



----- LangChain -----

DocGPT

專案介紹

講解人：許恆修

平台：Udemy

章節目標

Key Point

重點

本章節中，在帶領同學認識現有文件問題，以及如何整合 LangChain 與 LLM 進行互動，用以提升對文件資料的理解。過程中，將帶領學員實現 DocGPT，作為本課程案例。

專案目標

本次專案旨在讓學員清楚了解到 LangChain 在文件資料情境上的問題認識，並結合手上工具 LangChain 整合 LLM，最後實作出一個 DocGPT 應用程式，提供使用者用自然語言方式與文件資料進行對話。

實作案例

本章節中，學生將熟練利用 pdf 檔案存放至 Google 雲端硬碟，並且結合 LangChain 提供的 各式 `document_loader` 讀取檔案後，跟 OpenAI 換取 Embedding，最後透過 Chain 來進行問答。

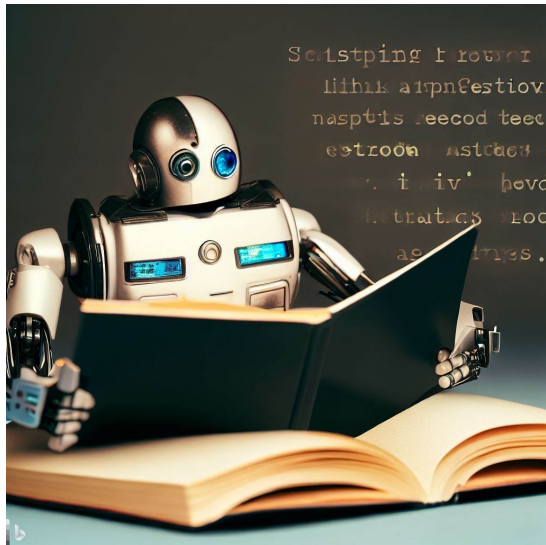
LangChain

問題介紹



「書，代表想說的話

通過閱讀書本，人們可以學習新知識，了解不同的觀點和想法，並在這些基礎上進行自己的思考和推理。書本還可以啟發人們的創造力和想像力，激發他們對未來的探索和思考。但是，閱讀可能需要較長的時間來瞭解和吸收內容，而當今快節奏的生活方式和工作節奏讓人們時間更緊迫。閱讀似乎是更加耗時的，且某些內容可能比較難以理解，尤其是在面對專業術語和學術文獻時，將導致人們會感到挫敗。



改善問題後的三個優點

總的來說，DocGPT 可以是一個強大的工具，可以幫助您節省時間、提高理解力、並提高工作效率。如果您正在尋找一種從 PDF 文件中獲取更多資訊的方法，那麼本次專案實作是一個不錯的選擇。

節省時間

DocGPT 可以快速且輕鬆地總結和提取 PDF 檔案中的關鍵信息，節省了大量的時間和精力。這對學生、專業人士和需要閱讀和分析大量 PDF 檔案的任何人都非常有幫助。

提高理解程度

DocGPT 可以通過提供總結、解釋和回答問題的方式，幫助您更好地理解 PDF 檔案的內容。這對於複雜或技術性文件特別有幫助。

更高生產力量

DocGPT 可以幫助您提高效率，讓您有更多時間專注於其他任務。這對於忙碌的專業人士尤其有幫助，他們需要在短時間內完成大量工作。

LangChain

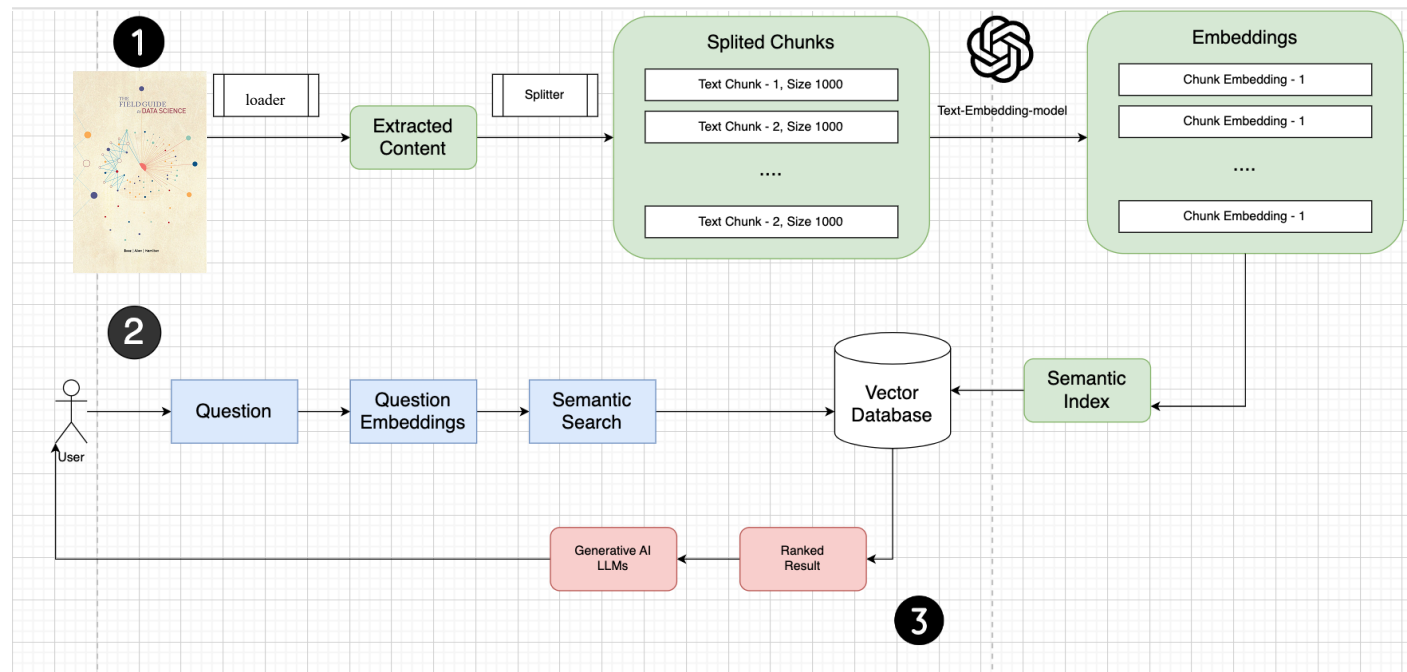
系統架構



梳理整體系統技術

總共有以下步驟：

1. 提取出 書本資料，將提取出的內容，切割成 N 份小Chunk，接著用 OpenAI Text-Embedding 模型將Chunk 向量化。
2. 當使用者提出問題時，向量化使用者問題，並計算與與 Vector Base 中向量相近的內容。
3. 最後提取出最相近的內容，再透過大型語言模型 LLM 回答使用者問題



LangChain

「
終章
」

感謝參與

