1. 研究計畫內容（以 10 頁為限）：
2. 摘要

有鑑於網路媒體已成為人們生活中不可分割的一部份，人們往往會利用各種類型之數據分析網站來獲取相關資訊，根據我們的觀察，從數據分析網站中只能取得統計數據結果，且統計結果不準確，也無法得知詳細內容，資料來源可靠度也無保障，若能從較多民眾使用之論壇中取得資料，並對資料進行分析及顯示來源，則可以大幅提高資料可靠度，因此本計畫擬定設計一網站，透過爬蟲技術取得各大論壇的文章及評論等，利用BERT(Bidirectional Encoder Representations from Transformer)自然語言模型對文本進行分類，分為中立、負面、正面之三大面向，並以視覺化圖表顯示結果，且為了可以在各種設備上進行使用，本次專案以 RWD(Responsive Web Design)響應式網頁的方式進行開發。本網頁將提供聲量分析、評論觀看等功能供使用者操作，本計畫預計能達成提供聲量資訊給意見領袖改善自身規劃，以及提供企業資訊以幫助決策之成效。

1. 研究動機與研究問題
2. 研究動機

隨著FinTech的發展愈漸成熟，分析商業活動背後數據以及發掘潛在商機，已經成為企業在市場中競爭、生存的一大關鍵(Sivarajah et al., 2017)。有涉足電子商務之眾多企業無論是公關、行銷等方面，都需要掌握目前聲量以及聲勢風向，用以發掘潛在市場和提高競爭力，而品牌形象、網路口碑、來源的可信度，都是影響消費者購買決定之重要因素(López & Sicilia, 2014)。

對於人類而言，資料量過於龐大時，即便收集到大量資料，也難以在短期時間內吸收、消化成為有效資訊，這些片段資料也容易造成決策者進行過於武斷的決策，影響決策結果，導致金錢、時間成本提升，將錢投入行銷黑洞之中(van der Borgh et al., 2020)，因此如何快速將龐大數量資料進行統整、分析並加以分類在決策當中扮演極其重要的角色。

綜上述原因，電商相關企業求拓展市場、提高競爭力、創造品牌口碑，必須培養人才，而培養專業數據分析人才之長期人力以及時間成本極為昂貴，勢必有求於專業之數據分析企業，因此，本專案計畫期望利用爬蟲技術抓取網路上各大論壇之資料，並以龐大數量之資料集結合深度學習，配合BERT模型進行分類，最後以圖形化之視覺型圖表結合資料來源相關資訊進行呈現，用以增加資料來源之可信度，可使涉足電商相關業務之企業能夠掌握更全面、無死角的資訊，以利於決策，此外，網紅也能透過本專案系統了解自身評價，進行未來規劃。

1. 研究問題

經過我們對市面上數據分析以及資訊整合網站之觀察，我們發現現有網站中存在著兩項問題(1)缺乏可靠的資料來源(2)缺乏實際情緒面向

* 在網路溫度計網站[[1]](#footnote-1)中，有進行情緒面向之判斷，且有進行各種人物之主題分類（如
* Youtuber、立法委員、人氣歌手等等），針對這些人物利用AI人工智慧進行情緒面向之分析，然而此網站中針對這些社群關注之評論與文章面向比率，中立的則數占比非常高，平均超過八成，此外，這些評論以及文章並無標示資料來源，無法得知這些資料是否可信。我們認為透過此專案計畫之自然語言模型BERT進行模型的訓練，可以達到情緒面向更高的準確率，並且在視覺化圖表呈現時一併顯示資料來源以提高可信度。
* 在KOL Rader網站[[2]](#footnote-2)中，儘管有詳細的篩選功能，也有追蹤目標人物的互動、觀看、粉絲成長率，然而，負面行銷也會造成上述各指標之成長(Stevens, 2009)，而當目標人物之各種指標為受負面行銷而增加，會造成使用者之決策錯誤的情況發生，當未考慮到是否為負面新聞、行為所造成的曝光時，若電商企業只因上述三類指標而選擇該目標人物，可能導致已建立之品牌口碑遭受打擊。數據平台作為協助決策的重要角色，必須耗費大量時間抓取與人物相關之大量評論文本，我們認為僅針對這些文本進行關鍵字的探索和分類並不足以輔助決策，本研究計畫將此人物相關之大量文本進行情緒分析，以加強也輔助決策之能力，幫助企業挑選KOL時，能夠降低更多的風險以及不確定性。

1. 研究目的

根據研究動機與問題，本計畫的目的如下所示：

* 整合網路評論並且加以分析，同時進行比較以及分類，隨時隨地皆可進行網路聲量以及關鍵字分析等等之觀察，並且透過模型改善決策者因片面資訊而造成之行銷方面錯誤決策，降低消費者之消費風險與不確定性。

本系統之特色茲說明如下

1. 網路與文章之蒐集自動化，減少資料蒐集耗費之成本

由於民眾熟悉之社群媒體平台皆不一，民眾之文章以及評論會分散於各種社群媒體平台，若透過人工進行資料的蒐集，需要耗費大量時間以及人力到各個社群媒體進行蒐集，本專案計畫透過網路爬蟲技術，並定期進行各社群媒體平台之文章、評論爬蟲，作為本系統以及模型更新之依據。

