

# Sentiment News Project Team3

學生:

A1105505 林彧頌

A1105507 蘇柏諺

A1105521 黎子歲

A1105523 巫柔筠

A1105524 吳雨宣

教授:

黃健峯 教授

## Table of Contents

<b>1. MULTI-SOURCE CRAWLING.....</b>	<b>2</b>
<b>2. REAL-TIME DATA COLLECTION.....</b>	<b>4</b>
<b>3. TOPIC-SPECIFIC CRAWLING .....</b>	<b>6</b>
<b>4. SENTIMENT ANALYSIS .....</b>	<b>9</b>
<b>5. DYNAMIC AND ADAPTIVE CRAWLING .....</b>	<b>10</b>
<b>6. DATA STORAGE.....</b>	<b>11</b>
<b>7. VISUALIZATION.....</b>	<b>28</b>
<b>心得 .....</b>	<b>34</b>
<b>分工 .....</b>	<b>36</b>

## 1. Multi-source Crawling (如有需要雲端資料庫請私訊林或頌提供)

github，或參考.sql 的資料庫)

我們爬取的新聞網站有「中時」、「自由時報」、「tvbs」，以及為了讓新聞來源更豐富，我們使用 API(<https://newsapi.org/>)來爬取更多新聞來源，請注意我們為了幫助資料庫可以快速爬取，並平衡資料庫，因此我們有設置當日新聞最多爬取三則，因此新聞量不符合實際網頁的結果，是我們僅爬取三筆的結果，但這三筆皆是經過篩選後確定為該類別的新聞。

	中時			自由時報			tvbs			API		
時間	最多爬到 2 年前 (2022/10/24)			最多只能爬 500 頁 (有些只能爬不到 2 年，有些最久可爬到 2004 年)			最多可以爬到 295 頁 (2003/6/3)			最多爬到 1 個月前		
資料 筆 數	H	St	Sp	H	St	Sp	H	St	Sp	H	St	Sp
	5068	11903	15895	15268	19188	13894	15952	16013	30422	237	373	351
	Total:32866			Total:48350			Total:62387			Total:961		
	全部爬蟲量:144564											
新聞 來 源 數	1 間			1 間			1 間			24 間		

(H:Health 的資料筆數/St:Stock 的資料筆數/Sp:Sport 的資料筆數)

根據上方表格可知，通常當一間新聞網的新聞量越大，相對的能爬取到的時間跨度越小(當然也有例外狀況，但是是少數，像是 tvbs 是因為他的 sport 數較多)。

(1) 中時:

➤ User-Agent 模擬:

程式碼使用一組常見的 User-Agent 列表，每次請求隨機選擇一個 User-Agent，以模擬不同裝置和瀏覽器的訪問，防止因固定 User-Agent 而被偵測為爬蟲。

➤ 隨機延遲:

在每次請求之間加入隨機延遲 (1-3 秒)，可以避免持續、高頻率的請求，降低被反爬蟲機制察覺的可能性。

➤ 防止 Webdriver 被檢測:

透過 Chrome WebDriver 中的 `execute_cdp_cmd` 方法，將 `navigator.webdriver` 屬性設定為 `undefined`，這可以掩蓋 WebDriver 的使用，避免被網站偵測到是自動化工具。

- 多組選擇器(CSS\_SELECTOR):  
使用 `safe_find_element` 函數來指定多組選擇器，確保即使網站隨機更改結構，也可以找到目標元素，增加爬蟲的穩定性和兼容性。
- 重試機制:  
若在預設的等待時間內無法加載目標元素，會進行最多 3 次重試，以應對網站的臨時延遲或加載問題。
- 每次抓取完一頁後，要關閉瀏覽器並重新打開網頁，如果直接從已開啟的網頁換頁，會被該網頁阻擋。

## (2) 自由時報:

- 延遲請求:  
在爬取每個頁面之後，加入 `time.sleep(1)`，讓每次請求之間稍作等待，避免過度頻繁請求觸發網站的 `rate-limiting`
- 例外處理與重新載入:  
使用 `WebDriverWait` 和 `TimeoutException` 來應對載入超時的情況。當網頁沒有在指定時間內載入完畢，會捕獲 `TimeoutException`，並重新嘗試載入頁面，用以應對因為伺服器延遲或載入緩慢導致的失敗。
- 過濾重複連結:  
因為我們在這間新聞的爬取方式是先將該網頁上的所有連結存入 `news_url_list`，但網頁上可能會出現重複的新聞連結，因此我們建立 `seen_hrefs` 集合儲存已添加的網址，避免重複爬取同一篇新聞，這可以減少重複請求並提高效率。

## (3) tvbs:

- 延遲請求:  
使用 `time.sleep(2)` 在每頁爬取後進行延遲，避免過於頻繁的請求，從而減少被認為是異常行為的可能性。
- 多重選取器容錯:  
使用多個選取器查找資料，例如時間和摘要的元素定位，這樣即使網站改變結構，也能增加抓取成功的機會。
- 異常處理:  
使用 `try-except` 機制來捕獲錯誤並忽略單篇文章的解析錯誤，確保爬蟲能繼續執行而不會因為小問題而停止。

## 2. Real-time Data Collection

我們用了第一題所提及的種種方式處理反爬蟲機制，觀察每間新聞網可爬取的最長時間，並盡可能爬到最多的資料量，接著我們設計一個功能，讓使用者可設定從今日起追溯多久以前的新聞資料。

在儲存資料的部分，我們使用了 supabase，是因為它可以支援即時數據同步，讓組員們可以同時共享及操作資料庫。但因為免費版的用戶，單一資料庫的儲存上限為 500MB，而我們爬取的新聞量十分龐大，所以開了兩個資料庫來儲存所有資料。

這兩個資料庫分別命名為 “IR\_GenAI\_Team\_Proj1” 和 “IR\_GenAI\_Team\_Proj1\_second\_Database”，每個資料庫包含 12 個資料表儲存 3 個 topic 相關的新聞及情緒分析：

**資料表 stock:**

stockID	stock_name
股票代號	股票名稱

**資料表 stock\_news:**

stockID	title	date	content	source	id	url
股票代號	標題	日期	內容	來源	index	連結

**資料表 stock\_news\_API:**

stockID	title	date	content	source	id	url
股票代號	標題	日期	內容	來源	index	連結

**資料表 stock\_news\_sentiment:**

id	news_id	sentiment	emotion	star
index	新聞編號	情緒分數	情緒類型	星等

**資料表 health:**

healthID	health_name
keyword 代號	keyword 名稱

**資料表 health\_news:**

healthID	title	date	content	source	id	url
keyword 代號	標題	日期	內容	來源	index	連結

資料表 health\_news\_API:

healthID	title	date	content	source	id	url
keyword 代號	標題	日期	內容	來源	index	連結

資料表 health\_news\_sentiment:

id	news_id	sentiment	emotion	star
index	新聞編號	情緒分數	情緒類型	星等

資料表 sport:

sportID	sport_name
keyword 代號	keyword 名稱

資料表 sport\_news:

sportID	title	date	content	source	id	url
keyword 代號	標題	日期	內容	來源	index	連結

資料表 sport\_news\_API:

sportID	title	date	content	source	id	url
keyword 代號	標題	日期	內容	來源	index	連結

資料表 sport\_news\_sentiment:

id	news_id	sentiment	emotion	star
index	新聞編號	情緒分數	情緒類型	星等

在資料表 stock\_news, health\_news, sport\_news 中，content 欄位因為包含大量文字且也屬於主鍵之一，在進行全文檢索時，易導致查詢效率降低。因此，這裡使用了 PostgreSQL 的 GIN (Generalized Inverted Index) 索引來加速這類查詢。

```
CREATE INDEX sport_news_content_gin_idx ON sport_news USING GIN
(to_tsvector(simple, content));
```

建立了一個名為 'sport\_news\_content\_gin\_idx' 的 GIN 索引，用於加速 'sport\_news' 表中 'content' 欄位的全文檢索。'USING GIN' 指定使用 GIN 索引類型，適合處理大量文本資料，並生成 inverted index 以加速關鍵字查找。'to\_tsvector(simple, content)' 將 'content' 欄位轉換為文本向量格式，'simple' 配置不區分詞幹、停用字，增加查詢的靈活性，適合快速搜

尋。

### 3. Topic-specific Crawling

本專題中，主要選定的主題為股市、健康與運動，這三主題可以說是現代人日常生活中最關心的三大議題，我們選定主題後，並不是直接將主題的名詞直接當作我們過濾新聞的主要指標，因為我們發現，如果僅單純定義某個主題的定義為關鍵字，雖然範圍很大可能以單一來看資料量多，但是該主題下許多小主題相加起來的資料量更多(如下圖比較所示)。



(由上述兩張圖可看出，設定出單一類別就會比大方向的資料來的多)

因此我們決定為每一個主題都定義十個子題，這些子題會跟主題有巨大的關聯，並且我們也利用 ChatGPT 來回答我們最好的十個子題做參考(如下頁圖所示)。



記憶已更新

我會建議你以以下10個子題進行健康主題的情緒分析：

1. 確診 - 涉及疾病的確診結果，通常帶有情緒波動。
2. 減重 - 關於體重管理，與健康生活方式相關，常引發正面或焦慮情緒。
3. 免疫力 - 討論提升或保持免疫系統健康的話題，特別是在流行病時期。
4. 營養 - 營養攝取和膳食平衡，對健康情緒的影響頗大。
5. 疾病 - 涵蓋不同疾病的討論，情緒上可能複雜多樣。
6. 病毒 - 尤其是疫情相關的病毒話題，通常帶有擔憂或焦慮的情緒。
7. 慢性病 - 長期病症的管理與影響，涉及患者和照護者的情感變化。
8. 健康飲食 - 營養均衡和飲食選擇的情緒分析，與生活品質緊密相連。
9. 癌症 - 這類話題對病患和家屬的情緒影響深刻，常帶有嚴肅的情感基調。
10. 醫療 - 涉及醫療品質、診斷準確性等，對患者的信任和滿意度有直接影響。

這些子題既涵蓋了具體健康問題，也涉及預防和生活方式等不同方面，對情緒分析的完整性和深度很有幫助。



然後我們也參考資料量相對多的中時，來判斷該子題的新聞量多寡，決定這些子題的可靠性與資料量，最後我們得出 3 個主題與各十個子題，最後有 30 個主題可以獲得，此 30 個主題請參閱第二節或第六節的資歷庫介紹。

以下表格將介紹各主題與各子題的類別含意：

#### Stock(股市)

統一	追蹤統一企業於市場的營運狀況與投資機會。
鴻海	了解鴻海供應鏈及全球市場的科技影響力。
台積電	探索台積電於半導體領域的競爭優勢。
廣達	分析廣達電子製造業務的成長趨勢與合作夥伴。
中華電	追蹤中華電在電信業的發展和技術創新。
聯發科	探究聯發科在半導體市場的產品競爭力。
長榮	觀察長榮於運輸和物流業的業績走勢。
富邦金	了解富邦金於金融保險市場的佈局策略。
長榮航	分析長榮航空的航線拓展及旅遊趨勢。
元大金	研究元大金於資產管理及金融市場的角色。

#### Health(健康)

確診	追蹤疫情確診數據以評估健康風險。
減重	探索減重方法及其對健康的影響。
免疫力	了解增強免疫力的方法和相關保健品。
營養	分析營養對健康的益處與飲食建議。
疾病	探討常見疾病的預防和治療方法。
病毒	研究病毒傳播途徑與防護措施。



慢性病	了解慢性病的成因及健康管理策略。
健康飲食	介紹健康飲食的要素與對身體的益處。
癌症	探索癌症的最新治療和預防進展。
醫療	追蹤醫療科技發展及對健康的影響。

#### Sport(運動)

賽事	追蹤各類運動賽事的賽程及結果。
球員	研究球員的表現和職業生涯發展。
聯賽	分析各大聯賽的競爭趨勢及影響力。
世界盃	探討世界盃比賽及球隊的參賽表現。
籃球	追蹤籃球比賽及其全球影響力。
奧運	研究奧運賽事及選手表現的亮點。
NBA	探索 NBA 的明星球員與聯賽動態。
MLB	了解 MLB 的賽季變化及選手動向。
棒球	研究棒球運動的發展和國際賽事。
中職	追蹤中華職棒的賽事及球員動態。

另外關於處理不相關的資料，我們主要透過三項技術來做文字過濾，第一項是於爬蟲中，我們使用關鍵字，於該網也搜尋相關新聞，第二項是利用 Gemini，使用以下的 prompt(會依照不同的主題做改變)，進行文字過濾，第三則是如果於 Gemini 的 API 不可用時，會改成利用 keyword 的方式檢查是否為該類別的文章，這樣可以有效的過濾非該類別或是廣告的資料。以下是 prompt:

```
prompt = f"""
以下是今天的新聞資訊。請根據這些資訊判斷其是否是股市新聞，如果不是請回答無關：

標題：
{title}

文章內容：
{news}
"""
```

```
prompt = f"""
以下是今天的新聞資訊。請根據這些資訊判斷其是否是健康類別的新聞，如果不是請回答無關：

標題：
{title}

文章內容：
{news}
"""
```

```
prompt = f"""
以下是今天的新聞資訊。請根據這些資訊判斷其是否是運動類別的新聞，如果不是請回答無關：

標題：
{title}

文章內容：
{news}
"""
```

## 4. Sentiment Analysis

情緒分析部分，我們採用 BERT 模型 "nlptown/bert-base-multilingual-uncased-sentiment" 來針對新聞文本進行自動評分。這個模型以多語言的 BERT 為基礎，專門訓練來處理多語言的情緒分析任務，能夠為每一篇新聞內容給出 1 到 5 顆星的評分。透過星級評分，我們可以進一步將情緒分類成三種主要類型：

