

一言不合就排課

范佳琦 田展禾 洪峻宸 莊維軒 盧奕潔 俞懿

查詢查詢查詢查詢查詢查詢查詢查詢查詢查詢

查詢

112查詢2查詢



查詢查詢

查詢查詢

▼ 查詢



Copyright © 2024 National Chengchi University. All Rights Reserved.

查詢查詢查詢查詢 (02)29387599查詢查詢 查詢查詢67599

研究目標



研究目標

- 網頁式的聊天機器人
 - 文字輸入、語音輸入



研究目標

- 網頁式的聊天機器人
 - 文字輸入、語音輸入
- 更精準的搜尋結果



研究目標

- 網頁式的聊天機器人
 - 文字輸入、語音輸入
- 更精準的搜尋結果
- 直接產生出課表



實作流程

查詢查詢查詢查詢查詢查詢查詢查詢查詢查詢

查詢

112查詢2查詢



查詢查詢

查詢查詢

▼ 查詢



Copyright © 2024 National Chengchi University. All Rights Reserved.

查詢查詢查詢查詢 (02)29387599查詢查詢 查詢查詢67599

資料處理

歷年課程基本資料收集

資料處理

歷年課程基本資料收集

1. 取得所有課程

資料處理

歷年課程基本資料收集

1. 取得所有課程
2. 取得所有課程評價網站

資料處理

歷年課程基本資料收集

1. 取得所有課程
2. 取得所有課程評價網站
3. 擷取評價資料

資料處理

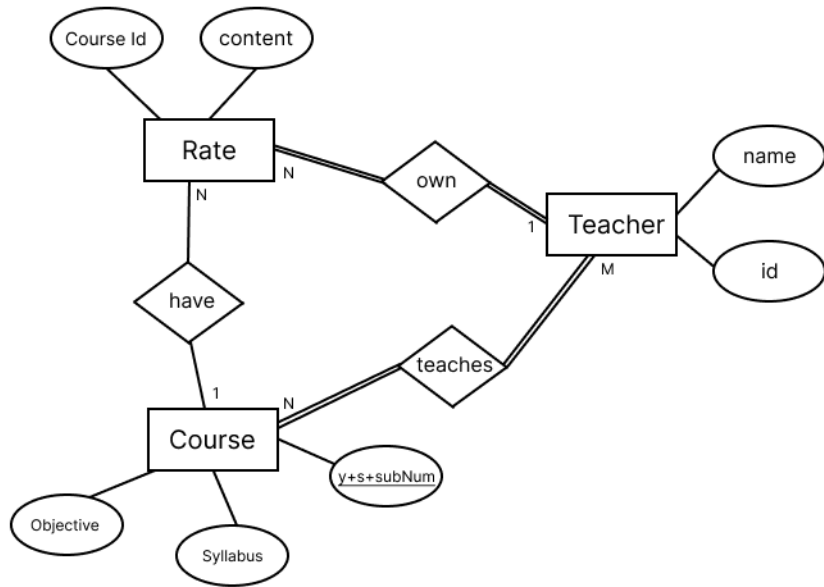
歷年課程基本資料收集

1. 取得所有課程
2. 取得所有課程評價網站
3. 擷取評價資料
4. 整理至資料庫

資料處理

歷年課程基本資料收集

1. 取得所有課程
2. 取得所有課程評價網站
3. 擷取評價資料
4. 整理至資料庫



資料處理

評價標籤化

- Zero-shot Classification
- 容易、有趣、輕鬆、高分、困難、沉悶、繁重、低分

id	s	name	time	unit	classroom	teacher	labels
1121000348011	1	管理學	三234	企管系		詳備註	無
1121000348041	1	管理學	一EFG	企管系	商館260105	林呈欣	容易,有趣,輕鬆,高分
1121000348061	1	管理學	一567	企管系	研究250301	張聰本	困難,有趣,繁重,高分
1121000348051	1	管理學	五D56	企管系	學思040205	林介勝	容易,有趣,輕鬆,高分
1121000348071	1	管理學	三D56	企管系	商館260205	陳怡安、黃國華	困難,沉悶,繁重,低分
1121000348001	1	管理學	三D56	企管系		詳備註	無
1121000348031	1	管理學	二EFG	企管系	商館260207	李璦淑	困難,有趣,繁重,高分
1121000348021	1	管理學	三D56	企管系	商館260313	傅浚映	困難,有趣,繁重,高分
1121000359051	1	統計學 (一)	四234	統計系	商館260311	薛慧敏	困難,有趣,繁重,高分
1121000359041	1	統計學 (一)	二D56	統計系	學思040103	陳怡如	困難,有趣,輕鬆,低分
1121000359061	1	統計學 (一)	四234	統計系	商館260106	黃子銘	無
1121000359011	1	統計學 (一)	二234	統計系	研究250103	楊素芬	困難,沉悶,繁重,高分
1121000359031	1	統計學 (一)	二D56	統計系	商館260311	薛慧敏	困難,有趣,繁重,高分
1121000359071	1	統計學 (一)	四D56	統計系	資訊140206	周珮婷	困難,有趣,繁重,高分
1121000359081	1	統計學 (一)	五D56	統計系	商館260106	張育瑋	容易,有趣,輕鬆,高分
1121000359201	1	統計學 (一)	二234	統計系	研究250105	張育瑋	容易,有趣,輕鬆,高分
1121000359221	1	統計學 (一)	四D56	統計系	研究250105	吳漢銘	困難,沉悶,繁重,高分
1121000360091	1	統計學 (二)	四D56	統計系	學思040103	薛慧敏	困難,有趣,繁重,高分
1121000359021	1	統計學 (一)	二234	統計系	大勇210103	劉惠美	困難,沉悶,繁重,高分
1121000359211	1	統計學 (一)	四D56	統計系	商館260306	陳立榜	困難,沉悶,繁重,高分
1121000219051	1	經濟學	四234	經濟系	大勇210201	李文傑	困難,有趣,輕鬆,高分
1121000219041	1	經濟學	五D56	經濟系	大智190305	陳心蘋	困難,沉悶,繁重,高分
1121000219111	1	經濟學	三D56	經濟系	資訊140201	彭喜樞	困難,沉悶,繁重,高分
1121000219561	1	經濟學	三D56	國貿系	炎成創新廳	郭炳伸	困難,有趣,繁重,高分