1. 情緒面向之分析，簡明之視覺化圖表

藉由網路爬蟲之龐大數量資料集結合深度學習，配合自然語言模型BERT進行分析判斷情緒面向，使用者能夠了解目前網路聲量以及風向狀況，以便改善決策。此外，本專案系統在以模型分類情緒面向之同時，將統計資訊以視覺化圖表方式作為傳達方式，提供簡單明瞭之視覺化圖表。

1. 資料來源透明化，供決策者即時掌握資訊

提供判斷情緒面向之文本來源，提高可信任性，使用者能自行檢視資料來源，本系統也

提供情緒面向之詳細評論查看功能，系統將詳細評論資訊進行分類，決策者能透過查看詳細評論來達到即時掌握資訊之功能。

1. 文獻探討與回顧
2. KOL形象

KOL指的是「Key Opinion Leader」，中文譯為「關鍵意見領袖」，代表在特定領域、議題、群眾中，擁有發言權以及強大影響力的人，無論是政治人物、名人明星、網紅等等，廣義來說都屬於KOL的範疇，而KOL涵蓋各種領域之人物，因此能與各種喜好、領域不同之民眾產生高度互動關係，可見KOL的影響力非常強大(Egger et al., 2016)。

而隨著現在自媒體的蓬勃發展，KOL除了受到族群的認同及尊重外，也能夠透過代言、合作等方式幫助企業、廠商在品牌形象以及行銷效果上帶來重要的提升，所以KOL對於自身的形象管理也顯得重要，企業透過名人來做代言，為商品背書的策略能夠創造出新品牌的價值，並延續產品的生命週期，且增進公司的收益(Ferrand & Pages, 1999)，並且當可信度高的代言人推薦，消費者會對該品牌的品質有信心(林陽助 et al., 2007)，同時代言人的真誠形象會正向影響消費者對於產品之態度(Dittmore et al., 2012)，透過上述的研究也能證實KOL的形象好壞都會對於企業或是消費者產生一定程度的影響。

1. 網路聲量

網路聲量是用來衡量一人物、議題、商品等在網路上的曝光度、討論度，是一種新興指標，利用文字探勘技術來計算特定關鍵字出現在網路文章、新聞引用、網友評論等等的頻率，並以人工智慧來協助分析社群網路大數據後所得出之產物。隨著手機、電腦的普及以及網路發達，在網路撰寫評論、發表意見已經習以為常，根據DATAREPORTAL的調查[[3]](#footnote-3)顯示全球的網路用戶每天有40%以上的時間都在使用網路，並且有3成以上的時間用於社群媒體，可得知網路和大眾的密切程度。

網路聲量正能夠代表KOL在現今大眾眼裡的形象優劣，在有關網路聲量的研究中發現，網路聲量確實能影響選舉，促使了投票率的提升，代表更多人透過社群媒體做政治參與的行為(葉易修 et al., 2019)，而這也表示網路聲量對於政治人物是有一定的重要價值，無法輕易忽視，參選人的品牌形象，對於社群媒體的討論區，如果無法獲得大多數使用者的喜愛，聲量也無法提升；另一方面，在補習教育產業中，老師品牌的網路聲量，也成為了影響營收的重要關鍵指標(吳淑如 et al., 2017)。

根據上述研究結果可以發現，網路聲量在現今的世代中，成為了大眾衡量一項品牌的參考指標，網路上的評論、文章等等對於KOL的網路聲量高低影響巨大，同樣的KOL的網路聲量高低也會連帶影響一個企業、廠商品牌的獲益和形象。

1. 網路論壇評論

網路論壇評論、相關文章等等都是能夠累積網路聲量的一些元素，而正向、負向、中立的網路論壇評論，正是影響網路聲量高低的一項關鍵，回顧先前針對網路論壇評論進行分析的文

獻(Zheng et al., 2020) ，其透過大量蒐集在線教育平台網路評論，利用BERT加上BiGRU(Bidirectional Gated Recurrent Unit)模型來設計FSS(Feedback System Structure)系統，