1. **負面**：如果模型給出的星級評分為 1 星或 2 星，表示新聞情緒較為負面，反映出該新聞可能包含負面的詞彙或情緒，例如批評、爭議等。
2. **中立**：評分為 3 星的新聞被認為是中立的，意味著內容可能是客觀描述，缺乏強烈的情緒色彩，或包含正負情緒較平衡的內容。
3. **正面**：當新聞的評分為 4 星或 5 星時，代表該新聞具備正面的情緒特徵，可能帶有支持性、樂觀的語氣或偏向積極的觀點。

為了統一數據的結構與便於後續分析，我們將這些情緒結果整理並儲存在 sentiment table 中。儲存的内容包含：

1. **情緒類型**：以 "positive"、"neutral" 和 "negative" 來區分每篇新聞的情緒性質。

2. **星級評分**：將原始星級評分 1-5 直接記錄在資料庫中，以便細分分析使用。

(後續第 7 節將介紹分析的結果)

## 5. Dynamic and Adaptive Crawling

針對 Dynamic crawling 與 Adaptive crawling 我們主要的操作如下：

### A. Dynamic Crawling 動態爬蟲

Dynamic Crawling 是一種靈活應對網頁變動的爬蟲策略，適合處理頻繁變動的網站。具體技巧如下：

- Selenium: 用於自動化瀏覽器行為，特別適合處理需要互動（如點擊、滾動、等待動態加載）的網頁。透過模擬瀏覽器行為，我們可以從網頁中獲取經 JavaScript 渲染的內容。
- BeautifulSoup: 配合 Selenium 抓取靜態內容，解析網頁 HTML 結構。BeautifulSoup 提供簡單直觀的接口，用來定位網頁中的特定元素，利於資料提取。
- API: 有些網站提供 API 以供程式化訪問資料。直接訪問 API 相比網頁爬取更加穩定和高效，可以避免不必要的渲染和頁面加載過程。
- User Agent: 設置合適的 User Agent 以模擬真實的瀏覽器訪問。不同的網站對於 User Agent 的要求可能不同，隨時切換適當的 User Agent 可以減少封鎖風險。
- Headers: 根據網站的要求自訂 Headers，這包括設定 Referer、Accept、Accept-Language 等，讓請求看起來像是從瀏覽器發出的。這樣可以增強訪問的真實性，避免被封鎖。
- 模擬人類行為: 觀察各網站的封鎖機制，根據網站的反爬行措施動態調整，例如隨機停頓（TIMER）來模擬正常瀏覽速度，並防止 Webdriver 被檢測，或是不要連續點過多網頁，而適當關閉與開啟網頁，避免不正常的過九訪問。
- TIMER: 動態設置隨機停頓時間，讓爬蟲行為更接近人類操作，減少網站封鎖的機率。
- 生成式 AI: 使用生成式 AI 輔助過濾不必要的新聞，且也可以加入生成式 AI 於給出網頁爬蟲的改良建議，想適架構改變時，適時將網頁資訊與目前的爬蟲，提供給生成式 AI 詢問修改方向。

### B. Adaptive Crawling 自適應爬蟲

Adaptive Crawling 則更著重於錯誤處理和動態適應不同網頁結構的改變。具體技巧如下：

- CSS Selector: 利用 CSS 選擇器定位網頁元素。當網站結構發生輕微變動時，可以透過調整選擇器來適應新結構，減少爬蟲中斷的可能性。
- 正規表達式處理結構錯誤: 針對結構錯誤或網頁資料的變動，使用正規表達式提取關鍵資料，避免依賴特定 HTML 結構，提高代碼的適應性。
- 類似元素辨識: 當結構變化時，透過觀察相似元素的標籤或屬性模式來找到相對應的資料。例如，如果某個元素的 class 名稱改變，可以根據其他不變的屬性辨識出目標資料。
- User Agent: 在 Adaptive Crawling 中持續更改 User Agent，應對網站的反爬機制。自適應爬蟲會根據不同的網站需求調整 User Agent，並在檢測到異常時更改其內容。
- 生成式 AI 輔助撰寫代碼: 生成式 AI 可協助撰寫調整過後的爬蟲代碼。例如，用生成式 AI 輔助過濾不必要的新聞，且也可以加入生成式 AI 於給出網頁爬蟲的改良建議，想適架構改變時，適時將網頁資訊與目前的爬蟲，提供給生成式 AI 詢問修改方向。
- 錯誤通知與自動重試: 當爬蟲遇到異常時會自動通知，並根據預設的重試策略重新執行爬取作業。這樣可以有效避免因暫時性錯誤而導致的數據缺失。

Adaptive Crawling 的核心在於適應結構變化和處理異常，透過結合正規表達式、AI 輔助代碼生成和自動錯誤處理機制，使爬蟲在面對動態變化的網頁時能夠保持穩定高效。

## 6. Data Storage

我們使用了兩個資料庫，分別命名為“IR\_GenAI\_Team\_Proj1”和“IR\_GenAI\_Team\_Proj1\_second\_Database”，每個資料庫包含 12 個資料表儲存 3 個 topic 相關的新聞及情緒分析：

**資料表 stock:**

stockID	stock_name
股票代號	股票名稱

sto...	stock_name
1216	統一
2317	鴻海
2330	台積電
2382	廣達
2412	中華電
2454	聯發科
2603	長榮
2618	長榮航
2881	富邦金
2885	元大金

資料表 stock\_news:

stockID	title	date	content	source	id	url
股票代號	標題	日期	內容	來源	index	連結

	sto...	title	date	content	source	url	id
	2330	台股電機 需求受惠	2024-10-22	隨全球AI晶片需求持續高漲，台股電（233 China Times		https://www.chinatimes.com/newspaper/	236
	2330	台股電 聯發科、中砂漲勢	2024-10-22	台股電於法說會釋出樂觀預期，多檔 China Times		https://www.chinatimes.com/newspaper/	237
	2330	拉抬電力解險 中小股接棒漲	2024-10-22	權王台股電經過上周五單日暴衝後，21日 China Times		https://www.chinatimes.com/newspaper/	238
	2330	嘉善科技建廠與需求 需求題材疊加	2024-10-21	隨著全球AI晶片需求持續高漲，台股電（2 China Times		https://www.chinatimes.com/realtimeev	239
	2330	台股電 聯發科、中砂漲勢	2024-10-21	10月ETF除息季，本周進入最後6個開獎， China Times		https://www.chinatimes.com/realtimeev	240
	2330	台股電 聯發科、中砂漲勢	2024-10-21	台股電17日舉行法說會，第三季稅後純益 China Times		https://house.chinatimes.com/202410210	241
	2454	《盤後解析》23600得而復失 專家：攻高！	2024-10-21	台股電(2330)休兵，台股未能站上23600點 China Times		https://www.chinatimes.com/realtimeev	242
	2454	《盤中解析》聯發科領軍衝鋒 IC設計族群	2024-10-21	聯發科(2454)領軍衝鋒，IC設計族群遍地 China Times		https://www.chinatimes.com/realtimeev	243
	2454	《盤中解析》股王接棒領攻 本週關	2024-10-21	美股震盪，台股盤中指數上漲逾100點！ China Times		https://www.chinatimes.com/realtimeev	244
	2454	《科技》聯發科天璣9400加持 智慧與JP C	2024-10-18	聯發科（2454）於本月初宣布推出最新的： China Times		https://www.chinatimes.com/realtimeev	245
	2454	《產業》聯發科、聯發科！IC設計族群	2024-10-18	「IC設計族群遍地開花」，於9月由經濟部及 China Times		https://www.chinatimes.com/realtimeev	246
	2454	《盤中解析》台股電創1100元新高 台股	2024-10-18	台股電(2330)ADR狂漲近10%，台股開盤 China Times		https://www.chinatimes.com/realtimeev	247
	2454	《盤後解析》台股電 聯發科、中砂漲勢	2024-10-17	台股今日開盤平穩，台股電(2330) China Times		https://www.chinatimes.com/realtimeev	248
	2454	聯發科 AI 小模型助邊緣運算	2024-10-15	隨AI模型供應發展，晶片大廠高通與聯發 China Times		https://www.chinatimes.com/newspaper/	249
	2454	《盤中解析》電子大廠聯發 台股站穩二萬	2024-10-15	美股昨夜由科技股帶頭上攻，四大指數全 China Times		https://www.chinatimes.com/realtimeev	250
	2454	1分鐘聯發科 台股電 台股電1800億	2024-10-15	小編今天（15）日精選5件不可不知的圈內 China Times		https://www.chinatimes.com/realtimeev	251
	2454	14檔電子股外資接棒 漲潮	2024-10-14	聯發科(2330)晶片將量產，超微近日也發布 China Times		https://www.chinatimes.com/newspaper/	252
	2454	CSP加速設計地位 台股、聯發科、世芯	2024-10-14	AI基礎設施布局加速，有別聯發大力推進 China Times		https://www.chinatimes.com/newspaper/	253
	2454	據盤中法：GB200出貨在即 供應趨緊	2024-10-14	市場分析：分析機構看好 Blackwell將為N China Times		https://www.chinatimes.com/newspaper/	254

(“IR\_GenAI\_Team\_Proj1” 以 id 排序 第 1 頁)

	sto...	title	date	content	source	url	id
	1216	https://cc.tvbs.com.tw/img/upload/2017/	2014-06-09	烏克蘭總統波羅申科，7日在國會宣誓就職	tvbs	https://news.tvbs.com.tw/world/534681	43133
	1216	https://cc.tvbs.com.tw/img/upload/2017/	2014-06-06	柯文哲今天公佈兩支國際競選廣告，其中一	tvbs	https://news.tvbs.com.tw/politics/534280	43134
	1216	就是不相聞！劉曉真質疑總統競選	2014-06-04	花貓雜誌會昨天進行最後競選，不過雜誌	tvbs	https://news.tvbs.com.tw/local/534011	43135
	1216	https://cc.tvbs.com.tw/img/upload/2017/	2014-06-03	康師傅和統一兩大食品巨頭，在大陸又掀起5	tvbs	https://news.tvbs.com.tw/world/533920	43136
	1216	中華國棒 / (快訊) 屏東下雨球場積水	2014-06-03	原訂今天晚上在屏東棒球隊舉行的中華國棒	tvbs	https://news.tvbs.com.tw/sports/533899	43137
	1216	https://cc.tvbs.com.tw/img/upload/2017/	2014-06-03	大陸池體市場向來是食品巨頭兵家必爭之地	tvbs	https://news.tvbs.com.tw/politics/533874	43138
	1216	《獨家》「翻白眼」爭議神隱 新人MV于	2014-05-30	統一獅啦啦隊Uni-girls成員于小文，日前訪	tvbs	https://news.tvbs.com.tw/entertainment/	43139
	1216	「會考變基測」 分區成績公布似PR值	2014-05-27	國中會考成績怎麼公布，今天終於拍板，是	tvbs	https://news.tvbs.com.tw/life/533119	43140
	1216	會考成績公布「全國一致」 分區「系統	2014-05-27	教育部終於拍板怎麼公布會考成績公布方式，	tvbs	https://news.tvbs.com.tw/life/533082	43141
	1216	https://cc.tvbs.com.tw/img/upload/2017/	2014-05-26	陷入國家分裂危機的烏克蘭，在上一週日舉	tvbs	https://news.tvbs.com.tw/world/533006	43142
	1216	求償郵務賠不起？ 恐人民提單3800萬	2014-05-26	北上面試卻在捷運車廂裡被偷走的二十六歲	tvbs	https://news.tvbs.com.tw/local/532988	43143
	1216	拒絕！就是拒絕 千萬獎金席身邊	2014-05-26	3-4月份的統一發票開獎了，你兌換了嗎？	tvbs	https://news.tvbs.com.tw/local/532916	43144
	1216	https://cc.tvbs.com.tw/img/upload/2017/	2014-05-25	統一獅啦啦隊Uni-girls成員于小文，日前在	tvbs	https://news.tvbs.com.tw/entertainment/	43145
	1216	https://cc.tvbs.com.tw/img/upload/2017/	2014-05-21	球賽開球嘉賓很特別，是來自韓國的景浩	tvbs	https://news.tvbs.com.tw/life/532445	43146
	1216	秒翻白眼惹惱！于小文PO照單傳道歉	2014-05-21	統一獅啦啦隊Uni-girls中的紅隊隊員于小文	tvbs	https://news.tvbs.com.tw/sports/532356	43147
	1216	「翻白眼」影響 Uni-girls粉絲合照變少	2014-05-20	統一獅啦啦隊Uni-girls中的隊員于小文，因	tvbs	https://news.tvbs.com.tw/life/532291	43148
	2330	台股電 美能出口華為 美：無法保證	2020-05-16	台股電宣布要在亞利桑那設廠，外界解讀	tvbs	https://news.tvbs.com.tw/world/1324968	43149
	2330	因冰淇與台結緣！亞利桑那州風氣超台	2020-05-16	晶廣代工廠台股電選擇在亞利桑那州設廠	tvbs	https://news.tvbs.com.tw/world/1324942	43150
	2330	台股電 美能出口華為 美：無法保證	2020-05-15	台股電創辦人張忠謀2019年底訪美時「當	tvbs	https://news.tvbs.com.tw/life/1324723	43151

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 最後一頁)

