




## e 手包辦新聞網

陳惠如\* 羅崇銘\*\*

### 【摘要】

本研究收集 2019 年至 2020 年間刊載於聯合報的所有新聞標題，經由斷詞、詞性、情緒分析及 NRC Emotion Lexicon 等技術，輔以網站開發，提供讀者能夠篩選各類新聞情緒之平臺—「e 手包辦新聞網」，囊括 11 種情緒分析：正面、中性、負面、憤怒、預期、厭惡、恐懼、喜悅、悲傷、驚喜與信任，期望能夠以客觀方式呈現與提醒讀者每篇新聞的情緒。

為讓閱讀體驗更具個人化特色，進一步規劃了會員制度、推播通知、推薦讀者感興趣的新聞情緒類別內容，以及運用文字雲、折線圖及長條圖等視覺化圖表，顯現標題情緒分析之結果，以利讀者掌握當今時下熱門事件，或對於新聞事件之情緒演變等資訊一目了然。

### 關鍵詞

新聞標題 情緒分析 情感分析 標題分類 新聞網站

---

\* 國立政治大學圖書資訊與檔案學研究所研究生  
E-mail: hbl3987@gmail.com

\*\* 國立政治大學圖書資訊與檔案學研究所副教授  
ORCID 0000-0001-5068-6002  
通訊作者 E-mail: buddylo@nccu.edu.tw

## 壹、前言

「秀才不出門，能知天下事」於現今社會再也非一檔難事，隨著網路媒體的崛起，各大新聞平臺提供即時、多元的時事動態，匯集來自國內外新聞資訊，從而提升人們新聞視野之廣度與深度，掌握全球脈動。再者，新聞產業競爭日趨激烈，眾多媒體力求生存，加上讀者對於標題具有負面偏好，因而造就媒體大量利用災難、血腥和煽情等負面新聞題材博取大眾關切，以創造點閱流量。

然而，對於人們易受負面新聞內容吸引眾說紛紜，荷蘭學者 Dijksterhuis 與 Aarts (2003) 以達爾文進化論學說分析解譯，認為人們對負面刺激事物的關注力要大於正面刺激，主要其因來自生物避免危難之本能，隨演化發展人們已能夠快速對任何刺激加以分類、評估其適當性，推斷是否需要迴避，而負面新聞的存在猶如提醒應迅速躲避潛在危機的跡象。

近年來，學者們亦積極投入與負面新聞相關研究，加拿大麥吉爾大學透過實驗證實人們具有負面偏好，由 100 名受試者選擇閱讀新聞內容，從中發現負面新聞比正面新聞的可能性高出 50%，意即人們對於負面新聞的感興趣程度遠大於正面新聞 (Trussler & Soroka, 2014)。《美國國家科學院院刊》更發表史上規模最大的比較心理生理研究，以社會科學的角度分析來自 17 個國家的 1,156 名受訪者，檢視其對負面新聞的傾向與現今負面新聞的氾濫間關聯性，可得知兩者間具有相當因果關係 (Soroka, Fournier, & Nir, 2019)。

倘若人們持續性曝露於負面新聞中，對於感情的認知越趨麻木、感受力下降，形成所謂的「同情疲乏」，焦躁枯萎的心靈讓人逐漸變得冷感，隨之而來的憤怒、絕望無助感導致摧毀同理心 (Pffifferling & Gilley, 2000)。更有醫學專家指出，災難性報導容易使人們抑鬱、恐懼、焦慮、濫用毒品，抑或是創傷後壓力症候群等心理失衡狀況產生 (Pfefferbaum et al., 2014)。相形之下，積極正向的新聞體現世界的美好，充滿溫暖與希望，而在 1 項正負面新聞之觀後情感差異研究，針對 307 名受試者進行調查，從中發現相較於閱讀負面新聞，閱讀正面新聞的受試者給予正向觀後回饋，如可懂得享受閱讀的樂趣、領略其中奧妙等，由此可知閱讀正面新聞能獲得高度

的正面影響 (McIntyre & Gibson, 2016)。

有鑑於此，負面新聞盛行，尤其在 COVID-19 疫情橫行之當今，於此氛圍下恐有礙讀者的身心健康。雖然正面新聞予以讀者正向影響力，但各大新聞網站鮮少對自家出品的新聞進行情緒分類，且民間成立的媒體時事分析網站多半只注重分析所得之數據，使得讀者仍無法經由現行平臺功能進行篩選新聞之情緒。為此，本研究蒐羅 2019 年及 2020 年聯合報之新聞標題等，並經由情緒分析、斷詞及詞性、NRC Emotion Lexicon 及協同過濾等技術，輔以網站開發，提供讀者能夠篩選各類新聞情緒之平臺—「e 手包辦新聞網」。此外，以期為讓閱讀體驗更具個人化特色，規劃會員制度、設計推播通知，推薦讀者感興趣的新聞情緒類別內容，更運用文字雲、折線圖及長條圖等視覺化圖表，顯現標題情緒分析之結果，以利讀者掌握當今時下熱門事件，亦能一目了然新聞事件之情緒演變等。

## 貳、材料和方法

本研究建立「e 手包辦新聞網」使用 CKIP Tagger 斷詞及詞性、TF-IDF、SnowNLP 情緒分析、NRC Emotion Lexicon、Bidirectional Encoder Representations from Transformers (BERT)、協同過濾、PostgreSQL 資料庫、AWS 及 Django 框架十項工具，使用工具順序流程如圖 1 所示。

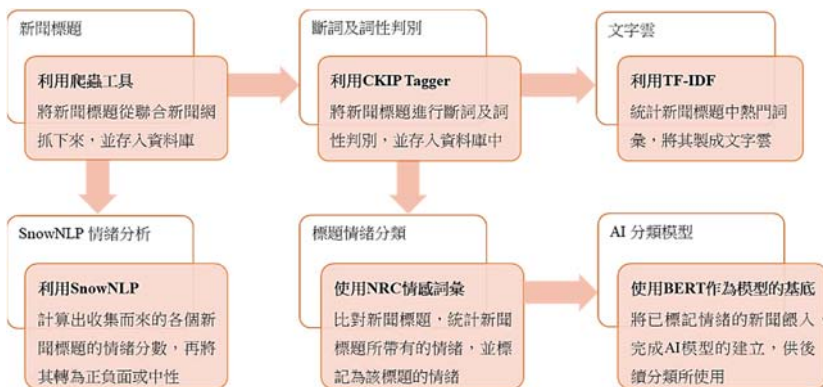


圖 1 技術架構與流程

資料來源：本研究整理

整體來說，「e 手包辦新聞網」之伺服器是以 Django 作為後台，首先透過聯合新聞知識庫收集新聞標題，進行新聞情緒分析，導入 PostgreSQL 資料庫中，並同時部署至 AWS 上，網頁前端部分採用 HTML、JavaScript、CSS，圖 2 為系統架構圖。