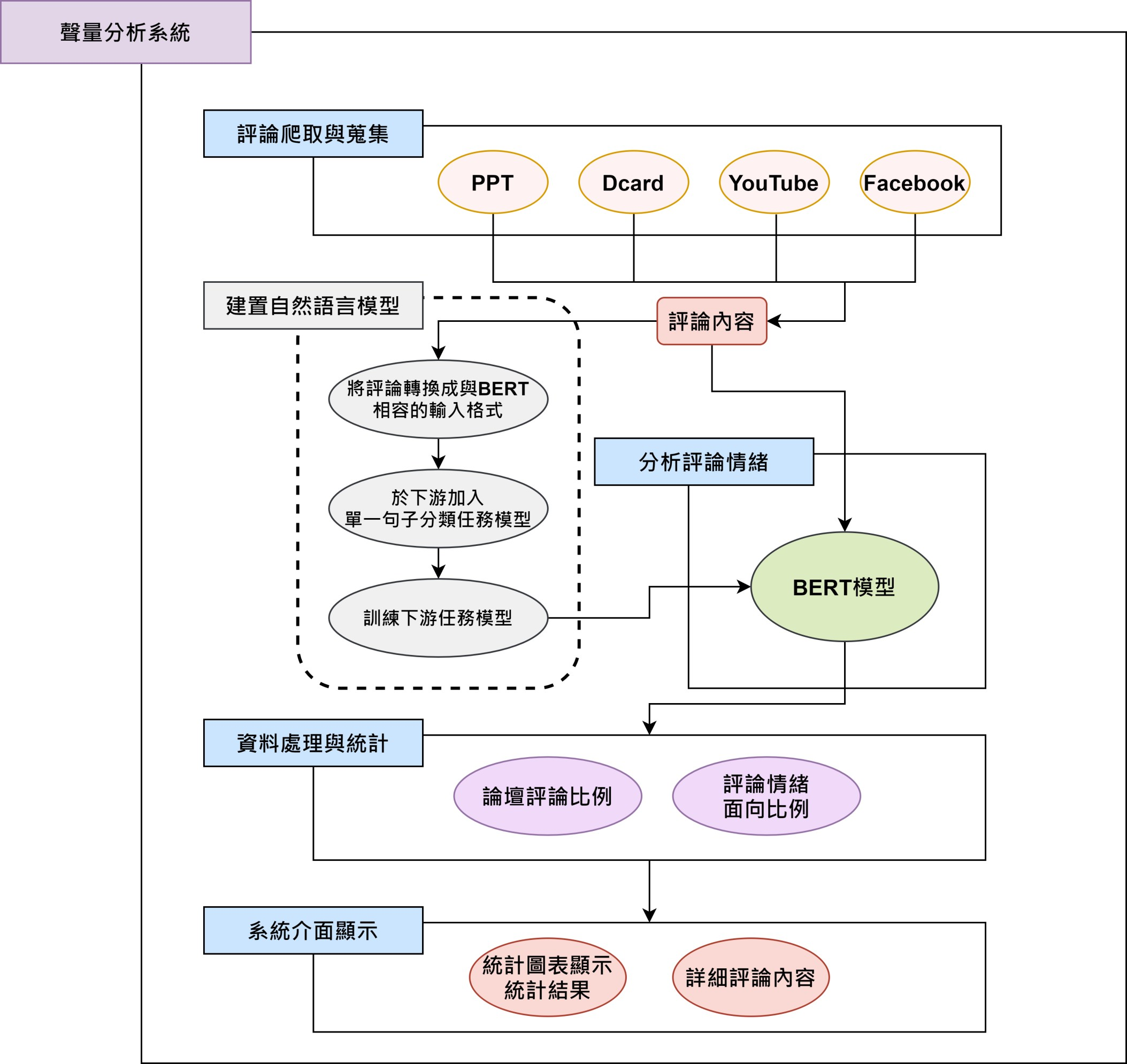
進行文本情緒分析，針對教育平台的評論去進行評論面向的分類，實驗結果指出相對於先前傳統機器學習方式，利用BERT和BiGRU模型的預測效果優異，準確率達到98.82%，提升

了21.54%，而另一項研究指出，利用卷積神經網路(Convolutional Neural Network, CNN)、長短期記憶神經網絡(Long Short-Term Memory, LSTM) 以及BERT來針對線上藥物評論進行分析，實驗結果說明使用BERT模型可獲得的最好結果(Colón-Ruiz & Segura-Bedmar, 2020)，上述研究結果證實 BERT 於現階段網路論壇評論分析之結果，預計都能達到不錯的效果。

1. 現有問題與預計解決方式

從文獻回顧中，我們得知先前已有一些針對網路論壇評論分析的類似研究，然而我們發現，針對現在越來越具有影響力的KOL其網路聲量分析的研究，除了網路溫度計、KOL Rader外，鮮少有人進行實作，而這兩個網站，在上述研究問題裡有提到，還是存在著一些缺點，所以本論文預計藉由BERT模型，來分析網路論壇上針對各個KOL的評論、文章等等，藉此解決上述研究問題，希望能夠藉由網路聲量的分析，來幫助這些意見領袖能夠改善自身規劃，或是讓企業、廠商能夠藉由這些KOL聲量分析的輔助，針對行銷策略進行調整。

1. 研究方法及步驟
2. 系統建置流程圖



圖表 1- 系統建置流程圖

本研究計畫的研究方法及步驟包括：(1)市場及使用者需求分析(2)計畫可行性分析(3)系統架構分析與設計(4)系統功能設計。

1. 市場及使用者需求分析

本計畫設計之聲量系統，從各大社群媒體平台、論壇，如PTT, Dcard, FB等等的貼文與留言中，蒐集並利用AI進行中文語意分析，分類為正面評論、負面評論、中立評論。使用者能觀看正面評論、負面評論、中立評論比率，透過這些比率可得名人、政治人物等等之聲