資料表 stock\_news\_API:

stockID	title	date	content	source	id	url
股票代號	標題	日期	內容	來源	index	連結

	stockID	title	date	content	source	url	id
	2330	台積電是否會想收購英特爾晶圓廠？	2024-10-17	台積電 CEO 魏哲家表示對收購英特爾晶圓廠	Techbang.com	https://www.techbang.com/posts/118975	551
	2330	台積電為了提升CoWoS封裝產能，不	2024-10-04	由於CoWoS需求旺盛，台積電還持續在全	Techbang.com	https://www.techbang.com/posts/118540	552
	2330	AMD 成台積電亞利桑那工廠第二大客戶，	2024-10-18	消息稱 AMD 成台積電亞利桑那工廠新客戶	Techbang.com	https://www.techbang.com/posts/118773	553
	2330	三星晶圓代工廠在製程方面與台積電展開	2024-10-19	三星晶圓代工廠在「工藝優勢」方面與台積電	Techbang.com	https://www.techbang.com/posts/118964	554
	2330	衛星廣播台積電全球布局，傳聞：在台擴	2024-10-10	從上帶視角看經濟！傳聞媒體近日透過通	Inside.com.tw	https://www.inside.com.tw/article/36452	555
	2330	魏哲家選 825 張台積電限制股票！市值新	2024-10-01	魏哲家等台積電高層主管持股應託限制	Inside.com.tw	https://www.inside.com.tw/article/36376	556
	2330	台積電法說：Q3 營率超標！ 單季 EPS 12.4	2024-10-17	台積電第三季 7 奈米以下營收比重達到 69	Inside.com.tw	https://www.inside.com.tw/article/36511	557
	2330	AI 晶片需求強！路透：台積電 Q3 獲利預	2024-10-14	根據 ISEG SmartEstimate 典型 22 位分析	Inside.com.tw	https://www.inside.com.tw/article/36475	558
	2330	台積電、三星導導在阿拉伯聯合大公國	2024-09-23	有關台積電與三星電子的討論一方面反映	Inside.com.tw	https://www.inside.com.tw/article/36279	559
	2330	27 歲、公司估值 140 億美元，Scale AI 要	2024-10-16	Scale AI 是一家數據標記公司，擁有 10 萬	Inside.com.tw	https://www.inside.com.tw/article/36491	560
	2330	台積電採用 NVIDIA cuLitho 平台 加速晶片	2024-10-09	台積電宣布採用 NVIDIA cuLitho 運算式微	Cool3c.com	https://www.cool3c.com/article/226430	561
	2330	台積電NRE耗電反應在高通Snapdragon 8	2024-10-01	2024年10月將是Android陣營2024年末至2	Cool3c.com	https://www.cool3c.com/article/225863	562
	2330	台積電、三星洽談在阿拉伯聯合大公國建	2024-09-23	華爾街日報報導，台積電、三星正與阿拉	Cool3c.com	https://www.cool3c.com/article/225132	563
	2330	傳聞「從空中看經濟」衛星廣播台積電全	2024-10-11	台積電全球布局曝光！衛星圖顯示，台積電	Thenewslens.com	https://www.thenewslens.com/article/24	564
	2330	台積電股價站上1085元創新高，「機國神	2024-10-19	台積電股價18日一度衝上新台幣1100元，	Thenewslens.com	https://www.thenewslens.com/article/24	565
	2330	「不敢教主」橫上財金教授：既然台積電	2024-10-04	台積電報酬比0050好，為什麼不買台積電	Thenewslens.com	https://www.thenewslens.com/article/24	566
	2330	台積電法說會：營運強盛超預期，CoWoS	2024-10-18	台積電昨天以亮麗的第3季業績與樂觀的聲	Thenewslens.com	https://www.thenewslens.com/article/24	567
	2330	AI股上漲，台積電重返千元，台股上揚32	2024-09-25	AI股、台股電今（25）強勢收盤，台股上	Thenewslens.com	https://www.thenewslens.com/article/24	568
	2330	解密台灣第二「砂礫」：台積電最先進工	2024-10-13	全球高階晶片製造龍頭台積電（TSMC）	Thenewslens.com	https://www.thenewslens.com/article/24	569

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 第 1 頁)

	stockID	title	date	content	source	url	id
	2881	廣邦金前9月大賺1212億 續登金控獲利三冠	2024-10-11	金控獲利出爐廣邦金控自結獲利，9月獲利	Yahoo Entertainment	https://tw.stock.yahoo.com/news/%E5% 713	713
	2881	盤人！國泰金、廣邦金獲利雙雙躍升「千	2024-10-11	盤人！國泰金（2882）、廣邦金（2881）	Yahoo Entertainment	https://tw.stock.yahoo.com/news/%E5% 714	714
	2881	鉅亨法說、Factset 最新調查：廣邦金(2881	2024-09-22	FactSet 最新調查：廣邦金(2881-TW)EPS	Yahoo Entertainment	https://tw.stock.yahoo.com/news/%E5% 715	715
	2881	台股四寶市值續到一家廣邦金 謝金河續	2024-09-26	台股站穩2萬2000點，但是台股四寶股價	Yahoo Entertainment	https://tw.news.yahoo.com/%E5%8F%B0 716	716
	2881	7家金控前3季獲利去年全年 廣邦金穩坐第	2024-10-11	（中央社紀實欄元報、謝方輝、張璟台北）	Yahoo Entertainment	https://tw.news.yahoo.com/7%E5%AE%B1 717	717
	2881	14家金控前3季獲利衝破5000億元！7家賺	2024-10-11	14家上市金控2024年9月自結獲利全數	Yahoo Entertainment	https://tw.stock.yahoo.com/news/14%E5% 718	718
	2881	廣邦金9月獲利79億創新高 奪獲利三冠王	2024-10-11	廣邦金控今（11）日公布獲利自結數，9月	Yahoo Entertainment	https://tw.news.yahoo.com/%E5%AF%B0 719	719
	2881	廣邦金前3季大賺1212億元創上次高 奪	2024-10-11	廣邦金控（2881）今（11）日下午公布9月	Yahoo Entertainment	https://tw.stock.yahoo.com/news/%E5% 720	720
	2881	台股四寶續到「一家廣邦金」？謝金河	2024-09-26	台股連開日線紅盤，今（27）日四大權值	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon 721	721
	2382	廣達9月營收月增6.8%	2024-10-08	廣達（2382）公布9月合併營收1,551.10	Technews.tw	https://finance.technews.tw/2024/10/08/ 722	722
	2382	AI 伺服器測試需求夯！廣達再追加 25.9 億	2024-10-04	代工大廠廣達近日代美國加州子公司 QMA	Technews.tw	https://technews.tw/2024/10/04/quantu- 723	723
	2412	中華電：低軌衛星 OneWeb 月底將 24 小時	2024-10-14	低軌衛星業者 OneWeb 訊號預計 10 月底可	Inside.com.tw	https://www.inside.com.tw/article/36477 724	724
	2412	2024年10月電信優惠方案整理：中華電信	2024-10-01	10月秋季到來，台灣3大電信業者中華電信	Cool3c.com	https://www.cool3c.com/article/225879 725	725
	2412	中華電信與AWS深化合作 通用生成式AI打	2024-10-16	中華電信將於台灣時間11/30在美國拉斯維	Thenewslens.com	https://www.thenewslens.com/feature/s- 726	726
	2412	中華電9月營收月增6.2%	2024-10-11	中華電（2412）公布9月合併營收194.86	Technews.tw	https://finance.technews.tw/2024/10/11/c- 727	727
	2412	中華電：OneWeb 月底完成 24 小時涵蓋全	2024-10-14	中華電信為強化網路韌性，積極布局低軌	Technews.tw	https://technews.tw/2024/10/14/by-the-e- 728	728
	2412	中華電攜手全球 IOWN 夥伴推動次世代寬	2024-10-10	中華電信於10月8日至11日參加在台北舉	Technews.tw	https://finance.technews.tw/2024/10/10/k- 729	729
	2412	中華電信導入美立信最新 5G 與 AI 節能技	2024-10-17	中華電信近日與美立信展開新一輪合作，	Technews.tw	https://technews.tw/2024/10/17/icht-erics 730	730
	1216	蘋果可能意外洩露M4 MacBook Pro規格	2024-10-21	過去幾週對蘋果公司來說並不好過，在該	Techbang.com	https://www.techbang.com/posts/119022 731	731

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 最後一頁)

資料表 stock\_news\_sentiment:

id	news_id	sentiment	emotion	star
index	新聞編號	情緒分數	情緒類型	星等

<input type="checkbox"/>	<b>id</b> int8 ▾	<b>news_id</b> int8 ▾	<b>sentime...</b> float8 ▾	<b>emotion</b> int8 ▾	<b>star</b> int8 ▾
<input type="checkbox"/>	7485	248	0.53250777721405	-1	1
<input type="checkbox"/>	7486	250	0.714459538459778	-1	1
<input type="checkbox"/>	7487	253	0.317470341920853	0	3
<input type="checkbox"/>	7488	254	0.275578439235687	-1	2
<input type="checkbox"/>	7489	276	0.279548972845078	1	5
<input type="checkbox"/>	7490	277	0.33317232131958	1	4
<input type="checkbox"/>	7491	278	0.26759260892868	-1	1
<input type="checkbox"/>	7492	279	0.303015768527985	1	5
<input type="checkbox"/>	7493	280	0.252116918563843	0	3
<input type="checkbox"/>	7494	281	0.246247783303261	0	3
<input type="checkbox"/>	7495	282	0.615129947662354	-1	1
<input type="checkbox"/>	7496	283	0.635010421276093	-1	1
<input type="checkbox"/>	7497	284	0.268846958875656	-1	1
<input type="checkbox"/>	7498	285	0.273285627365112	-1	1
<input type="checkbox"/>	7499	255	0.33317232131958	1	4
<input type="checkbox"/>	7500	256	0.279548972845078	1	5
<input type="checkbox"/>	7501	257	0.26759260892868	-1	1
<input type="checkbox"/>	7502	258	0.303015768527985	1	5
<input type="checkbox"/>	7503	286	0.38338366150856	-1	1
<input type="checkbox"/>	7504	287	0.308596044778824	1	5
<input type="checkbox"/>	7505	288	0.531269788742065	1	5
<input type="checkbox"/>	7506	289	0.282300621271133	1	4
<input type="checkbox"/>	7507	290	0.336022347211838	-1	2
<input type="checkbox"/>	7508	291	0.244144797325134	-1	1
<input type="checkbox"/>	7509	292	0.307010650634766	1	4
<input type="checkbox"/>	7510	302	0.278938353061676	1	4
<input type="checkbox"/>	7511	303	0.597020864486694	-1	1

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 第 1 頁)

<input type="checkbox"/>	<input type="text" value="id"/> id int8	<input type="text" value="news_id"/> news_id int8	<input type="text" value="sentime..."/> sentime... float8	<input type="text" value="emotion"/> emotion int8	<input type="text" value="star"/> star int8
<input type="checkbox"/>	48500	696	0.281964689493179	1	5
<input type="checkbox"/>	48501	697	0.547126233577728	1	5
<input type="checkbox"/>	48502	698	0.225262448191643	1	4
<input type="checkbox"/>	48503	699	0.239495068788528	-1	1
<input type="checkbox"/>	48504	700	0.353670656681061	1	4
<input type="checkbox"/>	48505	704	0.336090475320816	-1	2
<input type="checkbox"/>	48506	705	0.39715638756752	1	5
<input type="checkbox"/>	48507	706	0.588529169559479	-1	1
<input type="checkbox"/>	48508	707	0.309159696102142	-1	1
<input type="checkbox"/>	48509	711	0.508330702781677	1	5
<input type="checkbox"/>	48510	712	0.425418883562088	-1	1
<input type="checkbox"/>	48511	713	0.49403327703476	1	5
<input type="checkbox"/>	48512	714	0.45337387919426	-1	1
<input type="checkbox"/>	48513	715	0.316397190093994	1	5
<input type="checkbox"/>	48514	716	0.378156393766403	-1	2
<input type="checkbox"/>	48515	717	0.325916618108749	-1	1
<input type="checkbox"/>	48516	718	0.383041322231293	1	5
<input type="checkbox"/>	48517	720	0.30013719201088	1	5
<input type="checkbox"/>	48518	721	0.590233325958252	-1	1
<input type="checkbox"/>	48519	722	0.282662779092789	-1	1
<input type="checkbox"/>	48520	723	0.251511305570602	1	5
<input type="checkbox"/>	48521	724	0.279173970222473	-1	2
<input type="checkbox"/>	48522	725	0.288999050855637	1	5
<input type="checkbox"/>	48523	726	0.389515578746796	1	5
<input type="checkbox"/>	48524	727	0.261130183935165	0	3
<input type="checkbox"/>	48525	730	0.44001778960228	1	5
<input type="checkbox"/>	48526	731	0.368176996707916	-1	1

Page 411 of 411
100 rows
41,042 records

(“IR\_GenAI\_Team\_Proj1” 以 id 排序 最後一頁)

