

# 智慧企業整合 individual Project

主題：選舉票數即時回報系統

107034549 黃上科

## 摘要

台灣近年來由於商業媒體利用煽情、炒作與捏造新聞等違反媒體倫理手段增加收視率或報刊發行量使得媒體亂象氾濫，其中包含 2018 年 11 月 24 日的九合一大選，開票當晚，各大新聞台虛報選票的真實數據，對於此情況一般社會大眾卻毫不知情，是故，學生認為此情況亟需加以改變，媒體亂象造成大眾的誤會可能導致許多社會議題接踵而至。因此學生希望設計一個選舉狀況即時回報系統，藉此拋磚引玉，讓台灣的新聞環境做出改變。

## 目錄

1.前言.....	1
1.1 背景.....	1
1.2 動機.....	1
1.3 目的.....	1
2.研究方法.....	2
2.1 現況.....	2
2.2 問題分析.....	2
2.3 解決方法構想.....	2
2.4 python 爬蟲.....	3
3. 研究結果.....	5
3.1 網頁與資料庫的關係.....	5
3.2 操作介紹.....	6
4. 結論.....	6
5.參考資料.....	8

# 1.前言

## 1.1 背景

台灣近年來由於商業媒體利用煽情、炒作與捏造新聞等違反媒體倫理手段增加收視率或報刊發行量使得媒體亂象氾濫，其中包含 2018 年 11 月 24 日的九合一大選，開票當晚，各大新聞台虛報選票的真實數據。

## 1.2 動機

學生家庭於上述大選開票夜晚中，父母觀看著電視台直播開票情況，因為各自支持的黨派票數上下浮動不定，造成家裡關係緊張；然而，當晚學生拿著智慧型手機打開「中央選舉委員會」網站，仔細對照各個縣市候選人的票數，發現電視台與中央選舉委員會所報導之選舉投票數據大相逕庭。此情況亟需加以改變，因媒體亂象造成大眾的誤會可能導致許多不必要的社會議題接踵而至。

## 1.3 目的

因此學生設計一個選舉狀況即時回報系統，以 python 語言運用爬蟲方法，將中央選舉委員會的選舉投票數據依造台灣各縣市轉化成多個表格，提供網路新聞直播平台一清晰且即時回報的選舉票數，藉此拋磚引玉，望以網路新聞直播平台播放此系統即時選舉票數之方式，使台灣的新聞環境做出改變。

## 2. 研究方法

### 2.1 現況

中研院院士朱雲漢於著作「高思在雲」中指出：台灣自 1980 年代解嚴後，隨著蘋果、中國時報等報刊、新聞社的增多，各家公司為在有限資源如員工工時、報導內容蒐尋時間等資源中，盡可能極大化其公司收益而偏好使用較辛辣、聳動的標題。

### 2.2 問題分析

學生觀察家中長輩於選舉開票當晚的情況，發現長輩確實不喜歡使用只有文字的選舉票數即時回報頁面，反而除了舉選票數回報外，亦有主持人討論各大縣市市長候選人誰較可能勝出，則較討長輩喜愛。因此，現況的問題點如下所述：

1. 節目主持人討論各大城市市長候選人情況可能淪為商業媒體利用煽情、炒作與捏造新聞等違反媒體倫理手段的工具。
2. 新聞台所回報之選舉票數不具參考性之外，亦會煽動公民情緒。

### 2.3 解決方法構想

將上述提到的問題使用網頁設計加以改善，將中央選舉委員會網頁中，各縣市之候選人選舉得票數及各公投案之得票數，運用 python 語言以爬蟲方式，依照縣市或地區之地理關係，以不同表格加以呈現，而因為沒有節目主持人作為餘興，因此網頁中亦有娛樂功能，讓關心選票的公民在此選舉即時回報系統中，除了觀看真實數據外，亦可享受樂趣，結果頁面如下(圖 2.3)所示。



(圖 2.3)

## 2.4 python 爬蟲

由於上述解決方法的核心在於爬蟲程式的撰寫，因此本小節將一步步詳述此系統中，爬蟲程式的內容，步驟及內容如下：

### Step1. 引入欲爬蟲之對象

學生選擇 python 中的 requests 函數庫做為框架進行網頁爬蟲，爬蟲的對象則是中央選舉委員會的官方資料。

```
In [7]: import requests
res = requests.get('http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n6300000000000000.html')
#res.text

In [8]: from selenium import webdriver
driver = webdriver.Chrome()
driver.get('http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000100000000.html')
```

