

台北市長候選人臉書經營方式的影響分析

蔡芸琇¹、羊敏丹²、謝誌紘³、黃彥鈞⁴、廖達琪⁵

國立臺灣大學^{1, 2, 3, 4}、國立中山大學⁵

共同教育中心¹、新聞所²、植物所³、圖資系⁴、政治所⁵

助理教授¹、碩士生²、碩士生³、大學部畢業生⁴、教授⁵

摘要

隨著社群時代來臨，選舉的造勢工具也更加多樣化。無論是美國總統大選，川普與希拉蕊的推特經營戰略，抑或是 2014 年台北市長選舉，柯文哲憑藉臉書粉專經營，成功聚集人氣前進台北市府，皆可以看出社群媒體在選舉場域掀起不少波瀾。之於候選人而言，社群媒體的普及形同賦權，只要經營公開粉專，定期發文，與網友互動，即可以傳播自身所想要表達的意見。而大眾媒體如何傳播相關選舉資訊，代表了傳統媒體對於候選人的形象塑造。

在後真相時代，訊息的來源與管道更加混淆不分，選舉為民主國家必要之公共事務，候選人的形象如何被塑造，自然影響公眾的選舉意願與行為。也因此，探討候選人在社群媒體上的自我形塑，與傳統大眾媒體的描繪作比較，看出箇中差異與討論，即是本研究所關心並試圖去回答的議題。

而為了聚焦研究範圍與結果，本研究將以 2018 年台北市長選舉三位聲量最高的候選人，無黨籍參選人柯文哲、民進黨參選人姚文智，與國民黨參選人丁守中為研究對象；並嘗試從上述三位候選人本年度一到六月初的臉書粉專發文內容、網友回饋，與四大報社（在此代表傳統大眾媒體）的同期相關文本，做初步的文字探勘與視覺化呈現，希望能藉此看出候選人的關注議題與發文情緒之差異；接著剖析四大報（聯合報、中國時報、自由時報、蘋果日報）的各候選人報導文本，從所用字詞與文本情緒的比較，看出候選人自我塑造與大眾媒體的描繪有何不同？

關鍵詞：文字探勘 (Text Mining)、資料視覺化 (Data Visualization)、潛在狄利克雷分配模型 (Latent Dirichlet Allocation, LDA)、情緒分析 (Sentiment Analysis)。

壹、前言

自 2008 年至 2012 年間，美國總統候選人，歐巴馬在競選活動中大量使用資料探勘 (Data Mining) 技術，收集網路資訊，針對選民所認同的社會價值，進行了解與認識 (Mostafa, Mohamed M 2013)。這並不是競選資料探勘的首例，但卻對美國政治產生了關鍵性的影響。民主黨和共和黨的國家元首候選人，可透過政黨的全力支持，獲得最豐厚的數據分析資源。兩黨分別為各自的候選人提供完整且結構清晰的國家選民數據，以供競選決策使用 (Steinberger, Ralf 2012)。

網路資訊可分為，新聞媒體與社群網絡兩大類。新聞媒體又稱大眾傳媒，其經由電視、網路、報章等媒介傳播訊息，為一般民眾最容易接觸到的資訊媒介。一般民眾可以經由新聞媒體的報導追蹤候選人的行動，並經由媒體的描述認識候選人。而社群網絡主要作用是為一群擁有相同興趣與活動的人創建的網路虛擬社群 (Freeman, Linton C. 1978)，此類網站通常通過朋友，一傳十、十傳百、百傳千地把網絡展延開去，再透過文字探勘於這類的社群分析，就可取代過去採用人工問卷的方式，去探索出群眾目前的共識與想法 (Aggarwal, Charu C., and Haixun Wang. 2011)。

由於媒體營利性質的產業，為追求收視率或點閱率，少數媒體人會抓住候選人的語病，並放大渲染，藉此以吸引民眾。再者，目前媒體由少數集團所掌握，具有高度集中的情形，如此一來，可能導致新聞媒體被集團的意識形態所控制。而這些集團多數是具有高度政黨傾向的，新聞媒體可能會有使用新聞權的方式，進行攻擊或偏袒。而一般民眾若不經由查證或反思，很容易被媒體所左右。

當許多地方選舉的候選人，若不是主要國家政黨的候選人，甚至尚未獲得足夠支持，以滿足政黨數據收集標準的候選人，都處於數據劣勢。因此，這類的候選人就會選擇社群網絡來當成發聲管道，目前在台灣最為廣泛使用的社群網站是臉書 (Facebook)。因此，本研究藉由收集媒體之報導內容與候選人的臉書粉絲專頁，進行交叉比對析，希望能夠了解，沒有獲得主要政黨支持的候選人，是如何透過臉書經營來獲得選民的支持。

本研究以 2018 年台北市長選舉，以國民黨、民進黨，以及無黨籍三方政黨代表，挑選無黨籍參選人柯文哲、民進黨參選人姚文智，與國民黨參選人丁守中為研究對象。收集自今年一月到六月，三位候選人的臉書粉絲專頁發文內容、網友回饋，與代表傳統大眾媒體的四大報社 (聯合報、中國時報、自由時報、蘋果日報)，之同期相關文本，並將相關分析結果呈現在 (https://dppss90008.shinyapps.io/r_shiny_new/) 視覺化平台中。我們結合文字探勘以及資料視覺化技巧呈現數據，將繁雜的資料簡化成為易於吸收的內容 (Yang, YunYun, et al. 2008)，讀者可透過我們的視覺化平台，自行探索

本文所提及的各種分析現象。希望能藉此看出候選人的關注議題與發文情緒之差異。剖析四大報的各候選人報導文本，從所用字詞與文本情緒的比較中，看出候選人自塑形象與大眾媒體的描繪是否有不同。

貳、透過網頁爬蟲進行文本資料收集

為獲取候選人的發文內容、網路聲量相關資料以及其在傳統媒體的相關報導。本研究利用網路爬蟲技術（web crawler）針對台北市候選人進行網路資料的收集。所謂網路爬蟲技術，是一種自動化的網路檢索和擷取網頁資訊的程式。使用者可以經由爬蟲程式，將網站內容存取在用戶端，並經由資料清理（data cleaning）的方式收集到使用者想要的資料。

本研究從兩大媒體的來源獲取資料，其一為社群媒體：臉書（Facebook），其二為傳統媒體：四大報。臉書為目前主流的社群媒體，更是候選人經營自己的主要管道之一。候選人經由臉書粉絲專頁的經營，發布文章藉此來宣傳自己，讓一般民眾可以經由粉絲專頁認識此候選人的價值觀、候選人對於時事的評論與政見。另外，民眾也可以對於文章內容進行留言與使用臉書表情符號進行迴響。但由於臉書使用反爬蟲的機制，本研究無法使用一般爬蟲的程式爬取資料。因此，本研究使用臉書授權提供的 Graph API（<https://developers.facebook.com/tools/explorer/>）。進行候選人之發文文章內容的擷取。另外也可經由 API 收集分享數、按讚數及其他臉書表情符號進行後續分析。本研究收集無黨籍參選人柯文哲、民進黨參選人姚文智，與國民黨參選人丁守中從 2018 年 1 月至 2018 年 6 月中之臉書資料（柯文哲：138 筆，丁守中：265 筆，姚文智：243 筆），其爬取結果如（圖一）所示。

