

PSABot-程式麻瓜私人家教

專為自學者設計的資料整合聊天機器人

小組成員：邵安祺,梁晏慈,林湘羚,謝宛蓉

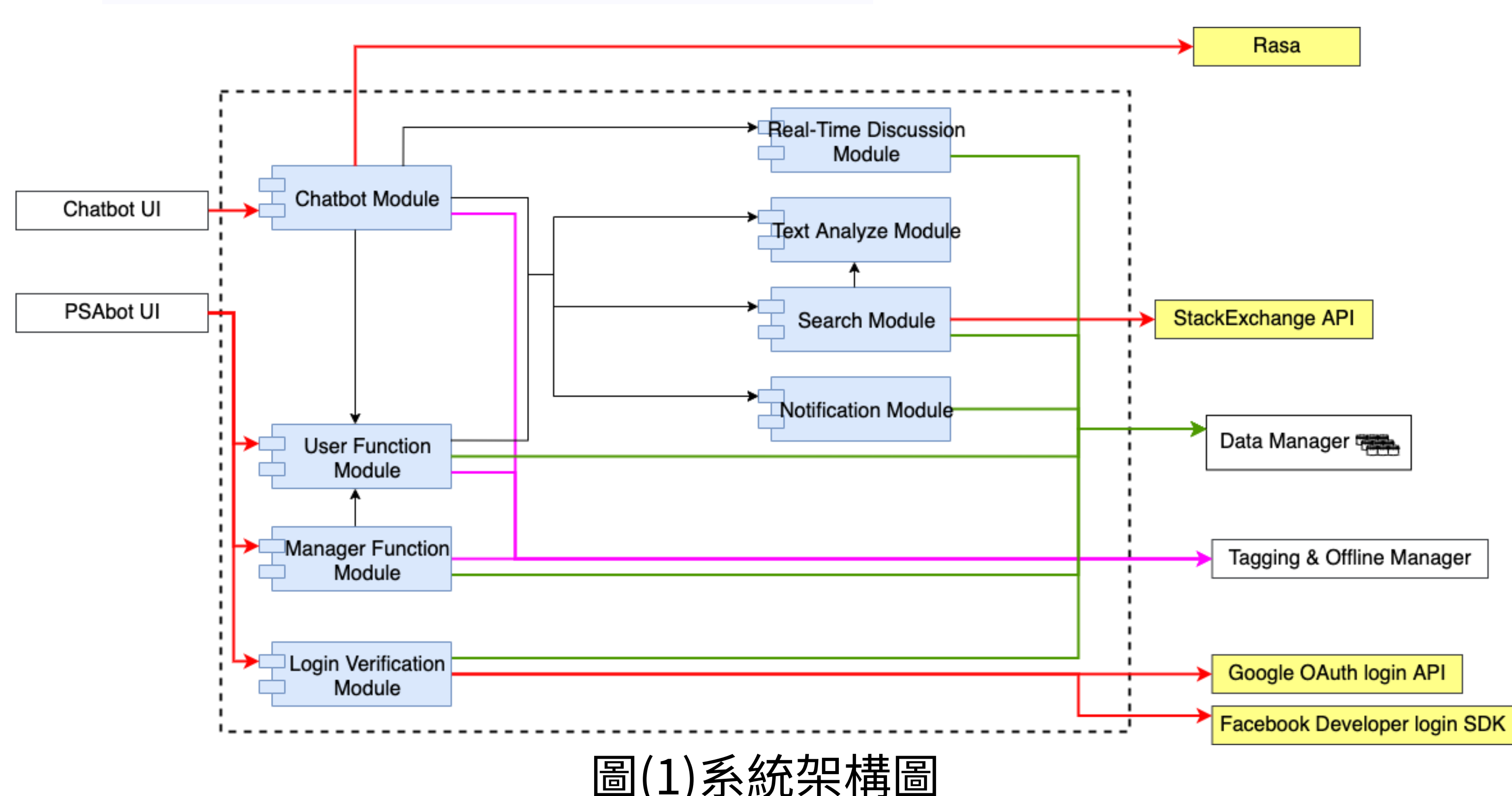
指導教授：馬尚彬教授

研究動機與目的

程式能力為現今人們重視的技能之一，程式設計學習者也急遽增加中。學習者在自學過程中，往往需耗費大量時間瀏覽網路資料以尋找問題解答，除了導致學習效率不彰之外，也容易有無法獲得正確資料之情況。

