

R 軟體與 Facebook API - Text Mining

江奕 副統計分析師

生統 eNews 於上一期中，帶大家認識了 R 軟體入門的相關基本應用與資料型態介紹等功能，希望大家對這套軟體已能有更深入的了解，又到了出刊的時間，本期 eNews 將進一步教大家如何使用 R 軟體串接 Facebook 所提供的 API，擷取任何您有興趣的社群資料用於分析上。

想撈取 Facebook 上的資料，得先取得通過 Facebook 的權限認證機制，以下幾個步驟教大家如何順利建立一個 Facebook App。

步驟一：

進入 <https://developers.facebook.com/> FB 開發人員的網頁，並於右上方我的應用程式中，點選新增應用程式。

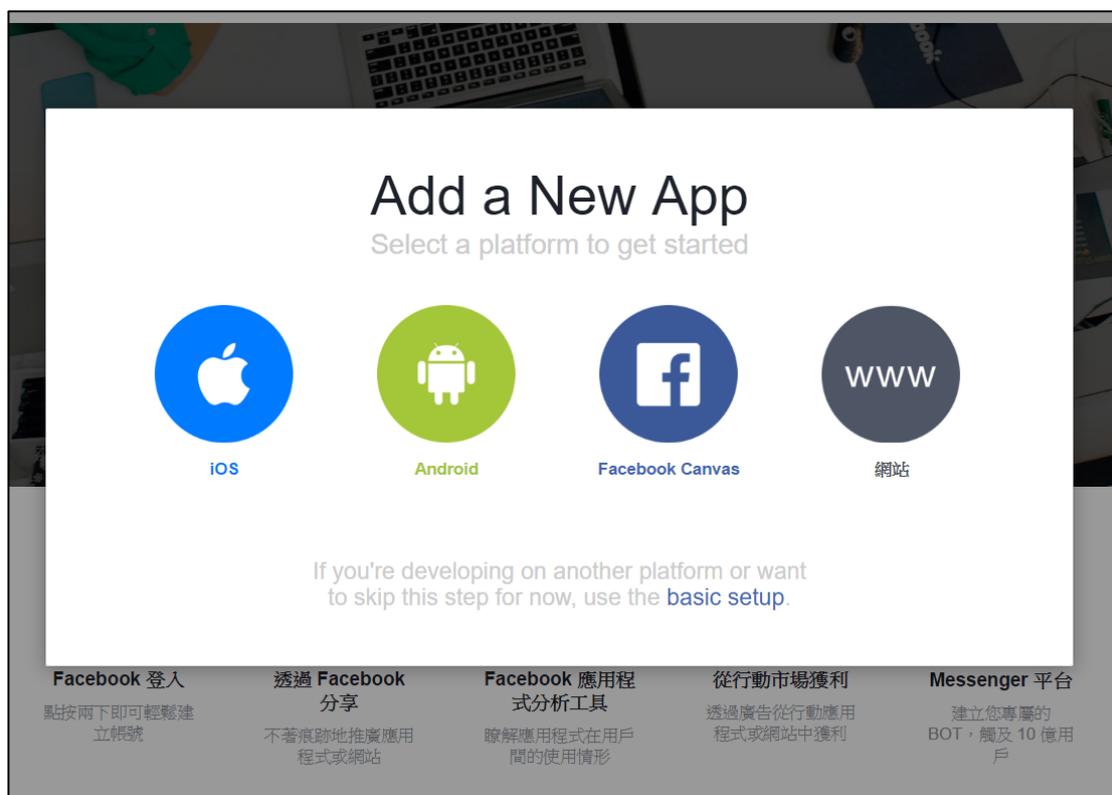


The image shows a screenshot of the Facebook Developers website. At the top, there is a navigation bar with the text "facebook for developers" and several menu items: "產品", "文件", "工具及支援", "最新消息", and "個案研究". A search bar and a user profile dropdown labeled "我的應用程式 JJ" are also visible. The main content area features a large banner with the text "緊密連結、放眼全球" and "全方位打造高人氣、高獲利應用程式的堅強陣容：Facebook". Below the banner, there are five icons representing different services: "Facebook 登入", "透過 Facebook 分享", "Facebook 應用程式分析工具", "從行動市場獲利", and "Messenger 平台". Each icon has a corresponding description below it.

