



基于langchain RAG问答应用实战

来自：AiGC面试宝典

宁静致远

2024年02月08日 10:04

一、前言

1.1 介绍

本次选用百度百科——藜麦数据 (<https://baike.baidu.com/item/藜麦/5843874>) 模拟个人或企业私域数据，并基于langchain开发框架，实现一种简单的RAG问答应用示例。

1.2 软件资源

- CUDA 11.7
- Python 3.10
- pytorch 1.13.1+cu117
- langchain

二、环境搭建

2.1 下载代码

```
$
```

2.2 构建环境

```
$ conda create -n py310_chat python=3.10 # 创建新环境
$ source activate py310_chat # 激活环境
```

2.3 安装依赖

```
$ pip install datasets langchain sentence_transformers tqdm chromadb langchain_wenxin
```

三、RAG问答应用实战

3.1 数据构建

藜麦数据 (<https://baike.baidu.com/item/藜麦/5843874>) 保存到 藜.txt 文件中。

3.2 本地数据加载

```
from langchain.document_loaders import TextLoader
loader = TextLoader("./藜.txt")
documents = loader.load()
documents
>>>
[Document(page_content='藜（读音lí）麦（Chenopodium\xa0quinoa\xa0Willd.）是藜科藜属植物。穗部可呈红、紫、黄，植株形状类似灰灰菜，成熟后穗部类似高粱穗。植株大小受环境及遗传因素影响较大，从0.3-3米不等，茎部质地较硬，可分枝可不分。单叶互生，叶片呈鸭掌状，叶缘分为全缘型与锯齿缘型。藜麦花两性，花序呈伞状、穗状、圆锥状，藜麦种子较小，呈小圆药片状，直径1.5-2毫米，千粒重1.4-3克。\\xa0[1]\\xa0\\n原产于南美洲安第斯山脉的哥伦比亚、厄瓜多尔、秘鲁等中高海拔山区。具有一定的耐旱、耐寒、耐盐性，生长范围约为海平面到海
```

拔4500米左右的高原上，最适的高度为海拔3000-4000米的高原或山地地区。藜麦富含的维生素、多酚、类黄酮类、皂苷和植物甾醇类物质具有多种健康功效。...

3.3 文档分割

文档分割，借助langchain的字符分割器，这里采用固定字符长度分割chunk_size=128

```
# 文档分割
from langchain.text_splitter import CharacterTextSplitter

# 创建拆分子器
text_splitter = CharacterTextSplitter(chunk_size=128, chunk_overlap=0)

# 拆分文档
documents = text_splitter.split_documents(documents)

documents

[Document(page_content='藜（读音lí）麦（Chenopodium\xa0quinoa\xa0Willd.）是藜科藜属植物。穗部可呈红、紫、黄，植株形状类似灰灰菜，成熟后穗部类似高粱穗。植株大小受环境及遗传因素影响较大，从0.3-3米不等，茎部质地较硬，可分枝可不分。单叶互生，叶片呈鸭掌状，叶缘分为全缘型与锯齿缘型。藜麦花两性，花序呈伞状、穗状、圆锥状，藜麦种子较小，呈小圆药片状，直径1.5-2毫米，千粒重1.4-3克。', metadata={'source': './藜.txt'}),
Document(page_content='藜麦是印第安人的传统主食，几乎和水稻同时被驯服有着6000多年的种植和食用历史。藜麦具有相当全面营养成分，并且藜麦的口感口味都容易被人接受。在藜麦这种营养丰富的粮食滋养下南美洲的印第安人创造了伟大的印加文明，印加人将藜麦尊为粮食之母。美国人早在80年代就将藜麦引入NASA，作为宇航员的日常口粮，FAO认定藜麦是唯一一种单作物即可满足人类所需的全部营养的粮食，并进行藜麦的推广和宣传。2013年是联合国钦定的国际藜麦年。以此呼吁人们注意粮食安全和营养均衡。', metadata={'source': './藜.txt'}),
Document(page_content='繁殖\n地块选择：应选择地势较高、阳光充足、通风条件好及肥力较好的地块种植。藜麦不宜重茬，忌连作，应合理轮作倒茬。前茬以大豆、薯类最好，其次是玉米、高粱等。', metadata={'source': './藜.txt'})
...]
```

3.4 向量化&数据入库

接下来对分割后的数据进行embedding，并写入数据库。这里选用m3e-base作为embedding模型，向量数据库选用Chroma

```
from langchain.embeddings import HuggingFaceBgeEmbeddings
from langchain.vectorstores import Chroma

# embedding model: m3e-base
model_name = "moka-ai/m3e-base"
model_kwargs = {'device': 'cpu'}
encode_kwargs = {'normalize_embeddings': True}
embedding = HuggingFaceBgeEmbeddings(
    model_name=model_name,
    model_kwargs=model_kwargs,
    encode_kwargs=encode_kwargs,
    query_instruction="为文本生成向量表示用于文本检索"
)
```

```
# load data to Chroma db
db = Chroma.from_documents(documents, embedding)

# similarity search
db.similarity_search("藜一般在几月播种？")
```

3.5 Prompt设计

prompt设计，这里只是一个prompt的简单示意，在实际业务场景中需要针对场景特点针对性调优。

```
template = '''
    【任务描述】
    请根据用户输入的上下文回答问题，并遵守回答要求。

    【背景知识】
    {{context}}

    【回答要求】
    - 你需要严格根据背景知识的内容回答，禁止根据常识和已知信息回答问题。
    - 对于不知道的信息，直接回答“未找到相关答案”

    -----
    {question}
'''
```