X	time	post	share	like	love	haha	sad	wow	angry
1	2018-06-16T02:00:45+0000	捏麵人初體驗！捏個小叮噹 <f0><U+009F><U+0098><U...	54	1258	25	16	0	1	1
2	2018-06-15T07:03:13+0000	今天適逢警察節，先向各位人民保姆說聲「辛苦了！」。對...	73	2191	20	2	0	0	2
3	2018-06-14T10:19:21+0000	同仁說要去運動酒吧看世界盃，順便舉行辦公室的聚餐，雖...	60	2976	27	24	1	1	0
4	2018-06-14T02:22:54+0000	臺北市是六都人口高齡化最嚴重的城市，未來我們的長輩們...	113	1331	15	2	0	0	1
5	2018-06-13T00:23:00+0000	護航吳音寧 也是總統亂政 蔡英文總統自己二年任期表現都...	224	1556	2	73	11	11	152
6	2018-06-12T01:19:28+0000	#丁守中 #拚台北驕傲 #內湖垃圾山 #城市之恥 內湖科技園...	42	1246	14	13	0	3	2
7	2018-06-10T07:05:13+0000	#丁守中 #拚台北驕傲 #清除內湖垃圾山 今天一早風和日麗...	33	1541	23	2	0	2	0
8	2018-06-10T01:42:57+0000	齊（騎）心協力，雙北大勝利。	62	2133	53	3	0	3	0
9	2018-06-09T08:10:05+0000	#丁守中 #拚台北驕傲 今早連裡兩場大型活動為黨員慶生，...	227	6080	75	9	0	1	3
10	2018-06-08T13:46:03+0000	#丁守中 #拚台北驕傲 #團結 #勝選 今天非常高興能與馬前...	59	1605	22	6	0	1	0
11	2018-06-08T10:00:42+0000	今晚八點，請鎖定中天52台新聞深喉嚨，我們幾位市長參選...	162	2442	27	9	0	3	0
12	2018-06-07T08:51:34+0000	市府欠公車司機加班費 凸顯的問題 昨天林瑞圖議員在質詢...	60	656	2	51	2	17	68
13	2018-06-06T07:04:41+0000	#丁守中 #拚台北驕傲 下次市長，別選柯文哲再浪費四年 柯...	102	2305	28	35	1	2	18
14	2018-06-04T10:35:53+0000	沒錯，各位，先前以為可以喬借黨部11樓當競選總部。但因...	68	2335	26	19	3	4	5

圖一、臉書爬文情形。使用 Graph API 獲得各候選人發文文章時間、文章內容、分享數、按讚數及其他表情符號內容。

傳統媒體為一般民眾最容易接觸到的資訊媒介。傳統媒體經由網路或電視傳播媒介

對於候選人進行資訊的傳播。一般民眾可以經由媒體的追蹤得知相關於候選人的訊息。由於多數媒體具有高度政治傾向且可能為了提升收視率，對於候選人可能有不恰當的報導。因此，本研究希望能夠藉由收集到媒體報導的資料，了解媒體對於候選人描繪的候選人形象，並調查媒體是否對於特定候選人有所偏頗。本研究使用 R 語言中的 Rvest 套件 (<https://cran.r-project.org/web/packages/rvest/index.html>) 以及 Rselenium (<https://cran.r-project.org/web/packages/Rselenium/index.html>) 套件進行資料的爬取。其中，我們收集了四大報（聯合報、中國時報、自由時報、蘋果日報），自 2018 年 1 月至 2018 年 6 月，柯文哲、丁守中與姚文智之新聞資料，其結果如（圖二）所示。

X	time	title	post
1	2018-06-01 11:58	呂秀蓮勸市民選「有品味的市長」 柯文哲：還可以啦	前副總統呂秀蓮日前拋出Bye bye民進黨的論調，被外界解...
2	2018-05-31 17:19	訪察高雄被爆吃魚翅宴？吳音寧：基於理念我沒有吃	北農總經理吳音寧送菜風波剛落幕，又遭爆料去魚翅餐廳。...
3	2018-05-30 23:51	柯文哲喊起薪50K較合理 北市府約聘竟有人僅23K	台北市長柯文哲曾為年輕人低薪抱屈，認為4萬5至5萬元較...
4	2018-05-30 14:15	是否和柯文哲整合 蔡英文未回應	民進黨今天下午舉行中執會，討論台北市長人選徵召案，姚...
5	2018-05-29 17:19	都是孩子都要愛 伊甸攜手陳佩琪關懷慢飛天使	台北市長柯文哲夫人陳佩琪今以醫師身份出席伊甸基金會舉...
6	2018-05-29 09:49	民進黨推出姚文智 蔡正元神分析：目的不是為了勝選	前國民黨政策會執行長蔡正元。聯合報系資料照分享 faceb...
7	2018-05-28 16:05	影／綠營推姚文智出戰 丁：非最強投手教練團打假球？	民進黨選對會決議推派同黨立委姚文智出征台北市長選戰，...
8	2018-05-28 14:22	冷眼集／分裂的一方敗多贏少 喊棄保也沒有用	民進黨選對會今建議提名立委姚文智為台北市長參選人，該...
9	2018-05-27 20:22	民進黨台北市黨部主委改選 陳正德勝出	民進黨今改選縣市黨部主委、全國黨代表及各縣市黨代表，...
10	2018-05-27 11:54	許忠信參選台南市長 金恆煒：台南不要第二個柯文哲	台聯前立委許忠信昨天在台南宣布參選市長，今天北上召開...