### Step2. 選擇頁面中之標籤

將網頁中的標籤利用迴圈的方法找出。

```
from bs4 import BeautifulSoup
soup = BeautifulSoup(driver.page_source, 'lxml')
links = soup.select('td a[href^=../TC/n63000]')
```

```
domain = 'http://vote.2018.nat.gov.tw/pc/zh_TW'
```

```
for ele in links:
    print(domain + ele.get('href').strip('.'))
```

```
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000000000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000010000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000020000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000030000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000040000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000050000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000060000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000070000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000080000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000090000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000100000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000110000000000.html
http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000120000000000.html
```

### Step3. 抓出網頁原始資料

抓到未經整理過的，原始的資料。

```
In [271]: "-----用Pandas找表格共同資料-----"
import pandas as pd
dfs = pd.read_html('http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000100000000.html')
votes = dfs[1]
votes
```

```
Out[271]:
```

	0	1	2	3	4	5	6
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	註記	號次	姓名	性別	得票數	得票率	推薦之政黨
2	NaN	1	吳萼洋	男	428	0.40%	無
3	NaN	2	丁守中	男	46381	43.36%	中國國民黨
4	NaN	3	姚文智	男	17387	16.25%	民主進步黨
5	◎	4	柯文哲	男	42356	39.60%	無
6	NaN	5	李錫錕	男	420	0.39%	無
7	投開票所數	已送/應送:	117/117	NaN	NaN	NaN	NaN

#### Step4. 整理原始資料使之較具可讀性

為了提高可讀性，要對未整理的資料進行修整，修整結果如下。

```
res = requests.get('http://vote.2018.nat.gov.tw/pc/zh_TW/TC/n63000000100000000.html')
soup = BeautifulSoup(res.text, 'lxml')
area = soup.select_one('b').text.replace('\xa0', '-')
votes['投票地區-投票種類-補充說明'] = area.strip('-')
votes
```

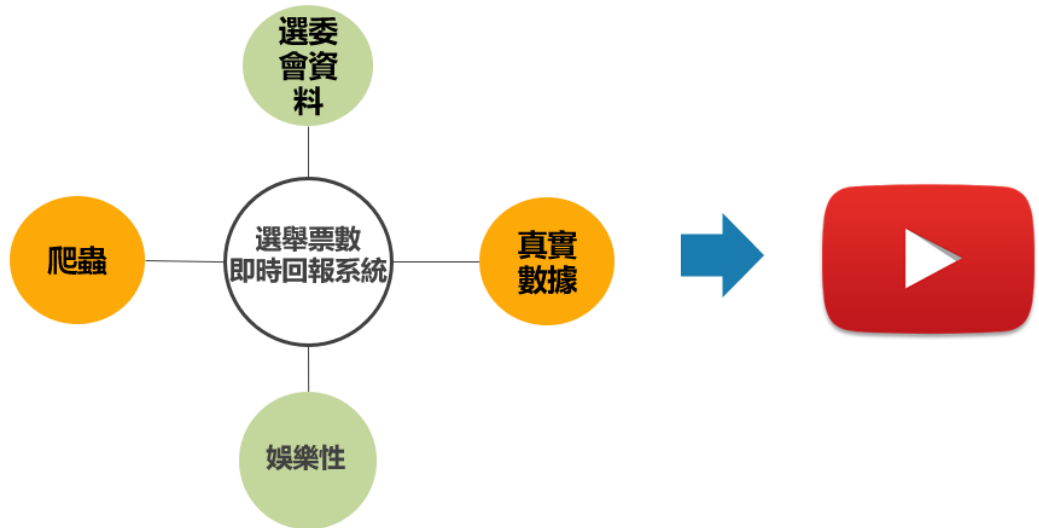
```
def getRef(url):
    res = requests.get(url)
    soup = BeautifulSoup(res.text, 'lxml') #讀檔案進入python
    dfs = pd.read_html(res.text)
    votes = dfs[1]
    votes.columns = votes.loc[1]
    votes.drop([0,1,7],inplace=True)
    votes = votes.replace(np.nan, ' ', regex=True) #"-----引進numpy清除空格-----"

    area = soup.select_one('b').text.replace('\xa0', '-')
    votes['投票地區-投票種類-補充說明'] = area.strip('-')

    return votes
```

