

# 九合一大選市長選舉-關鍵字分析

## 文字探勘導論-期末報告

巨資四A 張智淵 巨資四A 彭英致 巨資四B 陳思緯

### 目標

- 爬取PTT各候選人文章，初步統計聲量
- 藉由與候選人相關的新聞資料，篩選出關鍵詞彙，包含與候選人相關的議題、政治事件等
- 用Label Studio將收集到的資料作情感分類，統計正負情緒篇數

### 一、資料蒐集、統計聲量

- 資料來源: PTT八卦版(Gossiping)、政黑板(HatePolitics)
- 時間範圍: 2022整年(到九合一大選前)
- 關鍵字: 候選人名字

### 二、斷詞、關鍵字視覺化

- 陳時中
  - 從陳時中的文字雲中可以看出再跟他有關的內容當中，較常出現的便是蔣萬安、跟黃珊珊這兩名對手這也很正常他們是陳時中最主要的對手，而在圖中還可以看出與他相關的內容當中也經常出現柯文哲我們認為這是因為他的對手黃珊珊是柯文哲的副市長並且關係也十分熱絡，還有他在之前擔任防疫指揮官時，經常與身為現任台北市長的柯文哲有口角上的衝突，因而結怨;在這次選舉中陳時中之前擔任指揮官時的舉措也被放大檢視對選舉造成很大的影響，在圖中可以看在他的許多決策當中最具爭議性的便是疫苗採購的部分。



## ● 蔣萬安

- 在這次選舉中蔣萬安的聲量雖然遠低於陳時中僅比黃珊珊稍高，但他在這次的選舉中民調一直以來都是第一或第二名而從他的文字雲中可以看出其中最明顯的亦是身為其對手的黃珊珊與陳時中，而柯文哲身為現任市長以及對手黃珊珊的上司也是相當常出現，而在圖中我們也發現周玉蔻這位媒體人亦是相當明顯，這可能跟周玉蔻前段時間猛烈攻擊蔣萬安的父親蔣孝嚴的緋聞案，甚至因此與張淑娟鬧上法庭引起廣大迴響。