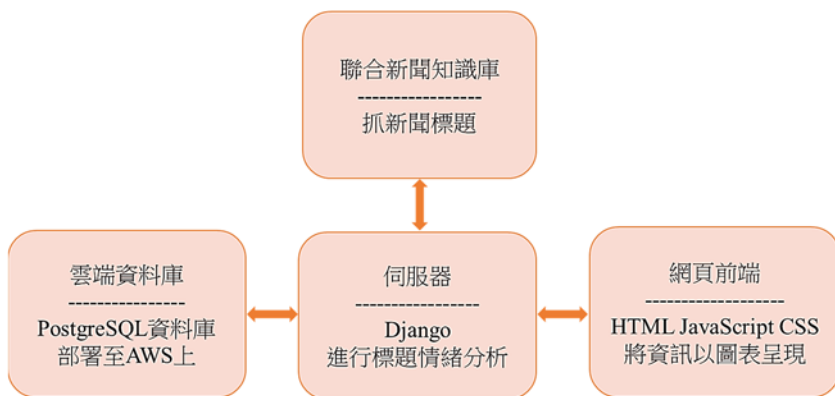


圖 2 系統架構圖

資料來源：本研究整理

## 一、聯合知識庫

本研究於聯合知識庫提供之全文報紙資料庫 (<https://udndata.com/ndapp/upoint/udntag/Index>) 中，蒐集 2019 年至 2020 年的新聞標題、版別、時間與網址，並針對其新聞標題進行 11 種情緒分析與分類，及分析疫情發生前後新聞標題所產生的情緒變化。

## 二、SnowNLP 情緒分析

SnowNLP 係為中文自然語言函式庫，有別於現今多數以處理英文內容為主流的函式庫，其具有多項功能，如執行中文斷詞、情感分析、文章分類、提取文章摘要及關鍵字等 (Chen, Chen, & Shi, 2018)。

因此，本研究利用 SnowNLP 情緒分析工具，以深度學習方法具體計

算得出各個新聞標題的情緒分數，再將其轉換為正負面或中性類別。情緒分數所得數值皆界於 0 與 1 之間，值越趨近於 0，即代表新聞標題情緒愈為負面，而越趨近於 1，情緒則愈為正向，且為使情緒類別客觀轉換，故本研究將上下限分別設定為 0.7、0.3，當值大於 0.7 標註為正面，小於 0.3 為負面，位於 0.3~0.7 區間則為中性，如表 1 所示。

表 1

新聞標題之情緒分數與類別對照

新聞標題	情緒分數	情緒類別
日本真子公主婚事 父親點頭	0.965915979903163	正面
游抗拒 調職被壓下？	0.71919406198646	正面
拜登經濟、發言團隊 女力當家	0.567473321641477	中性
明年起 大陸禁止固體廢物進口	0.422209272932613	中性
大陸網絡安全產業 規模逾 1700 億	0.410760879751891	中性
疫情效應 數位詐欺激增	0.285396690869622	負面
印尼移工 暫停入境 2 周	0.019731129276034	負面

資料來源：本研究整理

### 三、CKIP Tagger 斷詞及詞性

CKIP Tagger—中文斷詞暨實體辨識系統，係由中研院中文計算語言研究小組所發布的開源中文處理工具，套件的功能包含中文斷詞、詞性標注與實體辨識，且語料庫之一為中研院所建構的現代漢語平衡語料庫，其收錄報紙等相關文檔記載臺灣語言規律、語法基礎及寫作手法等，對於本研究進行新聞標題斷詞，能使得所得結果更加適切（CKIP Lab 中文詞知識庫小組，2020）。

此外，表 2 呈現經中研院所測試結果，顯示 CKIP Tagger 成效已大幅超越其他斷詞工具，如舊版系統的 CKIPWS 以及被廣泛使用的結巴系統 Jieba-zh\_TW，尤其在斷詞的 f1-score 達 97.33%，詞性的準確率(accuracy) 亦達 94.59%，可得知為現行市面上表現最佳的開源繁體中文處理工具。因

此，本研究利用 CKIP Tagger 套件進行斷詞與詞性判別，並針對斷詞結果以「+」進行分隔，以利儲存於資料庫中，如表 3 所示。了解新聞標題的詞性組成，有助於在 TF-IDF 統計新聞標題中的熱門詞彙前，刪除連接詞等無意義的詞語，避免將其作為熱門詞彙出現於文字雲當中。

表 2

**CKIP Tagger 與其他斷詞工具效能測試結果**

斷詞工具	(WS) precision	(WS) recall	(WS) f1-score	(POS) accuracy
Ckip Tagger	97.49%	97.17%	97.33%	94.59%
CKIPWS	95.85%	95.96%	95.91%	90.62%
Jieba-zh_TW	90.51%	89.10%	89.80%	無資料

資料來源：*CkipTagger*, Peng-Hsuan Li, 2020, retrieved from <https://github.com/ckiplab/ckiptagger>

表 3

**新聞標題斷詞與詞性判別**

新聞標題	斷詞結果	詞性
日本真子公主婚事 父親點頭	日本+真子+公主+婚事+ + 父親+點頭	Nc+Nb+Na+Na +WHITESPACE+Na+VA Nb+VC+WHITESPACE+V
游抗拒 調職被壓下？	游+抗拒+ +調職+被+壓下+？	B+P+VC+QUESTIONCAT EGORY
拜登經濟、發言團隊 女力當家	拜登+經濟+、+發言+團隊+ + 女力+當家	Nb+Na+PAUSECATEGOR Y+ VA+Na+WHITESPACE+Na +VA
明年起 大陸禁止固體 廢物進口	明年+起+ +大陸+禁止+固體+ 廢物+進口	Nd+Ng+WHITESPACE+Nc +VE+Na+Na+VC
大陸網絡安全產業 規模逾 1700 億	大陸+網絡+安全+產業+ +規模 +逾+1700 億	Nc+Na+VH+Na+WHITESP ACE+Na+VJ+Neu
印尼移工 暫停入境 2 周	印尼+移工+ +暫停+入境+2+周	Nc+Na+WHITESPACE+VH C+VA+Neu+Nf

