

意言既出，駟馬難追：

政事新聞意見紀錄簿

The Political News Opinion Recorder

E14084117

F74082141

黃子峻
王浩

大綱

①

發想動機

②

使用情境

③

架構

④

技術介紹

⑤

SWOT 分析

1

發想動機

Motivation

每天的政治新聞資訊量 太龐大

熱門的新聞網站有 TVBS、中國時報
ettoday、民視、三立、風傳媒等等數家媒體
總合一算，每天的政治新聞動輒上百篇
閱聽者對於資訊量，難以全部應付



**以事件群、議題為單位，整合各家媒體所釋出的資訊便
可在較少的閱讀量下達到更全面的了解**

純文字文章的新聞 結構難以閱讀

傳統新聞的資訊表達方式
都只把新聞的背景介紹、人物意見
用整篇「文字文章」的方式呈現



以**視覺化人物發言**呈現、**整理單篇新聞內的文章**、
輔以**情緒分析**便可在較**清楚明瞭**的模式下，獲取資訊

不利追蹤政治人物長期 對於議題的發言

目前新聞多只以當日事件發展進行報導
並不利於讀者了解、追蹤
長期政治人物對於議題的意見



利用**新聞分群**、**人物判斷**，能夠在同一頁面，**貫**
穿時間軸，直接看出政治人物的意見變化

2

使用情境

Demo

新聞焦點



蔡英文宣布松機加開篩檢站 柯文哲:我事先知道且反對

總統蔡英文昨天前往桃園時宣布,三總會在松山機場的大客車停車場增加一個大型篩檢站,這個消息讓台北...



朱立倫道歉「與桃議員溝通不周」!仍堅持提名張善政選桃園

對於桃園市長提名,中國國民黨朱立倫主席今(21)日與黃健庭秘書長及黨務幹部共同召開會議,討論桃...



「我任內年年出席WHA」!馬英九嘆:一定要恢復兩岸協商

世界衛生大會(World-Health-Assembly,WHA)即將於5月22日到28日舉行,台...



陳時中投書立陶宛媒體 籲各界支持台灣參與WHA

立陶宛國家廣播電視(LRT)今天在網站刊登衛生福利部長陳時中的投書,內容分享台灣的「科技防疫」...



蔡英文宣布松機加開篩檢站 柯文哲:我事先知道且反對

總統蔡英文昨天前往桃園時宣布,三總會在松山機場的大客車停車場增加一個大型篩檢站,這個消息讓台北市長柯文哲今天直言,「我事先知道,而且反對。」副市長黃珊珊也說,昨天才知道松山機場設置篩檢站「不是車...

議題關鍵字

柯文哲 中央 侯友宜 台北市 陳時中

相關新聞列表

1 端午連假將至 國民黨智庫促政府及早防疫

國民黨智庫今天表示,即將到來的端午連假恐因返鄉群聚,造成確診增加讓疫情出現雙峰甚至多峰的現象...

2 考試院開設關懷中心 強化協助確診員工

考試院今天指出,因應疫情,已通令所屬部會,即日起擴大現有防疫機制功能,成立關懷中心,主動提供考...

3 柯文哲反對增設松機篩檢站 蘇貞昌喊話中央地方合作

行政院長蘇貞昌今天視察松山機場旁新設大型篩檢站整體規劃,同時也向雙北喊話,盼中央、地方合作打贏...

