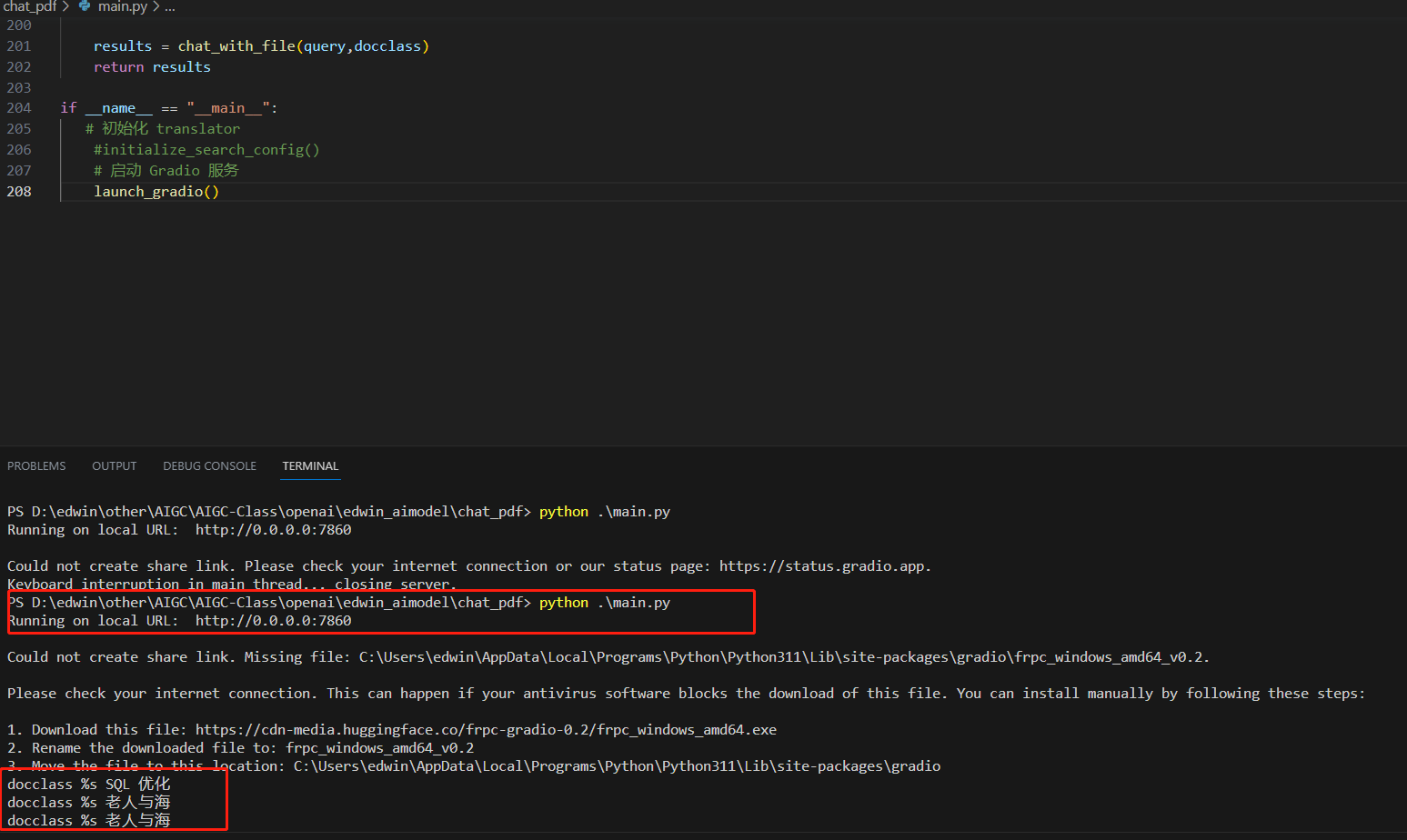
用LangChain+FASIS+Gradio构建本地PDF 智能问答

代码地址：<https://github.com/edwinjiang703/edwin_aimodel/tree/main/chat_pdf>

功能入口文件：main.py



一、工作流程

1、用户提交要处理的文档，该文档可以是PDF或图像格式

2、检测文件的格式，以便应用相关内容提取功能，利用Langchain.document\_loaders的UnstructuredImageLoader和UnstructuredFileLoader进行识别和提取

3、将识别的文档内容进行CharacterTextSplitter，然后进行OpenAIEmbeddings，最后存储在FASIS向量数据库中，

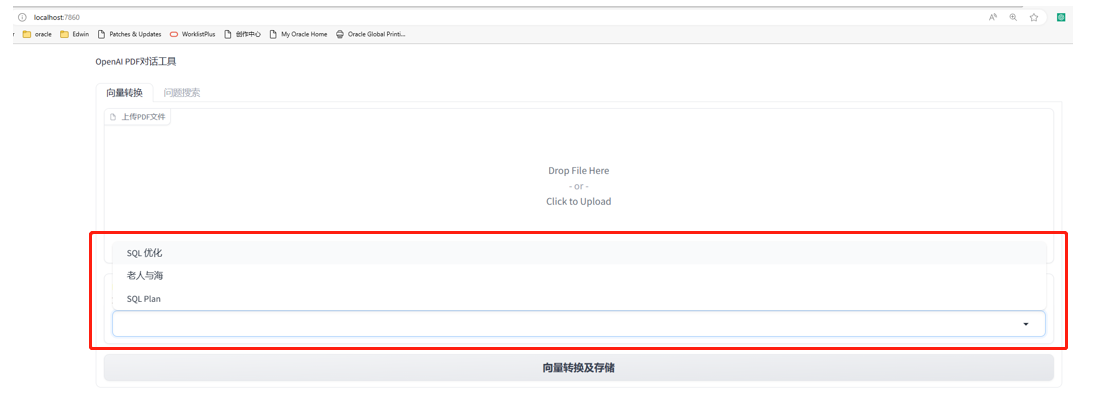
4、利用FASIS的similarity\_search功能根据界面的问题，在向量数据库中进行内容匹配，给出最佳的答案，如果没有找到合适的，会调用load\_qa\_chain，使用gpt-3.5-turbo模型进行寻找合适答案

5、界面用Gradio实现

二、功能界面截图：

1、上传文件

选择SQL 优化、老人与海、SQL Plan 三个类别，上传三个PDF文件。



2、PDF问答

* SQL优化：



* 谁是海明威？