資料處理

資料向量化及建立檢索器

text2vec-base-chinese : CoSENT model

向量化資料

原始資料

資料處理

資料向量化及建立檢索器

text2vec-base-chinese : CoSENT model

向量化資料



FAISS

原始資料

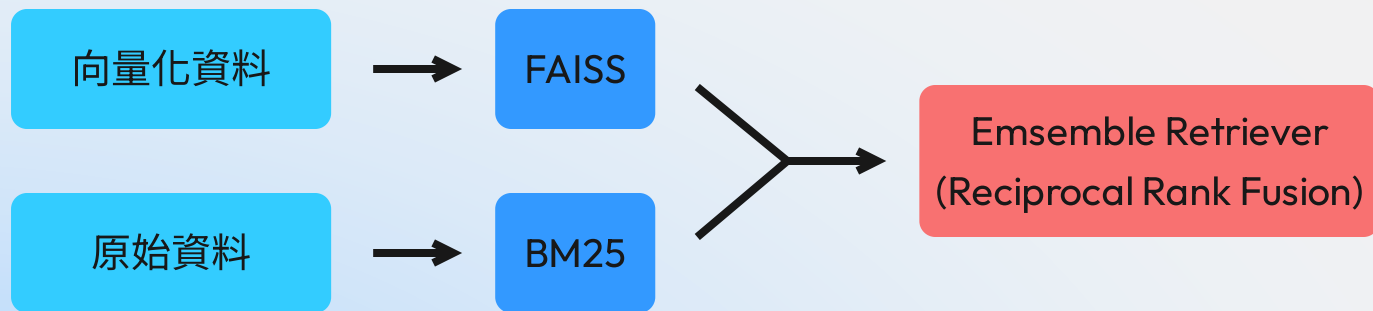


BM25

資料處理

資料向量化及建立檢索器

text2vec-base-chinese : CoSENT model



模型壓縮

Chinese-LLaMA-Alpaca-2-13B: 經中文finetune的llama2