資料表 health:

healthID	health_name
keyword 代號	keyword 名稱



health_name	health_name
1	確診
2	減重
3	免疫力
4	營養
5	疾病
6	病毒
7	慢性病
8	健康飲食
9	癌症
10	醫療

資料表 health\_news:

healthID	title	date	content	source	id	url
keyword 代號	標題	日期	內容	來源	index	連結

	healthID	title	date	content	source	url	id
	1	男子遊美國、印度染疫公務 台南確診第2例	2024-10-22	出國不只預防登革熱還要避免染上登公務！台南市處7月傳	China Times	https://www.chinatimes.com/realtimenev	1
	1	10月「國際乳癌防治月」 醫界人齊聲疾呼 守護健康與財富	2024-10-22	你知道粉紅絲帶象徵什麼嗎？這是乳癌防治運動的標誌，	China Times	https://www.chinatimes.com/newspaper	2
	1	新冠現登革熱病例 免費很快	2024-10-21	新北市本土登革熱個案之前分布於中和、新店，新冠19日	China Times	https://www.chinatimes.com/newspaper	3
	1	新冠出現登革熱1例、全市計67例！市府持續加強社區巡邏	2024-10-20	新北市新增1例本土登革熱病例，為新冠區40歲男，14日出	China Times	https://www.chinatimes.com/realtimenev	4
	1	流感疫情沒停歇！中市8人重症 7人仍住院救治	2024-10-20	近日天氣轉涼，各類病毒開始活躍，其中流感疫情沒有停	China Times	https://www.chinatimes.com/realtimenev	5
	1	乳癌5年存活率達9成 3招防護發	2024-10-20	乳癌位界我國女性癌症發生率首位長達19年，且人數持續	China Times	https://www.chinatimes.com/newspaper	6
	1	乳癌術後10年仍有復發風險 專家提醒定期追蹤3關鍵	2024-10-19	乳癌連續19年為我國女性癌症發生率首位，據統計，2022	China Times	https://www.chinatimes.com/realtimenev	7
	1	新北本土登革熱再 +1 市府加強巡邏查察	2024-10-16	新北市新增1例本土登革熱病例，為中和區41歲男，10月11	China Times	https://www.chinatimes.com/realtimenev	8
	1	年過50血癌高+糖尿病 醫提醒：2項象徵癌症王	2024-10-15	廣義王有癌王之稱，家醫科醫師韓士就指出，胰臟癌連續	China Times	https://www.chinatimes.com/realtimenev	9
	1	朱芯儀罹乳癌「靠復健夫妻第1個知道」 獲兩人疼愛：心意無價	2024-10-14	女星朱芯儀在2021年底確診乳癌第3期，歷經了18次化療，	China Times	https://www.chinatimes.com/realtimenev	10
	1	28歲女每天必喝鹽酥雞 1年發出大腸癌 醫揭3危險因子	2024-10-14	郵局地產的禍，大腸直腸外科醫師陳誠分享在開一天收	China Times	https://www.chinatimes.com/realtimenev	11
	1	羅東博愛醫院 奪冠癌症之王	2024-10-14	宜蘭羅東博愛醫院多年推動癌症防治有成，今年在癌症防	China Times	https://www.chinatimes.com/newspaper	12
	1	羅東博愛醫院舉辦癌症防治講座 獲社區民眾踴躍參加	2024-10-13	宜蘭羅東博愛醫院連續多年推動癌症防治有成，今年在	China Times	https://www.chinatimes.com/realtimenev	13
	1	手穿鎗傷以為肝病 一查竟大腸癌 2種人注意了	2024-10-12	一名40歲女子因手穿鎗傷，以為是肝病引發黃疸而求診，	China Times	https://www.chinatimes.com/realtimenev	14
	1	「馬鹿病毒」致死率近9成！疾管署升重安播提醒醫政	2024-10-11	近日「馬鹿病毒」肆虐非洲重安播，9月27日出現首例馬鹿	China Times	https://www.chinatimes.com/realtimenev	15
	1	新北本土登革熱再添1例 6旬婦染疫與高風險區有關	2024-10-10	新北市新增1例本土登革熱病例，為新店區65歲女，6日出	China Times	https://www.chinatimes.com/realtimenev	16
	1	台灣日本腦炎再增1例！安南6旬婦獲診 住家鄰近高風險環境	2024-10-10	日本腦炎的流行高峰期已過，甚至即將進入尾聲，但台南	China Times	https://www.chinatimes.com/realtimenev	17
	2	抗發炎、助減重 專家推6食材：地瓜葉CP值最高	2024-10-21	減重不只為了外表，更是為了健康，營養師李婉萍表示，	China Times	https://www.chinatimes.com/realtimenev	18

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 第 1 頁)

(“IR\_GenAI\_Team\_Proj1” 以 id 排序 最後一頁)

healthID	title	date	content	source	id	url
keyword 代號	標題	日期	內容	來源	index	連結

(“IR\_GenAI\_Team\_Proj1” 以 id 排序 第 1 頁)

<input type="checkbox"/>	heal... i... ▾	title varchar ▾	date date ▾	content varchar ▾	sou... varc... ▾	url varchar ▾	id int8 ▾
<input type="checkbox"/>	10	聞慶血小板應用，切入 700 多家醫療院所，協助修復組織	2024-09-22	思必瑞特生技以獨家	Technews.tw	https://technews.tw/2024/09/22/spirit-scientific_prp-therapy_plt-techn	337
<input type="checkbox"/>	10	日本山形大學 ABENICS 革新機器人關節，有望應用在醫療	2024-09-25	日本山形大學在 2021	Technews.tw	https://technews.tw/2024/09/25/abenics/	338
<input type="checkbox"/>	10	盧耀茂帶粵海省代表團 探討兩地醫療衛生合作空間	2024-10-17	醫務衛生局局長盧耀	Yahoo Entertainment	https://hk.news.yahoo.com/%E7%9B%A7%E5%AF%B5%E8%8C%82%E6%	339
<input type="checkbox"/>	10	醫健邁跨院健康紀錄功能下周一起新增大埔區三層醫療機	2024-09-25	醫衛局公布，醫健邁	Yahoo Entertainment	https://consent.yahoo.com/v2/collectConsent?sessionId=Lcc-session_ca	340
<input type="checkbox"/>	10	東區醫院 76 歲病人攝鼻胃照 醫醫醫生聯 X 光片判斷低	2024-10-10	【Yahoo新聞報道】東	Yahoo Entertainment	https://hk.news.yahoo.com/%E6%9D%B1%E5%8D%80%E9%86%AB%E9%	341
<input type="checkbox"/>	10	施政報告2024   醫管局歡迎有關醫療服務措施	2024-10-16	行政長官李家超發表	Yahoo Entertainment	https://hk.news.yahoo.com/%E6%96%BD%E6%94%BF%E5%A0%B1%E5%	342
<input type="checkbox"/>	10	林琳豪兒子斜坡每次檢查上千元 民建聯發布倡議書促政府	2024-10-06	兒童免疫力較低，除	Yahoo Entertainment	https://consent.yahoo.com/v2/collectConsent?sessionId=Lcc-session_et	343
<input type="checkbox"/>	10	中方向黎巴嫩提供緊急人道主義醫療物資援助	2024-10-07	中方向黎巴嫩提供緊	Yahoo Entertainment	https://hk.news.yahoo.com/%E4%B8%AD%E6%96%B9%E5%90%91%E9%	344
<input type="checkbox"/>	10	倡中醫師臨床考核及再實習 醫庫促擴醫療紀錄存取權	2024-10-07	【on.cc東網專訊】中	Yahoo Entertainment	https://hk.news.yahoo.com/%E5%80%A1%E4%B8%AD%E9%86%AB%E9%	345
<input type="checkbox"/>	10	施政報告 2024   200 元以上烈酒減九成稅 陳軾池、劉澤	2024-10-15	【Yahoo 新聞報道】	Yahoo Entertainment	https://hk.news.yahoo.com/%E6%96%BD%E6%94%BF%E5%A0%B1%E5%	346
<input type="checkbox"/>	10	施政報告2024   盧耀茂：檢討收費是善用醫療資源確切舉	2024-10-17	醫務衛生局局長盧耀	Yahoo Entertainment	https://hk.news.yahoo.com/%E6%96%BD%E6%94%BF%E5%A0%B1%E5%	347
<input type="checkbox"/>	10	公院連串醫療事故 議員促系統性檢討及改進	2024-10-10	【on.cc東網專訊】新	Yahoo Entertainment	https://hk.news.yahoo.com/%E5%85%AC%E9%99%A2%E9%80%A3%E4%	348
<input type="checkbox"/>	10	施政報告2024   康樂養老長者獲資助部分醫療費用	2024-10-21	政府明早推行「資助	Yahoo Entertainment	https://consent.yahoo.com/v2/collectConsent?sessionId=Lcc-session_ac	349
<input type="checkbox"/>	10	盧耀茂訪京與國家衛健委討論深化內地與特區跨境醫療創	2024-09-26	醫務衛生局局長盧耀	Yahoo Entertainment	https://hk.news.yahoo.com/%E7%9B%A7%E5%AF%B5%E8%8C%82%E8%	350
<input type="checkbox"/>	10	黃大仙醫院康復醫療護理部病房出現產後病毒肺炎症導	2024-10-02	東華三院黃大仙醫院	Yahoo Entertainment	https://consent.yahoo.com/v2/collectConsent?sessionId=Lcc-session_f2	351
<input type="checkbox"/>	10	政府為康樂院舍長者分擔部分醫療費 門診需多一年資助	2024-10-21	【Now新聞台】政府	Yahoo Entertainment	https://hk.news.yahoo.com/%E6%94%BF%E5%9C%9C%E7%82%BA%E5%	352
<input type="checkbox"/>	10	中國援助黎巴嫩緊急人道醫療物資運抵貝魯特機場	2024-10-21	中國援助黎巴嫩的緊	Yahoo Entertainment	https://hk.news.yahoo.com/%E4%B8%AD%E5%9C%98%E6%8F%B4%E5%	353
<input type="checkbox"/>	10	動植物公園11歲虎 楊潤雄：蘇隻頂龍獸子表現異常正接受	2024-10-20	香港動植物公園昨日	Yahoo Entertainment	https://hk.news.yahoo.com/%E5%8B%95%E6%A4%8D%E7%89%A9%E5%	354

←

Page 2 of 2

→

100 rows

118 records

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 最後一頁)

資料表 health\_news\_sentiment:

id	news_id	sentiment	emotion	star
index	新聞編號	情緒分數	情緒類型	星等

<input type="checkbox"/>	id int8 ▾	news_id int8 ▾	sentime... float8 ▾	emotion int8 ▾	star int8 ▾
<input type="checkbox"/>	1	1	0.471652418375015	-1	1
<input type="checkbox"/>	2	2	0.322035998106003	-1	1
<input type="checkbox"/>	3	3	0.423261284828186	-1	1
<input type="checkbox"/>	4	4	0.47499218583107	-1	1
<input type="checkbox"/>	5	5	0.29407611489296	-1	2
<input type="checkbox"/>	6	6	0.357157230377197	-1	1
<input type="checkbox"/>	7	7	0.338858067989349	-1	1
<input type="checkbox"/>	8	8	0.470773935317993	-1	1
<input type="checkbox"/>	9	9	0.309414684772491	1	4
<input type="checkbox"/>	10	10	0.371787697076797	1	4
<input type="checkbox"/>	11	11	0.540427923202515	-1	1
<input type="checkbox"/>	12	12	0.259319603443146	-1	1
<input type="checkbox"/>	13	13	0.240907609462738	-1	1
<input type="checkbox"/>	14	15	0.300990998744965	-1	1
<input type="checkbox"/>	15	14	0.303979247808456	-1	2
<input type="checkbox"/>	16	18	0.354494482278824	1	4
<input type="checkbox"/>	17	19	0.315445780754089	-1	2
<input type="checkbox"/>	18	20	0.319537907838821	-1	2
<input type="checkbox"/>	19	21	0.412672281265259	1	4
<input type="checkbox"/>	20	22	0.469074904918671	1	4
<input type="checkbox"/>	21	23	0.387570798397064	1	4
<input type="checkbox"/>	22	24	0.401530772447586	-1	1
<input type="checkbox"/>	23	25	0.406210094690323	-1	1
<input type="checkbox"/>	24	26	0.327337712049484	0	3
<input type="checkbox"/>	25	28	0.52362471818924	-1	1
<input type="checkbox"/>	26	29	0.608848452568054	-1	1
<input type="checkbox"/>	27	35	0.327268779277802	1	4

←

Page 1 of 292

→

100 rows

29,136 records

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 第 1 頁)

<input type="checkbox"/>	<span>id</span> int8	<span>news_id</span> int8	<span>sentime...</span> float8	<span>emotion</span> int8	<span>star</span> int8
<input type="checkbox"/>	29110	254	0.27349665760994	1	4
<input type="checkbox"/>	29111	256	0.37512218952179	1	4
<input type="checkbox"/>	29112	258	0.364371031522751	-1	1
<input type="checkbox"/>	29113	259	0.254133135080338	0	3
<input type="checkbox"/>	29114	260	0.324281275272369	-1	2
<input type="checkbox"/>	29115	240	0.335648238658905	-1	1
<input type="checkbox"/>	29116	241	0.33077335357666	1	5
<input type="checkbox"/>	29117	242	0.330199897289276	-1	1
<input type="checkbox"/>	29118	243	0.705542623996735	-1	1
<input type="checkbox"/>	29119	244	0.355169385671616	-1	1
<input type="checkbox"/>	29120	245	0.52687931060791	-1	1
<input type="checkbox"/>	29121	246	0.587608277797699	-1	1
<input type="checkbox"/>	29122	248	0.547461628913879	-1	1
<input type="checkbox"/>	29123	249	0.493502736091614	-1	1
<input type="checkbox"/>	29124	250	0.23548424243927	-1	2
<input type="checkbox"/>	29125	251	0.536998867988586	-1	1
<input type="checkbox"/>	29126	253	0.234903633594513	0	3
<input type="checkbox"/>	29127	255	0.509456157684326	-1	1
<input type="checkbox"/>	29128	261	0.387661963701248	1	4
<input type="checkbox"/>	29129	262	0.305821299552917	-1	2
<input type="checkbox"/>	29130	263	0.369618594646454	0	3
<input type="checkbox"/>	29131	264	0.327499121427536	-1	1
<input type="checkbox"/>	29132	265	0.271695286035538	0	3
<input type="checkbox"/>	29133	266	0.365827590227127	-1	1
<input type="checkbox"/>	29134	267	0.440816760063171	1	4
<input type="checkbox"/>	29135	268	0.334944188594818	0	3
<input type="checkbox"/>	29136	269	0.284579992294312	-1	1

