

人工智能实训项目功能描述报告

题目：基于 LangChain 的本地知识库问答系统

学生（学号） 班级，联系人方式如 email

姓名：黄亮铭

学号：2022155028

学院：计算机与软件学院

专业：软件工程（腾班）

班级：腾班

Email: 2306725926@qq.com

目录

1 项目状况简表.....	3
2 模块设计.....	5
3 努力方向.....	8
4 问题与解决思路.....	9
5 自我评价.....	9

1 项目状况简表

该项目为**基于 LangChain 的本地知识库问答系统**，主要具备以下功能：实现自然语言查询并提供精准回答、个性化知识库管理、多轮对话支持以及直观友好的用户界面。通过数据预处理、语义检索与生成模块的结合，系统能够高效解决用户的信息查询需求，特别是在专业领域的知识访问中表现出色。

本项目将采用面向对象设计的方法进行设计，严格遵循系统设计的核心原则。在设计过程中，我们将依据先前编制的开题报告，从中提炼出系统的核心功能模块。在此基础上，我们将精心设计每个功能模块的详细内容，并确保整个设计过程维持清晰的 **B/S 结构** 和 **MVC 架构**，以实现系统的高效性和可维护性。

本项目的架构采用 **MVC 架构**，将软件划分为**视图层**、**控制器层**和**模型层**三部分，达到简化设计、增强可维护性和提高扩展性的目的。

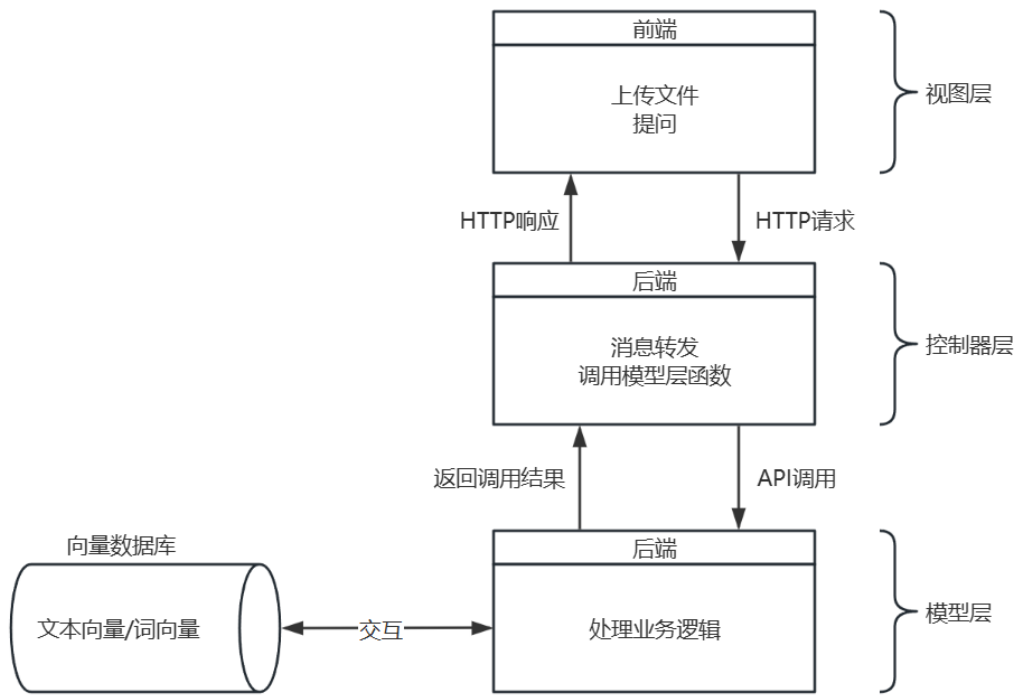


图 1 项目的简化 MVC 架构图

该项目的前端包含一个辅助模块：可视化模块，目的是为用户提供友好的交互体验。该项目的后端可以分为三个核心模块。

- **数据预处理模块：**负责完成数据清洗、分割以及语义嵌入生成与存储。
 - **数据检索模块：**实现基于语义相似度的高效匹配与多层筛选。
 - **内容生成模块：**通过 Prompt 模板设计和生成模型回答用户问题
- 项目的模块示意图如下所示。