資料來源：本研究整理

#### 四、NRC Emotion Lexicon

NRC Emotion Lexicon (EmoLex) 是經加拿大國家研究委員會 (National Research Council of Canada) 所建立，該情緒辭典描述英語單詞與情感間之關聯。NRC Emotion Lexicon 中包含 2 種情緒：正面與負面，以及 8 種基本情感：喜悅 (joy)、悲傷 (sadness)、憤怒 (anger)、恐懼 (fear)、信任 (trust)、厭惡 (disgust)、預期 (anticipation) 和驚喜 (surprise)。其中，8 種基本情感為採納心理學家 Robert Plutchik 所提出之理論「Plutchik's Wheel of Emotions」，如圖 3 所示，相鄰的情緒可互相組合，彼此間亦無明確界限，更表明 8 種情感存在著 4 個對立面，喜悅—悲傷、憤怒—恐懼、信任—厭惡和預期—驚喜 (Mohammad & Turney, 2013)。

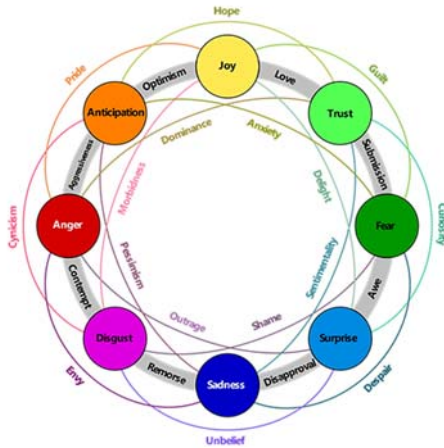


圖 3 Plutchik's Wheel of Emotions

資料來源：Plutchik's Wheel of Emotions, Wikimedia Commons, 2020, retrieved from [https://commons.wikimedia.org/wiki/File:Plutchik\\_Dyads.svg#/media/File:Plutchik\\_Dyads.svg](https://commons.wikimedia.org/wiki/File:Plutchik_Dyads.svg#/media/File:Plutchik_Dyads.svg)

雖然 NRC Emotion Lexicon 整體是基於英文資料所組成，恐因國情、文化等影響，涉及語言轉換時，容易造成些許差異，但研究顯示，大多數情感規範在橫跨語言仍相當穩定，故該辭典可擴充至 105 種語言，其中更包含繁體中文 (Mohammad & Turney, 2013)。

利用 SnowNLP 情緒分析工具所得情緒只有正負面或中性，倘若只將新聞標題情緒區分為 3 類恐過於偏頗，為避免單一情緒，將更精細情緒分類，本研究再採用 NRC Emotion Lexicon 對照新聞標題，以標記該新聞標題所帶有的情感。NRC Emotion Lexicon 共收錄 14,182 個英文詞彙，經審視刪除重複及不適合詞彙，剩 5,060 個單詞，再利用這 5,060 個單詞與新聞標題相互比對，進而統計出其隱含情緒，以利後續分析使用。

同樣一則新聞標題，可能涵蓋不同層面情緒，正負面情緒完全對立，而其餘八種情緒仍可同時並存，以「破產」為例，此一單詞為負面涵義，且概括憤怒、厭惡、恐懼及悲傷 4 種情緒，猶如驗證「Plutchik's Wheel of Emotions」理論，情緒相互拼湊，突破虛無之界限框架。

表 4

詞語與情感的關聯

英文	中文	正面	負面	憤怒	預期	厭惡	恐懼	喜悅	悲傷	驚喜	信任
aberrant	反常	0	1	0	0	0	0	0	0	0	0
admirer	愛慕者	1	0	0	0	0	0	0	0	0	0
armament	武器	0	0	1	0	0	1	0	0	0	0
bankruptcy	破產	0	1	1	0	1	1	0	1	0	0
barbarism	粗俗	0	1	0	0	0	0	0	0	0	0
barrier	屏障	0	1	1	0	0	0	0	0	0	0

資料來源：本研究整理

## 五、TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) 為一種常用於資訊檢索或文字探勘的加權技術，計算評估單詞在特定文件的集合或詞庫中的相關重要性 (Ramos, 2003)。

於公式(1)中，Term Frequency 為詞彙出現頻率，在特定新聞中，若詞語的 TF 值愈高，代表其重要性越高；公式(2)，Inverse Document Frequency 為逆向檔案頻率，由總新聞數目除以特定新聞中該詞語的數目，再將得到的商取以 10 為底的對數，此一目的為降低常見但不具意義的詞語，例如：「的」、「了」或「我們」等。公式(3)中，TF-IDF 透過 Term Frequency 與



Inverse Document Frequency 之數值相乘所得，最終過濾常見並保留重要的詞語。本研究利用 TF-IDF 統計新聞標題中的熱門詞彙，並將熱門詞彙以文字雲的方式呈現。

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

$$idf_i = \lg \frac{|D|}{|\{j:t_{j \in d_j}\}|} \quad (2)$$

$$tfidf_{i,j} = tf_{i,j} \times idf_i \quad (3)$$

## 六、Bidirectional Encoder Representations from Transformers (BERT)

BERT 是由 Google 提出，用於自然語言處理 (NLP) 的預訓練技術。BERT 的架構精神為雙向 Transformer 的 Encoder，BERT 主要的應用會是以 Transfer learning 後接著 Fine tune model (Devlin, Chang, Lee, & Toutanova, 2019)。本研究使用 BERT 作為模型的基底，並使用已完成標記的新聞標題訓練及建立 AI 模型，以利後續分類使用。

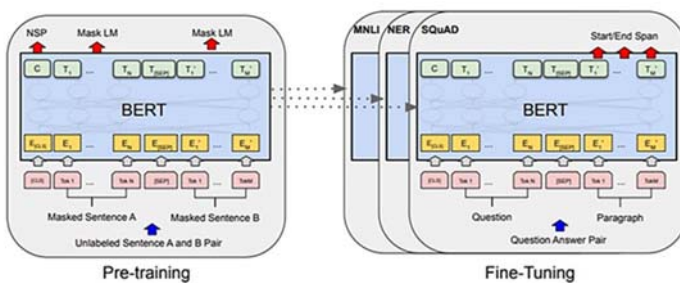


圖 4 兩階段遷移學習在 BERT 下的應用：

使用預先訓練好的 BERT 對下游任務做 fine tuning

資料來源：*Bert: Pre-training of deep bidirectional transformers for language understanding*, by Devlin, J., Chang, M. W., Lee, K., & Toutanova, K., 2019, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, p. 4173

## 七、協同過濾

協同過濾 (collaborative filtering) 是建構推薦系統中被廣泛使用的方法，透過使用者的已知偏好、歷史行為數據預測其他使用者對相同主題或產品的偏好程度，並將其推薦給使用者。以使用者為基礎的協同過濾，係指透過興趣愛好等計算使用者相似度推薦值，即為找尋與使用者具有相似偏好、興趣相同的另一使用者，兩者感興趣的內容可能相似，因而將他們分為同類，再對目標使用者的興趣進行預測，產生推薦結果 (Su & Khoshgoftaar, 2009)。

