

聞到有先後：基於多源文本的段落生成

以及資料加值系統實作

ChronoNews: Multi-Source

Textual Paragraph Generation

and Value-Added System Implementation.

指導教授：高宏宇

專題成員：陳冠廷、鄭宇辰

開發工具：ChatGPT, Python

測試環境：Google Colab

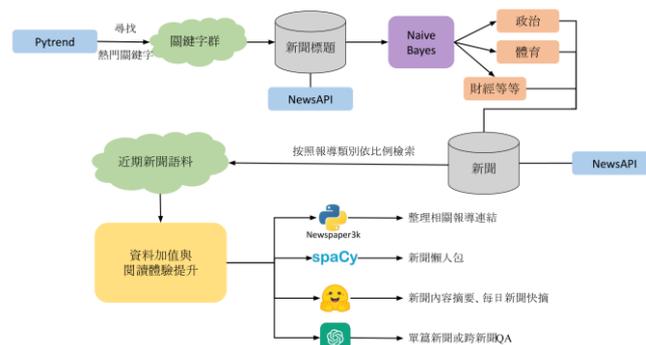
一、簡介：

近年來，台灣人越來越難從新聞中獲得國際新聞。我們希望開發一個整合平台，帮助大家更輕鬆的接觸國外的新聞資訊。

我們以 Python 為主要開發語言，透過 PyTrend、NewsAPI 等套件擷取每日熱門關鍵字及其近日相關新聞（First Layer）。使用 Kaggle 資料集訓練 NLTK 的 Multinomial NB Classifier；將新聞標題分類，並根據類別給予不同比例來源的新聞（Second Layer）。如關鍵字出現某演員，近期報導多針對其社會事件官司，則“buzzfeed”，“ign”等娛樂平台的新聞會較“BBC”，“CNN”的新聞還少。

我們以 Second Layer 文本製作 corpus，以 SpaCy 識別 corpus 的命名實體。根據《懶人包超文本敘事設計分析：順序與結構元素的討論》（陳雅惠，2018）所列舉的懶人包結構樣式，我們將人／組織的行為（plot）透過 Hugging Face, OpenAI 提供的套件及相關工具重新生成標題，並且針對某熱門關鍵字的 corpus 進行摘要。最後將摘要及翻譯（使用 Googletrans 套件）以及相關連結並陳。

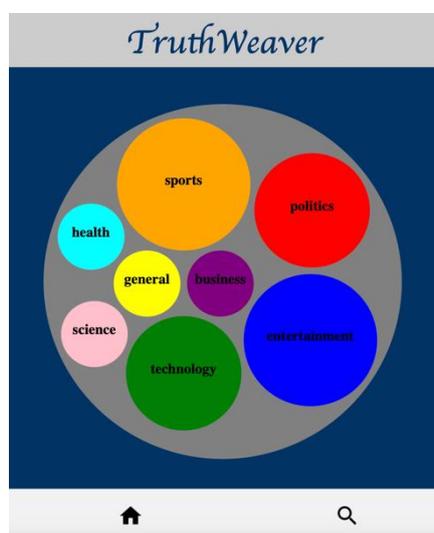
透過 ChatGPT 協助製作呈現網站，以 Flask 作為後端，html、css、js 作為開發語言，具有響應式網站的設計，讓使用者在不同裝置上都有良好閱覽體驗。



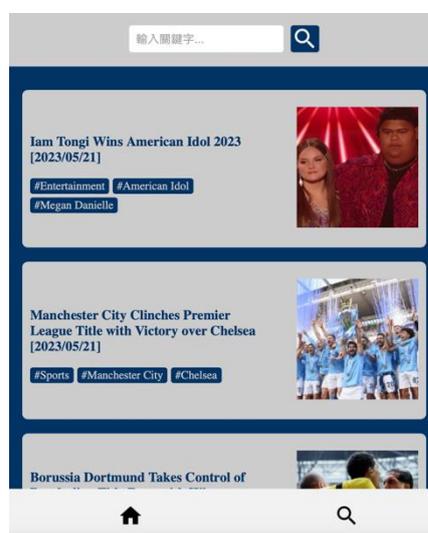
圖一：系統架構

二、測試結果：

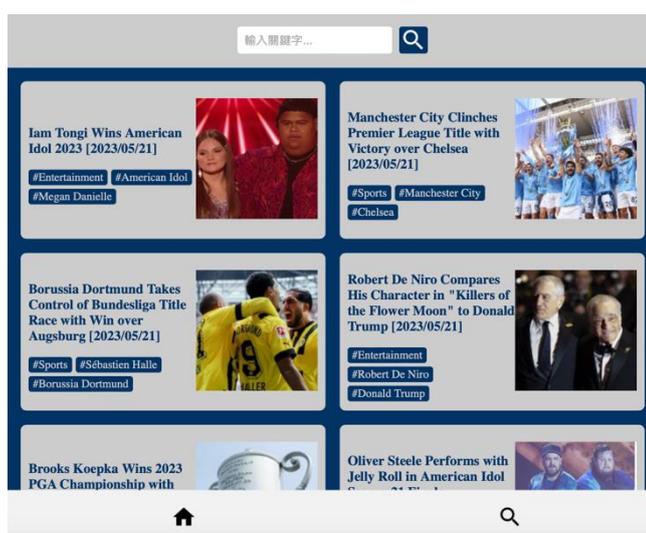
- 首頁的新聞類別能夠正確顯示，並且透過圖形大小可視化，使用者可以清楚看出今日熱門關鍵字在各類別的比例大致如何。
- 搜尋頁的標題以及 tag 出自於 SpaCy、Hugging Face 以及 OpenAI 套件。
- Multinomial Naive Bayes Classifier 可以透過 First Layer 的標題正確判斷新聞內容應該歸屬於哪一個類別。
- 網站介面可以符合各種比例的視窗、裝置，讓使用者獲得最舒適的操作體驗。
- 歡迎大家來現場看看我們的懶人包內容～



圖二：系統首頁



圖三：系統搜尋頁



圖四：系統搜尋頁（演示 RWD 效果）