塗鴉牆及基本資訊蒐集而來，用來推估使用者之興趣資料較充裕；反之從奇摩拍賣蒐集之文字廣告，僅有標題、摘要和網址 3 部分(圖 5.15)，資訊有限，無法擷取足夠關鍵字，因此在這裡把 Yahoo API 傳入參數 maxnum 等於 5，代表最多只從廣告內容擷取 5 個關鍵字。因為廣告關鍵字的不足，所以還要經過適當的關鍵字擴充程序，才可以擷取到比較多詞彙做為廣告特徵檔，以增加廣告和使用者興趣關鍵字特徵檔交集的機會。

本系統先前示範推薦給使用者 Nancy 的廣告(圖 5.16)為 ABC 日本與學校、汎佳旅行社-日本、宜蘭民宿-進站，冬遊宜蘭樂慢活，為依照相似度以及分數的結果取得前三名的廣告，會顯示在粉絲團推薦系統及朋友推薦系統頁面上。原本僅使用 Yahoo 關鍵字 API 所取得的廣告關鍵字、經過 4.4.2 方法詞彙擴充後的關鍵字、相似度與分數之資訊彙總於表 5.5。

[ARC日本語學校](#)

ARC有長期和短期留學課程。擁有大學院昇學準備班商務日語班等。可以線上申請。

[tc.arc-academy.net](http://tc.arc-academy.net)

[汎佳旅行社-日本](#)

20年日本旅遊之領航者，主導日本主流商品、專辦日本團體行程，提供高品質服務！

[www.fctour.com.tw](http://www.fctour.com.tw)

[宜蘭民宿-進站，冬遊宜蘭樂慢活](#)

有山、有水、有美食、有溫泉，宜蘭7家用心民宿熱情邀約，感受自在慢活幸福時光！

[www.bnbstation.tw](http://www.bnbstation.tw)

圖 5.16 推薦給 Nancy 之廣告

表 5.6 詞彙擴充之推薦廣告結果

廣告標題	原本的關鍵字	擴充後之關鍵字	相似度	分數
ARC 日本語學校	arc,擁有,日本語,日本語學校,留學	學生, 日本, 維基百科, 日本語, 留學, 日語, 詞典, 國語詞典, 教育, 大學, arc, 高等教育, 日本語學校, 維基, 奧克蘭, 江戶時代, 擁有, 第一次	0.031	1
汎佳旅行社-日本	日本, 汎佳, 汎佳旅行社, 領航領航者	證券交易所, 日本, 領航, 汎佳旅行社, 汎佳, 國立故宮博物院, 日本自衛隊, 日本職棒, 領航者, 日本首相	0.017	0.55
宜蘭民宿-進站，冬遊宜蘭樂慢活	宜蘭, 宜蘭民宿, 慢活, 時光, 民宿	渡假民宿, 時光, 自由時報電子報, 行政院農業委員會, 交通部觀光局, 慢活, 民宿, 宜蘭民宿, 蘭陽平原, 宜蘭	0.017	0.55



## 6. 結論與未來研究方向

### 6.1 結論

本研究主要的貢獻有以下幾點：

1. 從 Facebook 的使用者所訂閱的粉絲團資訊得到使用者興趣關鍵字特徵檔之構想。
2. 取得粉絲團清單中粉絲團互相的關係及共同的特性。
3. 利用正規化概念分析及興趣演算法從使用者粉絲團清單挖掘出隱含的使用者興趣關鍵字特徵檔。先透過正規化概念分析建構出使用者所訂閱的粉絲團清單(usr\_likes 物件)與粉絲團的關鍵字之 Formal Context，再把原本的 Formal Context 圖形化成 Concept Lattice。隨後利用興趣關鍵字演算法挑選出不會太過一般化和特定化的概念(Concept)之內涵(Intent)做為使用者的興趣關鍵字特徵檔。
4. 提出了粉絲興趣關鍵字特徵檔的應用方向，例如推薦廣告、推薦粉絲團、推薦朋友等，每種應用方向都以使用者興趣關鍵字特徵檔作為出發點，再加以些微的變化。例如推薦廣告，為了要讓廣告能和使用使用者興趣關鍵字特徵檔能有更多的重合，因此在廣告部分做了詞彙擴張的動作；推薦粉絲團則是相似度利用使用者興趣關鍵字特徵檔來計算，再結合傳統協同式過濾推薦(Collaborative Filtering)演算法做粉絲團的推薦；推薦朋友則考量了實務與行銷面，讓使用者能先做性別的過濾。

### 6.2 未來研究方向

在此綜合未來可以加入，以讓預估使用者興趣能夠更準確之研究方向：

#### 1. 關鍵字的擷取

本研究為使用 Yahoo 的關鍵字擷取 API，但 Yahoo 關鍵字 API 所擷取的關鍵字有太過特定的問題，例如「1111 人力銀行」如可斷為「1111」及「人