我們建置了一個結合聊天機器人與程式語言討論平台的Web應用系統稱之為PSAbot (Programming Solving Assistant Bot)。希望透過「PSAbot」能幫助學習者查詢答案以及取得重點關鍵字，減少查詢時間，並提供即時討論功能以提升學習效率。

系統架構



使用者操作介面：

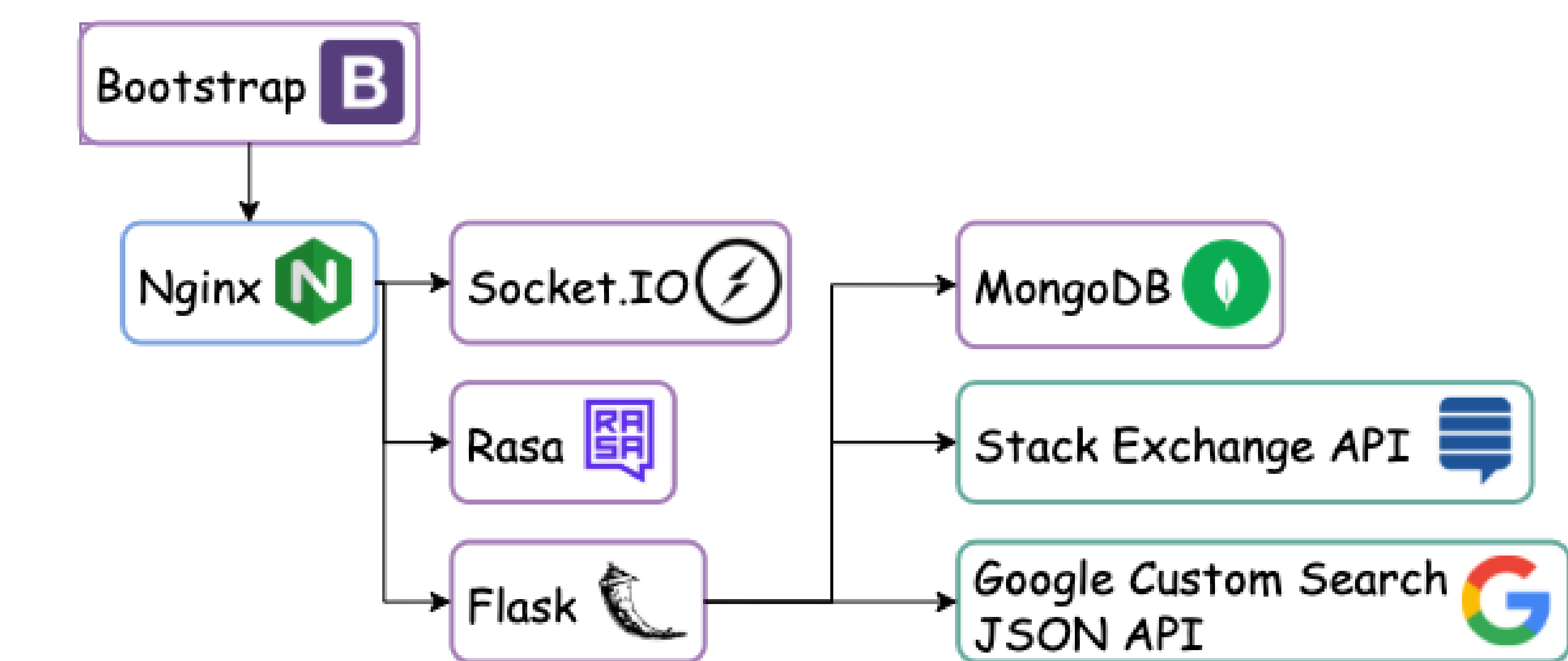
- PSAbot UI - 所有聊天室以外之使用者操作介面
- Chatbot UI - 聊天機器人之使用者操作介面。

