

顧客評分回饋之生成式旅遊推薦客服機器人

Generative Tourism Recommendation Chatbot Based on Customer Feedback and Ratings

指導教授: 李國川 老師

組員: 溫芷妘 林愉庭 劉衣庭

國立聯合大學 資訊工程學系

苗栗市南勢里聯大2號

{U1024006,U1024010, U1024027}@o365.nuu.edu.tw

摘要

隨著 AI 技術快速發展，智慧客服系統成為企業數位化轉型的重要工具。針對現有客服系統在數據隱私與功能方面的限制，本研究設計了一套基於語言模型的智慧客服系統。系統使用內部資料，能避免機密外流，並透過模型生成回覆，提供即時、準確的回答與圖片展示，提升使用者體驗，還加入了評分機制讓使用者為模型的回覆打分數。

為提高回答的準確性，本研究採用 RAG 技術，實現語言模型與資料庫檢索的高效結合。研究方法涵蓋 Selenium 進行網頁資料自動化抓取、Whisper 模型進行音訊轉錄、Breeze 模型提取景點關鍵字，以及 MongoDB 進行圖片存儲與管理。系統功能包括智能問答、訊息評分、管理者操作介面及多模態資訊呈現。整體設計不僅提升了服務效率還增加了互動性。

關鍵詞：RAG、語言模型、Selenium、Whisper、Breeze

摘要

With the rapid advancement of AI technology, intelligent customer service systems have become a key tool for enterprises undergoing digital transformation. To address the limitations of existing systems in terms of data privacy and functionality, this study designed an intelligent customer service system based on language models. By

utilizing internal data, the system prevents sensitive information from leaking while generating responses through the model to provide real-time, accurate answers along with image displays, enhancing the user experience. Additionally, a rating mechanism was integrated, allowing users to evaluate the model's responses.

To improve response accuracy, this study adopted Retrieval-Augmented Generation (RAG) technology, achieving an efficient integration of language models with database retrieval. The research methodology includes automated web data scraping using Selenium, audio transcription with the Whisper model, keyword extraction with the Breeze model, and image storage and management via MongoDB. System features include intelligent Q&A, response rating, an administrator interface, and multi-modal information presentation. The comprehensive design not only enhances service efficiency but also fosters greater interactivity.

關鍵詞：RAG、Language Model、Selenium、Whisper、Breeze

一、研究背景與目標

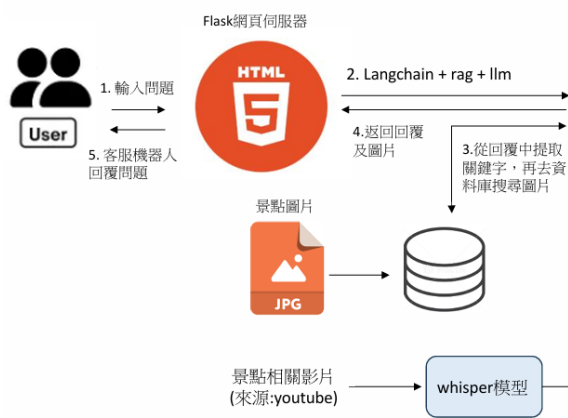
現有多數客服系統依賴人力或第三方服務，存在人手部組或數據隱私風險，特別是在處理企業內部敏感資訊時，極需要一套專屬且安全的解決方案。為此，本研究旨在設計一套基於語言模型的智慧客服系統，以保障數據安全，實現高效客戶服

務。

該系統具備內部部署特性，確保數據僅在內部網絡中處理，避免外洩。透過類似 ChatGPT 的語言模型，即時回應客戶查詢，並採用網頁架構設計以支持功能擴展，滿足多元需求。同時，系統能替代部分人工客服，處理標準化問題，降低人力成本，提升資源效率。最終，本研究致力於構建穩定、安全且功能強大的智慧客服解決方案，助力企業數位化轉型，增強競爭力。

