人工智慧期末作業2

安裝套件

\*\*輸入

**!**pip install langchain

**!**pip install openai

**!**pip install google-search-results

**!**pip install unstructured

**!**pip install 'chromadb>=0.3.5'

**!**pip install pinecone-client

**!**pip install youtube-transcript-api

**!**pip install pytube

**!**pip install pdf2image

**!**pip install pytesseract

**!**pip install tiktoken

*# Python要3.9以上*

\*\*輸出

一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 紙張 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

\*\*輸入

*# 先去https://serpapi.com/申請API KEY*

**import** os

os**.**environ["OPENAI\_API\_KEY"] **=** "你自己的Openai api key"

os**.**environ["SERPAPI\_API\_KEY"] **=** "你自己的Serp api key"

apikey **=** os**.**environ["OPENAI\_API\_KEY"]

serpapikey **=** os**.**environ["SERPAPI\_API\_KEY"]

\*\*輸入

**from** langchain.llms **import** OpenAI

llm **=** OpenAI(model\_name**=**"text-davinci-003",max\_tokens**=**1024)

llm("如何評價人工智慧？")

\*\*輸出

一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述一張含有 文字, 字型, 代數, 螢幕擷取畫面 的圖片

自動產生的描述

\*\*輸入

**!**wget https://github.com/shhuangmust/AI/raw/111-2/must.txt

\*\*輸出

一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述

\*\*輸入

**from** langchain.document\_loaders **import** UnstructuredFileLoader

**from** langchain.chains.summarize **import** load\_summarize\_chain

**from** langchain.text\_splitter **import** RecursiveCharacterTextSplitter

**from** langchain **import** OpenAI

**from** langchain.document\_loaders **import** TextLoader

*# 導入文字*

*# loader = UnstructuredFileLoader("lg\_test.txt")*

loader **=** TextLoader("must.txt")

*# 將文字轉換成Document物件*

document **=** loader**.**load()

print(f'documents:{len(document)}')

*# 初始化文字分割器*

text\_splitter **=** RecursiveCharacterTextSplitter(

chunk\_size **=** 500,

chunk\_overlap **=** 0

)

*# 切分文字*

split\_documents **=** text\_splitter**.**split\_documents(document)

print(f'documents:{len(split\_documents)}')

\*\*輸出

一張含有 文字, 字型, 白色, 設計 的圖片

自動產生的描述

\*\*輸入

llm **=** OpenAI(max\_tokens**=**1500)

*# 建立總結chain*

chain **=** load\_summarize\_chain(llm, chain\_type**=**"refine", verbose**=True**)

\*\*輸入

*# 執行總結的chain（只取前5段）*

chain**.**run(split\_documents[:5])

\*\*輸出

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 陳列 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述

\*\*輸入

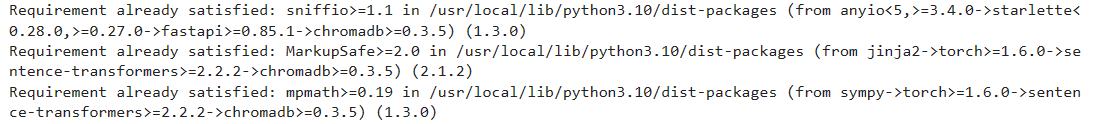
**!**pip install 'chromadb>=0.3.5'

\*\*輸出

一張含有 文字, 字型, 紙張, 螢幕擷取畫面 的圖片

自動產生的描述一張含有 文字, 字型, 紙張, 螢幕擷取畫面 的圖片

自動產生的描述一張含有 文字, 紙張, 字型, 文件 的圖片

自動產生的描述

#建立本機資料庫QA機器人

\*\*輸入

**from** langchain.embeddings.openai **import** OpenAIEmbeddings

**from** langchain.vectorstores **import** Chroma

**from** langchain.text\_splitter **import** CharacterTextSplitter

**from** langchain **import** OpenAI,VectorDBQA

**from** langchain.document\_loaders **import** DirectoryLoader

**from** langchain.chains **import** RetrievalQA

*# 載入資料夾中所有TXT檔案*

loader **=** DirectoryLoader('./', glob**=**'\*\*/\*.txt')

*# 將資料轉成document物佚，每個檔案會為作為一個document*

documents **=** loader**.**load()

*# 初始化載入器*

text\_splitter **=** CharacterTextSplitter(chunk\_size**=**100, chunk\_overlap**=**0)

*# 切割加载的 document*

split\_docs **=** text\_splitter**.**split\_documents(documents)

*# 初始化 openai 的 embeddings 物件*

embeddings **=** OpenAIEmbeddings()

*# 將 document 透過 openai 的 embeddings 物件計算 embedding向量資料暫時存入 Chroma 向量資料庫用於後續的搜尋*

docsearch **=** Chroma**.**from\_documents(split\_docs, embeddings)

*# 建立回答物件*

qa **=** VectorDBQA**.**from\_chain\_type(llm**=**OpenAI(), chain\_type**=**"stuff", vectorstore**=**docsearch, return\_source\_documents**=True**)

*# 進行回答*

*# result = qa({"query": "現行明新科技大學之校訓為何？"})*

result **=** qa({"query": "工專時期第3任校長是誰"})

print(result['result'])

\*\*輸出

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

\*\*輸入

print(result)

\*\*輸出

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

\*\*輸入

docsearch **=** Chroma**.**from\_documents(split\_docs, embeddings, persist\_directory**=**"./vector\_store")

docsearch**.**persist()

\*\*輸出



\*\*輸入

docsearch **=** Chroma(persist\_directory**=**"./vector\_store", embedding\_function**=**embeddings)

result **=** qa({"query": "工專時期第二任校長為何？"})

print(result['result'])

\*\*輸出



\*\*輸入

result **=** qa({"query": "校歌中，同建三民主義新中國的下一句是什麼？"})

print(result['result'])

\*\*輸出