模組元件：

- Text Analyze Module - 提供文字前置處理、區塊關聯度分析、關鍵字擷取、文章摘要等功能。
- Chatbot Module - 提供所有與Chatbot相關功能，包含Rasa Chatbot核心模組以及對Rasa發出需求之相關介面。
- Real-Time Discussion Module - 提供共同討論相關功能，推薦對應使用者、建立使用者群組以及使用FAQ Bot。
- Login Verification Module - 提供一般使用者Google、Facebook第三方登入方法。提供一組特定帳號密碼登入管理員模式。
- User Function Module - 提供一般使用者所有與討論平台相關功能。
- Manager Function Module - 提供管理員維護系統相關功能，包含調整自動更新相關參數、匯入FAQ資料、貼文管理。
- Search - 提供查詢系統內部、外部資料之功能。
- Tagging & Offline Manager - 管理本系統中定期資料擷取以及維護本系統之階層式標籤。

實作技術

- 使用Bootstrap 4進行網頁開發
- 使用Nginx進行網頁部署、反向代理與負載平衡
- 使用Flask進行後端開發並結合MongoDB資料庫
- 使用Google Custom Search JSON API及Stack Exchange API擷取Stack Overflow文章分析資料
- 使用Rasa機器學習框架建立聊天機器人
- 使用Socket.IO建立即時討論功能