蔡英文宣布松機加開篩檢站 柯文哲:我事先知道且反對

總統蔡英文昨天前往桃園時宣布,三總會在松山機場的大客車停車場增加一個大型篩檢站,這個消息讓台北市長柯文哲今天直言,「我事先知道,而且反對。」副市長黃珊珊也說,昨天才知道松山機場設置篩檢站「不是車來速」。蔡英文昨天下午視察桃園市八德區公所「防疫關懷包快易通」,辦理情形時宣布四大重點防疫工作,特別是為了滿足PCR篩檢量能的需求,也在執政團隊協調之下,現在台北基桃已經成立8個大型篩檢站。最快在今天,三總也會在松山機場的大客車停車場,增加一個大型篩檢站,提供篩檢、看診、領藥三合一服務。消息一出,柯文哲今天受訪對著現場媒體說:「我可以講實話嗎?我事先知道,而且反對。」他認為,不要為了說而說,因為三總總醫院名額466人、中山分院160人,現有應該要在現有的點提升能量,不是設立新的點。「不要為了說而說,醫學中心戰略預備部隊,不管三總跟台大都還沒開出來,有需要就加重,不是開設新的點,不要瞎指揮,造成地方困擾。」柯文哲說道。柯文哲說,中央設置新篩檢站,他們一定會去維持交通,但打仗是要有戰略,因為北市計畫防疫急門診擋在第一線,同時將住院標準嚴格、防疫專責旅館加重,加上65歲以下用車來速、65歲以上用防疫急門診分流,系統效率才能達到最高。他強調,北市有一定作戰計畫,但中央直接下令來,根本不知道「我們的作戰計畫」,因為三總門診量都還沒開出來,怎會叫人設立第三個點,這是戰術錯誤。站在一旁地黃珊珊也透露,醫福會執行長王必勝有問她要不要開設下一站(篩檢站),他們也回應當前要先量能是否有必要擴充,反倒是希望實驗室量能增加給北市,譬如台大400擴充至800或1000人次。黃珊珊表示,王必勝的確有詢問,只要中央設站「交通我們會幫忙」,因為中央要設站「我們也無法阻止」,但是昨天才知道松山機場篩檢站「不是車來速」。她強調,北市一直表示希望現有量能增加為優先,不要讓醫院疲於奔命。

蔡英文
2022-05-21



三總會在松山機場的大客車停車場增加一個大型篩檢站,這個消息讓台北市長柯文哲今天直言,「我事先知道,而且反對。」

查看過去意見



桃園市八德區公所
2022-05-21



四大重點防疫工作,特別是為了滿足PCR篩檢量能的需求,也在執政團隊協調之下,現在台北基桃已經成立8個大型篩檢站。

查看過去意見



柯文哲
2022-05-21



中央設置新篩檢站,他們一定會去維持交通,但打仗是要有戰略,因為北市計畫防疫急門診擋在第一線,同時將住院標準嚴格、防疫專責旅館加重,加上65歲以下用車來速、65歲以上用防疫急門診分流,系統效率才能達到最高。



柯文哲



2022-05-21 11:57



中央不知道北市計畫就直接一道命令下來,「不要瞎指揮,造成地方困擾」,且三軍總醫院及其分院量能都還沒開出來,叫人家設篩檢站沒道理。

2022-05-21 19:33



此舉是造成地方困擾,批評不要瞎指揮、亂指揮議員簡舒培反喻,最聾的就是你柯文哲,「你在此,北市府發言人陳智菡表示,盲目擴張PCR量題,這時分散人力新增篩檢站,就是在錯誤的時策,表面工夫虛應故事、累死三軍。