Facebook 登入	透過 Facebook 分享	Facebook 應用程式分析工具	從行動市場獲利	Messenger 平台
點按兩下即可輕鬆建立帳號	不著痕跡地推廣應用程式或網站	瞭解應用程式在用戶間的使用情形	透過廣告從行動應用程式或網站中獲利	建立您專屬的 BOT，觸及 10 億用戶

步驟二：

選取您所新增的平台形式，在這裡我們點選最右邊的 WWW(網站)



步驟三：

這邊讀者可以取一個任意的 App 的名稱後，點選下方建立新的 Facebook 應用程式編號(Create New Facebook App ID)。



步驟四：

輸入欲顯示的名稱(此處不得輸入含有 FB 等字樣的名稱)，聯絡用的電子郵件，於類別項目下拉選取您的用途。



The screenshot shows the 'Create New App' form in the Facebook Developer console. The title is '建立新的應用程式編號' (Create New App ID). Below the title is the instruction '開始將 Facebook 整合到你的應用程式或網站' (Start integrating Facebook into your app or website). The form has three main sections: '顯示名稱' (Display Name) with a text input field containing 'Rtext'; '聯絡電子郵件' (Contact Email) with a text input field containing '我們會使用這個名稱和你溝通跟應用程式有關的重要事宜'; and '類別' (Category) with a dropdown menu showing '教育' (Education). At the bottom, there is a checkbox for '一旦繼續，就代表你同意 Facebook 開放平台政策' (By continuing, you agree to the Facebook Platform Policy) and two buttons: '取消' (Cancel) and '建立應用程式編號' (Create App ID).

步驟五：

點選右上角的 **Skip Quick Start** 或從網頁右上方我的應用程式選取剛剛所建立的 App 進入主控板。



步驟六：

點選左方灰色區塊主控板，右方則會顯示 API 版本、應用程式編號，應用程式密鑰。



步驟七：

點選應用程式密鑰右方顯示按鈕(可能須重新輸入 FB 密碼)。將此應用程式編號與應用程式密鑰先記錄下來，後續將用於撈取 Facebook 上的資料。



以上完成了 Facebook APP 的申請，接著我們利用應用程式編號以及密鑰就可利用 R 軟體來取用 Facebook 上的資料囉！

開啟我們的 R 軟體後，我們必須先安裝 R 套件”Rfacebook”用來連結 Facebook，安裝套件程式碼：install.packages("Rfacebook")。

```
1 install.packages ( "Rfacebook" )
2 library ( Rfacebook )
3 fb_oauth = fbOAuth ( app_id = "6344xxxx ... x",
4                       app_secret = "9cf4xxx ... x" )
```

程式碼第 1-2 行中分別為安裝 Rfacebook 套件以及載入套件，接著 3-4 行使用 fbOAuth 函數將步驟七的應用程式編號與應用程式密鑰分別輸入至參數中，app_id =”應用程式編號”，app_secret =” 應用程式密鑰”，執行程式後，會出現 Copy and paste into Site URL on Facebook App Settings: http://localhost:1410/ When done, press any key to continue...等訊息文字於 R 的 Console 中，接著將網址 http://localhost:1410/ 複製，開啟步驟六的主控板頁面，點選左方設定中的基本資料點取下方新增平台。



選擇網站。



將網址 `http://localhost:1410/` 輸入於網站網址中。



回到 Console 介面按下任意鍵後，會有以下訊息：

Waiting for authentication in browser...

Press Esc/Ctrl + C to abort

Authentication complete.

Authentication successful.

此時會自動開啟瀏覽器顯示 Authentication complete. Please close this page and return to R.