圖(2)系統實作技術圖

系統特色



Chatbot搜尋協助

Bot分析使用者問題並協助使用者於Stack Overflow討論平台進行搜尋並提供重要關鍵字。



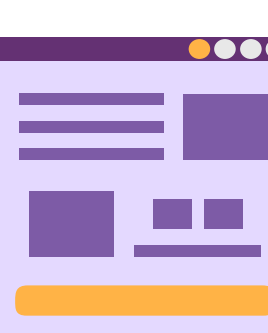
線上共同討論

透過本系統的推薦人演算法，邀請適合的使用者一同討論問題。



Stack Overflow文章整合分析

利用自然語言分析進行文章摘要以及區塊整合分析，以減少使用者瀏覽文章時間。



學習討論平台

提供討論平台基本功能及學習紀錄，標籤技能積分可用以檢視使用者學習狀況。

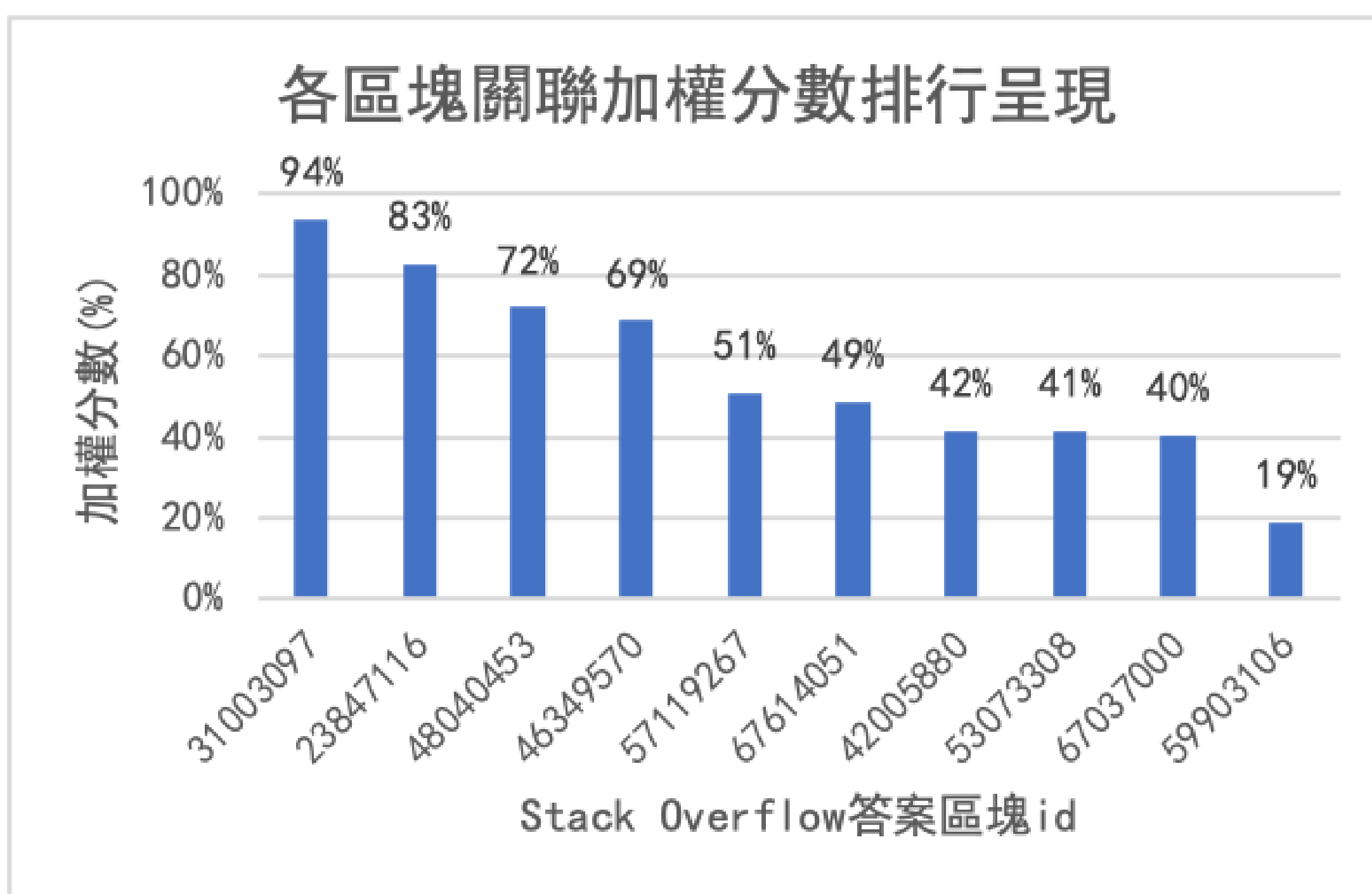
核心功能

Stack Overflow答案區塊分析

- 透過Google Custom Search JSON API以及Stack Exchange API抓取目標Stack Overflow網頁資料
- 進行文本前置分析，包含斷詞、詞性還原、POS挑選以及停用詞過濾
- 提取各文章答案區塊進行LDA主題模型訓練
- 利用LDA模型預測使用者問題並計算與各區塊之關聯加權分數
- 區塊分析範例測試案例與結果如下圖(3), (4)所示

Google Custom Search JSON API搜尋結果(前五名):
https://stackoverflow.com/questions/24722212/python-cant-find-module-in-the-same-folder
https://stackoverflow.com/questions/33862963/python-cant-find-my-module
https://stackoverflow.com/questions/38135119/python-cant-find-local-module
https://stackoverflow.com/questions/37233140/python-module-not-found
https://stackoverflow.com/questions/54314354/cannot-find-module-when-importing-from-project-in-python
目標問句: can not find module in local file