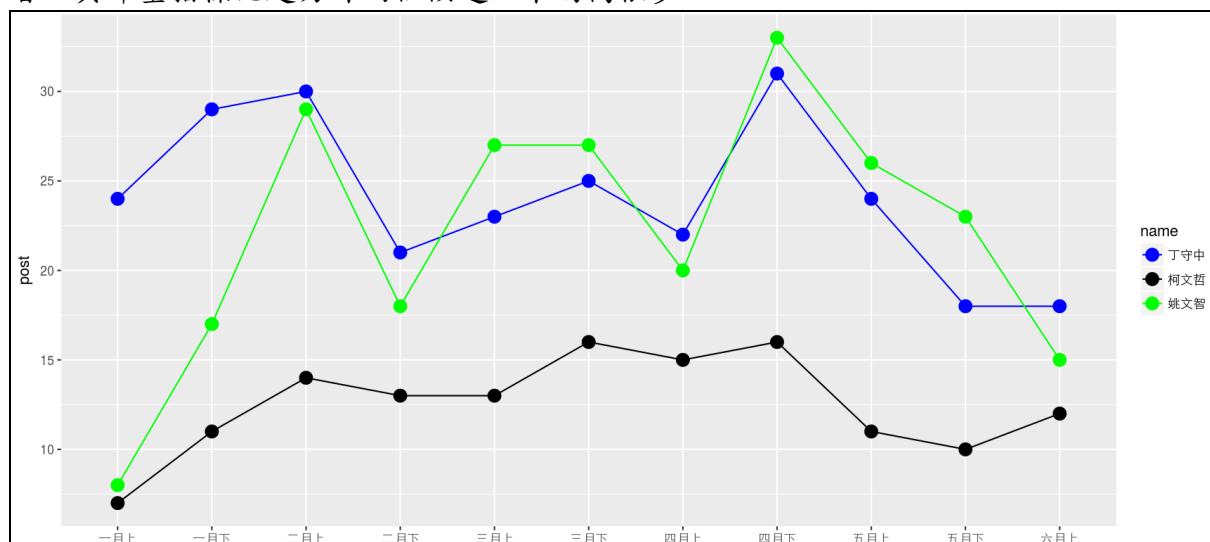
圖二、新聞媒體爬文情形。以柯文哲為例，此結果為爬取柯文哲在聯合報新聞網之標題與新聞文章內容。

參、網路聲量指數的建構與應用

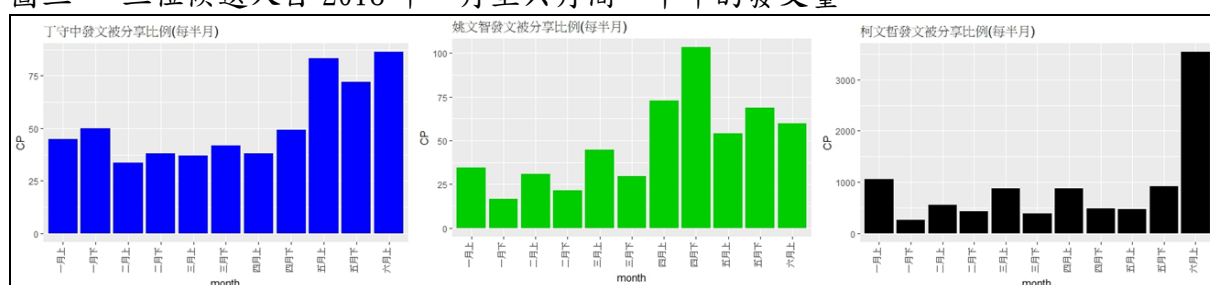
為了解各候選人發文情況，及其網路聲量，本研究對於文章臉書粉專發文量進行分析。對於各候選人的發文量進行比較。本研究以每半個月為間格統計，各候選人的發文量，可以觀察到丁守中與姚文智的發文量相當，而柯文哲的發文量較少，大略為丁守中與姚文智的一半，可見丁守中與姚文智比起柯文哲更致力於臉書上的經營（圖三）。

然而，發文的數量並無法代表此候選人臉書經營的成功與否。成功的臉書經營者，在於其粉絲專頁成功吸引群眾進去瀏覽。本研究使用臉書特有的按讚與分享機制，去判斷吸引網路群眾的情形，我們認為，若群眾認同此候選人，獲得越高按讚數量，與分享數量的候選人，其網路聲量越高。為了判斷出候選人的網路聲量高低，我們參考性價比（cost - performance ratio, CP 值），提出了聲量指標，進行比較。此聲量指標，定義為，發文數量為分母，文章的分享數量或按讚數量為分子，算出的數值意義代表此候選人發文的成本效益程度，越高的數值代表文章的網路聲量越高。聲量指標比也就是所謂的 CP 值，越高的 CP 值代表此商品獲得的效益程度。消費者總是希望以越便宜的價格獲得效能越好的商品。而候選人就類似於消費者，希望能使用越少的文章就可以獲得越多

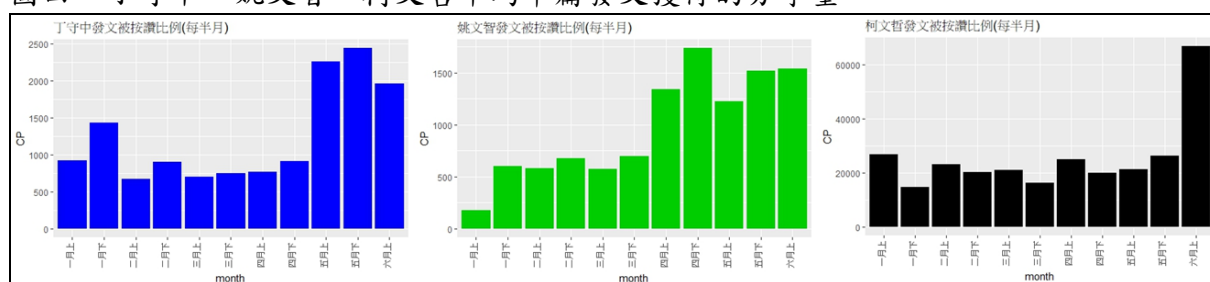
的迴響。本研究使用每半個月份量的文章進行分析，從趨勢上可看到此三個候選人的聲量指標皆有越來越高的情形。從（圖四）與（圖五）中可觀察到，柯文哲平均單篇發文獲得的分享量與按讚量（柯文哲的聲量指標大約落在 1000~5000 之間）是另外兩個候選人遠遠不及的（丁守中與姚文智的聲量指標大約落在 50~100 之間）。再者，柯文哲的發文量比起另外兩個候選人還來的少，其更可以凸顯出柯文哲是個成功的社群網路經營者，其聲量指標比起另外兩位候選人來的高很多。