二、系統架構及介紹

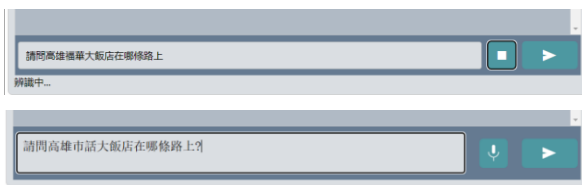
(一)系統架構圖



(二) 前端系統功能說明

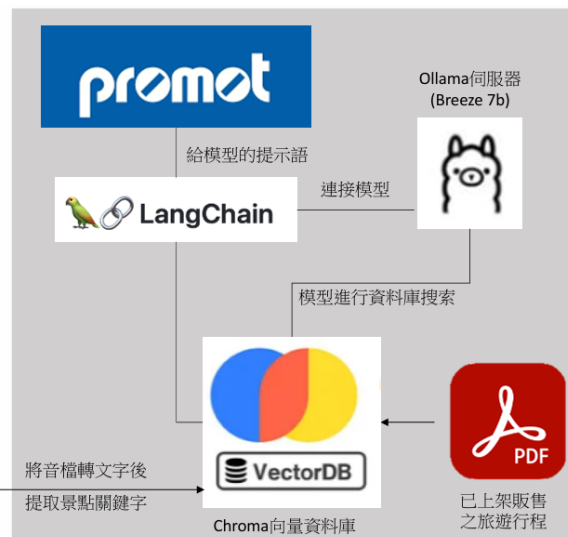
智能問答：

使用者透過文字或語音輸入問題，系統將問題傳送至後端伺服器，取得回覆。



訊息評分：

使用者可對系統生成的回覆進行評分，結果會存入資料庫以供後續的使用。



PDF 檔案管理：

管理者可檢視、刪除現有的上架行程檔案或景點相關文件，並可上傳新的檔案進行更新與管理。



行程回饋管理：

管理者可檢視使用者對行程的回饋資訊，包括個別星數的景點、搜尋內容、刪除無用回饋資料等，並能將有效回饋整合成景點資料輸出為 PDF 檔案。



影片搜尋：

管理者可以輸入 YouTube 的影片網址和自訂的檔案名稱，將關鍵字將匯出成 PDF 檔案。



(三) 後端系統功能說明

產生回覆：

透過模型產生回覆，並從回覆的文字中提取景點關鍵字並附上相應的圖片。

帳號管理：

管理者的密碼經由 bcrypt 加密儲存，確保即使資料庫外洩，也能有效降低資安風險，保障資料安全。

影片製作景點檔案：

後端會先使用 whisper 模型將影片的音檔存成文字檔，然後再使用 breeze 模型從文字檔中提取出景點關鍵字，並將關鍵字匯出成 PDF 檔案。

MongoDB 資料庫：

將圖片及管理者帳號儲存在 MongoDB 資料庫中。

三、專題技術

(一) Selenium

我們所使用的行程資料是透過 Selenium 網路爬蟲技術，從雄獅旅遊的官方網站中自動化抓取而來。整個資料抓取過程分為以下二個階段：

1. 全台灣行程網址的收集：首先，我們蒐集並提取雄獅旅遊網站中所有與台灣地區相關的行程頁面網址，確保涵蓋全台灣各縣市的旅遊行程資訊，且排除了離島的行程。
2. 深入行程頁面的資料提取：在獲得各行程的網址後，我們逐一訪問這些行程頁面，並提取以下關鍵資料，行程名稱、行程價格、行程內容、行程餐食和行程旅館。

(二) Fine-Tune

本次微調基底模型採用了聯發科的 MediaTek-Research/Breeze-7B-32k-Base-v1_0，這是一款由聯發科研發小組用 Mistral-7B Fine-Tune 出來的繁中開源大型語言模型 (LLM)。Breeze7B 系列具備卓越的繁體中文與英文處理能力，特別是在繁體中文知識表現上，與 GPT 不相上下。

為滿足特定應用需求，我們首先從網頁中收集了與行程、景點、餐廳及飯店相關的資料，經過整理後製作成專屬資料集，並將其上傳至 Hugging Face 平台以方便管理與重用。微調時，採用了 Hugging Face 社區提供的資料集格式進行處理，使用的資料集為自製資料集 y1203t/datasetinput，其中的欄位包含 instruction、input、output。