34.80.42.37:5000/dynamic/person/柯文哲/2022-05-30/0

2022-05-25 17:50

現在醫療量能還撐得住,如果確診數少,還可以適度放寬。

2022-05-25 17:50

「我也不要再當印度神童」,疫情會拖多久其實與政府控制有關,地方政府過去就在規畫線上教學,有意踩剎車,假設地方堅持實體上課,每日確診數就是5萬、10萬。

2022-05-25 17:50

開放政府、資訊公開透明大家就會跟著做,但可能是因為台灣政治習慣不說清楚、講明白,講得很隱晦,所以台灣戰略一直不太清楚。

2022-05-25 17:50

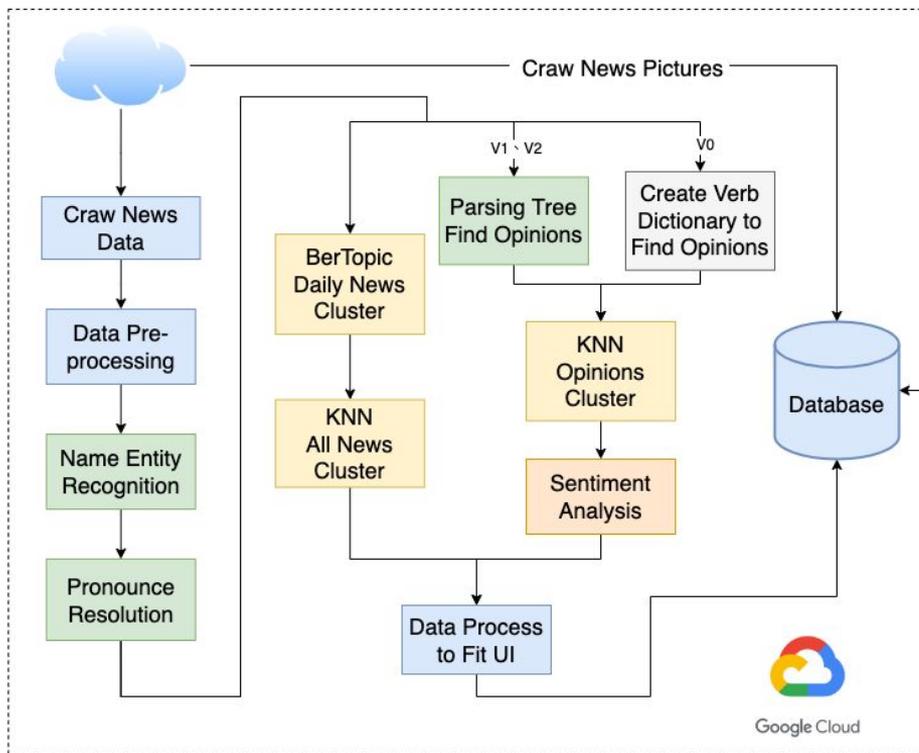
視訊診療還需要時間改善,目前是先求有再求好,北市府會盡快找出問題並修正。

3

架構

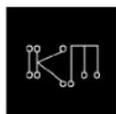
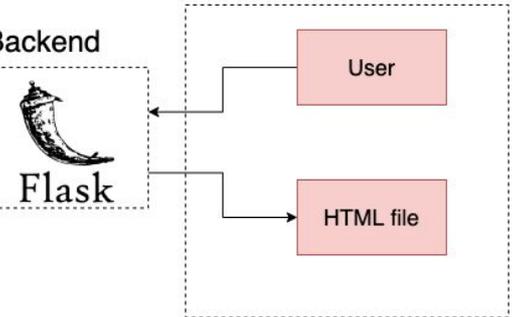
Architecture