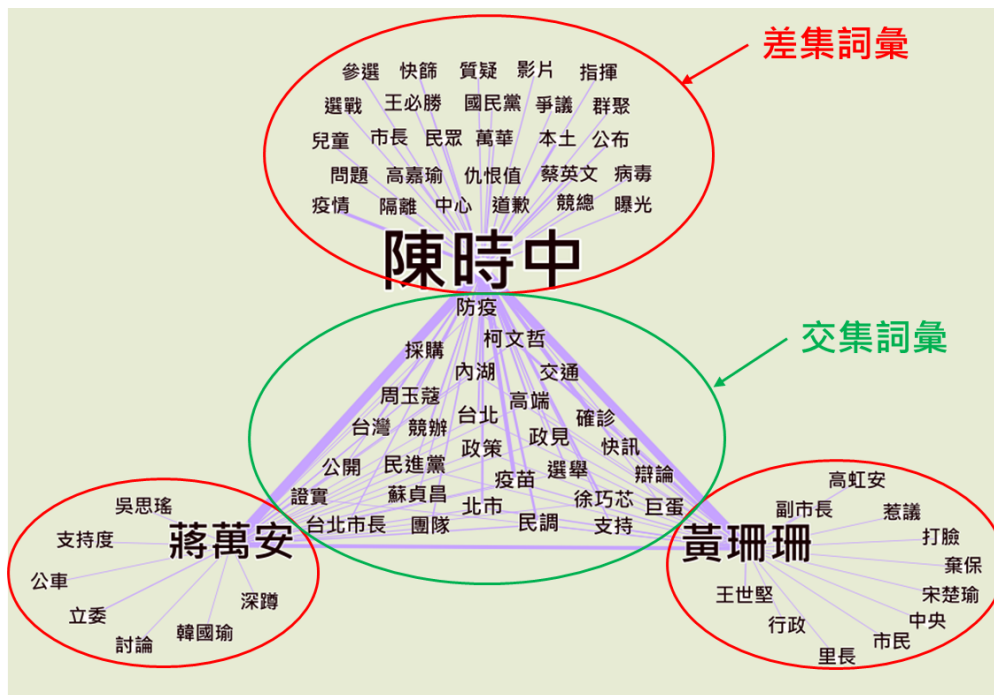
## ● 黃珊珊

- 黃珊珊在本次選舉中聲量最低民調大多是在第二三名徘徊，在她的文字雲中我們可以發現最明顯的是她的對手蔣萬安、陳時中以及她的上司現任市長柯文哲；而在圖中我們也能發現市議員王世堅也相當明顯，這是因為他在選舉前表示在陳時中與黃珊珊之間不知道要選誰可他身為民進黨員應該是要明確支持陳時中的，其中的關係耐人尋味；黃珊珊與陳時中也經常討論有關內湖交通的議題，因此內湖在圖中亦是相當明顯。



## ● 社群網路圖

- 為了觀察本次選舉中的三位候選人的關係以及特點，我們使用了社群網路圖來觀察並進行交集和差集以期更加具體的了解三位候選人，從圖片中可以看出在本次選舉的討論中許多詞彙都同時與兩位以上的候選人有關，我們從中選擇幾個來進行探討，例如：高端、內湖、交通、柯文哲、周玉蔻、巨蛋，而在圖中我們也以差集的視覺化處理來讓一些平常沒在關注政治的人們可以迅速理解以及區分這些候選人們各自不同的議題，也能從中看出這些候選人的分別與哪些議題有較強的連結。



### 三、關聯規則、情緒分類

#### ● 關聯規則

- 將斷詞後的結果利用關聯規則透過關鍵字之間的連結，了解與候選人有關的重要新聞。

下面三張圖依序為陳時中、蔣萬安、黃珊珊候選人帶入關聯規則模型後的結果。apriori演算法的參數設定上，min\_support分別設定0.001(陳時中)和0.01(蔣萬安、黃珊珊)。透過分析出的關鍵字之間的關聯，去找出相對應的重要新聞。

#### ● 陳時中

- 勘災穿皮鞋被批 陳時中：不是服裝比賽該「撩落去」就「撩」
- 藍白猛攻「公廁偷窺片」 陳時中：光責備不會讓人當選
- 重提器捐案柯文哲想罵髒話 陳時中：不要這麼容易生氣
- 侯市府拿中央替恩恩案背書？陳時中：已發公文 講清楚比較好
- 馬斯克提議把台灣成特別行政區 陳時中：嚴正抗議

## ● 蔣萬安

- TVBS封關民調／蔣萬安36%力壓陳時中25% 北市投票率預估6成1
- 高端疫苗爭議不休 蔣萬安促公開採購資料
- 蔣萬安展決心辭立委...高虹安是否跟進？ 競辦回應了
- ET民調 / 蔣萬安支持度41.5%居冠領先陳時中8.3個百分點

## ● 黃珊珊

- "防疫旅館"誰發想？全國首間業者為黃珊珊背書
- 內科與內湖不同 黃珊珊：內科交通問題才是問題
- 北市長辯論／大巨蛋3市長誰負責？黃珊珊：國民黨的弊案 柯文哲善後
- 民進黨指陸擋我買疫苗 黃珊珊：為何蔡總統出面就行？

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
12120	(罵, 陳時中, 想, 柯文哲)	(髒話)	0.00	0.00	0.00	1.00	866.40
12099	(陳時中, 屍體, 踏)	(陳水扁, 具)	0.00	0.00	0.00	0.83	722.00
9398	(行政區)	(馬斯克, 特別)	0.00	0.00	0.00	1.00	722.00
11069	(皮鞋)	(勘災, 穿, 陳時中)	0.00	0.00	0.00	1.00	722.00
12224	(推, 公廁)	(治, 換, 馬桶, 陳時中)	0.00	0.00	0.00	0.83	722.00