llamacpp: Port of Facebook's LLaMA model in C/C++

3.141592	3.141592	3.141592	3.141592
----------	----------	----------	----------

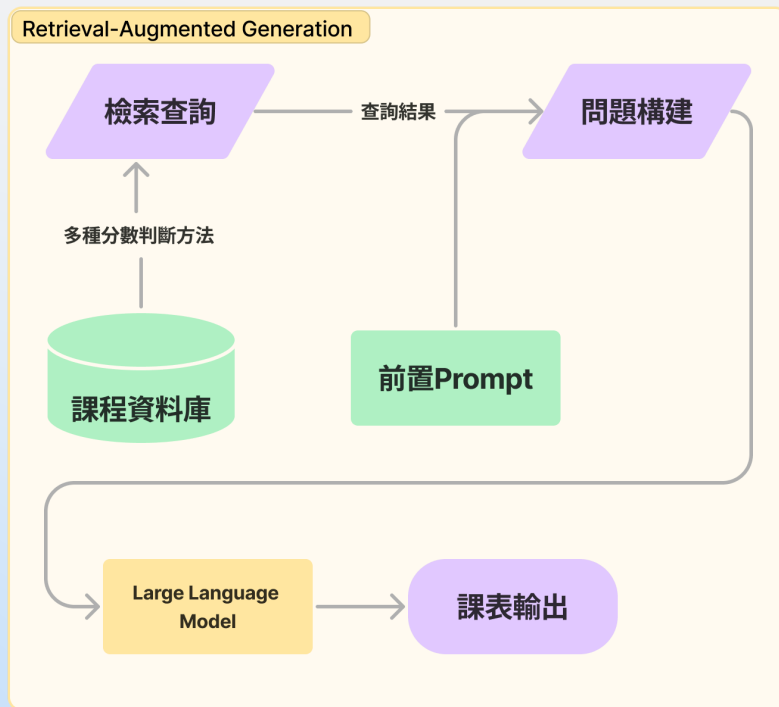
[illegible]

模型壓縮

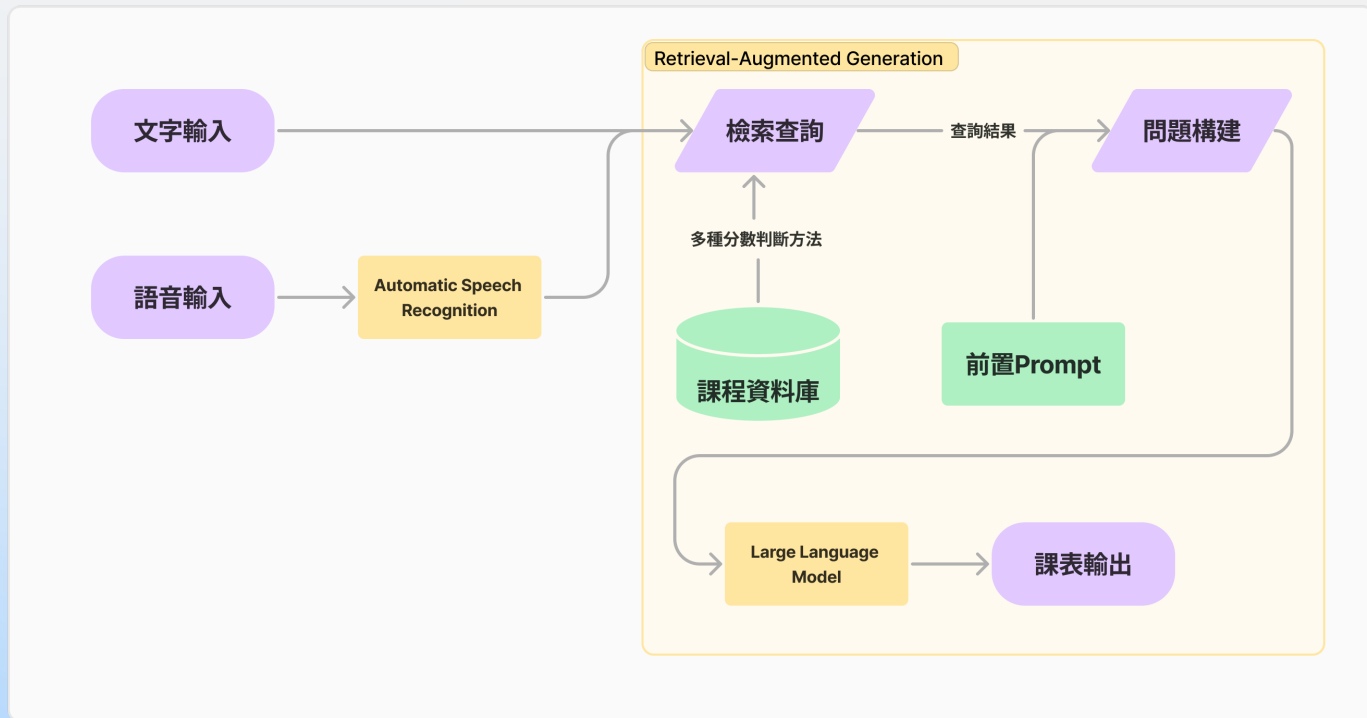
硬體設備：NVIDIA RTX 3090 (24G VRAM)