圖三、三位候選人自 2018 年一月至六月間，半年的發文量。



圖四、丁守中、姚文智、柯文哲平均單篇發文獲得的分享量。

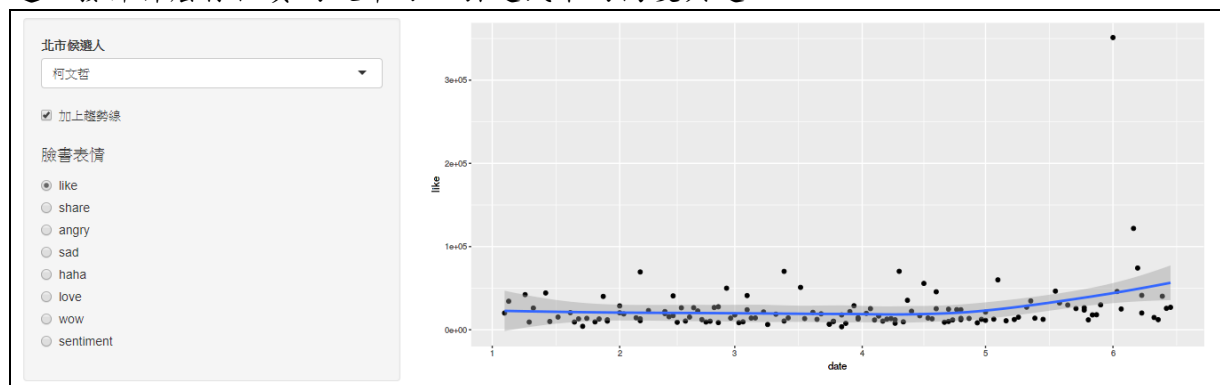


圖五、丁守中、姚文智、柯文哲平均單篇發文獲得的按讚量。

肆、非結構化資料探索之文字探勘過程

本研究針對各文章之表情符號繪製散佈圖，我們發現有不少為離群值的資料。其離群值可能是代表，該篇文章特別引起到民眾的關注。以柯文哲為例，在按讚散佈圖中有一筆資料按讚數來到三十五萬，進一步觀察此筆資料的內容，為柯文哲正式宣布，臺北

市政府正式脫離千億債務俱樂部，如（圖六）與（圖七）所示。可利用此方式，知道候選人發佈什麼樣性質的文章可以引起民眾的高度興趣。



圖六、柯文哲文章之表情符號（like）時間對數量的散佈圖。

北市候選人

何文哲

顯示項目

5

從大排到小

臉書表情

like

share

angry

sad

haha

love

wow

sentiment

date	post	like
2018-06-01T02:44:53+0000	從今天起，臺北市政府正式脫離千億債務俱樂部。我剛上任時，接下了1468億元的債務，過了3年半還了差不多520億元，讓債務減至948億元，人均負債也從上任時的5.4萬，下降至3.5萬，如果今年還有歲入增加，大概可以還到540億元。這是地方自治史上民選首長的最高還債紀錄，但其實我們只是每天認真工作，該做的做，不該做的不要做，到年底就會發現剩下很多錢。就像追加預算，以前都沒在控管，現在通過要進市長專案報告，後來幾乎沒什麼在追加。世大運開閉幕我們花3.6億，8年前同一個場地的聽奧花了5.4億，世大運的規模還是聽奧的好幾倍。這幾年我們每天操、每天盯，消耗預算的毛病也少了。現在債務還了，效率提高了，我們也有更多的財力解決問題，像市場改建、公辦都更、東區門戶計畫等等，現在都有足夠的財力進行。我們改變的現在，是下一代的未來。一個負責任的政府，不應該債留子孫，這就是4年來臺北市政府所秉持的信念。#嚴守財政紀律 #進步價值光榮城市	351160
2018-06-06T07:06:58+0000	我太太很有個性，她本來就是一個獨立的人，所以提告跟寫臉書也不需要我同意。這件事情要從去年世大運結束後說起，從那時候開始，一直有關於我的不實新聞出來，每天晚上的政論節目再來討論這些假新聞。抹黑抹紅整整快要一年了，要回應，也不知如何回應，不回應，變成是默認。舉例來說，世大運閉幕時的黑衣人，沒收場內觀眾的臺灣旗幟，但場內維安是憲兵和國安局負責的，這些都不是臺北市長直接指揮的，結果政論節目持續罵了一個多月。之後國安局發了一篇新聞稿，說明場內維安的憲兵是國安局指揮的，和北市府警察局無關，但之前罵成這樣，一篇小小的聲明根本沒人看。還有節目拿張照片，把黃向<U+7FA3>議員的哥哥黃向成當成辛旗（胡錦濤的助理），影射我去中國大陸做什麼賣臺事業，其實當時我還在臺大醫院當醫生，實在沒時間也沒能力搞什麼兩岸事務。甚至還有節目拿出一張5百萬的小吃店收據，指控我們選舉帳目不清，最後才說那個是道具，根本提不出任何證據。最誇張的是，有節目說「郭（郭台銘）柯配」要參選2020年總統副總統，我到目前為止沒有跟郭台銘見過面，怎麼會有這種新聞？面對假新聞、真指控，我們也有發聲明澄清，但人家根本不理會，繼續在節目上說謊造謠。這次講到我們家的財務，不小心被陳佩琪看到，只能算電視臺運氣不好，因為我們家的錢都是她在管的，所以才會跳出來，連很久沒開的臉書都重啟了。我知道她很生氣，這幾年當市長，我學到「生氣時不要做決定」，所以我會再跟她溝通。當然如果她最後做了什麼決定，我會完全的尊重她。	121823
2018-04-01T02:08:50+0000	#0401	87000
2018-06-07T13:07:22+0000	最近，很多人關心國中小加裝冷氣的議題。事實上，一個政策應該要有通盤的考量和長期的規劃，所以要抗暑，裝冷氣只是最後的手段。節能減碳、綠色環保是政策的基本原則，這幾年我們在每一間學校的屋頂設置植栽和小田園，加裝走道噴霧和屋頂灑水，教室也有排氣風扇、遮陽板。此外，市府也已經和建築師公會簽約，委託他們對本市236間學校作體檢，有充分的資料，我們才能進一步作節能減碳的整體規劃。明年我們也已經編列各級學校隔熱屋頂的預算，其他的陸續規劃中。至於冷氣不是	74195