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
327	(採購)	(蔣萬安, 疫苗)	0.01	0.03	0.01	0.92	29.81
323	(蔣萬安, 採購)	(疫苗)	0.01	0.03	0.01	0.92	29.81
115	(採購)	(疫苗)	0.01	0.03	0.01	0.92	29.81
385	(辭)	(立委, 蔣萬安)	0.03	0.04	0.03	1.00	25.82
384	(立委, 蔣萬安)	(辭)	0.04	0.03	0.03	0.82	25.82
383	(辭, 蔣萬安)	(立委)	0.03	0.04	0.03	1.00	25.82
386	(立委)	(辭, 蔣萬安)	0.04	0.03	0.03	0.82	25.82
183	(立委)	(辭)	0.04	0.03	0.03	0.82	25.82
182	(辭)	(立委)	0.03	0.04	0.03	1.00	25.82
272	(同)	(蔣萬安, 台)	0.02	0.03	0.01	0.63	20.32
268	(蔣萬安, 同)	(台)	0.02	0.03	0.01	0.63	20.32
42	(同)	(台)	0.02	0.03	0.01	0.63	20.32
331	(支持度)	(民調, 蔣萬安)	0.01	0.05	0.01	0.69	13.21
328	(支持度, 蔣萬安)	(民調)	0.01	0.05	0.01	0.69	12.93
133	(新)	(民調)	0.03	0.05	0.02	0.69	12.93
124	(支持度)	(民調)	0.01	0.05	0.01	0.69	12.93
336	(蔣萬安, 新)	(民調)	0.03	0.05	0.02	0.68	12.70
339	(新)	(民調, 蔣萬安)	0.03	0.05	0.02	0.65	12.48



	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
114	(旅館)	(防疫)	0.01	0.03	0.01	1.00	37.72
303	(旅館)	(防疫, 黃珊珊)	0.01	0.03	0.01	1.00	37.72
301	(旅館, 黃珊珊)	(防疫)	0.01	0.03	0.01	1.00	37.72
345	(綠)	(藍, 黃珊珊)	0.02	0.02	0.01	0.59	27.86
344	(藍, 黃珊珊)	(綠)	0.02	0.02	0.01	0.65	27.86
245	(內湖)	(交通, 黃珊珊)	0.03	0.02	0.01	0.58	27.50
11	(內湖)	(交通)	0.03	0.02	0.01	0.58	27.50
242	(內湖, 黃珊珊)	(交通)	0.03	0.02	0.01	0.58	27.50
243	(交通)	(內湖, 黃珊珊)	0.02	0.03	0.01	0.70	27.50
10	(交通)	(內湖)	0.02	0.03	0.01	0.70	27.50
240	(交通, 黃珊珊)	(內湖)	0.02	0.03	0.01	0.70	27.50
171	(藍)	(綠)	0.02	0.02	0.01	0.62	26.53
170	(綠)	(藍)	0.02	0.02	0.01	0.59	26.53
343	(綠, 黃珊珊)	(藍)	0.02	0.02	0.01	0.59	26.53
346	(藍)	(綠, 黃珊珊)	0.02	0.02	0.01	0.62	26.53
266	(巨蛋, 黃珊珊)	(大)	0.01	0.04	0.01	1.00	24.82
268	(巨蛋)	(大, 黃珊珊)	0.01	0.04	0.01	1.00	24.82
51	(巨蛋)	(大)	0.01	0.04	0.01	1.00	24.82
289	(擋, 黃珊珊)	(疫苗)	0.01	0.04	0.01	0.79	20.58

## 四、結論

- 經由這次的專題報告的製作當中我們可以藉由這次的專題報告來幫助一些平常沒在關注政治的人們可以透過關鍵字更加迅速理解跟這些候選人相關的議題。
- 在本次的專題中我們從PTT政黑板、PTT八卦版爬取資料來觀察，從中我們對此次的專題報告所有爬取的資料進行正負向情緒分析發現P/N比(正面情緒篇數/負面情緒篇數)為0.454，我們推測可能是因為在今次台北市市長選舉，不管是哪位候選人都是以抨擊對手來做為主要手段，而在其中尤以陳時中的負面情緒篇數最多。
- 我們推測此次的九合一選舉結果可能受到外溢效果影響，不管是台北市的市長選舉影響了其餘縣市、還是新竹市的市長選舉影響了台

北市等的其餘縣市可能都受到外溢效果的影響，所以根據這點我們覺得這個現象是我們以後可以繼續深度探討與研究的方向。