量，利用這些資訊來提供KOL改善自身規劃。此外，使用者除了觀看比率外，也可以從這些評論，往下篩選在某議題的聲量，進一步擬定改善或是經營計畫。此外，使用者也能夠觀看各評論的詳細內容，可以透過這些內容更細節的去理解聲量走勢的原因。

1. 計畫可行性分析
2. 技術可行性

* 資料擷取：關於資料擷取之部份，我們預計會使用Python裡面之套件如Selenium[[4]](#footnote-4)、Request[[5]](#footnote-5)等套件來進行網路上文章、留言的擷取，以此來達成本次專案擷取文章所需的技術。
* 關於AI判定部分：我們預計使用Google提出用以處理自然語言處理的BERT技術，運用BERT訓練我們所要的模型，得到訓練完的成果後，其可把我們從網路上取得的文章、留言進行分類判讀，將其分類為正面、反面和中立，以此來達到本次專案AI判定所需之技術。
* 關於資料存取之部分：我們預計使用PostgreSQL建置資料庫管理系統來存取和會員相關之資料，將會員資料存入資料庫中，利用PHP程式語言連結資料庫，以此來達成本次專案存取資料之所需技術。

總和以上觀點可得知，本次專案之技術對於現今的科技發展狀況而言，屬於可完成之範疇，故技術可行性高。

1. 系統架構分析與設計

本研究計畫所開發之聲量分析系統為針對特定目標人物進行相關網路評論分析，此系統主要可以分為以下三項範疇：(1)評論資料爬取與蒐集：利用Python提供的Selenium自動化框架來進行網路各大社群論壇上文章、評論的擷取，將針對特定人物蒐集到的評論資料整理為結構化資料並存入資料庫中；(2)建置與訓練語意分析之模型：此研究將會利用Pytorch、Tensorflow、Keras的BERT技術，運用BERT技術訓練我們所需要的模型，期可將從網路上取得之文章、評論進行情感分類判斷，達成此研究所需之AI判定技術；(3) 系統介面與功能之開發：本系統會以Codeigniter此PHP Framework同時結合前端語言來完成系統之呈現，鑒於其免費、可擴展且快速的MVC架構，同時內建簡潔、輕量的函式庫，因此適合本系統此種類型的小型應用。利用此架構開發出會員資料管理、人物之網路評論分析、人物相關議題之進階分析等功能。

1. 評論資料爬取與蒐集

本系統主要為爬取各大社群論壇的評論資料，此類型之網站常使用JavaScript/AJAX等動態載入資料的技術以及含有搜尋、登入等機制的網頁，因此選用Selenium作為本系統之評論爬蟲框架。本計畫主要爬取知名社群論壇，包括：Facebook、Dcard、PTT。針對社群論壇中，標題或文章在單位時間內含有特定人物之關鍵字的相關留言進行爬取，以固定一周作為時間單位存入資料庫，以便於日後進行統計及評論分析。

1. 建置與訓練語意分析模型

為了提高語意分析模型的準確及正確性，以改良研究問題所述網站之情緒分析功能，必須將不符合該目標人物以及無意義之評論排除，因此利用模型判斷標題與評論之關聯性為第一個任務，而第二個任務便是將目標評論進行情緒面向分析，而將上述無意義之留言排除，才能確保第二項任務之準確以及正確性。以PTT文章[[6]](#footnote-6)，並以館長為目標人物作舉例：