圖(3)範例問句與欲分析之文章



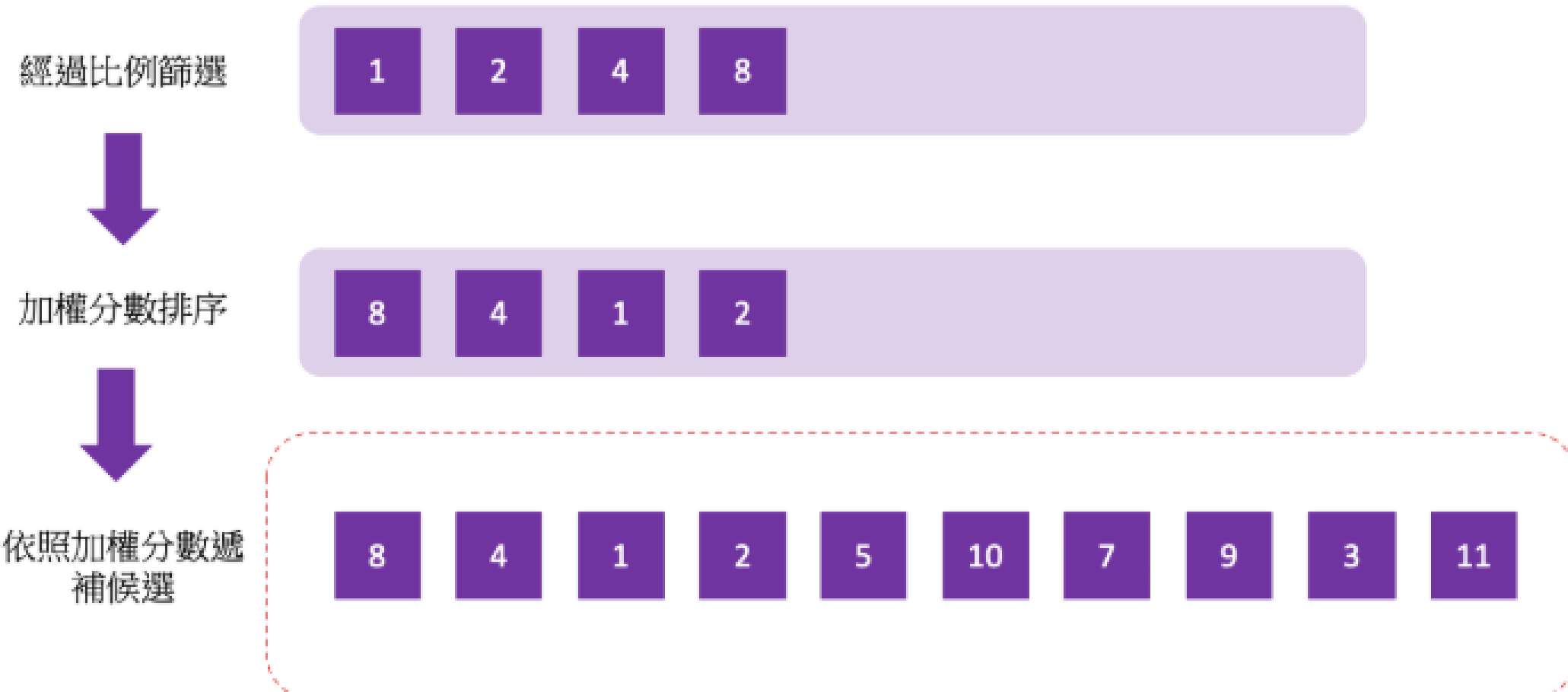
圖(4)各個區塊分析結果之分數與排行

共同討論推薦人演算法

- 篩選持有技能與問題標籤匹配數大於1/2的使用者。
- 利用標籤層級權重對技能積分加權以進行排序。
- 若推薦人選不足十位，於剩餘候選人依技能積分補足。
- 分批發出邀請，每隔一分鐘發出一批通知。
- 演算法範例以及實際操作如下圖(5),(6)所示

案例問題	如何利用scikit-learn計算兩文本間的相似度？					標籤層級	加權分數	案例			
相關標籤	Python, NLP, scikit-learn					1	1	Python			
						2	3	NLP			
						3	5	scikit-learn			
User#	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10	User11
技能標籤與權分	Python: 10 NLP: 5 scikit-learn: 3	Python: 12 NLP: 8 Jedis: 3	Python: 30 Web frameworks: 20 Flask: 10	Python: 20 NLP: 10 NLTK: 8	Python: 70 Data processing: 20 Tensorflow: 10		Python: 46 standard libraries: 30	Python: 25 NLP: 20 GUI: 23 spacy: 5 Wac: 7	Python: 33 JGU: 23 Wac: 7	Python: 49 Game development: 15 Pygame: 5	Python: 9
正配標籤加權分數	10*1+5*3+3*5=40	12*1+8*3+2*4=30	30*1+30	20*1+10*8+3=50	70*1+70		46*1+46	25*1+20*8+3=85	33*1+33	49*1+49	9*1+9
正配標籤比例	1	2/3	1/3	2/3	1/3	0	1/3	2/3	1/3	1/3	1/3

圖(5)推薦人範例設計



圖(6)演算法實際排序流程與結果

開發流程

- 本系統依循軟體工程開發模式進行系統規劃。
- 使用專案規劃軟體trello記錄與進行開發。
- 撰寫需求文件、設計文件以及測試文件以記錄專案規劃歷程。
- 根據軟體功能需求規劃77個測試案例進行需求測試，通過74項，2項未通過者為回傳時間延遲導致，1項為特殊字元差異所導致。

開發文件

PSAbot
SlidePSAbot
軟體文件PSAbot
Demo