

意言既出，駟馬難追：政事新聞意見記錄簿

The Political News Opinion Recorder

指導教授：高宏宇教授

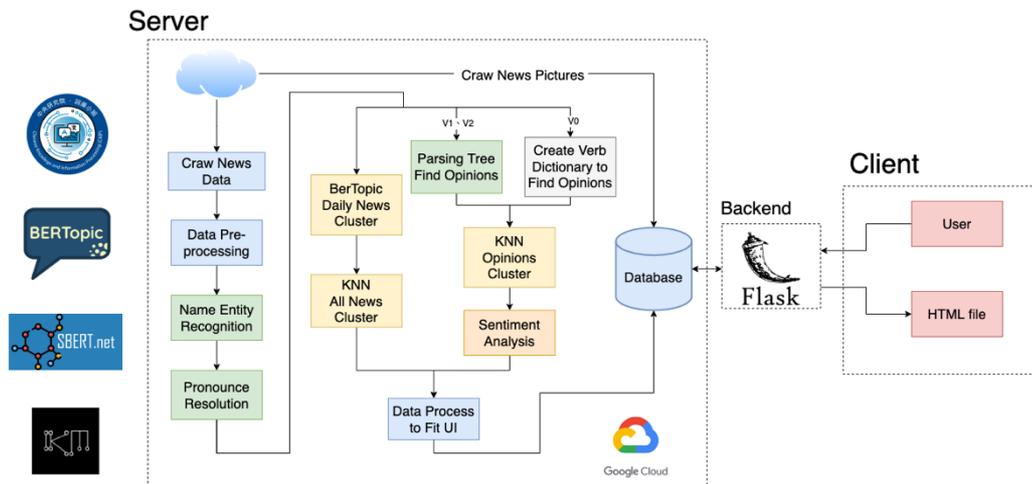
專題成員：黃子峻、王浩

開發工具：Python、Flask

測試環境：Windows10

一、簡介：

我們以事件群、議題為單位，整合各家媒體所釋出的資訊，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，最後利用議題分類、人物判斷，能夠在同一頁面，貫穿時間軸，直接看出政治人物長期的意見變化。



1. 我們先爬下多家網路媒體的新聞，進行前處理，把多餘的空格去除，方便後續任務的執行。
2. 執行 Name Entity Recognition，找出文章中的專有名詞，並去除時間、數量值等等比較沒有利用價值的專有名詞，以利後續我們統計出分群關鍵字。
3. 我們根據剛剛得到的專有名詞，自行建立一個人名字典，把文章中裡面出現的人物單名根據字典做統一的代換。
4. 接著任務分成兩部分，一個為分群任務，另一個為意見擷取。
5. 分群任務我們先利用 BERTopic 對文章進行分群，但因為我們的時間維度要做到長期的議題追蹤，所以我們再根據已經分完群的當日議題，使用 sentence-transformer 以 ckiplab/bert-base-chinese 為 embedder 再做一次 KNN 分群，如此一來，我們就可以知道哪幾天的哪些事件是屬於同一個議題，做到跨天的分類。
6. 在意見提取的部分，我們使用兩種方式做驗測，第一個是使用 CKIP

CoreNLP Toolkit Parsing Tree；另一個為利用自行建立的動詞字典，再搭配 Name Entity Recognition 找出動主詞配對。

7. 為了避免意見重複出現在同一篇文章中，我們接著一樣再使用第五點提到的 KNN 分群，把相似度高的意見去除。
8. 意見提取完後，我們以 NTUSD 情緒字典 比較我們擷取出來的意見，作為判斷正負情緒的依據。
9. 最後我們根據需要的 UI 模式，調整檔案形式，並自行設定自動化爬取、以 Flask 為框架，將系統架設在 Google Cloud Platform 以供閱聽者使用。

二、測試結果：

	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	動詞往前的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)

The screenshot displays the ettoday news website interface. At the top, there's a navigation bar with '政事新聞意見紀錄簿' and buttons for '上一天新聞' and '下一天新聞'. Below this is the '新聞焦點' section with three featured articles: '免費快篩對象擴大!', '布林肯證實中對台灣迫 外交部:美方事前向台灣簡報', and '快訊/立法院三讀《鐵路法》修正 鬆綁土地、加重違規刑責'. The main content area is divided into '新聞分類 · 分類列表' and '新聞新聞概覽'. The '新聞新聞概覽' section features a large article titled '台鐵邁向公司化 立委:行車安全為最高指導原則'. This article discusses the transition of the Taiwan Railway Corporation to a public company and emphasizes that railway safety is the highest guiding principle. It mentions that the立法院 (Legislative Yuan) has passed the revised Railway Law, which aims to streamline land use and increase penalties for violations. The article also notes that the government will continue to support the railway's operations and ensure its financial stability. Below the main article, there are sections for '相關新聞列表' and '人物意見 · 議題追蹤', featuring profiles and opinions of key figures like 曾銘宗 and 陳樹德.