標題：館長是真的差還是被黑

評論1 ：館長是最頂的（與標題相關－正向評論）

評論2 ：真的差才有人要黑（與標題相關－負向評論）

評論3 ：b17去哪了（與標題無關）

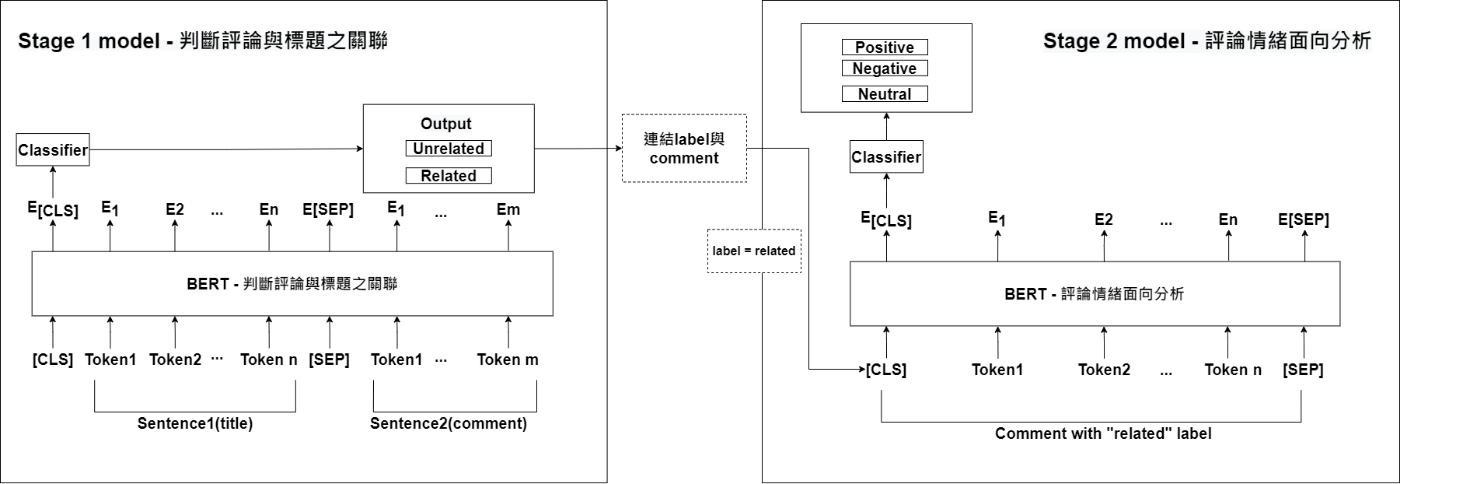
三則評論都會輸入至判斷評論與標題關聯之模型，進行判斷後，評論1、2皆與標題相關，而評論3則不相關，因此只將評論1、2會輸入至評論情緒面向分析之模型進行情緒面向分析，而評論1為正向、評論2為負向。

此兩項下游任務將會使用BERT fine-tunning完成，採用監督式學習的方式，下游任務部分，本計劃選擇使用分類模型BertForSequencClassification，以bert-base-chinese版本進行訓練，在BERT模型上層加入classifier進行句子分類，以下為欲達成目標之兩項下游任務。

1. 判斷標題與評論是否具關聯性
2. 判斷與標題相關評論之情緒面向

在資料集的部分，只採用長度在512字數內之資料，將爬取之文章分為標題以及評論，並將其轉為tokens, segments, masks tensors，作為第一部份之模型輸入。我們預計利用K-Fold Cross-Validation的方式進行處理，首先將資料拆分為9:1的比例，由90%的資料作為訓練集，10%做為測試集，將上述占9成之訓練集資料切分為5等分，每次訓練時將每等分資料分為訓練集和驗證集，其兩者比例為4:1，而每次訓練的驗證集資料皆不重複，總共訓練五次，最後將這五次的loss加總平均以作為最後的結果。

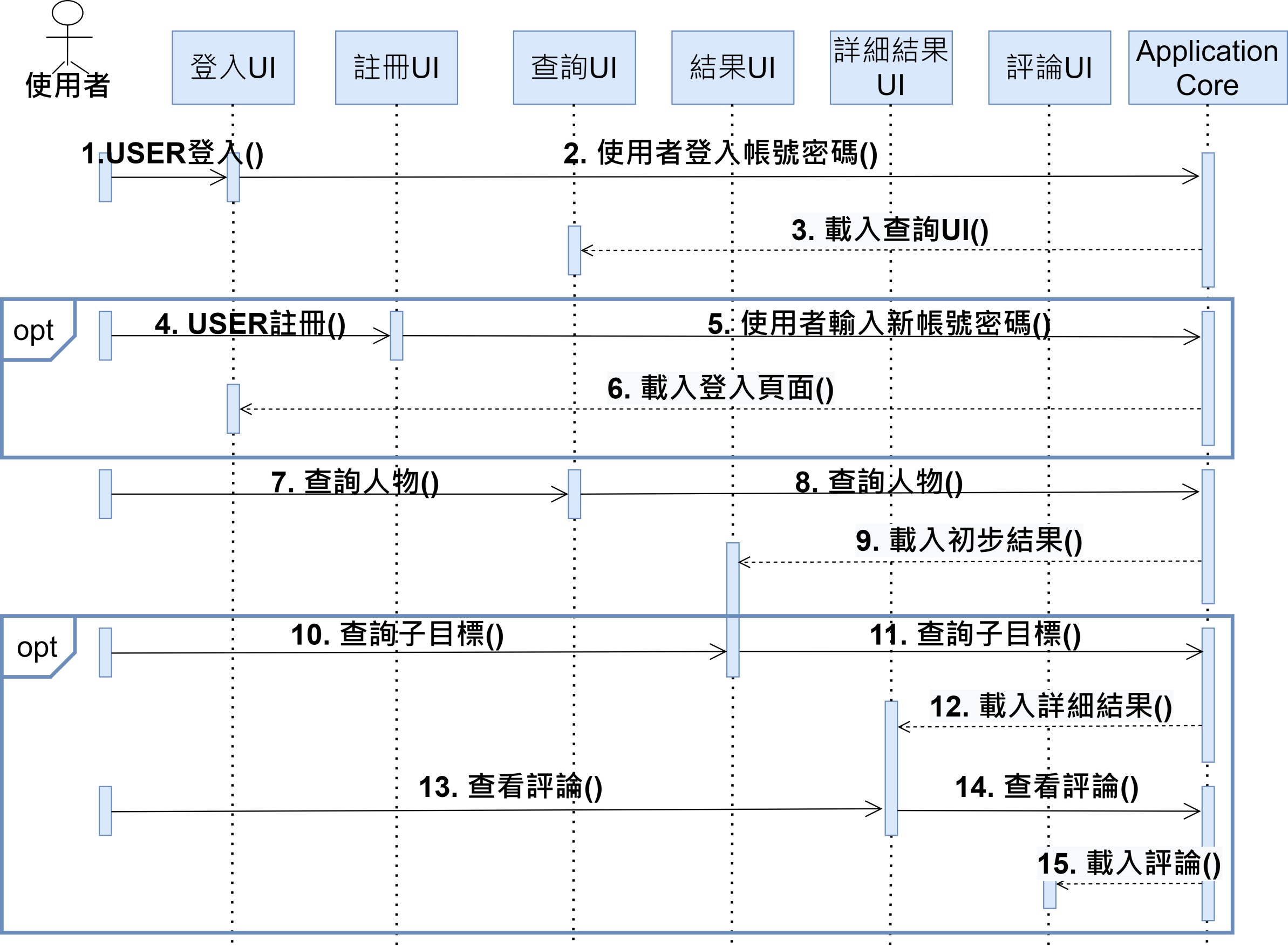
依下圖表2，將標題Token、評論Token與CLS, SEP標籤輸入至標題與評論是否具關聯性的BERT模型，將輸出之CLS Embedding通過Classifier區分為related, unrelated label，再將related label和對應評論做結合，僅將label為related之評論輸入至判斷與標題相關評論之情緒面向之BERT模型，而此模型輸出之CLS Embedding會通過Classifier區分三大情緒面向（正面、負面、中立）。