1. 7B：13G VRAM，答非所問
2. 13B：約24G VRAM，推估效果最好
3. 13B 8bit：18G VRAM，在能力與記憶體用量之間取得較好的平衡

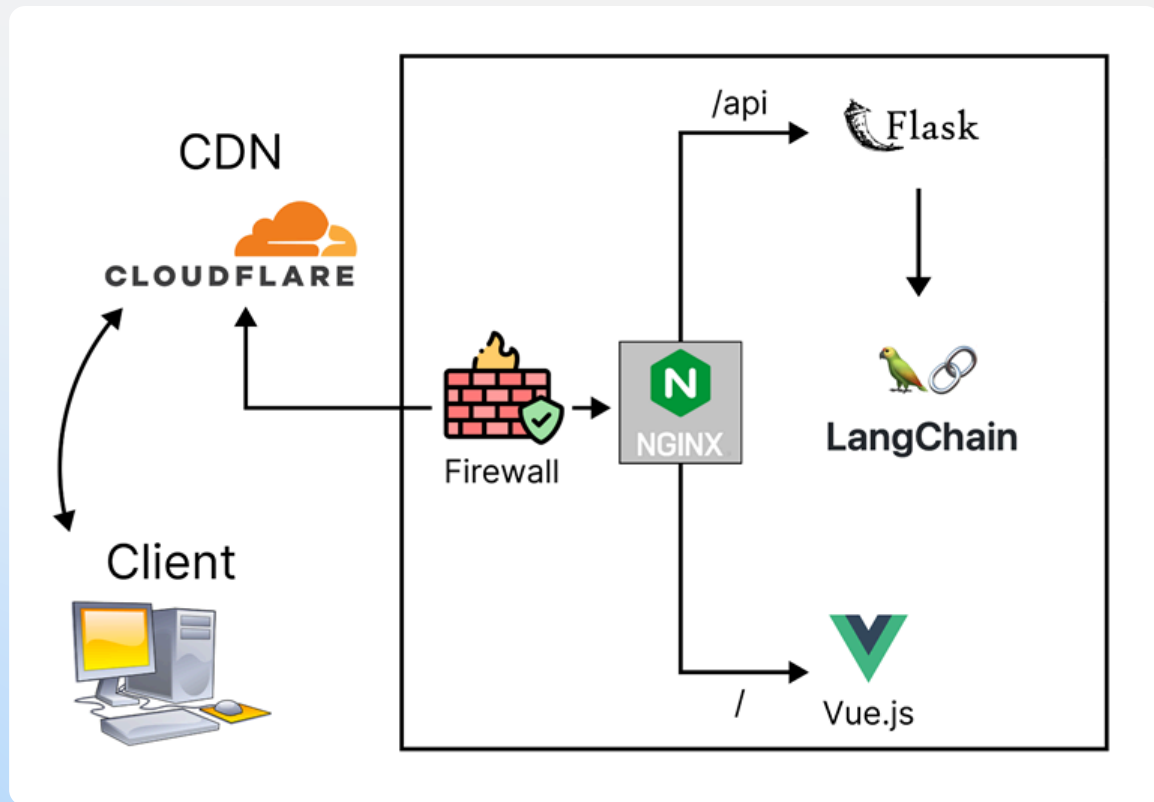
建構RAG





建構服務



建構服務



研究成果

 Course LangChain 



我想修經濟學，請幫我排課表！

B

根據所提供的信息，以下是您可以修的經濟學相關課程：

課程名稱	上課時間	老師名稱	上課教室
經濟學	星期三 9:0-12:0	胡偉民	
個人經濟學	星期五 13:0-16:0	孫士勝	
數理財務	星期一 14:0-17:0	黃台心	
國際經濟	星期六 9:0-12:0	廖仁哲	
金融學	星期三 13:0-16:0	林遠澤	
公共政策分析	星期二 14:0-17:0	廖郁賢	
數位經濟	星期四 9:0-12:0	陳心鵬	