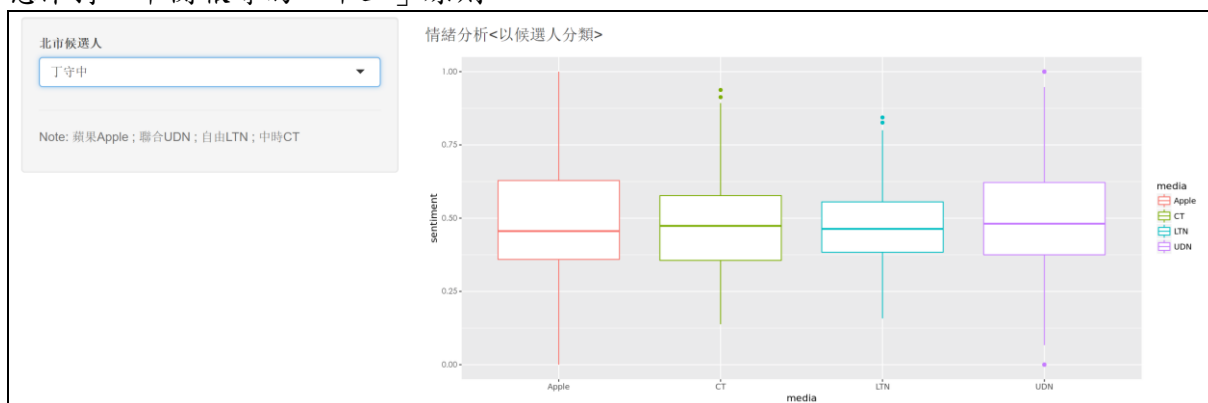
圖七、離群值的發生，來自柯文哲正式宣布，臺北市政府正式脫離千億債務俱樂部。

一、情緒分析

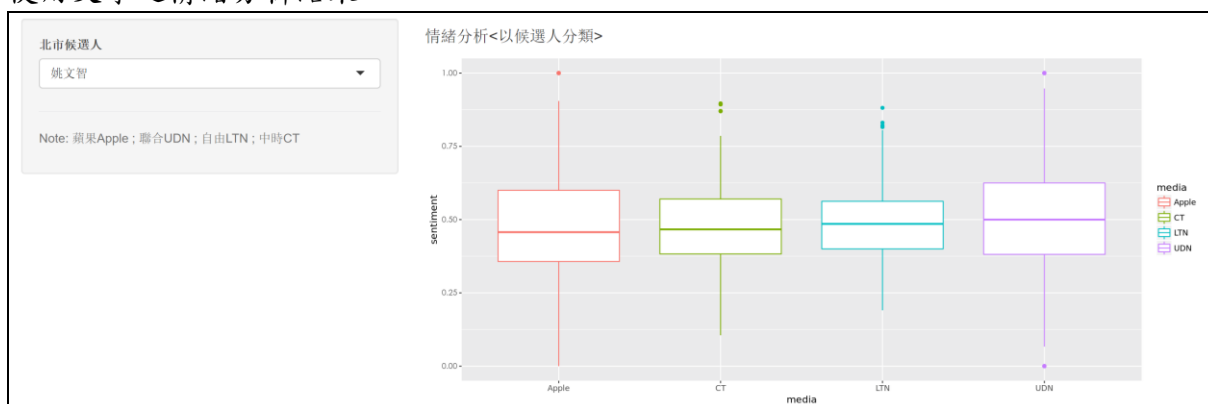
本研究從傳統媒體所獲取之新聞報導資料皆為「非結構化數據」。在中文環境中，我們使用 jiebaR (<https://cran.r-project.org/web/packages/jiebaR/index.html>) 將文本進行斷詞，並轉置成可分析使用之單詞庫。針對此單詞庫，使用國立臺灣大學情緒詞典 NTUSD (<http://academiasinicanlplab.github.io/>)，進行剖析，將正反用語挑選出來進行統計，計算此文章正面用語數量與負面用語數量進行評分。其評分方式如下公式，利用分數是否大與 0.5 分決斷此文章的傾向 (Wang, Shih-Ming, and Lun-Wei Ku 2016)。

$$\text{sentiment score} = \frac{\text{Positive words}}{\text{Positive words} + \text{Negative words}} \times 100\%$$

為針對文本資料並提取其價值，並研究其內容，本研究使用文字探勘進行資料的探索，情緒分析是一種文字探勘的其中一種應用情境，嘗試從文本內容辨識出正面與負面單詞，利用兩極性的言詞，判斷說話者或發文者的態度 (Ortigosa, Alvaro, José M. Martín, and Rosa M. Carro 2014)。在四大報內容分析中，可利用情緒分析觀察，媒體是否對於特定候選人有偏頗的評價。從（圖八）至（圖十）可觀察出，四大報對於三位候選人的用字遣詞，整體來說，四大報對三位候選的正負面用字分布趨近一致，正面用語與負面用語數量平均落於 0.5，在今年六月前，仍是選戰初期，不難看出四大報皆想維持，平衡報導的「中立」原則。