本研究主要收集讀者感興趣的新聞情緒，並將讀者進行分群，進而透過協同過濾法，依目標讀者的新聞情緒傾向進行預測，以建立推薦內容。

## 八、PostgreSQL 資料庫

PostgreSQL 是 1 套企業資料庫系統，1986 年由加利福尼亞大學柏克萊分校電腦科學系基於 POSTGRES 4.2 所研發的物件關聯式系統，其功能包含具有多版本並行控制 (MVCC)、時間點資料 (PITR)、表格空間 (Tablespaces) 與線上熱備份等 (PostgreSQL, 2021)。

## 九、AWS

亞馬遜雲端運算服務 (AWS)，是由亞馬遜所建立的雲端運算平臺 (Amazon Web Services, 2021)。本研究利用 Amazon Relational Database Service，建立執行 PostgreSQL 資料庫環境並連接到資料庫。

## 十、Django 框架

本研究由 HTML、JavaScript、CSS 作為「e 手包辦新聞網」網頁前端，搭配 PostgreSQL 資料庫，使用 Django 建立網頁。Django 網頁應用框架為 MTV 軟體設計模式，包含 4 個部分：網址 (URLs) 為入口或訪問內容的路徑；視圖 (View) 以接收請求或數據庫的內容，並回傳模板進行展示；儲存數據的模型 (Models)，以及做為端點之模板 (Templates)，以提供最終內容的形式 (Rubio, 2017)。

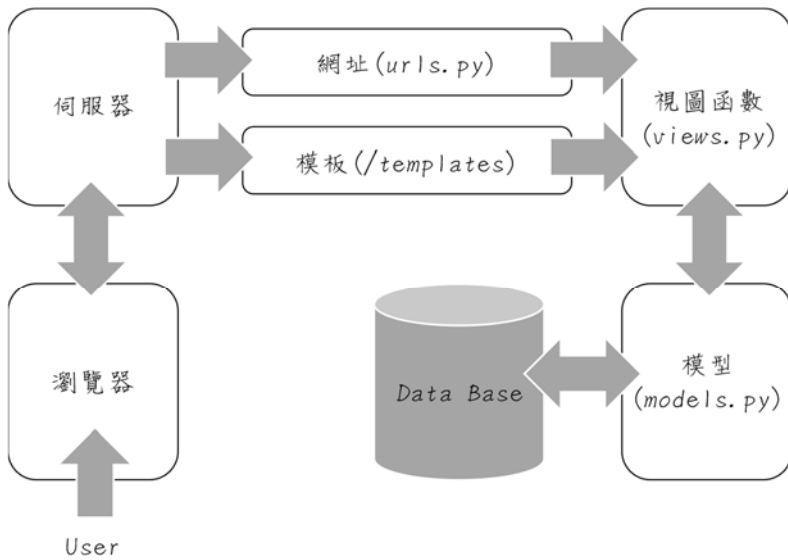


圖 5 Django 之 MTV (Model-Template-View) 架構

資料來源：本研究整理

本研究以「e 手包辦新聞網」(<http://210.240.245.72:8000/index/>) 展現研究成果，將從聯合知識庫中獲取的大量新聞標題進行情緒分析，並藉由分析結果呈現於平臺網站中，網站頁面分為「主頁」、「搜尋」、「最新熱門」、「統計資料」與「關於我們」，以下將依序介紹說明。

## 一、主頁

### (一) 會員註冊與登入

使用者可於主頁進行註冊以加入「e 手包辦新聞網」會員，並於註冊頁面中同時設定自身偏好、情緒與版面，若註冊頁面尚未填寫完畢，系統將立即提醒使用者填寫該欄位。此外，主頁左上角顯示使用者登入狀態、登出與更改偏好設定之選項。



圖 6 會員註冊與登入

資料來源：本研究整理

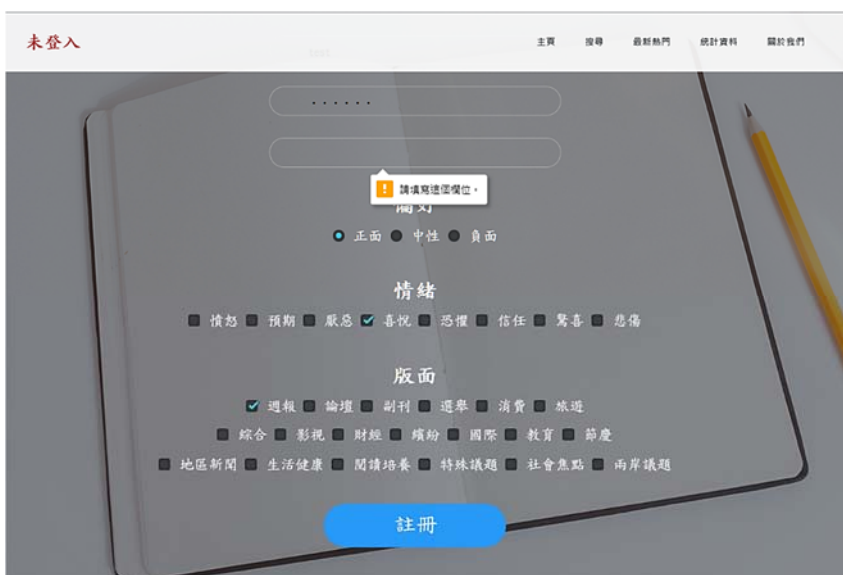


圖 7 個人偏好設定

資料來源：本研究整理

## (二) 個人化推薦

擁有「e 手包辦新聞網」會員之使用者，可享有網站提供之推薦服務，幫助使用者排除不想要的情緒、資訊等，且偏好設定不論何時皆可更改，隨時呈現使用者想看的內容。例如，將個人偏好設定為「偏好：正面」、「情緒：喜悅」與「版面：週報」，網站所提供之個人化推薦結果如圖 8 所示。

