作業內容:爬蟲及數據分析練習

目的:分析FaceBook好友數以及一年之內推文數之間的相關係數

學習內容:

1.使用爬蟲與FB api互動

2.編寫class

3.使用各類module

4.處理list, tuple, dict的資料

5.進行資料清理和分析

操作:

1.執行” static\_computing”檔案

2.至https://developers.facebook.com/tools/explorer 取得token (須包含使用者 post及 friends權限)

3.輸入token，計算自己與朋友一年內的發文數以及相關係數

說明:

一

FB\_parser:負責連結網頁取得好友數和一年之內的貼文數

有兩個field, 九個function，包含:

token:由使用者輸入

data:自動計算自己和好友一年內的貼文數及好友數，存成list

建構子:

\_\_init\_\_(self, token):輸入token

三個class method

get\_currentTime(self):取得現在的時間並加以處理成list輸出

inlastOneYear(self, dateInput):判斷輸入的日期是否在一年之內，回傳boolean值

getPage(self, url):輸入網址，轉成json檔案回傳

五個 method

get\_posts\_date(self,myData):輸入json，將檔案中posts的日期轉成list回傳

next\_page(self, myData): 取得下一頁的網址

get\_friend\_number(self, id = **'me'**): 取得自己或是朋友的好友數，預設值為自己，可以輸入好友的id

get\_numbers(self, id = **'me'**):取得自己或好友的推文數和好友數，回傳tuple

get\_friend(self, id = **"me"**):取得自己或好友的朋友的id，回傳list

static\_computing:負責進行數學計算

有兩個field, 五個function，包含:

X,y由輸入兩筆資料

mean(self, x):計算平均

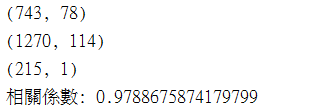
de\_mean(self, x):計算標準差

variance(self, x):計算變異數

dot(self, x, y):計算內積

correlation(self, x , y):計算相關係數

執行結果:



問題:

目前各功能大致執行順暢，但FB有權限問題僅能抓到三到四位朋友的id，故資料量不足，且若有人發文數量過大時有可能會有bug，未來希望能找到方法取得所有朋友的資料並進行分析，以取得更準確的數據