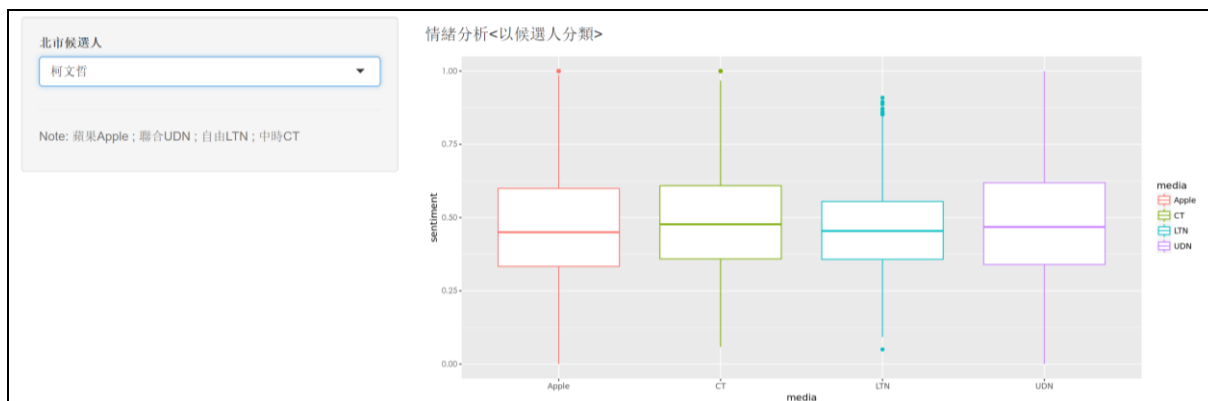


圖八、四大報（聯合報、中國時報、自由時報、蘋果日報）分別對於丁守中的報導文章，使用文字之情緒分析結果。

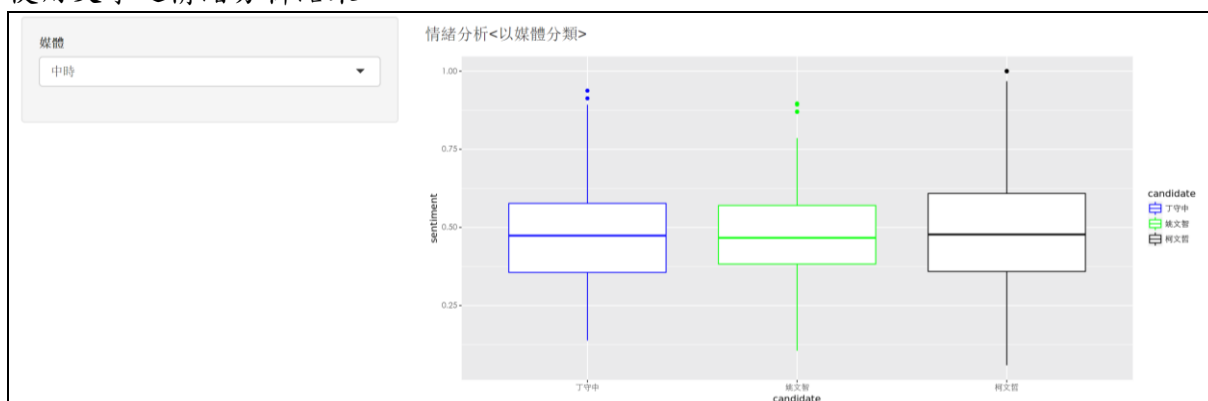


圖九、四大報（聯合報、中國時報、自由時報、蘋果日報）分別對於姚文智的報導文章，使用文字之情緒分析結果。

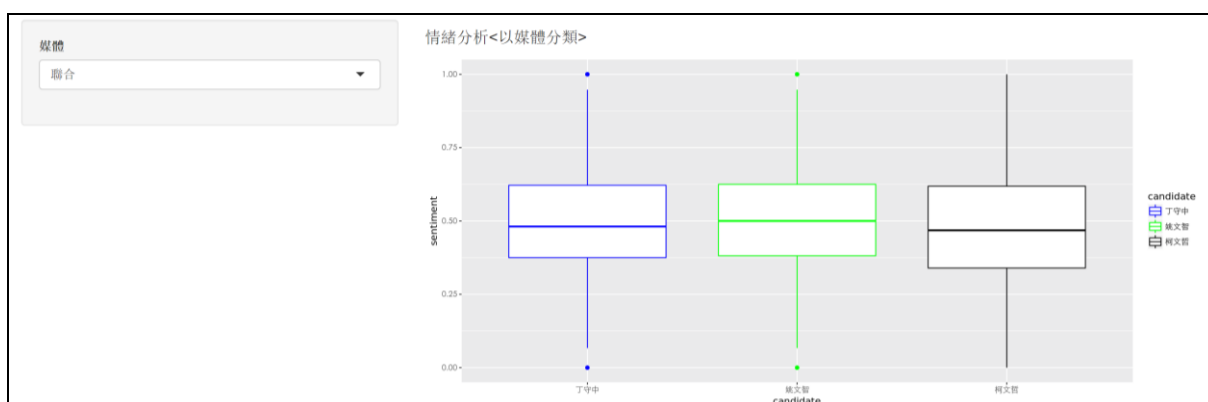
普遍的研究結果顯示，聯合報與中國時報，是屬於國民黨的統派報紙(林炳宏 2003) 自由時報是屬於民進黨的獨派報紙(林麗雲 2008)，若從這個假設出發，聯合報與中國時報，應該會盡量少負面描述丁守中，而自由時報應該會使用較多負面文字，來描述支持統派的候選人。雖然（圖八）至（圖十）的視覺化結果，四大報的用字分布非常接近，但是，換個角度觀察，從（圖十）至（圖十四）仍可看到四大報報導三位候選人，些微的不同處。觀察（圖十一），發現中國時報，在對三位候選人的進行描述時，相對使用較少的負面字描述丁守中，有符合我們一開始的假設。



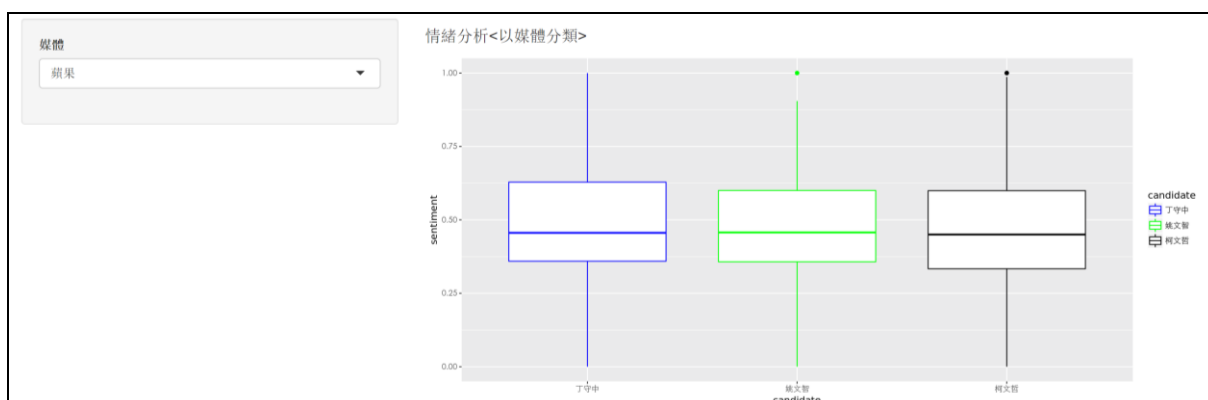
圖十、四大報（聯合報、中國時報、自由時報、蘋果日報）分別對於柯文哲的報導文章，使用文字之情緒分析結果。



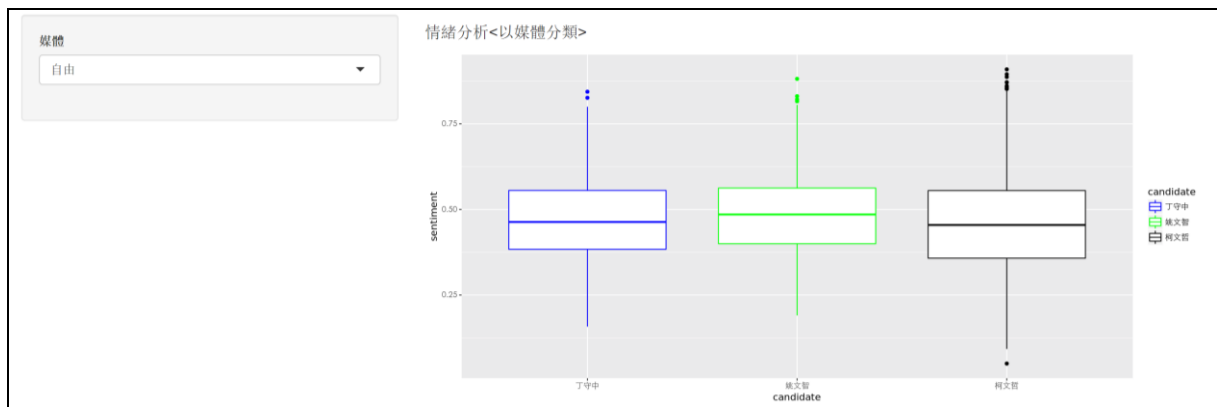
圖十一、中國時報分別對於三位候選人的報導文章，使用文字之情緒分析結果。



圖十二、聯合報分別對於三位候選人的報導文章，使用文字之情緒分析結果。



圖十三、蘋果日報分別對於三位候選人的報導文章，使用文字之情緒分析結果。



圖十四、自由時報分別對於三位候選人的報導文章，使用文字之情緒分析結果。

但（圖十二）與（圖十三）顯示，聯合報與蘋果日報，反而對三位候選人的用字遣詞較一致。觀察（圖十四），發現自由時報對於丁守中與姚文智，使用較少的負面描述，卻很明顯特別針對柯文哲有較多負面用字，這是非常有趣的結果，顯示民進黨在今年五月宣布，年底首都之戰，民進黨確定不再禮讓台北市長柯文哲，今年五月五日，聯合報及刊登出相關報導（<https://udn.com/news/story/11311/3124633>），柯文哲與民進黨決定「分手」後，自由時報六月就已開始醞釀整個選戰輿論風向。