Server



Client

Backend



4

技術介紹

Skills Introduction

CKIP Lab 中文剖析

基本上句子自動剖析利用

語法規律和

斷詞後的文本做比對,

找出可能的短語結構,

由於**存在歧義的短語**結構,

因此, 如何利用結構出現的**機率**及

檢測結構中**詞與詞之間搭配的合理性**

成為解決結構歧義的方法。

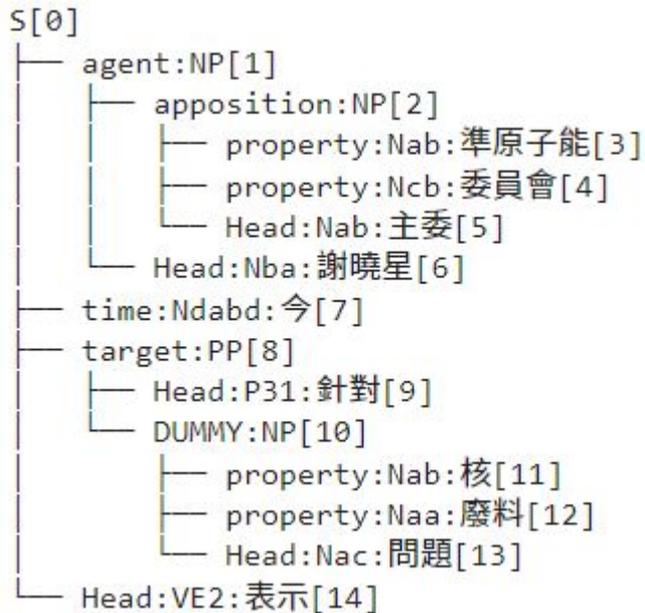
句法抽取

統計相關的**規則機率**、**中心語機率值**及**中心語與搭配語機率**, 作為剖析器歧義結構挑選與機率統計的依據

結構剖析

除了利用規則機率外並**考慮詞彙搭配的合理性**作為歧義結構評估的方法

中文剖析樹 Parsing Tree



語意角色

用於**修飾**物體名詞、修飾事件動詞的角色，有少數幾個用來標記語法功能，共60個。

- agent** – 表事件中的肇始者，動作動詞的行動者
- apposition** – 表物體的同位語，即指涉相同的物體
- target** – 述詞內容表達的對象或是轉移的方向

詞類分析

每一個詞類都會代表的一定的語法意義

- NP** – 名詞組
- Nab** – 實體可數的非集合名詞，即為個體名詞
- Nba** – 正式專有名稱，包括人名、歷史事件等等
- VE2** – 二元述詞，一個主語一個句賓語

語意角色

agent

表事件中的肇始者，動作動詞的行動者

詞類分析

VE2

二元述詞

以主事者 (agent) 為主語，終點 (goal) 為句賓語，語意多為表語言行為之述詞

VC

動作單賓述詞

需要兩個參與論元；在語法上，這兩個參與論元都是名詞組，分別充當述詞的主語和賓語

VE11

動作句賓述詞

接句子論元的動作述詞，語意多為語言行為或表思維活動。需要兩個或三個論元來滿足其語意

Case study

version 0

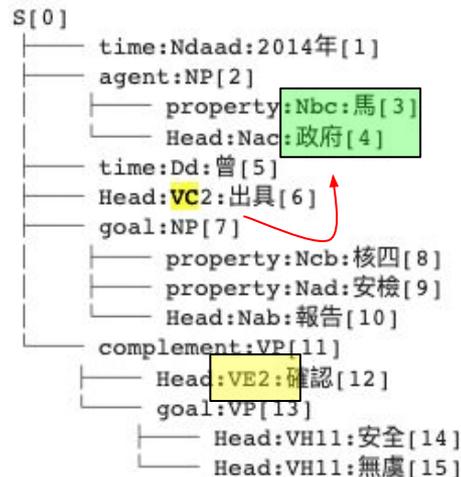
<

version 1

<

version 2

原能會表示，作為獨立的安全管制機關，對核四公投「沒有預測立場，也會尊重公投結果」



動詞字典往前的
第一個專有名詞

Parsing Tree VE2
一對一結構

Parsing Tree VE2 + VC + VE11
且複雜結構

分析數據

	配對方式		數據結果		
	動詞選擇	主詞選擇	precision	recall	f1
version 0	手動建立動詞字典	動詞往前之特定條件下的第一個專有名詞	0.689	0.521	0.575
version 1	Parsing Tree VE2	與動詞屬同一句子的 agent, 且結構為一對一	0.679	0.518	0.569
version 2	Parsing Tree VE2 + VC + VE11	與動詞屬同一句子的 agent, 且結構不限	0.609	0.699	0.639

(驗證數據: 100篇手動標記 ettoday 新聞意見)