Type here



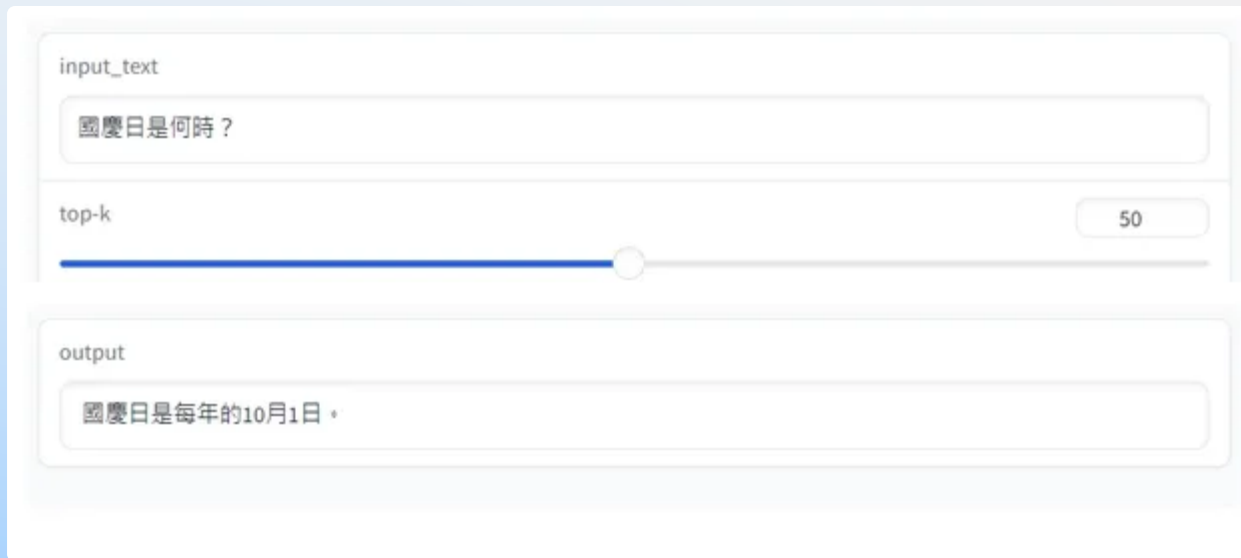
問題與討論

模型的其他選擇

- mlc1lm, llama, llama2
- 由於訓練資料為大陸的資料，用語時不出現簡體及大陸用語
- 以我們的資料微調模型，但因為硬體問題沒有嘗試

模型的其他選擇

- mlc-llm, llama, llama2
- 由於訓練資料為大陸的資料，用語時不出現簡體及大陸用語
- 以我們的資料微調模型，但因為硬體問題沒有嘗試



input_text

國慶日是何時？

top-k

50

output

國慶日是每年的10月1日。

評價標籤優化

- 問題：大量冗詞及口語用字，導致標籤分數不準確
- 解法：摘要評價後再進行分類評分
- 目前沒有找到合適的方法與模型

檢索優化

- text2vec-base-chinese 效果不甚理想
- 借助 **Reciprocal Rank Fusion** 結合 BM25 的綜合分數
- 希望未來有其他更好的開源模型

強化否定詞判定

- 否定敘述會被拿去做搜尋
- 需要理解句子的真正含意
- 在輸入後列舉出在選課情境中常見的情況，並進行特殊情況判斷，將句子改寫為正面敘述。
- 可以使用 Knowledge Graph 去真正的理解語意並判讀

未來展望

未來展望

- 服務分散式架構
- 提高評價分數的準確性
- Prompt優化
- 反義詞判斷與處理優化
- 以使用者回饋微調模型

QA 時間