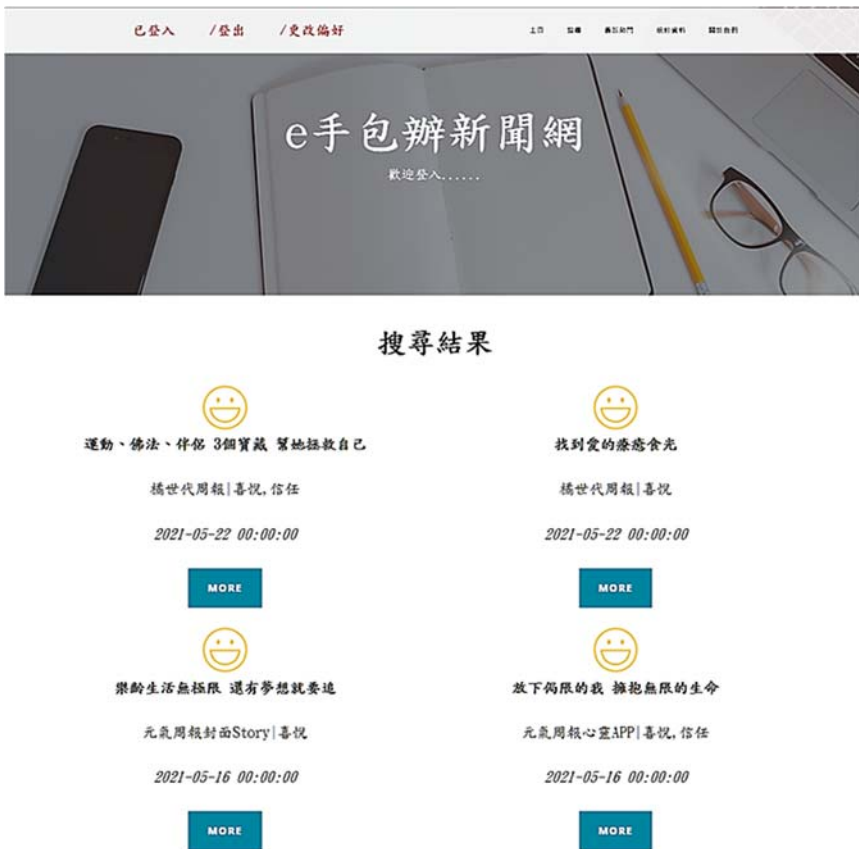


圖 8 個人偏好設定結果

資料來源：本研究整理

## 二、搜尋

使用者可於搜索框內輸入關鍵字或透過下方之「偏好」、「情緒」及「版面」進行重點勾選，兩者亦可相互結合，搜尋結果以標題與連結呈現。例如，輸入「關鍵字：臺灣」並勾選「偏好：負面」與「情緒：憤怒」，其搜尋結果如圖 9 所示。



### 搜尋結果

 全球都喊危險，只有台灣社會無感 焦點 憤怒, 恐懼, 罵評 2021-05-09 00:00:00 <a href="#">MORE</a>	 中資假冒外企 竊台灣高科技回傳陸 社會 憤怒, 厭惡 2021-05-08 00:00:00 <a href="#">MORE</a>
 陸艦闖台灣和與那國島間 兩岸國際 憤怒 2021-05-02 00:00:00 <a href="#">MORE</a>	 輕視台灣人才 廣西推進四心工程 兩岸動態 憤怒, 罵評 2021-04-27 00:00:00 <a href="#">MORE</a>

圖 9 搜尋結果

資料來源：本研究整理

### 三、最新熱門

使用者可透過「大家都在搜」及「精選熱門」檢視熱搜比對結果。「大家都在搜」為 Google 當日熱門關鍵字搜尋排行，而「精選熱門」係依據詞頻計算聯合報新聞標題之熱門詞彙排行，即代表使用者可了解 Google 搜尋之用戶關注傾向與聯合報報導傾向之異同。



圖 10 最新熱門

資料來源：本研究整理

#### 四、統計資料

向使用者提供新聞情緒分析資料的查詢，可於下拉式選單中設置數據的範圍，如情緒、版面、數量和日期，輸出結果亦可選擇以文字雲、折線圖、圓餅圖或折線圖呈現。



您可以在此繪製您的個人圖表

圖 11 統計資料

資料來源：本研究整理

##### (一) 文字雲

以文字雲呈現熱門關鍵字之出現次數多寡與情緒傾向。





圖 12 文字雲

資料來源：本研究整理

## (二) 折線圖

使用者可自行設定座標軸（X、Y 軸）與折線之數值，以更符合使用者的個人需求。若選擇 X 軸為月份，Y 軸為新聞數量（則），折線所呈現之新聞情緒，且將游標移至折線上，可查看各節點資訊，如圖 13 中，四月份新聞標題情緒為喜悅的數量共 1,097 則。此外，使用者亦可於右方資訊欄點擊增減，以比較重點情緒，圖 14 以點選驚喜、預期、喜悅與恐懼為例。