圖表 2- 模型架構圖

1. 系統介面與功能之開發

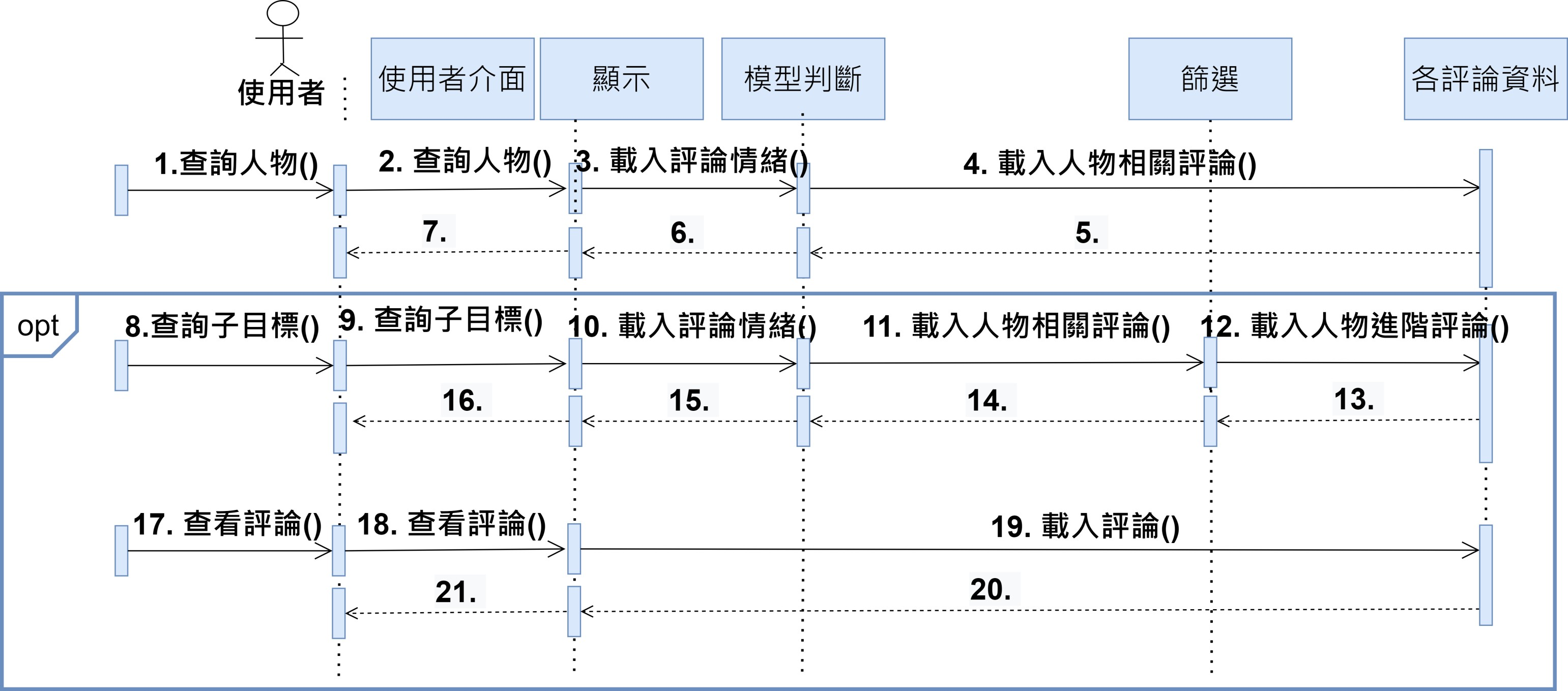
此系統之系統介面循序圖如下圖表3所示，將系統介面分為六個，分別為登入、註冊、查詢、結果、詳細結果、評論介面。使用者一開始會進入登入介面，登入成功後會載入查詢介面。此外，使用者可以選擇進入註冊介面，當使用者註冊成功則進入登入頁面讓使用者登入。登入後，使用者可以於查詢介面輸入想要查詢的人物，系統會根據查詢的人物回傳初步的結果，其包含此人物之正面、中立、負面評論的比率，除此之外，使用者可選擇進行人物之相關議題分析之查看。