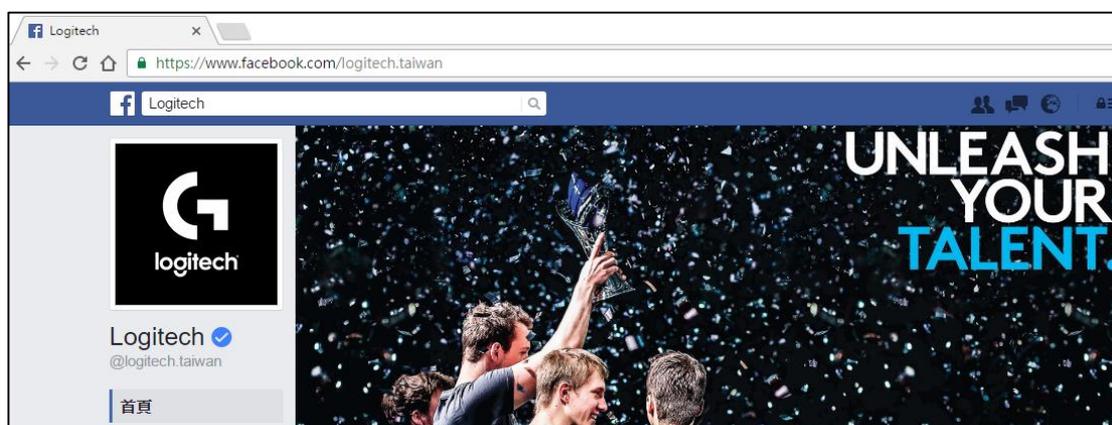
此處若有預先登入非申請 Facebook API 的使用者帳號可能會出現錯誤訊息，請將此帳號登出後，重新登入已申請過 Facebook API 的帳號即可。

```
5 dir_fb_oauth = "d:/fb_oauth"  
6 save(dir_fb_oauth, file = dir_fb_oauth )  
7 load( dir_fb_oauth )
```

程式碼第 5 行為欲存取的路徑，第 6 行使用 save 將 Facebook 存取權儲存起來，後續可以直接使用第 7 行程式 load 函數碼載入 fb_oauth 即可。接下來，可以尋找您感興趣的粉絲專業來擷取資訊囉！假設感興趣的是北醫粉絲版，就可將網址列後面的 ID：269261497877 複製。



這邊順帶提醒大家，ID 部分也有可能像羅技電子是英文 ID：logitech.taiwan



```

8 fb_page = getPage ( page="logitech.taiwan", token =
9 fb_oauth, n = 600 )
10 names( fb_page )

```

我們這邊使用羅技電子來做後續的範例操作，程式第 8 行使用 getPage 函數抓取，page 參數 ID 為 logitech.taiwan，token 參數中為剛剛所建立的認證，n 參數為擷取前 600 篇文章，讀取後可使用第 9 行查看 fb_page 有哪些名稱。簡述幾個較常使用到的資料變數。

"message"：官方貼出的文章。

"created_time"：官方貼文的文章時間

"likes_count"：每篇文章的按讚數

"comments_count"：每篇文章的留言數

"shares_count"：每篇文章的分享數

```

11 #查詢前 5 篇官方的貼文：
12 fb_page$message[ 1:5 ]

```

利用\$字符號與[]指標系統，可撈出官方所發表的文章。

以下為 R 的輸出結果

- [1] "這人生，美好的派對生活無「線」展開\n#Logitech
- [2] "【#羅技不藏私 黑暗領域技術的秘密】 \n 外出拜訪客戶
- [3] "【2016 台灣電競節】 三大亮點！
- [4] "為什麼羅技要稱它旗艦？\nG900 鼠王背後的秘密！
- [5] "【羅技 M720 Triathlon 多工滑鼠開放預購、(〇´▽`〇)ﾉ】

```

13 #查詢前 15 篇留言數目：
14 fb_page$comments_count[ 1:15 ]