← Page 292 of 292 → 100 rows 29,136 records

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 最後一頁)

資料表 sport:

sportID	sport_name
keyword 代號	keyword 名稱

<div>🔍 spo... i... ▼</div>	sport_name varchar
1	賽事
2	球員
3	聯賽
4	世界盃
5	籃球
6	奧運
7	NBA
8	MLB
9	棒球
10	中職

資料表 sport\_news:

sportID	title	date	content	source	id	url
keyword 代號	標題	日期	內容	來源	index	連結

<div>🔍 spo... i... ▼</div>	<div>🔍 title varchar ▼</div>	<div>🔍 date date ▼</div>	<div>🔍 content varchar ▼</div>	<div>🔍 source varchar ▼</div>	<div>url varchar ▼</div>	<div>id int8 ▼</div>
<input type="checkbox"/>	1	2025台新女子路跑 即日起報名	2024-10-22	台新銀行連續8年冠名贊助「台新女子路跑	China Times	https://www.chinatimes.com/newspaper/ 1
<input type="checkbox"/>	1	台新女子路跑報名 首創數位組	2024-10-22	台新銀行連續八年冠名贊助「台新女子路跑	China Times	https://www.chinatimes.com/newspaper/ 2
<input type="checkbox"/>	1	霹靂舞「世界舞維」一紙賽場不知意仍榮冠	2024-10-21	台灣霹靂舞好手「Quake」張振國榮譽挑戰	China Times	https://www.chinatimes.com/realtimenev 3
<input type="checkbox"/>	1	波奇波奇支持台灣運動賽事 為選手助力挑	2024-10-21	台灣首家波奇磚創始品牌「PokePoke-波奇	China Times	https://www.chinatimes.com/realtimenev 4
<input type="checkbox"/>	1	台新女子路跑開放報名 呼籲全台女力站出	2024-10-21	TAISHIN WOMEN RUN 台新女子路跑今日	China Times	https://www.chinatimes.com/realtimenev 5
<input type="checkbox"/>	1	2024全民運動大禮券先登場 622名選手國內	2024-10-20	2024全民運動在屏東，總共有33個競賽項目	China Times	https://www.chinatimes.com/realtimenev 6
<input type="checkbox"/>	1	首屆中國信託杯 促兩岸棒球交流	2024-10-20	為促進兩岸基層棒球發展、深化體育公益	China Times	https://www.chinatimes.com/newspaper/ 7
<input type="checkbox"/>	1	西拉雅龜飛車進入第11年 籃球賽角對造型	2024-10-19	交通部觀光署西拉雅國家風景區管理處	China Times	https://trueiii.chinatimes.com/content/2 8
<input type="checkbox"/>	1	全國競技體操錦標賽首度進軍台東 體育署	2024-10-19	113年全國競技體操錦標賽19日在台東縣立	China Times	https://www.chinatimes.com/realtimenev 9
<input type="checkbox"/>	1	中國信託杯青少年棒球聯賽 大陸6地開打	2024-10-19	為促進兩岸基層棒球發展、深化體育公益	China Times	https://www.chinatimes.com/realtimenev 10
<input type="checkbox"/>	1	花蓮太平洋國際馬拉松開幕 212隊報名參賽	2024-10-18	由教育廳體育署、花蓮縣政府指導，花蓮	China Times	https://www.chinatimes.com/realtimenev 11
<input type="checkbox"/>	1	LOL世界賽「中韓超級對決」8強賽高潮戰	2024-10-18	LOL〈英雄聯盟〉2024世界賽8強賽事開打	China Times	https://www.chinatimes.com/realtimenev 12
<input type="checkbox"/>	1	海碩盃「相隔5年台北重燃戰火」巨型LED成	2024-10-18	度過新冠病毒疫情整整5年空窗期，2024台北	China Times	https://www.chinatimes.com/realtimenev 13
<input type="checkbox"/>	2	NBA「果然沒給頂薪 阿隆戈登打斷接受金	2024-10-22	正如外界傳言，金塊果然沒有拿出頂薪續	China Times	https://www.chinatimes.com/realtimenev 14
<input type="checkbox"/>	2	《時來運轉》NBA新球季東、西區精彩看	2024-10-22	NBA全新球季(2024-2025)10/23(三)分別由	China Times	https://www.chinatimes.com/realtimenev 15
<input type="checkbox"/>	2	MLB「魔法用盡了！大都會夢斷洛杉磯 阿	2024-10-21	紐約的大都會的驚奇之旅結束，在大谷翔平	China Times	https://www.chinatimes.com/realtimenev 16
<input type="checkbox"/>	2	NBA「3年3900萬美元！勇士續約穆迪 未	2024-10-21	根據（ESPN）權威記者蓋拉尼亞德先爆料	China Times	https://www.chinatimes.com/realtimenev 17
<input type="checkbox"/>	2	MLB「洋基隊友高城留下委託：給他7億美	2024-10-21	洋基前役與著索托（Juan Soto）約款達3	China Times	https://www.chinatimes.com/realtimenev 18
<input type="checkbox"/>	2	橫濱海灣壘球 運動部的能與不能	2024-10-20	中職台灣大賽精彩上演，創紀錄有超過4萬	China Times	https://www.chinatimes.com/opinion/20 19

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 第 1 頁)

	spo...	title	date	content	source	url	id
	5	NBA / 明對戰小飛俠 NBA等藝境林書豪	2012-02-10	台灣籃球好手林書豪，這幾天成為NBA的熱	tvbs	https://news.tvbs.com.tw/sports/8488	33243
	5	〈獨家〉臺灣燈會風光勝 民批評了又蓋了	2012-02-07	2012臺灣燈會，元宵歡樂風光閃耀，吸引	tvbs	https://news.tvbs.com.tw/life/8012	33244
	5	敗選路更寬？ 張蘇達、綠荷榜入署	2012-02-07	陳&#20914;新內閣又有人新列入，前	tvbs	https://news.tvbs.com.tw/politics/8031	33245
	5	https://cc.tvbs.com.tw/img/upload/2017/	2012-02-03	馬戲團的小丑可以拿5個小球輕鬆的在手上	tvbs	https://news.tvbs.com.tw/world/7797	33246
	5	挑戰「灌籃王」 8名校冠軍比劍奪	2012-02-02	打籃球最吸引大家目光的就是「灌籃」，	tvbs	https://news.tvbs.com.tw/life/7645	33247
	5	偷手機上被客人臉書 亂發謾罵訊息	2012-01-30	每天去台大校園籃球等運動的民眾，可要	tvbs	https://news.tvbs.com.tw/local/7344	33248
	5	https://cc.tvbs.com.tw/img/upload/2017/	2012-01-11	2011年奧比士名人榜中，香港天王劉德華	tvbs	https://news.tvbs.com.tw/local/5593	33249
	5	https://cc.tvbs.com.tw/img/upload/2017/	2012-01-05	美國一名男子，帶女友去看籃球賽，正當	tvbs	https://news.tvbs.com.tw/world/4996	33250
	5	球隊比賽失分 教練激動當場罵	2012-01-03	台南新榮中學籃球隊教練田本玉，向來以	tvbs	https://news.tvbs.com.tw/sports/4710	33251
	5	https://cc.tvbs.com.tw/img/upload/2017/	2011-12-27	美國職籃NBA今年被看好奪冠的奧克蘭	tvbs	https://news.tvbs.com.tw/sports/77220	33252
	5	https://cc.tvbs.com.tw/img/upload/2017/	2011-12-21	金正日突然過世，去年才公開發行的金正	tvbs	https://news.tvbs.com.tw/world/76648	33253
	5	https://cc.tvbs.com.tw/img/upload/2017/	2011-12-20	金正恩早年在瑞士留學，同學形容他是個	tvbs	https://news.tvbs.com.tw/local/76537	33254
	5	富承旭食療骨折 醫生建議開刀治療	2011-12-20	藝人富承旭之前因為打籃球，造成右手食	tvbs	https://news.tvbs.com.tw/entertainment/	33255
	5	樓上小孩打籃球太吵 隔半夜被敲門	2011-12-19	同樣大廈的住戶，竟相對開公堂，新北	tvbs	https://news.tvbs.com.tw/local/76358	33256
	5	打掃神醫打球 高中生涯女老師點頭	2011-12-18	不打掃就點頭，新北市一名高中女導師，	tvbs	https://news.tvbs.com.tw/life/76294	33257
	1	納達爾3盤解對手 輕鬆晉級澳網第2輪	2018-01-16	網球男單頭號種子納達爾（Rafael Nadal）	tvbs	https://news.tvbs.com.tw/sports/853156	33258
	1	平昌冬奥大和解！ 南韓邀北韓共舉半島	2018-01-10	而韓在板門店達成共識，北韓確定要參加	tvbs	https://news.tvbs.com.tw/world/850625	33259
	1	詹寶羅為首 27選手下週出征大馬羽球賽	2018-01-10	總獎金35萬美元的馬來西亞羽毛球大師賽	tvbs	https://news.tvbs.com.tw/sports/850596	33260
	1	對戰冠軍高手狂輸學生 爭議理事長搶世	2018-01-06	近年飛揚運動盛行，國內相關協會努力爭	tvbs	https://news.tvbs.com.tw/sports/847961	33261

（“IR\_GenAI\_Team\_Proj1” 以 id 排序 最後一頁）

資料表 sport\_news\_API:

sportID	title	date	content	source	id	url
keyword 代號	標題	日期	內容	來源	index	連結

	spo...	title	date	content	source	url	id
	1	Sony 在台推出OLED電競螢幕INZONE M1	2024-10-08	Sony宣布新一代旗艦級電競螢幕INZONE！	Cool3c.com	https://www.cool3c.com/article/226322	173
	1	Sony INZONE 電競產品線推出OLED電競	2024-09-25	Sony電競產品組INZONE公布兩款電競	Cool3c.com	https://www.cool3c.com/article/225405	174
	1	瑞士18歲單車手比賽身亡，UCI賽事安全受	2024-09-28	18歲瑞士單車手Muriel Furrer在瑞士舉行的	Thenewslen.com	https://www.thenewslen.com/article/24/	175
	1	大巨蛋演唱會、運動賽事來了！信用卡商	2024-10-07	台北大巨蛋開放舉行棒球賽後，奧杰倫和	Technews.tw	https://finance.technews.tw/2024/10/07/	176
	1	總會稱亞洲龍舟錦標賽帶動九龍東經濟	2024-10-10	第15屆亞洲龍舟錦標賽今日起一連三日在	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	177
	1	因應「三項唐人賽事」早前中環灣仔封路	2024-10-13	運輸署表示，因應「三項唐人賽事」早前	Yahoo Entertainment	https://hk.news.yahoo.com/%E5%98%A0	178
	1	英荷互贈節節慶 舉辦多項賽事賀國慶	2024-09-29	【Now新聞台】國慶前，英荷互贈節節	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	179
	1	機場三跑道工程賽事報名額額兩倍 抽籤	2024-09-30	香港國際機場第三跑道擴闊快1月底啟用，	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	180
	1	打擊非法串流轉播收費足球賽事 IT界	2024-09-23	【en.cc東網專訊】香港海關近日接獲版權	Yahoo Entertainment	https://hk.news.yahoo.com/%E6%89%93	181
	1	羅之英決定不就被拒參加學界足球賽事	2024-09-23	仁道醫院羅之英紀念中學提交學界足球賽	Yahoo Entertainment	https://hk.news.yahoo.com/%E8%91%A3	182
	1	海關引用「傳播權利」拘兩男 疑涉及	2024-09-24	海關打擊以非法串流技術向公眾傳播足球	Yahoo Entertainment	https://hk.news.yahoo.com/%E6%85%B7	183
	2	再度裁員！TikTok 進一步轉向 AI 審核內	2024-10-13	2 名知情人士告訴路透社，大馬有 700 多	Inside.com.tw	https://www.inside.com.tw/article/36466	184
	2	FIFA賽程排太滿，球員傷病頻傳，工會	2024-10-15	上週末歐洲足球再次迎來國際比賽週，然	Thenewslen.com	https://www.thenewslen.com/article/24/	185
	2	NFL傳奇球員阮達特：從難民與蝦農之子，	2024-09-28	阮達特（Dat Nguyen）是第一位打進美國	Thenewslen.com	https://www.thenewslen.com/article/24/	186
	2	不晉傳國的一擊！歐錦球員「在前輩引	2024-09-28	若把「讓空軍」概念套入便可知，前輩球	Thenewslen.com	https://www.thenewslen.com/article/24/	187
	2	國際七人制明年3月啟德體育園舉行 球	2024-09-30	「香港國際七人制球賽」落實明年3月底在	Yahoo Entertainment	https://hk.news.yahoo.com/%E5%9C%8E	188
	3	NBA、NFL、MLB等美國各種職業運動	2024-10-01	美國並沒有聯邦法律有規定運動賽事「必	Thenewslen.com	https://www.thenewslen.com/article/24/	189
	3	FIFA賽程排太滿，球員傷病頻傳，工會	2024-10-15	上週末歐洲足球再次迎來國際比賽週，然	Thenewslen.com	https://www.thenewslen.com/article/24/	190
	4	英格蘭國家隊新任新領隊Thomas Tuchel，	2024-10-16	英格蘭國家足球隊新領隊確認，將由德國	Thenewslen.com	https://www.thenewslen.com/article/24/	191