圖表 3- 介面循序圖

而此系統之應用程式核心循序圖如下圖表4所示，圖中有三個控制物件，分別為顯示、模型判斷以及篩選，還有一實體物件為各評論資料。當使用者想要查詢某人物之正向、中立、

負面評論時，系統會先回傳此人物之相關評論資料，再透過模型去判斷評論的面向屬性。此外，使用者可選擇進行人物之相關議題分析之查看。系統將會從各評論的資料透過篩選，找出符合相關議題之資料，最後利用模型判斷出評論之情緒。



圖表 4- 應用程式核心循序圖

1. 系統功能設計

依下圖表5區分為下列功能：

1. 會員資料管理

* 會員忘記密碼功能

當會員進行帳號登入時，若忘記原本的帳號密碼時，會傳送驗證碼至先前註冊帳戶時輸

入的電子郵件，驗證成功後，使用者可以輸入新的密碼，再次驗證後可以重新設定該會員的密碼。

* 歷史紀錄功能

會員可以對之前查看過之歷史紀錄進行查看，可以觀看到先前所查詢目標之評論比例，也能查看先前進行進階查詢後的結果。

* 我的最愛功能

會員可以對想持續關注的目標進行追蹤，將目標加入至最愛名單中，之後可以直接查看最愛清單。

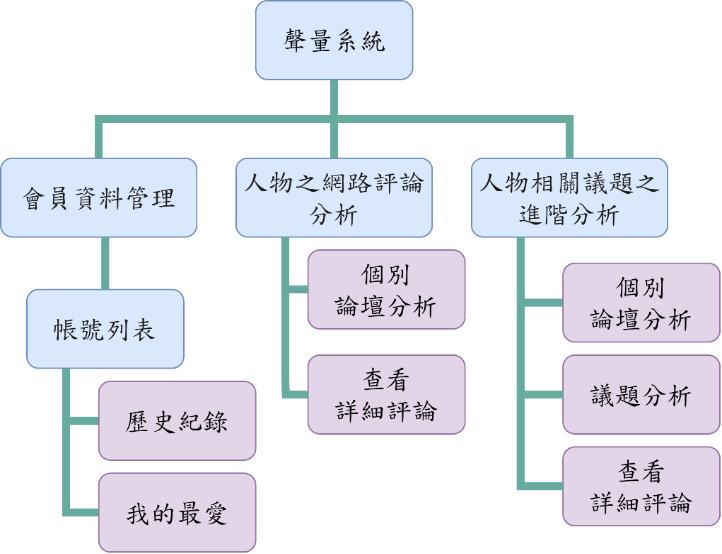
1. 人物之網路評論分析

本系統會對爬取的資料進行分析，利用BERT技術訓練人工智慧模型進行分析，對留言、評論進行情緒面向分析，最後將這些統計的結果以圓餅圖的方式呈現於網頁上，而使用者可以查看個別論壇之數據分析以及詳細評論內容。

1. 人物相關議題之進階分析

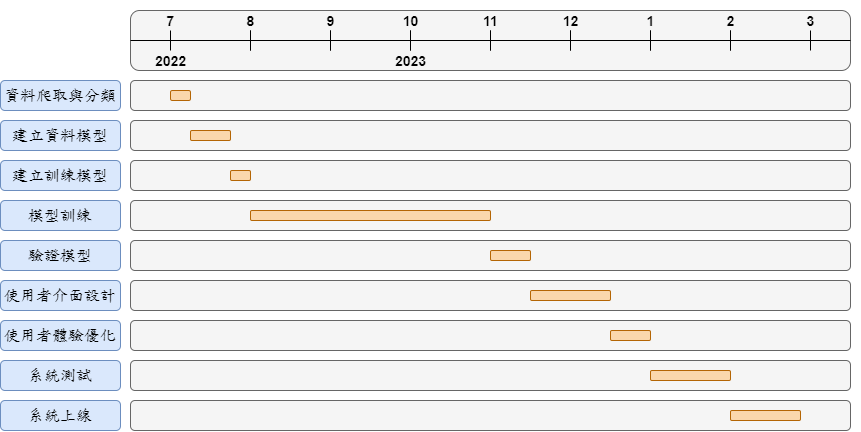
使用者可以針對原本的查詢人物之相關議題進行分析，擷取此人物之關鍵字評論，同樣利用訓練好的BERT模型對符合條件的目標進行語意的情緒分析，將這些評論、留言分類為正負、中立三面向，並且依照不同的論壇來源進行分類，將統計的結果，並以圓餅圖呈現總