```

以下為 R 的輸出結果

- [1] 0 1 0 1 4 0 2 0 1 2 1 1 15 21 2

```

15 # 單純取出粉絲頁的第 14 篇文章：
16 post = getPost ( post = fb_page$id[ 14 ],
17 token = fb_oauth )
18 # 顯示前 20 筆回應
19 comments = post$comments$message
20 comments[ 1:20 ]

```

以下為 R 的輸出結果

```

[1] "小羅對不起大家！(跪主機板 名單已補上 QwQ"
[2] ""
[3] ""
[4] "誰中啊？000"
[5] "我也來卡位 XD"

```

```

21 # 顯示前 6 筆按讚人資訊
22 head( post$likes )

```

以下為 R 的輸出結果

	from_name	from_id
1	林 xx	1395956xxxx
2	陳 xx	1004927xxxx
3	許 xx	3143544xxxx
4	Lee xx	1015455xxxx
5	葉 xx	1164919xxxx
6	Sylvia xx	1586205xxxx

```

23 messageTXT = fb_page$message[ 1:600 ]
24 dirDATA = "路徑"
25 write( messageTXT, file = dirDATA,
26        ncolumns = length( messageTXT ), sep="," )

```

程式碼 23-26 行將官發 PO 的 600 篇文章寫出到您的裝置中，待後續使用。

```

1 library( tm )           # Text Mining Package
2 library( Rwordseg )     # 中文斷詞
3 library( tmcn )        # 處理中文字的輔助套件
4 library( wordcloud )   # Word Clouds
5 library( colorspace )  # Color Space Manipulation

```

使用 R 軟體進行文字探勘前，需先安裝些許套件，其中包含，tm、Rwordseg、tmcn、wordcloud、colorspace 等，其中 Rwordseg 套件需安裝含有 Java 環境的作業系統中才能正常操作，若 R 軟體執行載入套件過程出現相關錯誤，請讀者前往 http://java.com/zh_TW/download/ 下載適合自己系統的版本進行安裝並執行 library(rJava)載入套件。程式碼 1-5 行為載入所需套件指令。

```

6 #Representing and computing on corpora. - tm
7 FbData = Corpus( DirSource( "路徑", encoding = "big5" ) )
8 #Interface to apply transformation functions to corpora. - tm
9 FbData = tm_map( FbData, stripWhitespace )
10 FbData = tm_map( FbData, removePunctuation )
11 FbData = tm_map( FbData, removeNumbers )
12 FbData = tm_map( FbData,
13 function( word ){ gsub( "[A-Za-z0-9]", "", word ) } )
14 FbData = tm_map( FbData, PlainTextDocument )

```

程式碼第 7 行將剛剛所存取的 600 篇文字檔使用 Corpus 函數讀取至 R 軟體(tm 套件)。9-14 行主要為資料的整理，包含清除多餘的空白鍵(Strip extra whitespace from a text document. Multiple whitespace characters are collapsed to a single blank.)，清除標點符號(Remove punctuation marks from

a text document.)，清除數字(Remove numbers from a text document.)以及英文大小寫等等，最後創建為一純文字檔(Create plain text documents.)。

```

15 StopWords = stopwordsCN( )
16 FbData = tm_map( FbData, removeWords, StopWords )
17 FbData2 = tm_map( FbData, ontent_transformer( segmentCN ),
18 nature = TRUE, returnType = 'tm' )
19 FbData3 = Corpus( VectorSource( FbData2 ) )
    
```

