**LLM Q&A Search System**

**1. Code flowchart:**

**图示

描述已自动生成**

**2 Technical Description:**

**This project is implemented using Python. The language model used is metallama/Llama2-7b chat hf (Llama2) provided by HuggingFace. The model adopts a prompt based approach for inference, directly generating output results without the need for additional fine-tuning. The preprocessing of documents, text segmentation, and information retrieval are all implemented by calling the API provided by LangChain. The reason why the model was not fine tuned is that the requirements of this task do not involve the necessity of fine-tuning, and I did not collect additional datasets for fine-tuning.**

**4 how to run：**

**Just need to click and run line by line in the order of the code on Google Colab.**

**Attention: Before running, you need to log in to your HuggingFace account in the Colab environment. You only need to provide read-only permission to run normally without any additional settings.**

**5 show the result:**

**图形用户界面, 应用程序

描述已自动生成**