本次微調基於 unsloth 框架進行。該框架專為高效微調設計，支援多種量化技術和模型加速特性，能夠顯著降低顯存需求並提升訓練效率。此外，我們使用了 Hugging Face 提供的 SFTTrainer 工具簡化

微調流程。

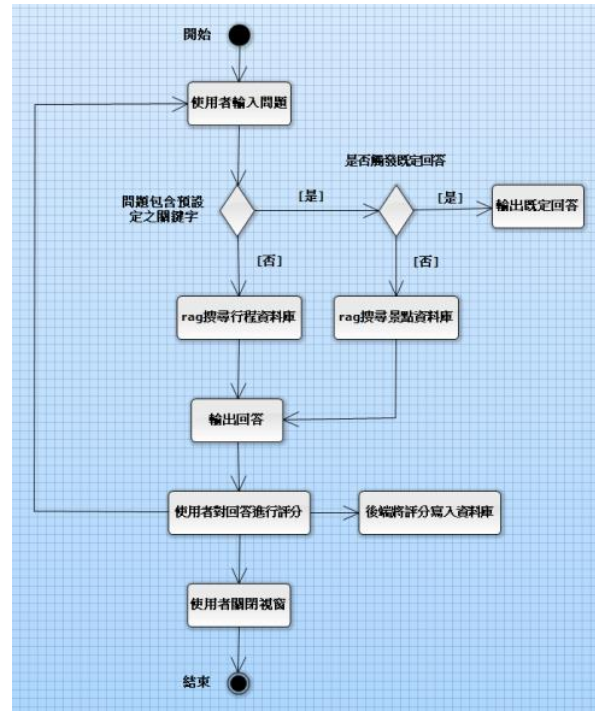
微調設置中，LoRA 的配置使用 rank 為 16，dropout 設置為 0；在訓練參數方面，最大序列長度設置為 2048，每次訓練的批次大小為 2，並採用累積梯度的方式，步數設定為 4，優化器選用的是 adamw_8bit。最後，我們將模型儲存於本地資料夾，並可透過 Hugging Face 提供的量化格式 GGUF 儲存成 q4_k_m.gguf 的格式，進一步壓縮儲存模型的空間需求，使其更容易用於各種環境。

(三) 影片處理

對 YouTube 影片處理設計了一套完整的自動化系統，通過結合多種先進技術實現從影片到結構化旅遊資訊的轉換，具體步驟如下：

1. 連接前端與接收影片網址：系統首先通過前端接收使用者提供的 YouTube 影片網址，利用後端技術實現影片下載的自動化操作。
2. 影片音檔下載與轉錄：將 YouTube 影片的音訊提取為 MP3 格式，並儲存於後端伺服器進行後續處理。並採用 Whisper 語音轉錄模型將音檔(MP3)轉錄為文本(TXT)。在文本生成後，進一步清理過程中可能出現的重複內容。
3. 文本處理與生成景點列表：使用 FastLanguageModel (MediaTek-Research/Breeze-7B-FC-v1_0) 對文本進行處理，提取出與旅遊景點相關的關鍵資訊，生成簡化的景點列表(TXT)。
4. 格式轉換與回傳：將生成的景點列表(TXT)轉換為 PDF 格式，並通過系統回傳至前端，提供使用者直接下載的功能。

(四) 訊息處理



上圖是處理客服機器人對話的程式流程圖。為了提升回答的準確度，我們搭建了 RAG (Retrieval-Augmented Generation) 系統，使 LLM 模型在回答使用者問題前，能先檢索 PDF 資料進行參考。

首先，透過 PyMuPDFLoader 文件載入器讀取 PDF 文件，接著使用 Text Splitter 將文件中的文字內容分割成多個小型區塊 (chunk)，以避免單個文件資訊量超過 LLM 的 token 限制。接下來，使用 Embedding 將這些分割後的文字區塊轉換為向量，並將處理完成的向量資料儲存至 ChromaDB。透過這套流程，不僅確保了資料檢索的效率與準確性，還進一步強化了模型的回應能力，提供更精準的答案給使用者。