（“IR\_GenAI\_Team\_Proj1” 以 id 排序 第 1 頁）

<input type="checkbox"/>	<div><div></div><div>spo...</div><div>...</div></div>	<div><div></div><div>title</div><div>varchar</div></div>	<div><div></div><div>date</div><div>date</div></div>	<div><div></div><div>content</div><div>varchar</div></div>	<div><div></div><div>source</div><div>varchar</div></div>	<div><div></div><div>url</div><div>varchar</div></div>	<div><div></div><div>id</div><div>int8</div></div>
<input type="checkbox"/>	5	優惠送贈送門票兌換券 新北市國道大客車	2024-10-20	為推廣籃球體育活動，新北市立圖書館與4	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	326
<input type="checkbox"/>	5	U19籃球聯盟新球季熱鬧開打 傳統球隊與6	2024-09-27	記者陳建興 / 台北報導 由福康文教基金會7	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	327
<input type="checkbox"/>	5	U19籃球聯盟Eagles超級教練坐鎮指導 Thu	2024-10-06	記者陳建興 / 台北報導 由福康文教基金會8	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	328
<input type="checkbox"/>	5	議長至3對3籃球賽 廣發武林帖	2024-10-07	記者林建興/台南報導 一一三年議長至三落	Yahoo Entertainment	https://tw.news.yahoo.com/%E8%AD%B0	329
<input type="checkbox"/>	5	「施顯宗、陳英緯」涉賄！ 籃協重懲「E	2024-10-11	在效力SBL彰化柏力力時期，傳出涉賄遭捕	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	330
<input type="checkbox"/>	5	讚好書，挺籃球 新北市圖書館送贈送門票賞	2024-10-21	(記者謝政傑新北報導)圖書館送你贈送門票賞	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	331
<input type="checkbox"/>	5	籃球 / 登峰造極激戰點燃戰火 南山燒8	2024-09-25	2024與廣發登錄連場青年籃球邀請賽今日9	Yahoo Entertainment	https://tw.sports.yahoo.com/news/%E7%	332
<input type="checkbox"/>	5	蘇尚博奪籃球「細哥寶心」引一萬人燈孔9	2024-10-09	蘇尚博奪籃球「細哥寶心」引一萬人燈孔9	Yahoo Entertainment	https://tw.news.yahoo.com/%E8%94%A7	333
<input type="checkbox"/>	5	打籃球忘讓「凱羅勝」找上門！警告警署1	2024-10-02	據記得面、高中時期，每到下課就會看到-	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	334
<input type="checkbox"/>	5	HBL註冊人數破萬 體育署盼基層籃球共同1	2024-09-27	(中央社記者陳春傑台北27日電) 113學年1	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	335
<input type="checkbox"/>	5	涉賄永不掉用！籃協嚴重懲 施顯宗、陳5	2024-10-11	中華民國籃球協會今(11日)發出公告，指	Yahoo Entertainment	https://tw.sports.yahoo.com/news/%E6%	336
<input type="checkbox"/>	5	3×3台北狂賽籃球聯賽第8 創下歷史最佳1	2024-09-25	[NOWnews今日新聞]國際籃球(FIBA) 3×	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	337
<input type="checkbox"/>	5	跳脫韓式球風改走歐洲風範籃球 勇士許E	2024-09-27	體育中心 / 黃真超報導台灣職籃P. LEAGUE	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	338
<input type="checkbox"/>	5	籃球) 爭權判決落敗 戰神只在意協助	2024-10-12	首場TPBL首輪熱身賽遭到台灣永登賽豹射1	Yahoo Entertainment	https://consent.yahoo.com/v2/collectCon	339
<input type="checkbox"/>	5	U19籃球聯盟 / 俱樂部VS球隊提出新火花	2024-10-21	[NOWnews今日新聞]由福康文教基金會主	Yahoo Entertainment	https://tw.sports.yahoo.com/news/u19%E	340
<input type="checkbox"/>	5	【籃球專電 / 李亦伸】長攝連投籃董事長衛1	2024-10-19	台灣職籃TPBL、PLG兩聯盟統合不成，兩1	Yahoo Entertainment	https://tw.sports.yahoo.com/news/%E3%	341
<input type="checkbox"/>	5	PLG開幕戰26日和平籃球線登場 廣邦勇E	2024-10-14	體育中心 / 黃真超報導台灣職籃P. LEAGUE	Yahoo Entertainment	https://tw.sports.yahoo.com/news/plg%E	342
<input type="checkbox"/>	10	中職總冠軍賽10/19起獨家文峰：賽程、睇1	2024-10-15	今年中職總冠軍賽「台灣大賽」由中信兄弟1	Thenewslens.com	https://www.thenewslens.com/article/24:	343
<input type="checkbox"/>	10	中職季後挑戰賽：統一獅2勝0敗晉級季後1	2024-10-13	中職季後挑戰賽由統一獅率先取得2勝0敗	Thenewslens.com	https://www.thenewslens.com/article/24:	344

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 最後一頁)

資料表 sport\_news\_sentiment:

id	news_id	sentiment	emotion	star
index	新聞編號	情緒分數	情緒類型	星等

<input type="checkbox"/>	<input checked="" type="checkbox"/> id int8	<input type="checkbox"/> news_id int8	<input type="checkbox"/> sentime... float8	<input type="checkbox"/> emotion int8	<input type="checkbox"/> star int8
<input type="checkbox"/>	6138	1	0.373622298240662	1	5
<input type="checkbox"/>	6139	2	0.358152985572815	1	4
<input type="checkbox"/>	6140	3	0.354607582092285	1	4
<input type="checkbox"/>	6141	4	0.399910032749176	1	4
<input type="checkbox"/>	6142	5	0.444306015968323	1	5
<input type="checkbox"/>	6143	6	0.440266907215118	1	5
<input type="checkbox"/>	6144	7	0.382041305303574	1	4
<input type="checkbox"/>	6145	8	0.401376128196716	1	4
<input type="checkbox"/>	6146	9	0.335866928100586	1	4
<input type="checkbox"/>	6147	10	0.372006595134735	1	4
<input type="checkbox"/>	6148	11	0.464778572320938	1	5
<input type="checkbox"/>	6149	12	0.384326547384262	1	4
<input type="checkbox"/>	6150	13	0.311268389225006	1	5
<input type="checkbox"/>	6151	14	0.492012947797775	-1	1
<input type="checkbox"/>	6152	15	0.376480549573898	0	3
<input type="checkbox"/>	6153	16	0.473212897777557	1	4
<input type="checkbox"/>	6154	17	0.295672386884689	0	3
<input type="checkbox"/>	6155	18	0.268503755331039	0	3
<input type="checkbox"/>	6156	19	0.405090361833572	1	4
<input type="checkbox"/>	6157	20	0.317654937505722	-1	1
<input type="checkbox"/>	6158	21	0.315959542989731	1	4
<input type="checkbox"/>	6159	22	0.310174763202667	1	4
<input type="checkbox"/>	6160	23	0.315281599760056	0	3
<input type="checkbox"/>	6161	24	0.396711558103561	1	4
<input type="checkbox"/>	6162	25	0.396110713481903	1	4
<input type="checkbox"/>	6163	26	0.308847904205322	0	3
<input type="checkbox"/>	6164	27	0.382041305303574	1	4

Page 1 of 332
 100 rows
 33,195 records

( “IR\_GenAI\_Team\_Proj1” 以 id 排序 第 1 頁)



<input type="checkbox"/>	<input checked="" type="checkbox"/> id int8 ▾	news_id int8 ▾	sentime... float8 ▾	emotion int8 ▾	star int8 ▾
<input type="checkbox"/>	39306	317	0.372452944517136	-1	1
<input type="checkbox"/>	39307	318	0.783632874488831	-1	1
<input type="checkbox"/>	39308	319	0.445605546236038	-1	1
<input type="checkbox"/> <input checked="" type="checkbox"/>	39309	320	0.725204467773438	-1	1
<input type="checkbox"/> <input checked="" type="checkbox"/>	39310	321	0.807688236236572	-1	1
<input type="checkbox"/>	39311	322	0.345719546079636	-1	1
<input type="checkbox"/>	39312	323	0.505402505397797	-1	1
<input type="checkbox"/>	39313	325	0.39782452583313	1	4
<input type="checkbox"/>	39314	326	0.28055140376091	-1	1
<input type="checkbox"/>	39315	327	0.268504470586777	1	4
<input type="checkbox"/>	39316	328	0.280975252389908	1	4
<input type="checkbox"/>	39317	329	0.271582156419754	1	4
<input type="checkbox"/>	39318	330	0.519515454769135	-1	1
<input type="checkbox"/>	39319	331	0.423464059829712	-1	1
<input type="checkbox"/>	39320	332	0.391009002923965	1	5
<input type="checkbox"/>	39321	333	0.474461257457733	-1	1
<input type="checkbox"/>	39322	334	0.282698750495911	0	3
<input type="checkbox"/>	39323	335	0.302771985530853	0	3
<input type="checkbox"/>	39324	336	0.689548075199127	-1	1
<input type="checkbox"/>	39325	337	0.249305218458176	1	4
<input type="checkbox"/>	39326	338	0.341130912303925	1	4
<input type="checkbox"/>	39327	340	0.346127569675446	1	4
<input type="checkbox"/>	39328	342	0.383467644453049	1	4
<input type="checkbox"/>	39329	339	0.393517076969147	-1	1
<input type="checkbox"/>	39330	341	0.381512761116028	-1	1
<input type="checkbox"/>	39331	343	0.431507736444473	1	5
<input type="checkbox"/>	39332	344	0.315193623304367	1	4

(“IR\_GenAI\_Team\_Proj1” 以 id 排序 最後一頁)

展示 databse 瀏覽、搜尋功能(以 health\_news 為示範):

(1) 資料以 date 排序

SELECT \*

FROM health\_news

ORDER BY date ASC;

	#	he..	...	o	Title vachar		d..	o	Content vachar		source vachar		uri vachar	id	in8
	3				六宅宅神護身功 不僅靈驗電流				台中有一處60歲的民眾，十多年前幫了萬功德，好難得！ tvb				https://news.tvbs.com.tw/file/412668	15488	
	7				標案與SBS談 衛局署：目前並無SARS				健明醫務機構SARS院內數部都，不過這並不代表不影響！ tvb				https://news.tvbs.com.tw/file/412412	16306	
	87				87號宅神護身功 情況危急恐染病				11名醫護醫務SARS通報病患，隨後護理和山醫院院方！ tvb				https://news.tvbs.com.tw/file/412382	16305	
	7				傳聞11名SARS病患 已悉數痊癒				健明醫務通報的11名SARS病患，又說2名病患為可被傳染！ tvb				https://news.tvbs.com.tw/file/412330	16304	
	3				上海體自 假裝看電視抗肺炎				上海市的SARS反應雖然比較輕，不過上海市政府在防疫！ tvb				https://news.tvbs.com.tw/file/412042	15457	
	3				引用陳炳成醫師的話說 到底誰上癮				陳次雄指出月份到國美為藥劑品及癌症治療的！ 監製大！ tvb				https://news.tvbs.com.tw/file/411388	15466	
	3				溪洲連日新風 以自強免疫力防護				溪洲科學最近成功研發了一種可以促進身體免疫力對抗疫！ tvb				https://news.tvbs.com.tw/file/410430	15455	
	2				陸用醫學研究報告 半年成功改善				陸用醫學研究報告從今年1月1日起以來已經過去半年的研究！ tvb				https://news.tvbs.com.tw/file/410106	21344	
	2				減低藥物副作用 專治神經痛與骨質增生				巴西小兒專家：完全不知如何抵抗和外為了平衡障礙！ tvb				https://news.tvbs.com.tw/file/409904	21346	
	2				運動到重傷 與孩童小孩吃中餐				拼個吃大餐，真真真正正運動不到對兒童增加兒童量！ tvb				https://news.tvbs.com.tw/file/408647	21342	
	2				影到腳膝跳發 花樣年華年滿30公斤				有一名26歲的男小，連續兩年性肥胖，體重增加到128公斤！ tvb				https://news.tvbs.com.tw/file/408130	21332	
	2				影到中醫師康力 強調體能訓練				彰化傳導第30多名中醫師，每天都會幫幫在一場比賽！ tvb				https://news.tvbs.com.tw/file/407673	15454	
	2				T-RUSH新生派運動 為全民健康				配合韓二家支持「新生派運動」，活力十足的T-RUSH特別！ tvb				https://news.tvbs.com.tw/file/407568	21331	
	2				周樹松當身 海軍軍重30公斤				想讓南部地區的女性現在有一樣健康的康復，唐蓮娜！ tvb				https://news.tvbs.com.tw/file/406871	21330	
	2				計畫計養生 成功減重拾得美人歸				估臺市有一位重量168公斤的胖姐曾典文，在女模特的型！ tvb				https://news.tvbs.com.tw/file/406081	21329	
	2				(短跑+長跑、民轉轉) 瘦到下部 跟壯健保持保衛				方念慈：「這種花是花，面對困難的時候是難，瘦弱的弱！ tvb				https://news.tvbs.com.tw/topic/34A04567	15463	
	2				練和瘦月重400公斤的中興才？				Q：她跟練也們瘦後10月時瘦出的新聞，容易瘦小或瘦！ tvb				https://news.tvbs.com.tw/file/404334	21328	
	2				還原黑龍虎 傳遞地下沒殺光				你知道中國歷史上的最後一條黑龍虎嗎，也曾經遺失黑龍！ tvb				https://news.tvbs.com.tw/file/403996	15428	
	2				阿爾斯利達300公斤 豪氣一飛				26歲的軒行市民張惠香，在同阿爾斯利的協助下嘗試去！ tvb				https://news.tvbs.com.tw/file/403556	21327	
	3				減肥成功 減肥人球數量減半				中社最熱門的行動雜誌就於雜誌報導，不過對於想減肥的人！ tvb				https://news.tvbs.com.tw/file/402029	16302	
	3				小耳耳症穴按摩 痛癢免痛苦				耳耳症穴按摩了可以減輕，還可以預防耳朵病，保健！ tvb				https://news.tvbs.com.tw/file/400886	15426	
	2				經絡針灸重46公斤 移入臺灣傳授				進一步認為之針，呼吸器通供生命的品產者，上午在才！ tvb				https://news.tvbs.com.tw/file/400357	21321	
	2				花30萬減肥失敗 代餐減肥還下來				想要快速減肥的民眾要注意了！如果一個星期能減掉2公斤！ tvb				https://news.tvbs.com.tw/file/399500	21325	
	2				Lagerfeld 04零果 與眾不同最性感				時尚界大師拉格斐德，在巴黎國際春季春夏季的時裝，展現！ tvb				https://news.tvbs.com.tw/file/398439	21324	
	2				拍攝影展 200萬藝術價值夜雨				台灣每年生最大的城市三億五千萬年來，作為月射，肉！ tvb				https://news.tvbs.com.tw/file/398289	21325	
	2				89公斤小胖打排球 兩月減重3公斤				亞洲區排球錦標賽，跟隊裏內的傳統勁，也有許多小胖仔！ tvb				https://news.tvbs.com.tw/file/394579	21323	
	2				紅包會合社團聚餐 強化社區防疫部署				這樣天天無慮，自己就是身邊的朋友是否已經有人都了！ tvb				https://news.tvbs.com.tw/file/393934	15460	