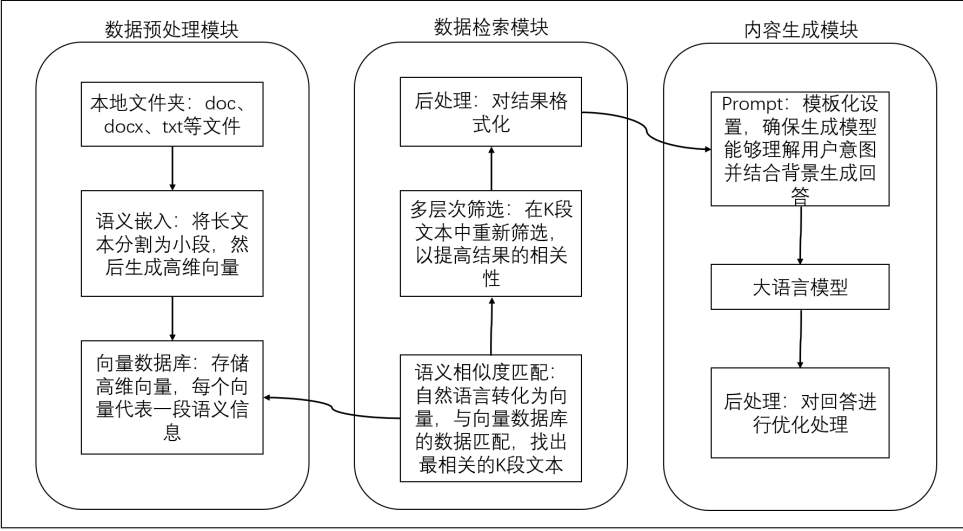


图 2 项目模块图

本项目的亮点及创新点有：①检索增强生成架构；②个性化知识库管理；③多轮对话支持；④基于 LangChain 的灵活性。我将项目的亮点及创新点总结如下表格所示。

亮点/创新点	内容描述	作用
检索增强生成架构	结合生成式大语言模型与本地知识库检索,通过语义匹配提升回答的准确性与可信度	提高回答在专业领域中的可靠性,适用于医疗、法律等知识密集型领域
个性化知识库管理	支持用户根据需求定制自己的知识库,系统会基于用户的私人知识库提供个性化回答	满足个性化需求,增强系统对不同应用场景的适应能力
多轮对话支持	系统具备上下文跟踪能力,能够处理用户的连续问题并保持语义一致性	适合复杂问题的多步解答,提升用户的连续使用体验
基于 LangChain 的灵活性	使用 LangChain 技术支持多种数据源整合,实现复杂问答逻辑处理并提升智能化水平	提升数据处理效率,支持多格式文档并生成语义嵌入

图 3 亮点/创新点表格

截至目前，项目完成了预处理模块和检索模块的核心功能开发，生成模块的基础设计已实现，界面设计已初步完成，整体进度良好，后续工作将重点优化系统性能和用户体验。

我将各个功能模块的原设计功能和目前完成进度总结如下表所示。

功能模块	原设计功能	完成进度
数据预处理模块	数据清洗、文本分割、语义嵌入生成与存储	已全部完成
数据检索模块	基于语义相似度的快速检索、多层次筛选和后处理功能	已完成检索逻辑的核心功能开发
内容生成模块	使用 LangChain 提供的库进行问题设计和回答生成	完成模板设计,生成功能调试中
UI 界面	直观友好的用户界面,支持知识库更新及对话管理	完成基本界面设计