我們的 PDF 文件分為「旅遊行程」和「推薦景點」兩種類型，並根據分類分別存放於對應的資料夾中。為了更好地應對不同需求，我們建立了兩個獨立的向量資料庫，分別對應這兩種類型的內容。此外，我們還精心設置了關鍵字觸發機制，能根據使用者的提問自動啟動對應的回應邏輯，確保提供準確且貼合需求的資訊服務。

(五) 資料庫使用

本系統的資料庫設計針對不同資料類型進行了專屬化處理，確保數據的安全性與管理效率。以下是針對帳號密碼、回饋以及景點圖片的處理方式：

1. 帳號密碼存儲：為保障用戶帳號的安全性，系統在儲存密碼時採用了 bcrypt 加密技術。bcrypt 的核心特點在於其內建鹽值機制（salt）和多次迭代運算，這使得密碼即使在資料庫遭到外洩後，也難以被破解。同時，由於 bcrypt 的設計特性，還能有效抵禦暴力破解攻擊，進一步提升資料安全性。

2. 回饋資料管理：使用者對行程進行評分後，系統會將回饋資訊按照所涉及的景點分別儲存。這樣的存儲方式有助於管理者快速檢索並分析使用者對不同景點的反饋，進而改進行程設計和服務內容。這些回饋數據還可以作為模型優化的重要參考，支持未來系統的進一步提升。

3. 景點圖片存儲：系統採用了 MongoDB 的 GridFS 庫來處理大文件，特別適合儲存和檢索大規模的景點圖片。首先，通過 Python 程式連線至伺服器，並使用 GridFS 建立資料庫專門存儲圖片檔案。這種方法不僅便於管理大量圖片數據，還能確保圖片的高效讀寫性能，滿足前端快速加載的需求。

四、分析與結論

比起 ChatGPT 的回覆，本系統的回答更加細緻，不僅能提供詳細的解答，還結合了評分機制與圖片展示功能，為使用者帶來更豐富且互動性更強的使用體驗。透過這些特點，本系統在功能和設計上充分考量了實際應用需求，展現了優於通用語言模型的應用價值。

本系統採用的智能問答功能，不僅能回應一般性問題，還能在解答過程中提取相關景點關鍵字，並附加對應圖片，讓使用者在獲取文字資訊的同時，能夠直觀地

了解內容細節。此外，評分機制的加入使得系統能根據使用者的回饋持續優化答案質量。這一設計不僅提升了系統的準確性，也加強了使用者參與感，形成良性循環。

與此同時，系統還運用了 RAG（Retrieval-Augmented Generation）技術，讓模型能在回答前檢索內部資料庫中相關的 PDF 文件作為參考，進一步提高了回覆的專業性與可靠性。這種方式結合了語言模型的生成能力與資料庫檢索的精準性，填補了通用語言模型無法直接處理專屬資料的空缺。此外，圖片的自動生成與整合，使得回答不僅是文字層面的補充，更是一種多模態的資訊呈現方式，進一步提升了用戶體驗。

綜合來看，本系統憑藉細緻的回答、互動性的設計與多模態呈現，在智慧客服的應用中展現了顯著的競爭優勢，為企業提供了數位化轉型的重要支持。

五、參考文獻

[1] “Selenium,”

<https://ustrustcorp.com/python-selenium/?srsltid=AfmBOopIQgnRpgolB31GWfSTv30NiDCeJOCTuOWsLjaT8w5Q94sKdYG>

[2] “Fine – Tune,”

<https://medium.com/@huangyihe/%E5%BE%AE%E8%B0%83llama-3-1-%E7%94%A8%E7%A5%9E%E5%99%A8unslloth-59e7355feeaf>

[3] “Whisper,”

<https://gsyan888.blogspot.com/2023/02/openai-whisper-ipynb.html>

[4] “RAG,”

<https://medium.com/@cch.chichieh/rag%E5%AF%A6%E4%BD%9C%E6%95%99%E5%D%B8-langchain-llama2-%E5%89%B5%E9%80%A0%E4%BD%A0%E7%9A%84%E5%80%8B%E4%BA%BA1>

lmd6838feb8c4

[5] “bcrypt,”

<https://ithelp.ithome.com.tw/articles/10196477>

[6] “景點圖片儲存,”

<https://www.delftstack.com/zh-tw/howto/mongodb/store-images-in-mongodb/>