(2) 找出資料表 `health_news` 和 `health_news API` 中的 `source`

SELECT DISTINCT source

FROM health news

UNION

SELECT DISTINCT source

FROM "health news API";

source
"ltn"
"Yahoo Entertainment"
"Thenewslens.com"
"Gamme.com.tw"
"Jandan.net"
"Cnbeta.com.tw"
"Douban.com"
"Secretchina.com"
"Technews.tw"
"Ithome.com.tw"
"Techbang.com"
"China Times"
"Inside.com.tw"
"Cnblogs.com"
"tvbs"

(3) 資料以 id 做升序排序

```
SELECT *  
FROM health_news  
ORDER BY id ASC;
```

	id	title	date	content	source	url	id
	1	男子從泰國、印度與孟加拉國回台後確診2例	2024-10-22	出國不只預防登革熱還要避免上公共場所！台北市府7月發	China Times	https://www.chinatimes.com/realtimenews/	1
	1	10月「國際乳癌防治月」 醫界呼籲民眾定期篩檢	2024-10-22	你知道紅絲帶象徵什麼嗎？這是乳癌防治運動的標誌，	China Times	https://www.chinatimes.com/news/	2
	1	新莊登革熱病例 全屬移居	2024-10-21	新北市本土登革熱個案之前分布於中和、新莊，新莊19日	China Times	https://www.chinatimes.com/news/	3
	1	新莊出現登革熱病例，全市計49例！市府持續加強社區巡邏	2024-10-20	新莊市新增1例本土登革熱病例，為新莊區40歲男，14日出	China Times	https://www.chinatimes.com/realtimenews/	4
	1	京都府府法務廳：中市8人重症 7人仍在住院	2024-10-20	近日死傷消息，各報新聞紛紛報導，其中京都府沒有傳	China Times	https://www.chinatimes.com/realtimenews/	5
	1	乳癌6年存活率達8成 3級防復發	2024-10-20	乳癌位居我國女性癌症發生率首位長達19年，且人數持續	China Times	https://www.chinatimes.com/news/	6
	1	乳癌術後10年仍有復發風險 專家提醒應定期追蹤	2024-10-19	乳癌復發率仍為我國女性癌症發生率首位，據統計，2022	China Times	https://www.chinatimes.com/realtimenews/	7
	1	新北本土登革熱再 +1 市府加強社區巡邏	2024-10-16	新北市新增1例本土登革熱病例，為中和區41歲男，10月11	China Times	https://www.chinatimes.com/realtimenews/	8
	1	年過50血癌案 醫師警告：2級惡性腫瘤	2024-10-16	陳姓王有德王之稱，高醫科醫師陳士軒指出，陳姓患	China Times	https://www.chinatimes.com/realtimenews/	9
	1	朱芯德獲乳癌 醫家提醒要定期追蹤 應用人體乳、心電	2024-10-14	女星朱芯德在2021年發現乳癌3期，歷經了8次化療，	China Times	https://www.chinatimes.com/realtimenews/	10
	1	28歲女病友必讀 1年變出大腫瘤 醫提3危險因子	2024-10-14	都是危險的癌，大腸直腸癌和腸胃癌雖然分享在一天	China Times	https://www.chinatimes.com/realtimenews/	11
	1	腸胃癌醫學會 專科醫師提醒	2024-10-14	腸胃癌醫學會提醒多年後復發癌症有處，今年在癌症	China Times	https://www.chinatimes.com/news/	12
	1	腸胃癌醫學會提醒 癌症預防從日常 應注意飲食安全	2024-10-13	腸胃癌醫學會提醒多年後復發癌症有處，今年在癌症	China Times	https://www.chinatimes.com/realtimenews/	13
	1	手舉重以海肝病 一直擴大腫瘤 2級惡性	2024-10-12	一名40歲女子因手舉重，以為是肝病引起腫瘤而求診，	China Times	https://www.chinatimes.com/realtimenews/	14
	1	「腸胃癌」致死率近5成！ 癌醫警告：應定期追蹤	2024-10-11	近日「腸胃癌」話題持續關注，9月27日出現首例	China Times	https://www.chinatimes.com/realtimenews/	15
	1	新北本土登革熱再增1例 6個病例與登革熱區無關	2024-10-10	新北市新增1例本土登革熱病例，為新莊區66歲女，6日出	China Times	https://www.chinatimes.com/realtimenews/	16
	1	台南日本腦炎再增1例！ 安南6旬婦確診 住泰租屋近菜園	2024-10-10	日本腦炎的流行期已過，甚至到連人馬，但台南	China Times	https://www.chinatimes.com/realtimenews/	17
	2	抗酸、助消化 專家提醒：吃胃藥CP值	2024-10-21	胃酸不只為了外食，更為了健康，胃酸過多時常	China Times	https://www.chinatimes.com/realtimenews/	18
	2	吃開胃藥肌肉 醫提醒應定期追蹤 醫師提醒	2024-10-20	有些減肥的人，完全不吃開胃藥導致水化食物，不過，	China Times	https://www.chinatimes.com/realtimenews/	19
	2	「腸胃白質」 腸胃癌發達 醫警告：應定期追蹤	2024-10-20	想要減重，蛋白質必須跟蹤，不過，減重醫師提醒	China Times	https://www.chinatimes.com/realtimenews/	20
	2	肺炎、紅斑環狀不問「腸胃癌」 醫提醒應定期	2024-10-19	肺炎、紅斑環狀是腸胃癌的藥品，則從腸胃	China Times	https://www.chinatimes.com/realtimenews/	21
	2	肺炎、紅斑環狀不問「腸胃癌」 醫提醒應定期	2024-10-19	肺炎、紅斑環狀是腸胃癌的藥品，則從腸胃	China Times	https://www.chinatimes.com/realtimenews/	22
	2	肺炎、紅斑環狀不問「腸胃癌」 醫提醒應定期	2024-10-19	肺炎、紅斑環狀是腸胃癌的藥品，則從腸胃	China Times	https://www.chinatimes.com/realtimenews/	23
	2	肺炎、紅斑環狀不問「腸胃癌」 醫提醒應定期	2024-10-18	肺炎、紅斑環狀是腸胃癌的藥品，則從腸胃	China Times	https://www.chinatimes.com/realtimenews/	24
	2	肺炎、紅斑環狀不問「腸胃癌」 醫提醒應定期	2024-10-18	肺炎、紅斑環狀是腸胃癌的藥品，則從腸胃	China Times	https://www.chinatimes.com/realtimenews/	25
	2	肺炎、紅斑環狀不問「腸胃癌」 醫提醒應定期	2024-10-17	肺炎、紅斑環狀是腸胃癌的藥品，則從腸胃	China Times	https://www.chinatimes.com/realtimenews/	26
	2	肺炎、紅斑環狀不問「腸胃癌」 醫提醒應定期	2024-10-17	肺炎、紅斑環狀是腸胃癌的藥品，則從腸胃	China Times	https://www.chinatimes.com/realtimenews/	27

除了展現資料庫的索引功能以外，我們也時做了介面，讓使用者可以直接選擇主題、新聞來源、日期、情緒分析(negative、positive、neural、all)來決定要顯示的新聞資訊，當然也可以都不要選擇，那麼就會顯示全部，只是資料量很多，因此會需要跑很久的時間，因此建議，僅使用特定範圍較佳，如下圖所示：

新聞公司

China Times

主題

stock

日期

2024/10/09

情緒

All

開始

Title	Date	Content	Source	Emotion	URL
台中勤美 洲際酒店 10 / 23起 迎賓 打造 全新住宿 體驗	2024- 10-09	【旅遊經 洪書瑱報導】根據台灣議題研究中心資訊，台中在六都觀光好感度拿下冠軍，其次為高雄、台北、新北、台南、桃園，而台中的觀光蓬勃，也吸引星級飯店及飯店品牌進駐，目前台中的星級飯店，包括：李方艾美、林酒店、麗寶福容、台中裕元花園、台中日月千禧、台中金典、長榮桂冠、台中福華、永豐棧、台中大毅老爺等，現在民眾到台中遊玩，又有新飯店可選擇了，新開幕酒店為IHG洲際酒店集團直營管理的旗下	China Times	1	Link

維護職 安、母性 勞工獲肯 定 長榮航 空領雙項 殊榮	2024- 10-09	長榮航空參加桃園市政府舉辦的113年桃園市推行職安衛與健康促進績優單位及人員頒獎典禮，榮獲「推行職業安全衛生優良家族」與「母性健康守護聯盟模範事業單位」雙項肯定，今日由孫嘉明總經理親自出席接受桃園市政府蘇俊賓副市長頒獎。	China Times	1	Link
運價修 正... 長榮9 月營收失 守500億	2024- 10-09	受到國際海運市場運價修正的影響，長榮9月合併營收446.37億元，失守500億元，月減16.97%，但年增82.41%；第三季合併營收1,527.92億元，是2022年第四	China Times	0	Link

## 7. Visualization

在視覺化的部分，我們主要針對情緒分析與網路爬蟲皆有做出使用者介面，以下是網路爬蟲的使用者介面，可以透過選取欲爬的新聞網(中國時報、TVBS、LTN、API)與選擇要爬取的主題(Stock、Helth、Sport)與選取爬取的頁數(此頁數會從最新新聞往後爬取)，最後動態更新到資料庫！

新聞公司

中國時報

主題

股市

頁數

輸入頁數

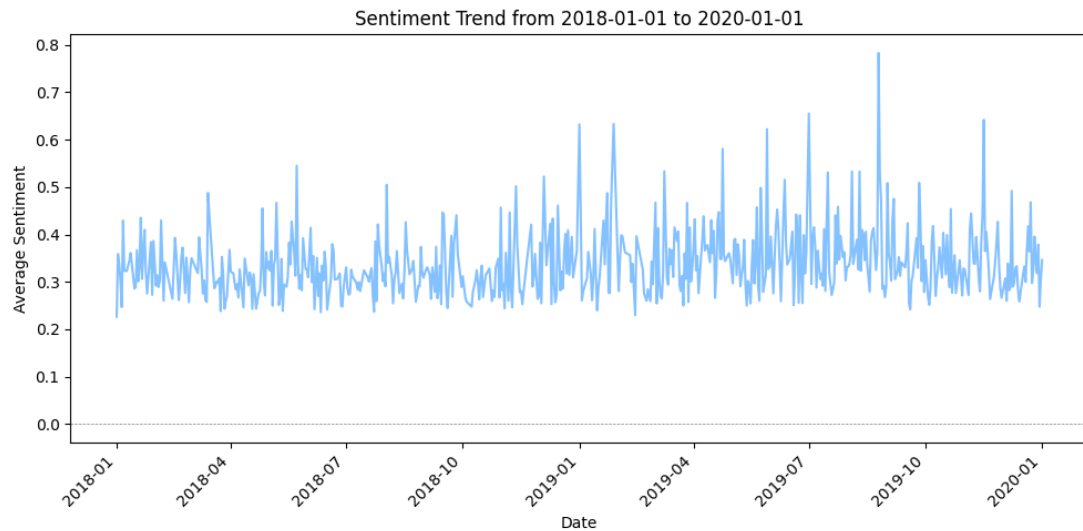
開始

在情緒分析視覺化的部分，我們在前端使用三個圖表呈現。使用者在網頁上選擇主題 (Health、Sport、Stock)、新聞來源及時間區間後，按下確認即可即時生成以下三張圖表：圖一為情緒分布圖，顯示正面、中立和負面情緒數量

