

將 QA(程式跑出來的那個)上傳至網站 使用 QA 與 RAG(網站的回應)做比對

QA->RAG 的參考答案 (需 100%正確)

下載雲端共用的檔案 (可改檔名)

原本只找教務處&學務處 後面全抓了 500 多個 PDF(ALL)

```
# 生成整個PDF的問題
✓ async def generate_questions_for_pdf(client, pdf_text, num_questions=30):
✓     prompt = f"""根據以下國立臺北商業大學法規內容，產生 {num_questions} 個問題：
    法規內容：
    {pdf_text['content'][:15000]} # 限制文本長度以防止超出模型的上下文限制

    請產生的問題必須符合以下要求：
    1. 問題必須直接從提供的法規內容中獲取答案，不得包含法規之外的資訊
    2. 問題應該具有實用性，即能夠幫助使用者理解和應用此法規
    3. 完全使用繁體中文，絕對不可使用簡體中文
    4. 問題的答案必須能夠從法規內容中明確找到對應條文

    請列出 {num_questions} 個問題：
    """
```

一個 PDF 產生 30 個問題 (下面為提示詞/要求)

對產出之問題進行篩選(本方案、本法規 什麼方案/法規?)

```
# 驗證問題
async def validate_questions(client, questions):
    validated_questions = []

    for q in questions:
        # 使用問題和上下文的前15000個字符以避免超出模型限制
        prompt = f"""以下是國立臺北商業大學法規內容與相關問題，請進行嚴格評估：

        法規內容：
        {q['context'][:15000]}

        問題：
        {q['question']}

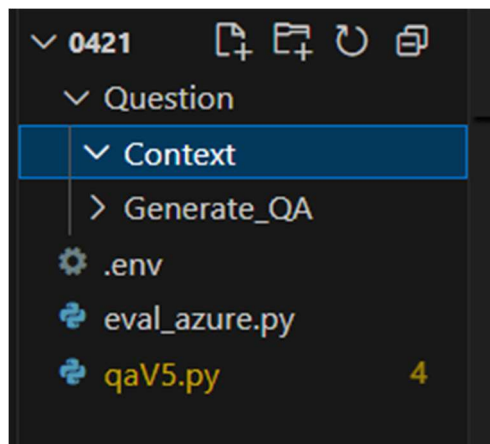
        請執行以下評估任務：
        1. 請直接從法規內容中引用相關條文來回答問題
        2. 回答必須完全基於提供的法規內容，不要添加未在法規中提及的資訊
        3. 回答必須具體明確，避免模糊或一般性陳述
        4. 請標明引用的具體條款編號(如第X條第Y款)
        5. 回答格式應為：「根據《法規名稱》第X條規定，...」
        6. 若問題無法從所提供的法規內容中找到足夠資訊回答，請回答「根據提供的法規內容無法回答此問題」
        7. 必須使用繁體中文回答，絕對不可使用簡體中文
        """
```

產出答案

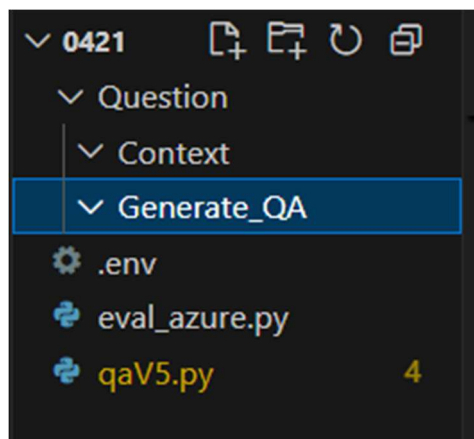
就算系統篩過 還是會出現奇怪的東西 需再進行篩選

思考一下如何使模型產出對上要求的回答

主要修改「生成整個 PDF 的問題」位置的提示詞



要跑的檔案放這(一次放一份、法規)



產出結果會在這()

```
PS C:\Users\lin20\Downloads\0421\env> cd Scripts
PS C:\Users\lin20\Downloads\0421\env\Scripts> activate
activate : 無法辨識 'activate' 詞彙是否為 Cmdlet、函數、指令檔或可執行程式的名稱。請檢查名稱拼字是否正確，如果包含路徑的話，請確認路徑是否正確，然後再試一次。
位於 線路:1 字元:1
+ activate
+ ~~~~~
+ CategoryInfo          : ObjectNotFound: (activate:String) [], CommandNotFoundException
+ FullyQualifiedErrorId : CommandNotFoundException