力銀行」更能讓使用者興趣特徵檔更加一般化，特徵檔之間重複機會可以增加，相似度值也可以提升。

且關鍵字擷取是整個分析非常重要的一部分，可以左右推薦準確度，因此以後可以加入其他更精確關鍵字擷取模組取代原本的 Yahoo 關鍵字 API。

## 2. 關鍵字的詞彙擴張

本研究在廣告推薦部分有為每則廣告做關鍵字的擴張，但其實使用者興趣關鍵字特徵檔也可以做詞彙擴張的動作。另外詞彙擴張方式，可以結合其他詞彙擴張的方式或是擴張後再加以人工篩選，讓使用者興趣關鍵字特徵檔能夠更準確。

## 3. 依據使用者訂閱的粉絲團多寡來決定 K 值

本研究的 K 值取 10，也就是無論使用者訂閱多少粉絲團，系統取的概念(Concept)數皆為 10 個。但在實際應用上可以依據使用者訂閱粉絲團的數量來決定 K 值，訂閱較多粉絲團的使用者可以給予較高的 K 值，訂閱較少的使用者則給予較低之 K 值，彈性的決定 K 值較能取得正確數量的概念，得到較精確的使用者興趣關鍵字特徵檔。

## 4. 加入 Facebook 的其他資訊

除了粉絲團塗鴉牆內容及基本資訊外，也可以加入一些使用者註冊時自願填寫的資訊，以讓預測使用者興趣的資訊能夠更充足。

## 5. 粉絲團重要度

因為每個粉絲團型態不一，有些粉絲團主題明確，有些粉絲團則是以分享資訊居多，例如分享特賣消息的粉絲團，每篇文章的內容可以都大相逕庭，此類型的粉絲團如果能先被過濾掉的話，不但 Concept Lattice 的節點能變少，更重要的是可以增加建置使用者興趣關鍵字特徵檔的準確度。

因此在進行正規化概念分析之前可以先進行粉絲團內部文章相似度計算，如果粉絲團內部文章相似度高，代表此粉絲團有在討論較特定的主題，

可以把此粉絲團列入後續之計算；如果粉絲團內部文章相似度低，代表此粉絲團張貼之訊息主題可能較散亂，則予以剔除在後續之分析。

## 6. 其他的應用方向

雖然本研究舉出三種應用使用者興趣關鍵字特徵檔的方式，但如果能準確挖掘出使用者的興趣，應可做更多方向的應用，也是未來可以繼續發展研究的課題。



## 參考文獻

- [1] Balabanovic, M., Shoham, Y. (1997), Fab: Content-Based Collaborative Recommendation, Communications of the ACM 40(3):66-72.
- [2] Breese, J. S., Heckerman, D. and Kadie, C. (1998), Empirical analysis of predictive algorithms for collaborative filtering, Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence.
- [3] Carpineto, C., Romano, G. (2000), Order-theoretical ranking, Journal of the American Society for Information Science 51(7):587-601.
- [4] Ganter, B., Wille, R. (1999), Formal Concept Analysis. Mathematical Foundations, Springer Verlag.
- [5] Herlocker, L. J., Konstan, A. J. and Ried, T. J. (2002), Empirical analysis of design choices in neighborhood-based collaborative filtering algorithm, Information Retrieval 5(4): 287-310.
- [6] Linden, G. D., Jacobi, J. A. and Benson, E. A. (2001), Collaborative recommendations using item-to-item similarity mappings, US Patent 6266649.
- [7] Priss, U. (2006), Formal concept analysis in information science, Annual Review of Information Science and Technology 40:521-543.
- [8] Sullivan, D. O., Wilson, D. C. and Smyth, B. (2002). "Improving case-based recommendation: A collaborative filtering approach", Proceedings of the Sixth European Conference on Case Based Reasoning: 278.
- [9] van der Merwe, D., Obiedkov, S. and Kourie, D. (2004), AddIntent: A new incremental algorithm for constructing concept lattices, in Lecture Notes In Artificial Intelligence 2961.

- [10] Wille , R.(1982), Restructuring lattice theory: an approach based on hierarchies of concepts, In: I. Rival (ed.): Ordered sets: 445 – 470.
- [11] Zhang , Y., Feng, B. (2008). Tag-based user modeling using formal concept analysis, Proceedings of Computer and Information Technology, IEEE Press, :485 – 490.
- [12] ARO Report, Jan. 2010. Available at:  
[http://tw.emarketing.yahoo.com/ems2/superior/performance/performance\\_03.php?menu\\_tree=D&menu\\_sub\\_tree=a](http://tw.emarketing.yahoo.com/ems2/superior/performance/performance_03.php?menu_tree=D&menu_sub_tree=a)
- [13] CheckFacebook.com Available at: <http://www.checkfacebook.com/>
- [14] CRM2 Available at: <http://crm2.tw/?p=964>
- [15] INSNA: What is Social Network Analysis Available at:  
<http://www.insna.org/sna/what.html>
- [16] Formal concept analyse Available at:  
[http://www.fca.radvansky.net/photogallery.php?album\\_id=2](http://www.fca.radvansky.net/photogallery.php?album_id=2)
- [17] Wikipedia Available at:  
<http://zh.wikipedia.org/wiki/%E5%8D%94%E5%90%8C%E9%81%8E%E6%BF%BE>