3.6 RetrievalqaChain构建

这里采用ConversationalRetrievalChain，ConversationalRetrievalQA chain 是建立在 RetrievalQAChain 之上，提供历史聊天记录组件。如下面定义了memory来追踪聊天记录，在流程上，先将历史问题和当前输入问题融合为一个新的独立问题，然后再进行检索，获取问题相关知识，最后将获取的知识和生成的新问题注入Prompt让大模型生成回答。

```
from langchain import LLMChain
from langchain_wenxin.llms import Wenxin
from langchain.prompts import PromptTemplate
from langchain.memory import ConversationBufferMemory
from langchain.chains import ConversationalRetrievalChain
from langchain.prompts.chat import ChatPromptTemplate, SystemMessagePromptTemplate,
HumanMessagePromptTemplate

# LLM选型
llm = Wenxin(model="ernie-bot", baidu_api_key="baidu_api_key", baidu_secret_key="baidu_secret_key")

retriever = db.as_retriever()
memory = ConversationBufferMemory(memory_key="chat_history", return_messages=True)
qa = ConversationalRetrievalChain.from_llm(llm, retriever, memory=memory)
qa({"question": "藜怎么防治虫害？"})
>>>
{'question': '藜怎么防治虫害？',
 'chat_history': [HumanMessage(content='藜怎么防治虫害？'),
  AIMessage(content='藜麦常见虫害有象甲虫、金针虫、蝼蛄、黄条跳甲、横纹菜蚜、蔊蓄齿胫叶甲、潜叶蝇、蚜虫、夜蛾等。防治方法：可每亩用3%的辛硫磷颗粒剂2-2.5千克于耕地前均匀撒施，随耕地翻入土中。也可以每亩用40%的辛硫磷乳油250毫升，加水1-2千克，拌细土20-25千克配成毒土，撒施地面翻入土中，防治地下害虫。')],
 'answer': '藜麦常见虫害有象甲虫、金针虫、蝼蛄、黄条跳甲、横纹菜蚜、蔊蓄齿胫叶甲、潜叶蝇、蚜虫、夜蛾等。防治方法：可每亩用3%的辛硫磷颗粒剂2-2.5千克于耕地前均匀撒施，随耕地翻入土中。也可以每亩用40%的辛硫磷乳油250毫升，加水1-2千克，拌细土20-25千克配成毒土，撒施地面翻入土中，防治地下害虫。'}]
```

3.7 高级用法

针对多轮对话场景，增加 question_generator对历史对话记录进行压缩生成新的question，增加combine_docs_chain对检索得到的文本进一步融合

```
from langchain import LLMChain
from langchain.prompts import PromptTemplate
from langchain.memory import ConversationBufferMemory
from langchain.chains import ConversationalRetrievalChain, StuffDocumentsChain
from langchain.chains.qa_with_sources import load_qa_with_sources_chain
from langchain.prompts.chat import ChatPromptTemplate, SystemMessagePromptTemplate,
HumanMessagePromptTemplate

# 构建初始 messages 列表，这里可以理解为是 openai 传入的 messages 参数
messages = [
    SystemMessagePromptTemplate.from_template(qa_template),
    HumanMessagePromptTemplate.from_template('{question}')
]

# 初始化 prompt 对象
prompt = ChatPromptTemplate.from_messages(messages)
llm_chain = LLMChain(llm=llm, prompt=prompt)

combine_docs_chain = StuffDocumentsChain(
    llm_chain=llm_chain,
    document_separator="\n\n",
    document_variable_name="context",
)

q_gen_chain = LLMChain(llm=llm, prompt=PromptTemplate.from_template(qa_condense_template))

qa = ConversationalRetrievalChain(combine_docs_chain=combine_docs_chain,
                                question_generator=q_gen_chain,
                                return_source_documents=True,
                                return_generated_question=True,
                                retriever=retriever)

print(qa({'question': "藜麦怎么防治虫害?", "chat_history": []}))
>>>
{'question': '藜怎么防治虫害?',
 'chat_history': [],
 'answer': '根据背景知识，藜麦常见虫害有象甲虫、金针虫、蝼蛄、黄条跳甲、横纹菜蚜、蒾蓄齿胫叶甲、潜叶
  蝇、蚜虫、夜蛾等。防治方法如下：\n\n1. 可每亩用3%的辛硫磷颗粒剂2-2.5千克于耕地前均匀撒施，随耕地翻入
  土中。\n2. 也可以每亩用40%的辛硫磷乳油250毫升，加水1-2千克，拌细土20-25千克配成毒土，撒施地面翻入土
  中，防治地下害虫。\n\n以上内容仅供参考，如果需要更多信息，可以阅读农业相关书籍或请教农业专家。',
 'source_documents': [
  Document(page_content=' 病害：主要防治叶斑病，使用12.5%的烯唑醇可湿性粉剂3000-4000倍液喷雾防治，一般防
  治1-2次即可收到效果。\\xa0[4]\\xa0\\n虫害：藜麦常见虫害有象甲虫、金针虫、蝼蛄、黄条跳甲、横纹菜蚜、蒾蓄
  齿胫叶甲、潜叶蝇、蚜虫、夜蛾等。防治方法：可每亩用3%的辛硫磷颗粒剂2-2.5千克于耕地前均匀撒施，随耕地翻
  入土中。也可以每亩用40%的辛硫磷乳油250毫升，加水1-2千克，拌细土20-25千克配成毒土，撒施地面翻入土中，
  防治地下害虫', metadata={'source': './藜.txt'}),
  Document(page_content=' 中期管理\\n在藜麦8叶龄时，将行中杂草、病株及残株拔掉，提高整齐度，增加通风透
  光，同时，进行根部培土，防止后期倒伏。\\xa0[4]', metadata={'source': './藜.txt'})],
 'generated_question': '藜怎么防治虫害? '}
```