Suggestion [3,General]: 找不到 activate 命令，但它確實存在於目前的位置。Windows PowerShell 預設並不會從目前的位置載入命令。如果您信任這個命令，請改為輸入 ".\activate"。如需詳細資料，請參閱 "get-help about_Command_Precedence"。
PS C:\Users\lin20\Downloads\0421\env\Scripts> cd C:\Users\lin20\Downloads\0421
PS C:\Users\lin20\Downloads\0421> .\env\Scripts\activate
(env) PS C:\Users\lin20\Downloads\0421>
```

- Power shell(管理員身分執行)> Set-ExecutionPolicy RemoteSigned (才不會被擋權限)
- 終端機> .env\Scripts\activate (啟動虛擬機、使用 tab 會比較快)

```
(env) PS C:\Users\lin20\Downloads\0421> pip install -r requirements.txt
Collecting aiohappyeyeballs==2.6.1 (from -r requirements.txt (line 1))
  Downloading aiohappyeyeballs-2.6.1-py3-none-any.whl.metadata (5.9 kB)
Collecting aiohttp==3.11.13 (from -r requirements.txt (line 2))
  Downloading aiohttp-3.11.13-cp312-cp312-win_amd64.whl.metadata (8.0 kB)
Collecting aiosignal==1.3.2 (from -r requirements.txt (line 3))
```

將 requirements.txt 丟進資料夾

終端機> pip install -r requirements.txt

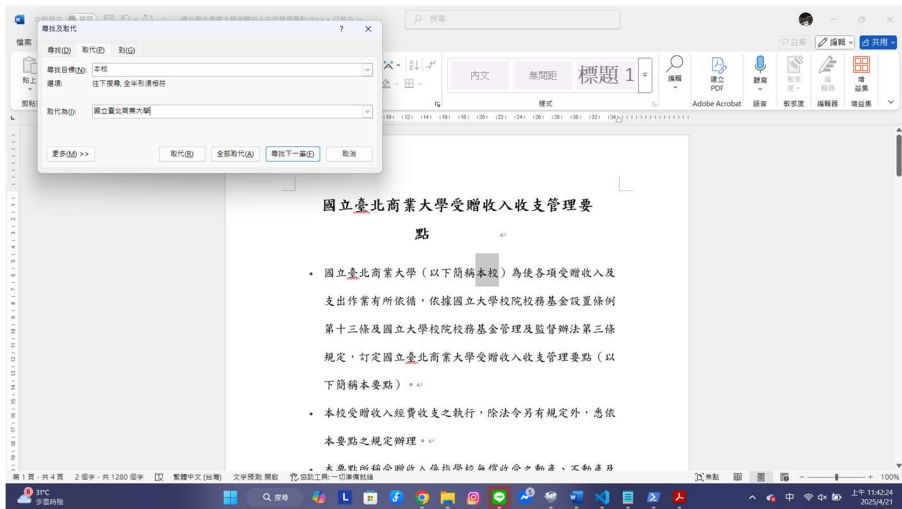
```
(env) PS C:\Users\lin20\Downloads\0421> python .\qaV5.py
將處理檔案：「國立臺北商業大學通識教育委員會設置要點」109年新版 (1).pdf
處理檔案：「國立臺北商業大學通識教育委員會設置要點」109年新版 (1).pdf
```

可以去幹別的事情了

```
115     # 為整個PDF生成問題
116     questions = await generate_questions_for_pdf(client, pdf_text, num_questions=30)
```

```
43     # 生成整個PDF的問題
44     async def generate_questions_for_pdf(client, pdf_text, num_questions=30):
45         prompt = f"""根據以下國立臺北商業大學法規內容，產生 {num_questions} 個問題：
46         法規內容：
47         {pdf_text['content'][:15000]} # 限制文本長度以防止超出模型的上下文限制
48
49         請產生的問題必須符合以下要求：
50         1. 問題必須直接從提供的法規內容中獲取答案，不得包含法規之外的資訊
51         2. 問題應該具有實用性，即能夠幫助使用者理解和應用此法規
52         3. 完全使用繁體中文，絕對不可使用簡體中文
53         4. 問題的答案必須能夠從法規內容中明確找到對應條文
54         5. 不能使用答：的方式來生成問題
55
56         請列出 {num_questions} 個問題：
57         """
```

每個人分配的資料夾都有 word 檔 對每個 word 進行資料前處理
再將 word 轉為 pdf 再丟給程式(只收 PDF)跑 QA



before



國立臺北商業大學受贈收入收支管理要點

- 國立臺北商業大學為使各項受贈收入及支出作業有所依循,依據國立大學校院校務基金設置條例第十三條及國立大學校院校務基金管理與監督辦法第三條規定,訂定國立臺北商業大學受贈收入收支管理要點。
- 國立臺北商業大學受贈收入經費收支之執行,除法令另有規定外,悉依國立臺北商業大學受贈收入收支管理要點之規定辦理。

after