二、詞頻文字雲

我們延續使用 jiebaR 文本斷詞後的單詞庫。針對此單詞庫，繪製出詞頻文字雲，可觀察三位候選人，最頻繁使用的單詞為何。心理學家福斯特的搜索模式認為，在心理詞典中，詞條的排列順序是從高頻詞到低頻詞，而詞的認知是通過搜索達到的。由於高頻詞比低頻詞較早被搜索，所以高頻詞比低頻詞被認知得更快(Churchill, Neil C., and William D. Bygrave 1990)。



圖十五、丁守中臉書粉絲專頁使用文字之詞頻統計，繪製出文字雲。

從（圖十五）與（圖十七）觀察發現，丁守中與姚文智，為了讓選民能高度認知到自己，才會反覆提起自己名字，藉以讓選民透過詞頻效應（word frequency effect）來記住自己，與加深選民對自己的印象（Taft, Marcus 1979）。但從（圖十六）卻看到，

柯文哲是唯一沒有把自己名字常掛在嘴邊的候選人，代表柯文哲非常有自信，自己已在選民認知中，超越了上述的心理學作用，反而是強化自己的政績，例如：進步、光榮、燈節、教育、公園、捷運、公車。



圖十六、柯文哲臉書粉絲專頁使用文字之詞頻統計，繪製出文字雲。



圖十七、姚文智臉書粉絲專頁使用文字之詞頻統計，繪製出文字雲。

三、潛在狄利克雷分配模型

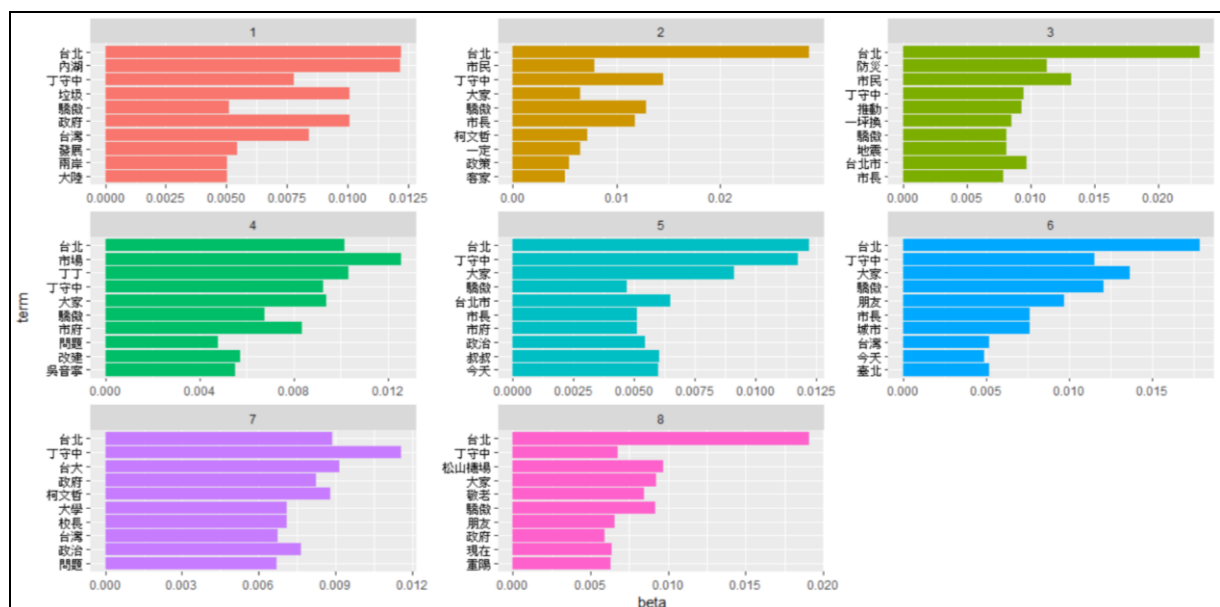
潛在狄利克雷分佈（簡稱 LDA），是一種主題模型，它可以將文本中的每篇主題，按照機率分布的形式歸納出結果（Blei, David M., Andrew Y. Ng, and Michael I. Jordan 2003）。LDA 是一種無監督學習算法，在訓練時，不需要手工標註的訓練集，只需要指定有多少種類的主題即可。LDA 可歸納出一個文本由多個主題組成，一個主題對應一個相關詞項集合，而 LDA 分佈描述了文本至主題層面的機率分佈，對於每一個主題，均可找出一些關鍵詞來描述該主題（Ramage, Daniel, Christopher D. Manning, and Susan Dumais 2011）。

將三位候選人的臉書粉絲專頁文本進行斷詞，並轉置成可分析之單詞庫，再使用 topicmodels (<https://cran.r-project.org/web/packages/topicmodels/index.html>) 的 R 語言套件包，分別針對三位候選人的臉書粉絲專頁文本，指定分類出八個主題，進

行 LDA 分析。從（圖十八）可明顯看出，由 LDA 歸納出柯文哲的八大主題，分別為：1. 世大運議題、2. 交通議題、3. 藝術議題、4. 教育議題、5. 國宅議題、6. 觀光議題、7. 選民向心力操作、8. 銀髮族議題。柯文哲仍是主打過去的執政績效為主，並在粉專中埋下了與選民向心力操縱有關的文字，例如：大家、價值、光榮、參與、進步。圍繞在大家一詞上，希望讓選民一同參與市政，與台北一同進步，產生台北價值。