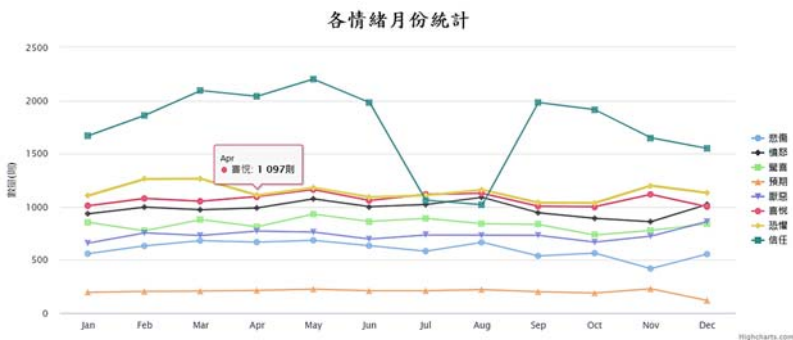


圖 13 折線圖一節點資訊

資料來源：本研究整理

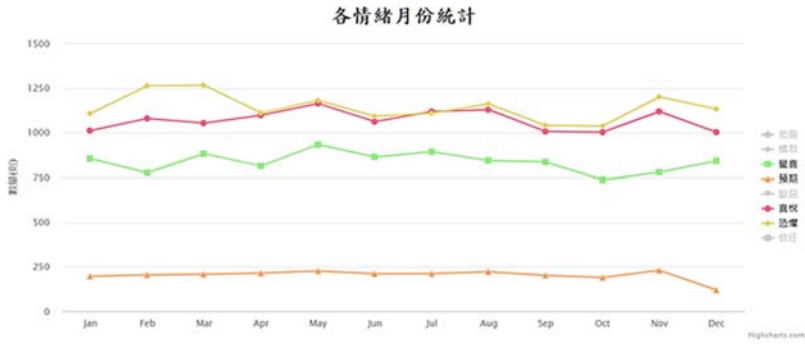


圖 14 折線圖—資訊欄

資料來源：本研究整理

### (三) 圓餅圖

使用者可透過圓餅圖了解某段時間內新聞版面或情緒數量佔比，亦可將游標移至不同扇區上方詳細查看其資訊。如圖 15 所示，為一個月內的各種新聞版面在整體中所佔之比例。

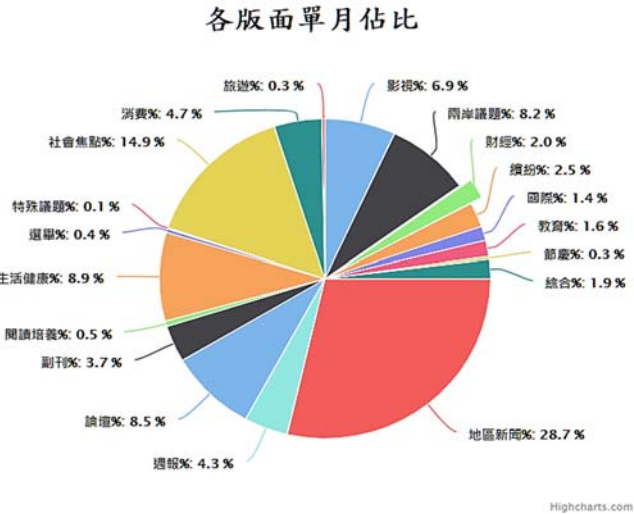


圖 15 圓餅圖

資料來源：本研究整理

#### (四) 長條圖

使用者可自行設定長條圖中座標軸（X、Y 軸）之數值，以顯示某段時間內的新聞情緒數量等，且將游標移至區塊上方可查看其詳細資訊。下圖為一年中各個月份的情緒數量之呈現。若選擇 X 軸為月份，Y 軸為新聞數量（則），長條圖將呈現各月份之新聞情緒數量，如圖 16 所示。再者，使用者亦可於下方資訊欄點擊篩選欲比較之新聞情緒，圖 17 以點選憤怒、恐懼、驚喜與信任為例。

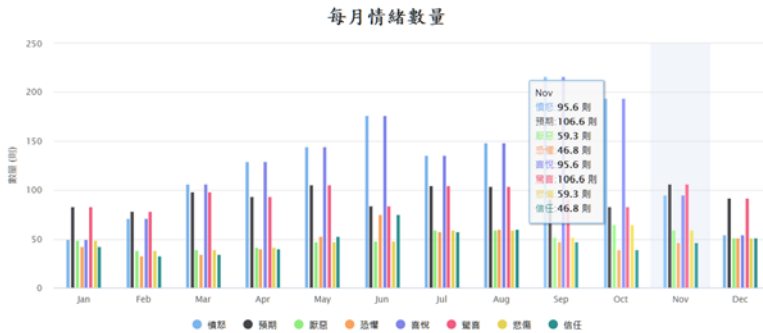


圖 16 長條圖－情緒數量

資料來源：本研究整理

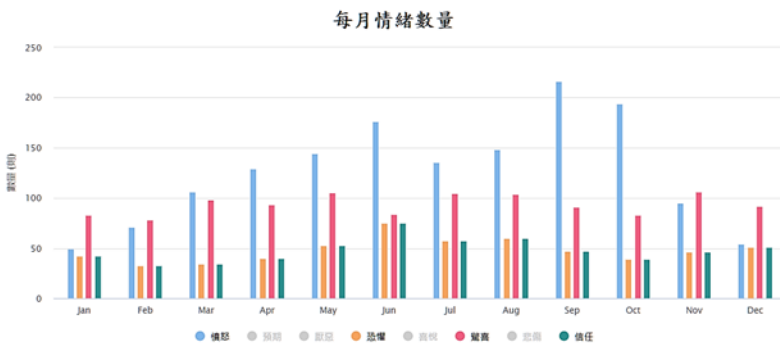


圖 17 長條圖－新聞情緒比較

資料來源：本研究整理

## 五、關於我們

對於「e 手包辦新聞網」網站進行簡介，並說明其分頁的功能概述、會員專屬等，如圖 18 所示。



圖 18 關於我們

資料來源：本研究整理

## 肆、討論

媒體興盛的時代，民眾可透過不同媒體獲取多面向的新聞報導，如一般新聞媒體網站，抑或是 Web 新聞聚合器，包含 Yahoo! News、Google News 等，且為維繫使用者顧客忠誠度，紛紛推出為使用者量身打造專屬的新聞個人化設定，在這之中，系統設計大多依據使用者興趣、點閱情況進而推薦特定主題之新聞，卻忽略使用者心理層面之因素，且現今人們對於資訊

中的情緒重視程度逐漸提高，使用者不再願意單方面接收主流媒體所輸出之資訊情緒。因此，本研究以「e 手包辦新聞網」提供使用者選擇偏好之新聞情緒的功能。

雖與一般新聞網相比，「e 手包辦新聞網」新聞資訊量相對缺乏，媒體資料仍具有擴充空間。但相較於市面上其他新聞分析網站，其所分析情緒、功能更加豐富多元。以島民衛星為例，其主要針對輿論與媒體進行正面、負面及中立 3 種情緒分析，且僅以一個月內排名靠前的新聞主題進行分析，其呈現結果相對單純。再者，主要以分析網路聲量為主的網路溫度計，其利用語意情緒分析各大討論區的評論，並以正負與中立面 3 種網路聲量區分文章情緒，而非新聞標題與其內容本身，恐受到大量評論或買留言的影響，導致分析結果為農場、釣魚標題等情況。情緒選擇與資訊量之比較如圖 19 所示，其中資訊量為不須透過該網站的檢索系統，即可看到新聞標題以及新聞內容，文章篇數也比較多。

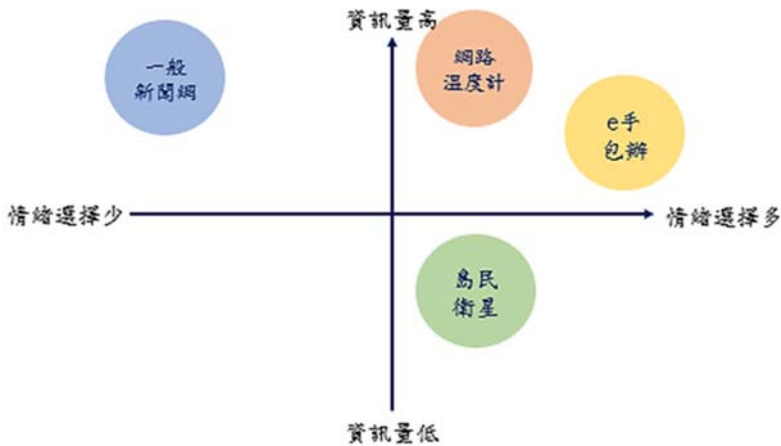


圖 19 情緒選擇與資訊量象限圖

資料來源：本研究整理

此外，先前研究顯示使用 BERT 分析中文語言能夠有更好的成果，BERT 語言模型透過對每個字進行編碼與解碼，減少了因斷詞成效的影響，使用預訓練的 BERT 語言模型更能有效的加速與提升分析任務。未來希望