	1	註記	號次	姓名	性別	得票數	得票率	推薦之政黨	投票地區-投票種類-補充說明
2			1	吳萼洋	男	428	0.40%	無	臺北市-市長選舉-候選人在松山區得票數
3			2	丁守中	男	46381	43.36%	中國國民黨	臺北市-市長選舉-候選人在松山區得票數
4			3	姚文智	男	17387	16.25%	民主進步黨	臺北市-市長選舉-候選人在松山區得票數
5	◎		4	柯文哲	男	42356	39.60%	無	臺北市-市長選舉-候選人在松山區得票數
6			5	李錫錕	男	420	0.39%	無	臺北市-市長選舉-候選人在松山區得票數

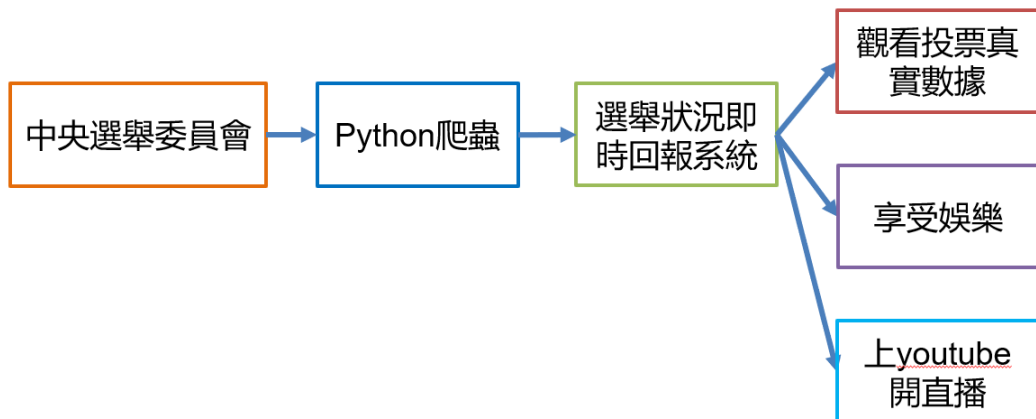




(圖 3.1)

### 3.2 操作介紹

若使用者想分享選舉投票情況的真實數據給予自己的朋友或社交平台上的粉絲，可以利用 youtube 進行登入後達成，若使用者覺得只觀看選舉數據過於枯燥，可以選擇三種網頁遊戲進行娛樂，遊戲共有：超級瑪利歐、中國軍棋、水果忍者。使用者整體流程如(圖 3.2)所示。



(圖 3.2)

## 4. 結論

經過改善後，此選舉狀況即時回報系統將涉嫌捏造新聞等違反媒體倫理手



段的相關可能加以阻絕，提供的數據可幫助了解使用者投票的真實情況而不受政黨、商業媒體所操控，降低公民情緒因媒體煽動而引起社會議題的可能性，例如：鄭捷北捷血案，使人們有不當的情緒(過度的仇恨、模仿)，造成社會二度傷害。系統下方亦提供連結能讓使用者依自己喜好選擇是否開設 youtube 直播新聞，讓更多使用者本身交友圈的朋友或社群粉絲接收選舉投票情況的真實數據，如此一來若有更多的使用者使用直播的方式傳播選舉投票情況，而不選擇觀看商業媒體的直播，將可使商業媒體檢討自我，讓台灣新聞環境做出改變。

## 5. 參考資料

Datacamp

<https://www.datacamp.com/>

利用單行 For 迴圈產生 List 資料組

<https://sites.google.com/site/ezpythoncolorcourse/listandsingleforloop>

中央選舉委員會

<https://www.cec.gov.tw/>

維基百科

<https://zh.wikipedia.org/zh-tw/Wiki>

高思在雲：一個知識份子對二十一世紀的思考，天下文化，朱雲漢，2015/01/28

Python x 網路爬蟲

<https://medium.com/dualcores-studio/python-x-%E7%B6%B2%E8%B7%AF%E7%88%AC%E8%9F%B2-c30ffda0ad78>

pandas 用法

<http://violin-tao.blogspot.com/2017/06/pandas-1-indexing.html>