程式碼第 15-16 行主要為清除停止詞，例如：既然，以及，因此..等等字詞清除(tmcn 套件)，17-18 行將文字檔進行中文斷詞的工作(Rwordseg 套件)，最後存為最後拿來分析的語料庫中。

```

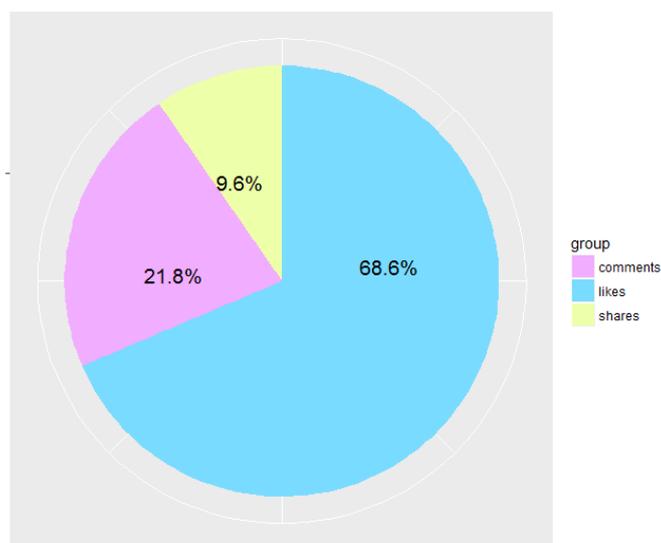
1 Fbcloud = segmentCN( FbData3[[1]]$content, nature = TRUE )
2 Fbcloud = unlist( Fbcloud )
3 noun = Fbcloud[ names ( Fbcloud ) == "n" ]
4 tab = table( noun )
5 Data = as.data.frame( tab[ tab >= 1 ] )
    
```

程式碼 1-5 行主要將資料取出只包含有名詞的文字，計算每個字詞出現次數並保留大於一次的部分。將其存為一資料框架變數。

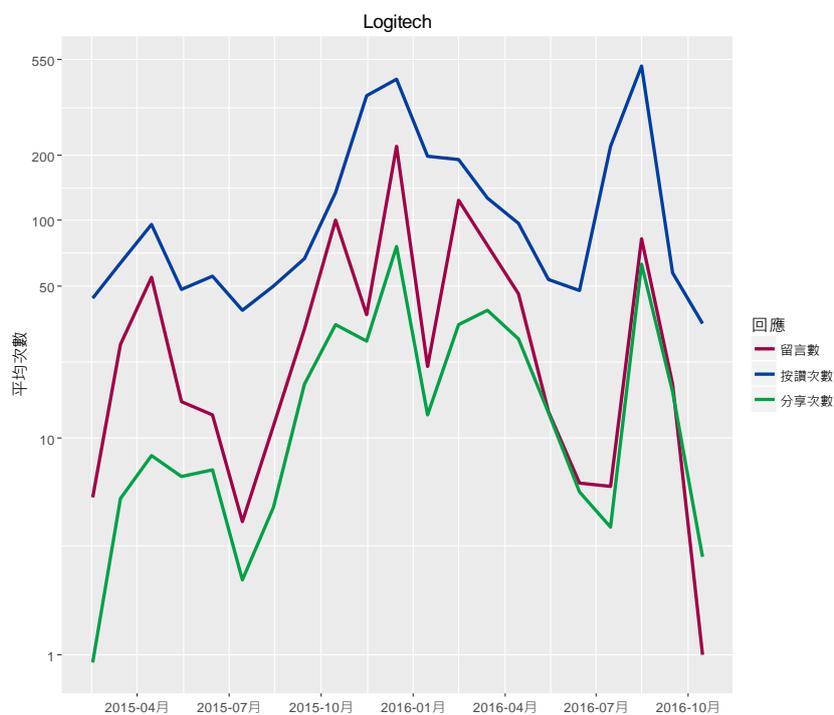
```

6 wordcloud(
7   words = Data$noun, freq = Data$Freq,
8   min.freq =8,
9   random.order = F, ordered.colors = T, scale=c(9,.8),
10  colors = rainbow_hcl( nrow( Data ) )
11 )
    
```


若想看回應留言數，按讚人數，分享人數的比例，可將 600 篇文章畫一圓餅圖。如下圖：



有興趣的讀者，也可以畫出月份對於留言數，按讚人數，分享人數的平均次數。如下圖：



上方的長條圖，圓餅圖，折線圖，都是使用 R 軟體中的 ggplot2 套件所畫出，有興趣的讀者可以參考網路上的資源進行實作看看！