能夠利用預訓練的 BERT 語言模型，來建立分析模型，以達到更好的分類結果 (Yao, Jin, Mao, & Luo, 2019)。

## 伍、結論

在無法得知讀者面對新聞的感受程度及思辨能力的情況下，顯然新聞情緒的分類重要性於現今社會中已不容小覷。為避免讀者閱讀負面新聞導致身心遭受不良影響，故本研究以情緒分析、斷詞及詞性、NRC Emotion Lexicon 等技術分析新聞標題情緒，並建立「e 手包辦新聞網」提供讀者選擇欲閱讀之新聞情緒文章，同時藉由視覺化統計圖表呈現特定時間區間之新聞情緒走向。

目前本研究僅分析單一獨立資料來源，其新聞特性或文筆撰寫上恐具有其侷限性，期許未來可擴及至其他新聞網，利用多種情緒分析，擷取相關特徵，以辨別假新聞、誘餌式標題的可能性，抑或是比對股市指數或社會案件數等外部資料。此外，亦能透過新聞標題模糊比對，將相似新聞標題整合，瞭解各家媒體對於同一新聞事件如何撰寫其標題，是否秉持公正客觀性，審查有無誇大事實之嫌疑，以減少因標題而造成的社會恐慌。

(接受日期：2021 年 11 月 15 日)

## 參考文獻

- Amazon Web Services (2021)。建立並連接到 PostgreSQL 資料庫使用 Amazon RDS。檢自 <https://aws.amazon.com/tw/getting-started/hands-on/create-connect-postgresql-db/>
- CKIP Lab 中文詞知識庫小組 (2020)。中文斷詞。檢自 <https://ckip.iis.sinica.edu.tw/project/ws>
- Chen, C., Chen, J., & Shi, C. (2018). Research on Credit Evaluation Model of Online Store Based on SnowNLP. *E3S Web of Conferences*, 53(3), 03039. doi:10.1051/e3sconf/20185303039

- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp.4171-4186). Minneapolis, Minnesota: Association for Computational Linguistics. doi:10.18653/v1/N19-1423
- Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological science, 14*(1), 14-18.  
doi: 10.1111/1467-9280.t01-1-01412
- Li, Peng-Hsuan (2020). *CkipTagger*. Retrieved from  
<https://github.com/ckiplab/ckiptagger>
- McIntyre, K. E., & Gibson, R. (2016). Positive news makes readers feel good: A “Silver-Lining” approach to negative news can attract audiences. *Southern Communication Journal, 81*(5), 304-315. doi:10.1080/1041794X.2016.1171892
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–Emotion association lexicon. *Computational Intelligence, 29*(3), 436-465.  
doi:10.1111/j.1467-8640.2012.00460.x
- Pfefferbaum, B., Newman, E., Nelson, S. D., Nitiéma, P., Pfefferbaum, R. L., & Rahman, A. (2014). Disaster media coverage and psychological outcomes: Descriptive findings in the extant research. *Current Psychiatry Reports, 16*(9), 464. doi:10.1007/s11920-014-0464-x
- Pfifferling, J.-H., & Gilley, K. (2000). Overcoming compassion fatigue. *Family practice management, 7*(4), 39-44.
- PostgreSQL (2021). *About*. Retrieved from <https://www.postgresql.org/about/>
- Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. *Proceedings of the First Instructional Conference on Machine Learning* (pp. 133-142).
- Rubio, D. (2017). *Beginning Django: Web application development and deployment with Python*. Berkeley, CA: Apress. doi: 10.1007/978-1-4842-2787-9
- Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences, 116*(38), 18888-18892. doi:10.1073/pnas.1908369116

- Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 1-19. doi:10.1155/2009/421425
- Trussler, M., & Soroka, S. (2014). Consumer demand for cynical and negative news frames. *The International Journal of Press/Politics*, 19(3), 360-379. doi:10.1177/1940161214524832
- Wikimedia Commons (2020). *Plutchik's Wheel of Emotions*. Retrieved from [https://commons.wikimedia.org/wiki/File:Plutchik\\_Dyads.svg#/media/File:Plutchik\\_Dyads.svg](https://commons.wikimedia.org/wiki/File:Plutchik_Dyads.svg#/media/File:Plutchik_Dyads.svg)
- Yao, L., Jin, Z., Mao, C., Zhang, Y., & Luo, Y. (2019). Traditional Chinese medicine clinical records classification with BERT and domain specific corpora. *Journal of the American Medical Informatics Association*, 26(12), 1632-1636. doi:10.1093/jamia/ocz164





## E-hand News

Hui-Ru Chen \*    Chung-Ming Lo \*\*

### [ Abstract ]

This research collects all newspaper titles published in the United Daily News from 2019 to 2020. Through word segmentation, part-of-speech tagging, sentiment analysis, NRC Emotion Lexicon and other technologies, supplemented by website development, it provides a platform for readers to filter various news sentiments: “E-hand News,” including 11 kinds of emotion analysis: positiveness, neutrality, negativeness, anger, anticipation, disgust, fear, joy, sadness, surprise and trust. It is expected to present and remind readers of the emotions of each news article in an objective manner.

In order to make the reading experience more personal, we have further planned the membership system, push notifications, recommended news sentiment categories that readers are interested in. In addition, we have adopted visual charts such as word clouds, line graphs and bar graphs to show sentiment analysis of newspaper titles. As a result, readers can easily grasp current popular events or the emotional evolution of news events at a glance.

---

\* Graduate Student, Graduate Institute of Library, Information and Archival Studies,  
National Chengchi University

E-mail: hbl3987@gmail.com

\*\* Associate Professor, Graduate Institute of Library, Information and Archival Studies,  
National Chengchi University

ORCID 0000-0001-5068-6002

Principal author for all correspondence E-mail: buddylo@nccu.edu.tw

## Keywords

News titles, Sentiment analysis, Emotion analysis, Title classification, News website

## [ Summary ]

### **Introduction**

“A scholar, without going out, knows the world’s affairs well” is not difficult in modern society. Along with the emergence of web media, various news platforms provide real-time and diverse news update and collect domestic and international news information to enhance the width and depth of people’s horizon and grasp global pulsation. Furthermore, due to the fierce competition in news industry and readers’ negative preference for titles, media, in order to survive, largely utilize negative news titles with disasters, bloodiness, and sensation to attract public concerns and create the click-through rate. In consideration of the prevalence of negative news, especially when COVID-19 epidemic is rampant, such atmosphere might harm readers’ physical and mental health. Although positive news would positively affect readers, major news web sites seldom precede sentiment classification of the news. Besides, private media news analysis web sites mostly stress on the analysis data that readers could not filter news sentiment from existing platforms.