評論的三面向比例，以長條圖顯示各大論壇的評論比例



圖表 5- 系統架構圖

本系統時程規劃如下圖表6：



圖表 6- 甘特圖

1. 預期結果

本計畫預計開發出一個能讓使用者針對特定人物、相關議題進行網路聲量查詢的整合性網頁平台。該網頁整合平台預期結果和功能架構如表格1所示。

˙ 讓企業能夠透過聲量查詢，來當作營運指導分針的輔助參考資料。

˙ 讓使用者能藉由系統所呈現之聲量更客觀了解事件樣貌。

˙ 訓練模型進行文本情感分析以節省人力成本。

|  |  |
| --- | --- |
| 選項 | 主要功能 |
| 登入頁面 | 1. 顯示登入資訊 2. 輸入登入資料 3. 檢測登入資料 |
| 註冊頁面 | 1. 顯示註冊資訊 2. 輸入註冊資料 3. 檢測註冊資料 |
| 聯絡我們頁面 | 1. 輸入意見資訊 |
| 查詢頁面 | 1. 顯示熱門搜尋人物 2. 查詢特定人物聲量 3. 搜索特定人物評論 |
| 進階查詢頁面 | 1. 顯示熱門搜尋人物 2. 查詢特定人物之相關議題聲量 3. 搜索特定人物與相關議題評論 |
| 結果頁面 | 1. 顯示特定人物聲量圓餅圖 2. 顯示各大論壇之人物聲量長條圖 3. 顯示特定人物照片 |
| 詳細結果頁面 | 1. 顯示各大論壇議題聲量長條圖 2. 顯示特定議題聲量圓餅圖 |
| 評論頁面 | 1. 顯示各大論壇3大面向詳細評論 |

表格 1

1. 參考文獻
2. Colón-Ruiz, C., & Segura-Bedmar, I. (2020). Comparing deep learning architectures for sentiment analysis on drug reviews. *Journal of Biomedical Informatics*, *110*, 103539.
3. Dittmore, S. W. (2012). *Impact of Perceived On-Field Performance on Sport Celebrity Source Credibility*.
4. Egger, C., First, W. S., Fischer, K., & Hartmann, G. (2016). *Identifying Key Opinion Leaders in Social Networks - An Approach to use Instagram Data to Rate and Identify Key Opinion Leader for a Specific Business Field*.
5. Ferrand, A., & Pages, M. (1999). Image management in sport organisations: the creation of value. *European Journal of Marketing*, *33*(3/4), 387–402.
6. López, M., & Sicilia, M. (2014). Determinants of E-WOM influence: The role of consumers’ internet experience. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(1), 28–43.
7. Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286.
8. Stevens, D. (2009). Elements of Negativity: Volume and Proportion in Exposure to Negative Advertising. *Political Behavior*, *31*(3), 429–454.
9. van der Borgh, M., Xu, J., & Sikkenk, M. (2020). Identifying, Analyzing, and Finding Solutions to the Sales Lead Black Hole: A Design Science Approach. *Industrial Marketing Management*, *88*, 136–151.
10. Zheng, J., Wang, J., Ren, Y., & Yang, Z. (2020). Chinese Sentiment Analysis of Online Education and Internet Buzzwords Based on BERT. *Journal of Physics: Conference Series*, *1631*(1).
11. 吳淑如. (2017). 品牌網路聲量應用於補習教育產業之研究.
12. 林陽助, 李光勳, & 李宜致. (2007). 廣告代言人與來源適配性對廣告效果影響之研究. *Journal of Data Analysis*, *2*(2), 17–54.
13. 葉易修. (2019). 透過社群輿情分析探討網路聲量與實際選票之關聯－以2018年高雄市長當選人韓國瑜為例.
14. 需要指導教授指導內容

本研究計畫之設計、構思、詳細功能開發與指導教授討論多次討論後，逐次修改至目前計畫型態，需指導教授指導之內容如下：

1. 討論研究計畫之架構與時間規劃，以及目標的訂定與可行性分析
2. 提供相關知識與技術協助，並教導文件之撰寫
3. BERT模型之建構與訓練
4. BERT模型實作文本情感分析任務相關技術指導
5. 提供網路爬蟲與與資料庫資料處理相關經驗與知識

1. https://dailyview.tw/ [↑](#footnote-ref-1)
2. https://www.kolradar.com/ [↑](#footnote-ref-2)
3. https://datareportal.com/reports/digital-2020-global-digital-overview [↑](#footnote-ref-3)
4. https://www.selenium.dev [↑](#footnote-ref-4)
5. https://docs.python-requests.org [↑](#footnote-ref-5)
6. https://www.ptt.cc/bbs/Gossiping/M.1631293582.A.EB8.html [↑](#footnote-ref-6)