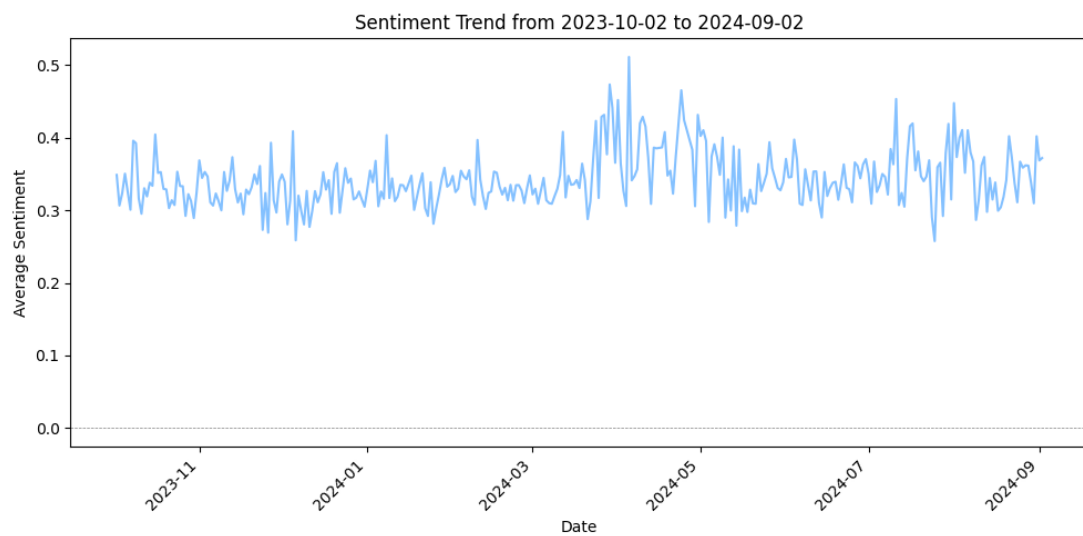
(positive、neutral、negative)；圖二為 star 數；圖三為時間序列圖，展示情緒隨時間的變化。



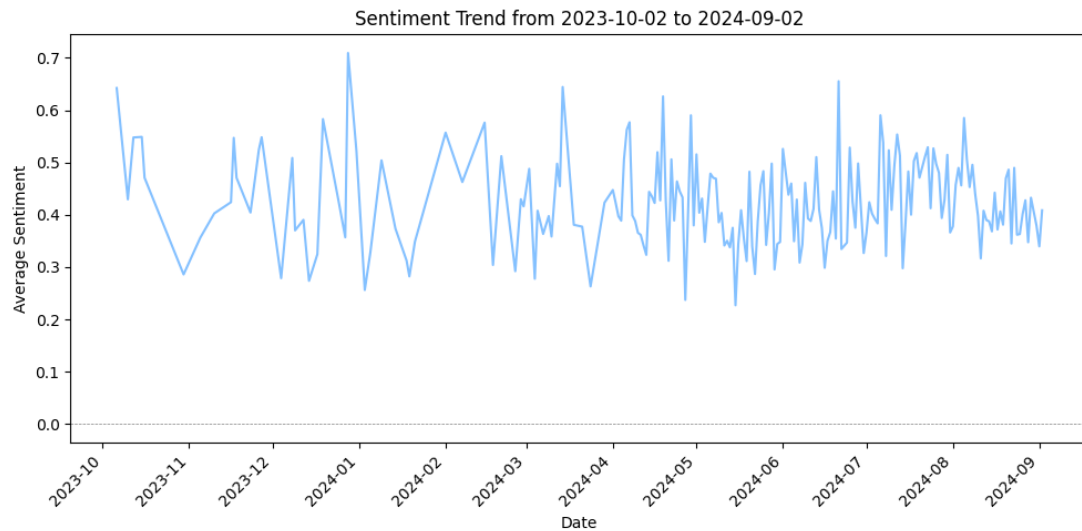
以下針對三個測試範例做說明：



這是 health 2018~2020，由此顯示人民對健康的相關議題愈來愈有信心，因為技術的上升，其中 2019 年是分界點，因為在疫情下的台灣，展現出極高的防禦心理，造成新聞多為正面報導。



由此圖可以顯示，由於近 2024 年奧運，因此有關於運動的相關議題，皆於 2024 開始情緒上漲，並於 2024 8 月屬於高峰值群。



股市方面，可以發現是屬於動盪較大的情緒指標，這會與當時刻的股市大盤走向有關，而最近因為戰爭、天災等等...因素，因此情緒指標走跌。

另外也有針對資料索引設計相關網頁，詳細說明請參考第六章 database 的索引資料庫部分：



新聞公司

China Times

主題

stock

日期

2024/10/09

情緒

All

開始

Title	Date	Content	Source	Emotion	URL
台中勤美 洲際酒店 10 / 23起 迎賓 打造 全新住宿 體驗	2024- 10-09	【旅遊經 洪書瑱報導】根據台灣議題研究中心資訊，台中在六都觀光好感度拿下冠軍，其次為高雄、台北、新北、台南、桃園，而台中的觀光蓬勃，也吸引星級飯店及飯店品牌進駐，目前台中的星級飯店，包括：李方艾美、林酒店、麗寶福容、台中裕元花園、台中日月千禧、台中金典、長榮桂冠、台中福華、永豐棧、台中大毅老爺等，現在民眾到台中遊玩，又有新飯店可選擇了，新開幕酒店為IHG洲際酒店集團直營管理的旗下	China Times	1	Link

而以下是網頁的其他網頁(包含首頁，與介紹如何使用):  
首頁:

## Sentiment News

融合大數據 Transformer 與 GenAI，我們為您  
提供最具價值的新聞情緒辨識。

### Crawl for news and twii.

自動化爬蟲!

Posted by *Team4* on Nov 3, 2024

關於:

## 關於我們

Sentiment News可自動生成相關新聞的情緒指標。

我們提供自動化網路爬蟲，提供多來源、自適應、動態的網路爬蟲，並可以動態儲存到對應的資料庫。

我們可以動態生成對應的情緒便是分數，並生成對應的可視化圖表

詳細內容，可點選以下的github鍵以了解更詳細的資訊。



## 心得:

A1105505 林彧碩

在本次的專題中，我主要負責分配工作、整合、web、word、ppt 與爬蟲主要的操作，我在本次的專題中，學習到很多，最主要於爬蟲的部份我們要學習如何可以不間斷的一直爬蟲，並使其可以自適應與動態爬蟲，在與老師對談的過程中，我了解到，動態爬蟲並非只單一指 selenium，而是有許多操作，我們在撰寫程式碼時，除了思考到要如何更加動態，也有去觀察不同網頁本身是因為怎麼的原因容易被擋，在過程中，我認為我最深刻的是，生成式 AI 的極限性，在過程中，有許多組向我詢問一些想法，因為我們這組是最早開始動工的，我認為如果設計一個生成式 AI 可以動態偵測網路架構並修改，是十分不切實際的，生成式 AI 僅能在小部分幫助很大，但當達到整個網頁架構時，他就會受限於字數限制、判斷能力，而無法給出最好的方法，我認為最好的做法，是在程式中，適時加入 GenAI 的判斷，像是針對非相關的新聞判斷，簡易判斷網頁結構的改變，與改善爬蟲時，向 GenAI 詢問等等...，這些我認為才是最好與 GenAI 溝通的方法，另外再寫網頁的部分，因為我希望可以達成可以長時間爬蟲、動態爬蟲、並且資料量大，因此我不追求時間最長(因為過長其實沒有用，以近期的資料愈好)，因此最長可以爬到 2003 年，此程式碼，我讓它架設在 server 上跑整整 2 星期，由此可見此程式是如此不被網站偵測為機器人，另外我們使用動態設定頁數的方式給使用者，也是為了讓使用者可以動態爬取尚未爬到的資料，最後在資料庫的部分，我們為了讓組員皆可以使用此 database 因此我使用 supabase 的雲端資料庫，但此資料庫不太方便公開給他人(僅可以加入特定使用者，因此我有用 .sql 的方式提供整個資料庫，如助教或老師需要觀看此資料庫，可以提供給我 github 帳號，我將提供整個資料庫給老師或助教看)。

以下將簡述一些爬蟲學習到的事:

在進行新聞資料爬取的過程中，我們先準備了一組常用的 User-Agent 列表，讓每次 request 都能隨機選用其中一個 User-Agent，藉此減少因使用固定 User-Agent 而被網站偵測為爬蟲的風險。此外，我們還利用了 Chrome WebDriver 的 execute\_cdp\_cmd 方法，把 navigator.webdriver 屬性設為 undefined，以隱藏 WebDriver 的使用。還有使用生成式 AI 與 css selector 等作適應性變化。甚至我們為了讓爬蟲更加多元，可以來源更多，因此我們還利用 API 的方法，還有許多沒敘述的方法請詳細參閱前面章節。

A1105524 吳雨宣

在爬取中時的新聞時，我們定義了一組常見的 User-Agent 列表，每次 request 時都隨機選一個 User-Agent，避免因為固定的 User-Agent 而被偵測為爬蟲。除此之外，還使用 Chrome WebDriver 中的 execute\_cdp\_cmd 方法，將 navigator.webdriver 屬性設定為 undefined，這可以隱藏 WebDriver 的使用。

本來想說既然中時用這些機制去解決被偵測為爬蟲的問題，同樣身為新聞網的自由時報應該也可以用這種方式吧，沒想到中時的爬蟲架構竟然不適用於自由時報。因為中時的部分，我們是直接在搜尋結果出來的那頁就直接抓取我們所需的新聞資料，發現自由時報不是用這種方式後，我們就改成在搜尋結果那頁先將所有的新聞連結都存起來，再透過新聞連結用 BeautifulSoup 的方式去抓取內文、日期等。而 TVBS 的部分雖然也是在搜尋結果那個頁面直接抓取需要的資料，但我是用 requests 和 BeautifulSoup 來進行網頁請求和解析。爬了三間新聞網後，發現爬蟲真的不是一件簡單的事，因為每個網頁的架構都不一樣，雖然生程式 AI 可以給我們大概的爬蟲架構，但很多細節都要人類自己去觀察修正，像是欲爬取資料的位置、網頁的架構等。這次的專題在爬蟲方面，我們真的花了很多心力，但也在這一次次的爬蟲中，對網頁架構和反爬蟲機制有更清楚的認識。

#### A1105507 蘇柏諺

我覺得這次專題實作讓我學到最多的是如何用 transformer 精準的判斷 sentiment，一開始我們使用 Gemini 等模型測試，但精準度、處理時間和對中文的理解都有很大的差異，特別是這次我們要處理幾十萬筆的新聞分析，所以實作時也要把龐大的資料量考慮進去，最後我們使用"nlptown/bert-base-multilingual-uncased-sentiment"模型，很順利的處理完所有新聞的情緒分析，讓我感到很有成就感，另外為了儲存資料並方便小組開發，我們學會使用 supabase 這個功能強大且好用的共用資料庫，也讓我覺得這次的實作受益良多。

#### A1105523 巫柔筠

這次的專題我負責的是，情緒分析的部分，我們一開始試了很多不同的方式跟模型，最後選擇的模型是"nlptown/bert-base-multilingual-uncased-sentiment"，這個模型會給星等，我們再利用五顆星等，去分成負面、中立、正面，我們也有嘗試做一些微調，但結果並沒有原本模型來的好，所以沒有更改太多參數。我主要是處理前端，可以用日期的區間、主題和來源去查詢統計的數據資料，一開始沒有注意到，因為我們的資料量龐大，所以用了兩個資料庫，這兩個資料庫的 id 會有重複，跑出來的結果會有誤，之後做了修正。這次的專題對於情緒分析又有更深入的了解，在前端跟資料的處理也學到了更多的技巧，這些經驗對於未來的專題都有很大的幫助。

#### A1105521 黎子歲

在期中專題我主要參與了情緒分析與視覺化的部分，目標是通過情緒分析來理

解新聞文章的情感傾向。這個專題實作讓我有機會深入認識 NLP 技術，使用工具來解析不同主題的情緒變化。我們選用了 transformers 庫中的預訓練模型 BERT 進行文本情緒分類。一開始，我們在幾個不同版本的 BERT 預訓練模型來選擇，最後選擇 nlptown/bert-base-multilingual-uncased-sentiment。原因是此版本已經基於龐大的資料集進行了預訓練，能更好的捕捉語言中的情緒表達特徵，且無須多做微調便能符合新聞情緒分析的情緒分類需求。視覺化的部分我們可以基於三種不同的資料組合生成圖表，多方呈現情緒變化和趨勢。

另外，此專題需要處理大量的新聞數據並進行情緒分析，我們選擇了 Supabase 來管理和儲存情緒分析結果。Supabase 是開源的 BaaS，它提供了方便的資料庫管理功能，讓我們能夠輕鬆儲存並查詢分析結果，同時可以透過 API 請求將情緒分析結果和前端的應用融合。

在前端方面，我們使用了 Flask 框架來構建 API 和管理資料流，使前端可以通過簡單的 API 請求即時獲取情緒分析結果。使用者能透過前端即時查看不同時間範圍、新聞來源的情緒趨勢，更直觀地理解情緒分析結果。

## 分工:

A1105505 林彧頤	新聞爬蟲設計(共參與 1,2,3,5,6,7 題)，主要爬蟲設計發想與實作，如何更加動態與具適應性，和如何結合生成式 AI，並將資料存於資料庫(資料庫設計)，最後設計網頁系統，供使用者自動化網路爬蟲、動態更新，並進行 index 功能，word 統整與撰寫、ppt 統整與撰寫，上台報告。
A1105507 蘇柏諺	新聞情緒分析(協助修改情緒分析)，word 撰寫、ppt 撰寫，上台報告。(共參與 4 題)
A1105521 黎子歲	新聞情緒分析(即時分析與搜尋)，並將資料存於資料庫，進行網頁撰寫，word 撰寫、ppt 撰寫，上台報告。(共參與 4,6,7 題)
A1105523 巫柔筠	新聞情緒分析(情緒分析與介面)，並將資料存於資料庫，進行網頁撰寫，word 撰寫、ppt 撰寫，上台報告。(共參與 4,6,7 題)
A1105524 吳雨宣	新聞爬蟲設計(共參與 1,2,3,5,6 題)，如何更加動態與具適應性，並將資料存於資料庫(資料庫設計)，word 撰寫、ppt 撰寫，上台報告。