For this reason, news titles of the United Daily News in 2019-2020 are collected for sentiment analysis, word segmentation and part of speech tagging, NRC Emotion Lexicon, and collaborative filtering, with the supplement of website development to provide a platform for readers filtering various news sentiments: “E-hand News”. Moreover, in order to present personal characters on the reading experience, a membership system is planned, push notification is designed, news sentiment categories in which readers are interested are recommended, and visual charts of word clouds, line graphs, and bar graphs are applied to display the title sentiment analysis result. It would benefit readers to

grasp current popular events and understand the emotional evolution of news event at a glance.

## **Material and Method**

News titles of the United Daily News in 2019-2020 are collected in this study. With SnowNLP sentiment analysis tool, deep learning is used for specifically calculating the sentiment scores of news titles, which are transformed into positive/negative or neutral categories. CKIP Tagger is further utilized for word segmentation and part of speech judgment to understand the part-of-speech composition of news titles. It would help delete meaningless words like conjunctions before calculating hot words in news titles in TF-IDF statistics and avoid using them as hot words in word clouds.

It would be biased to simply classify news titles sentiment into three categories. To avoid single sentiment, more detailed sentiment classification is required. NRC Emotion Lexicon contrast to news titles is applied to tag the emotion in news titles. NRC Emotion Lexicon contains two sentiments (positive, negative) and eight emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). Such eight emotions in NRC Emotion Lexicon are combined with the sentiments (positive, neutral, and negative) in SnowNLP sentiment analysis for sentiment analysis.

What is more, based on Bidirectional Encoder Representations from Transformers (BERT), tagged news titles training and established AI model are utilized for successive classification. News sentiment in which readers are interested are collected in this study for grouping readers. Collaborative filtering is used for predicting the news sentiment tendency of targeted readers to establish the recommended content. Besides, HTML, JavaScript, and CSS are used as the front end of “E-hand News”, PostgreSQL database is matched, and Django is eventually used for building webpages.

## **Result**

“E-hand News” (<http://210.240.245.72:8000/index/>) is used for presenting the research result in this study. Large amount of news titles in the knowledge bank of the United News are preceded sentiment analysis and the analysis results are displayed on the platform web site. The web pages covers home page, search, hot news, statistics, and about us.

## **Conclusion**

In the booming time of media, people could acquire multi-dimensional new reports through various media, such as general news media web sites or Web news aggregators, including Yahoo! News and Google News. To maintain users’ customer loyalty, exclusive new personal settings are created for users. The system design is mainly according to user interest and click situation to recommend news with specific topics; unfortunately, users’ psychological factors are ignored. Without knowing readers’ perception and speculation of news, the importance of news sentiment classification cannot be neglected in modern society. Besides, people are increasing the emphasis on sentiment in information and users are not willing to receive the sentiment of information output from mainstream media. For this reason, “E-hand News” is established in this study for readers selecting sentiment articles of news which they would like to read; meanwhile, visualized statistical charts would present the news sentiment trend in specific time intervals.

Nevertheless, there is merely an independent data source is analyzed in this study that the news characteristics and writing might be limited. It is expected to expand to other news webs and acquire relative characteristics with different sentiment analyses to distinguish false news and clickbait or to compare with external data of stock index or number of social cases.

## Romanized & Translated Reference for Original Text

- Amazon Web Services (2021)。建立並連接到 PostgreSQL 資料庫使用 Amazon RDS。檢自 <https://aws.amazon.com/tw/getting-started/hands-on/create-connect-postgresql-db/> 【Amazon Web Services (2021)。Jian li bing lian jie dao PostgreSQL zi liao ku shi yong Amazon RDS. Retrieved from <https://aws.amazon.com/tw/getting-started/hands-on/create-connect-postgresql-db/> (in Chinese)】
- CKIP Lab 中文詞知識庫小組 (2020)。中文斷詞。檢自 <https://ckip.iis.sinica.edu.tw/project/ws> 【CKIP Lab Chinese Knowledge and Information Processing (2020). *Word segmentation*. Retrieved from <https://ckip.iis.sinica.edu.tw/project/ws> (in Chinese)】
- Chen, C., Chen, J., & Shi, C. (2018). Research on Credit Evaluation Model of Online Store Based on SnowNLP. *E3S Web of Conferences*, 53(3), 03039. doi:10.1051/e3sconf/20185303039
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp.4171-4186). Minneapolis, Minnesota: Association for Computational Linguistics. doi:10.18653/v1/N19-1423
- Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological science*, 14(1), 14-18. doi: 10.1111/1467-9280.t01-1-01412
- Li, Peng-Hsuan (2020). *CkipTagger*. Retrieved from <https://github.com/ckiplab/ckiptagger>
- McIntyre, K. E., & Gibson, R. (2016). Positive news makes readers feel good: A “Silver-Lining” approach to negative news can attract audiences. *Southern Communication Journal*, 81(5), 304-315. doi:10.1080/1041794X.2016.1171892
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–Emotion association lexicon. *Computational Intelligence*, 29(3), 436-465. doi:10.1111/j.1467-8640.2012.00460.x
- Pfefferbaum, B., Newman, E., Nelson, S. D., Nitiéma, P., Pfefferbaum, R. L., &

- Rahman, A. (2014). Disaster media coverage and psychological outcomes: Descriptive findings in the extant research. *Current Psychiatry Reports*, 16(9), 464. doi:10.1007/s11920-014-0464-x
- Pfifferling, J.-H., & Gilley, K. (2000). Overcoming compassion fatigue. *Family practice management*, 7(4), 39-44.
- PostgreSQL (2021). *About*. Retrieved from <https://www.postgresql.org/about/>
- Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. *Proceedings of the First Instructional Conference on Machine Learning* (pp. 133-142).
- Rubio, D. (2017). *Beginning Django: Web application development and deployment with Python*. Berkeley, CA: Apress. doi:10.1007/978-1-4842-2787-9
- Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences*, 116(38), 18888-18892. doi:10.1073/pnas.1908369116
- Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 1-19. doi:10.1155/2009/421425
- Trussler, M., & Soroka, S. (2014). Consumer demand for cynical and negative news frames. *The International Journal of Press/Politics*, 19(3), 360-379. doi:10.1177/1940161214524832
- Wikimedia Commons (2020). *Plutchik's Wheel of Emotions*. Retrieved from [https://commons.wikimedia.org/wiki/File:Plutchik\\_Dyads.svg#/media/File:Plutchik\\_Dyads.svg](https://commons.wikimedia.org/wiki/File:Plutchik_Dyads.svg#/media/File:Plutchik_Dyads.svg)
- Yao, L., Jin, Z., Mao, C., Zhang, Y., & Luo, Y. (2019). Traditional Chinese medicine clinical records classification with BERT and domain specific corpora. *Journal of the American Medical Informatics Association*, 26(12), 1632-1636. doi:10.1093/jamia/ocz164