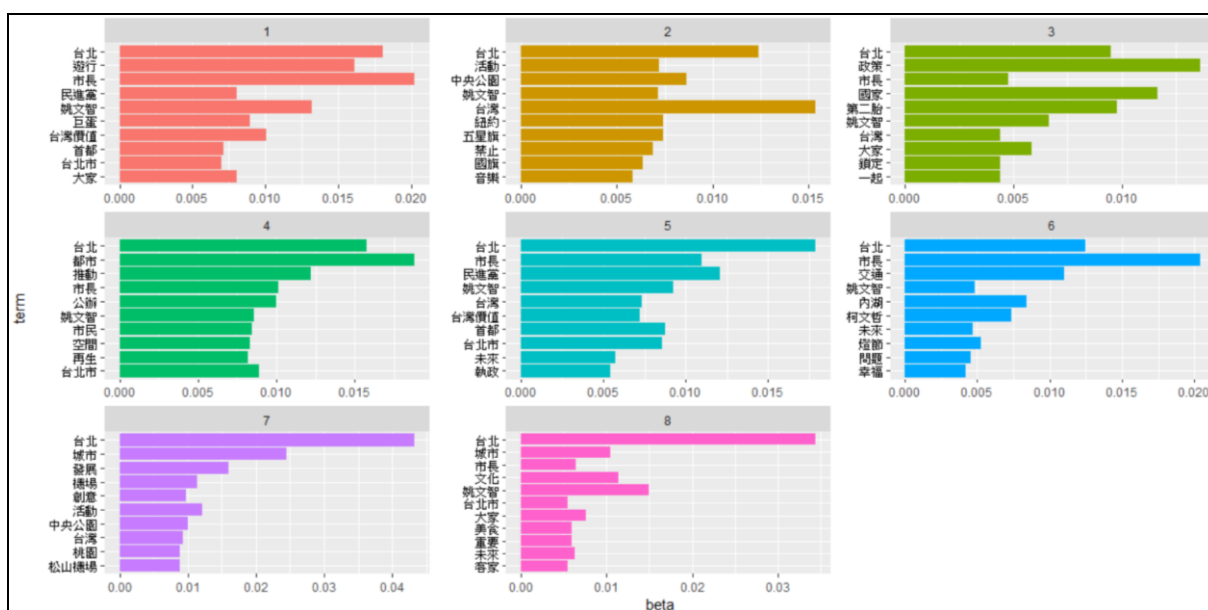
圖十八、柯文哲臉書粉絲專頁使用文字之詞頻統計，繪製出潛在狄利克雷分配模型。



圖十九、丁守中臉書粉絲專頁使用文字之詞頻統計，繪製出潛在狄利克雷分配模型。

從（圖十九）中明顯看出，由 LDA 歸納出丁守中的八大主題，分別為：1. 兩岸議題、2. 客家政策、3. 防災議題、4. 市場改建議題、5. 至 6. 較接近丁守中與選民的喊話、7. 台大校長一案、8. 銀髮族議題。丁守中在每一個主題都出現了自己的名字。並且提了和市政無關的台大校長爭議案，以及柯文哲與吳音寧之間尚未解決的爭議議題。關於強化

自己政見的部分，以兩岸、防災，以及銀髮族相關，稍微少了一些台北市民更想關注的進步、光榮、價值等面向，字面上的感受，較為保守求穩定。從（圖二十）中明顯看出，由 LDA 歸納出姚文智的八大主題，分別為：1. 巨蛋爭議、2. 兩岸議題、3. 生育政策、4. 空間再生、5. 台灣價值、6. 交通問題、7. 城市發展、8. 客家政策。姚文智也是同丁守中一樣，幾乎在每一個主題都出現了自己的名字。並且提了柯文哲與巨蛋一案尚未解決的爭議議題。關於強化自己政見的部分，以兩岸、生育，以及空間再生相關，有比較多使用台灣價值、幸福、發展，以及創意等求改革與轉型的氣氛營造，相對丁守中的保守，有較明顯的對台北市有更多創新期許。



圖二十、姚文智臉書粉絲專頁使用文字之詞頻統計，繪製出潛在狄利克雷分配模型。

伍、結論

本文研究從一開始就採用不同以往的研究步驟，從透過網頁爬蟲進行文本資料收集取代問卷與訪談記錄資料，接著定義網路聲量指數的建構與應用，來判斷候選人經營臉書粉絲專頁的成效，再透過非結構化資料探索之文字探勘過程，探討經營績效好的候選人，是如何建立受到選民青睞的議題，製造出受重視的討論熱度，進而提升網路聲量指標。發現柯文哲的成功在於，台北市的選民開始有自我意識，不再受制於，國民黨或民進黨一直以來慣用的族群對立選戰打法所影響。

透過臉書粉絲團的經營績效反映出，台北選民在乎的是，候選人實質在台北推動了何種真實的影響實績。柯文哲的發文，非常專注在市政經營上，所以，從統計分析結果發現，柯文哲完全沒有提到另外兩個對手的名字。但有趣的是，丁守中與姚文智，在各項文字探勘的結果中，都顯示出，反覆提到柯文哲的名字，也提出柯文哲處理得比較不

好的市政。這明白地應證了一句話，當你用一根手指指責別人的時候，其實有四根手指在指著自己。柯文哲的經營之道，就是非常功利且務實地去落實每一項任務，所以，不需要大量的發文，卻能很精準的發文，並且不去說對手的不是，也不去反駁對手的指責，這樣就事論事的風格，自然受到廣大會思考的選民喜愛。

參考文獻

1. Mostafa, Mohamed M. "More than words: Social networks' text mining for consumer brand sentiments." *Expert Systems with Applications* 40.10 (2013): 4241-4251.
2. Steinberger, Ralf. "A survey of methods to ease the development of highly multilingual text mining applications." *Language Resources and Evaluation* 46.2 (2012): 155-176.
3. Freeman, Linton C. "Centrality in social networks conceptual clarification." *Social networks* 1.3 (1978): 215-239.
4. Aggarwal, Charu C., and Haixun Wang. "Text mining in social networks." *Social network data analytics*. Springer, Boston, MA, 2011. 353-378.
5. Wang, Shih-Ming, and Lun-Wei Ku. "ANTUSD: A Large Chinese Sentiment Dictionary." *LREC*. (2016).
6. Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3.Jan (2003): 993-1022.
7. Ramage, Daniel, Christopher D. Manning, and Susan Dumais. "Partially labeled topic models for interpretable text mining." *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011.
8. Yang, YunYun, et al. "Text mining and visualization tools—impressions of emerging capabilities." *World Patent Information* 30.4 (2008): 280-293.
9. Ortigosa, Alvaro, José M. Martín, and Rosa M. Carro. "Sentiment analysis in Facebook and its application to e-learning." *Computers in human behavior* 31 (2014): 527-541.
10. Taft, Marcus. "Recognition of affixed words and the word frequency effect." *Memory & Cognition* 7.4 (1979): 263-272.
11. Churchill, Neil C., and William D. Bygrave. "The entrepreneurship paradigm (II): chaos and catastrophes among quantum jumps?." *Entrepreneurship Theory and Practice* 14.2 (1990): 7-30.
12. 林炳宏. "爭議性政治議題的新聞文本框架分析---以林毅夫返臺奔喪事件為例." *臺灣師範大學大眾傳播研究所學位論文* (2003): 1-175.
13. 林麗雲. "變遷與挑戰：解禁後的台灣報業." *新聞學研究* 95 (2008): 182-212.