單名代名詞 統一代換

利用 Name Entity Recognition 將

人物PERSON、**組織ORG**、**民族宗教政治團體NORP**加入候選列表，

因為我們的目標是代換第一個字相同的主詞，

所以其他沒有被列入這個列表的 person 等等 NER

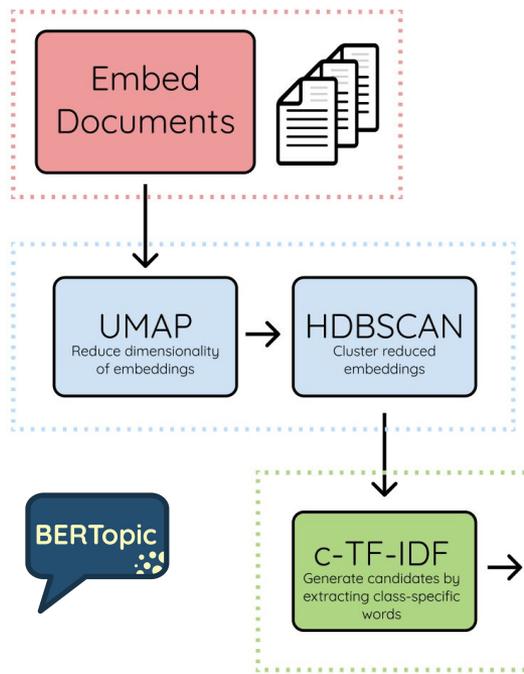
就會根據第一個字，被代換成候選列表內的字，

以達到統一化單名代名詞。

e.g. **蘇**今日在立法院表示...

如果文章內出現過多個**蘇貞昌**，那這句話就可以被帶換。

BERTopic 將新聞以事件分群



Step 1: 把政治新聞的標題文字用 BERT 模型轉為向量表示

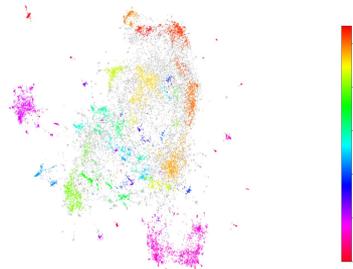
『刑事局緩頰稱「郭彥均對社會關心」 凌濤:打臉蘇貞昌』 [0.12, 0.34, ..., 0.56]

Step 2: 以 UMAP 算法將向量降維, 再用 HDBSCAN 分群

UMAP降維(neighborhood=15)

[0.12, 0.34, 0.56, 0.78, 0.90, ..., 0.12, 0.34, 0.56] [0.35, 0.34, 0.26, 0.18, 0.43]

HDBSCAN 分群



Step 3: 使用 c-TF-IDF 找出適合代表分群的詞彙

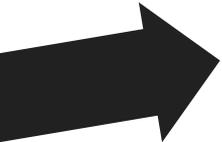
$$c-TF-IDF_i = \frac{t_i}{w_i} \times \log \frac{m}{\sum_j^n t_j}$$

關鍵詞: 疫情、蘇貞昌、刑事局.

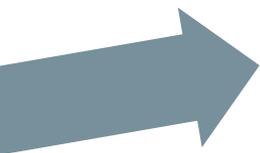
5

SWOT 分析

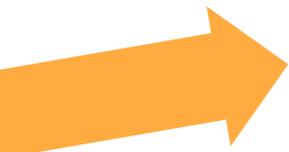
SWOT analysis



以議題分群減少新聞閱讀量



人物視覺化方式呈現新聞內容



追蹤政治人物長期的議題意見

SWOT 分析

STRENGTHS

節省大量閱聽者接收資訊的時間、
更清楚掌握長期政治人物的
議題意見變化



WEAKNESSES

意見擷取並非完全正確
演算法有進步的空間



在資訊爆炸的時代，
自動化萃取意見、
掌握跨時間的關鍵發言

OPPORTUNITIES

以分群功能來說
網路上有更具規模的
分群新聞網站

THREATS



謝謝聆聽！