

PBC Group30
股票新聞小幫手

股價趨勢新聞快搜

一、主題介紹

選題動機

股價波動，常為新聞事件影響的結果，常常新聞事件會影響股價起伏。因此我們希望透過程式，**先以折線圖呈現用戶欲了解的特定時段股價波動**，並快速找出造成股價**大漲大跌的新聞事件**，並藉由**點擊率排序**出最受關注的新聞，以進一步了解造成波動的相關原因，**省去搜尋及過濾新聞的時間**。同時提供我的最愛功能，讓用戶可以**更快速的觀察他們關注的股票動向**。

二、程式開發

<https://pse.is/JDDDA>

開發流程

Prototype 製作

修正原型，分部分寫程式功能

將功能統整至一個視窗介面

三、程式介紹

修正後整體架構

1. 輸入股價代碼
2. 網站爬蟲抓個股資訊
3. 繪製**互動式股價圖**
4. 依據股價圖輸入欲搜尋新聞的時間
5. 網站爬蟲抓新聞
6. 新聞篩選
7. 新聞呈現排序

Part1 程式碼介紹

爬股價，並繪製出互動式圖表

```
def crawl_finance(stock_id):
    import requests
    import datetime
    import pandas as pd
    import matplotlib.pyplot as plt
    now = int(datetime.datetime.now().timestamp())+86400
    url = "https://query1.finance.yahoo.com/v7/finance/download/" + stock_id + "?period1=0&period2=" + str(now) +
        "&interval=1d&events=history&crumb=hP2r0schx00"
    response = requests.post(url)

    with open("stockfile.csv", "w") as f:
        f.writelines(response.text)

    import pandas as pd
    fine_stockdata = pd.read_csv("stockfile.csv", index_col = "Date", parse_dates = ["Date"])
```

Crawl_finance :
到url上抓取資料，
寫入"stockfile.csv"

now :
代表當前的時間

url :
- 網站的股價資訊
- Stock_id, now
可以依據使用者輸入改變

pandas :
用pandas把資料
改成通用[Date]形式

Part1 程式碼介紹

```
def draw_figure(begin_year,last_year):
    import pandas as pd
    yearlist = []
    for i in range( int(last_year) - int(begin_year) + 1 ):
        yearlist.append(str(int(begin_year)+i))

    import csv
    list_ = []
    with open("stockfile.csv", "r") as f:
        tite = f.readline()
        a = csv.reader( f )
        for aline in a:
            if aline[0][:4] in yearlist:
                list_.append( aline )

    with open("chose_period.csv","w") as c:
        c.write(tite)
        for element in list_:
            for i in range(len(element)):
                if i == len(element) - 1:
                    c.write(element[i])
                else:
                    c.write(element[i]+",")
            c.write("\n")
```

draw_figure :

依照使用者輸入的內容,可以選取需要的股價時間區間讀取"stockfile.csv",找出開始與結束時間中的股價,並存成"chose_period.csv"

Part1 程式碼介紹

```
file = 'chosedata.csv'
pd = pd.read_csv(file)
y_values = pd['MKT']
x_values = pd['DATE']
x_list=[]
y_list=[]
for j in y_values :
    y.append(j)
for i in range(len(y)) :
    x.append(i)

import plotly
plotly.tools.set_credentials_file(username='yvonnemin', api_key='00B0RQFVhOE3lVt1Y5R7')

import plotly.plotly as py
import plotly.graph_objs as go

#Loading stock price(y) and time series(x)
trace = go.Scatter(
    x = x_list,
    y = y_list,
    mode = 'lines'
)

data = [trace] #Can draw more than one data series
py.iplot(data, filename='scatter-mode')
```

把chosedata中的
股價(MKT)跟時間(DATE)
抓出來設成X,Y軸

載入Plotly,用上述的x,y
畫出互動式圖表

Part2 程式碼介紹

用 selenium及pandas 抓網頁進行爬蟲, 抓取新聞

用tkinter做視窗

新增class:news及視窗 ReadTheNews

函式庫介紹

過程中遇到的問題

整合困難

撈到錯誤資料

網頁被整理，導致原程式碼不能用

未來展望

優化現有程式

- 1. 整合各部分功能至同一視窗
- 2. 功能尚未建立完整, 希望能發展成Prototype之呈現

程式功能擴充

- 擴充搜出連帶受影響的股價代碼, 找出有**相同漲幅趨勢、事件影響的股**, 以及受同樣新聞事件影響的股票, 以找出更全面的觀點去分析股票漲幅趨勢。如再擴充, 希望可以藉由找出**新聞影響股價波動的關聯, 預測股價趨勢。**