图 4 模块功能

2 模块设计

对于每一个模块，我将会从功能描述的角度描述其负责的任务，同时使用顺序图描述描述功能的顺序交互的过程。

2.1 数据预处理模块

数据预处理模块是项目的核心模块之一，旨在对用户上传的文件进行高效地处理，便于数据存储到向量数据库中，提高系统的灵活性和易用性。

2.1.1 功能描述

以下是数据预处理模块的功能描述：

- **数据清洗与分割：**接收用户上传的文档，同时对文档内容进行矫正，如去除空白行。然后将长文本内容分割成多个短片段，分割时需要确保每段具有完整语义，便于后续生成语义嵌入。
- **语义嵌入生成与存储：**首先利用嵌入模型 *shibing624/text2vec-base-chinese* 生成文本片段的高维语义向量，然后将生成的向量存储到向量数据库 *Chroma* 中，供检索模块调用。
- **动态更新：**提供增量式更新功能，允许用户上传新的文档以动态扩展知识库。并且，更新时系统会重新处理新数据并同步至向量数据库。

2.1.2 顺序图

数据预处理模块描述了整个模块流程中不同功能之间的交互。

以下是数据预处理模块的时序图，为了更加直观地体现交互，这里省略了中间的控制层。

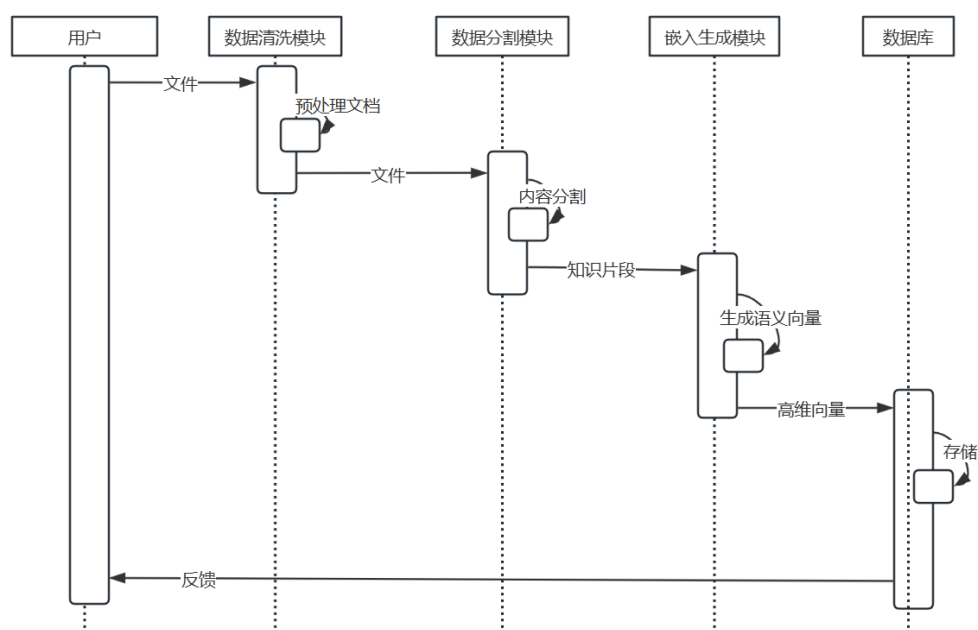


图 5 数据预处理模块顺序图

2.2 数据检索模块

数据检索模块是项目的核心模块之一。该模块的作用是根据用户提出的问题在向量数据库中寻找合适的片段，为后续的内容生成模块提供资料。

2.2.1 功能描述

以下是数据检索模块的功能描述：

- **语义相似度匹配：**接收用户输入的问题并将其转化为高维语义向量。然后在向量数据库中检索与该向量最相似的 K 个知识片段。
- **多层次筛选：**对检索出的 K 个知识片段进行进一步筛选，去除冗余或不相关片段，确保最终输出的片段与用户的问题高度相关。
- **检索后处理：**格式化检索结果，为后续的内容生成模块提供信息，如知识片段的内容和知识片段的来源。

2.2.2 顺序图

数据检索模块描述了整个模块流程中不同功能之间的交互。

以下是数据检索模块的时序图，为了更加直观地体现交互，这里省略了中间的控制层。

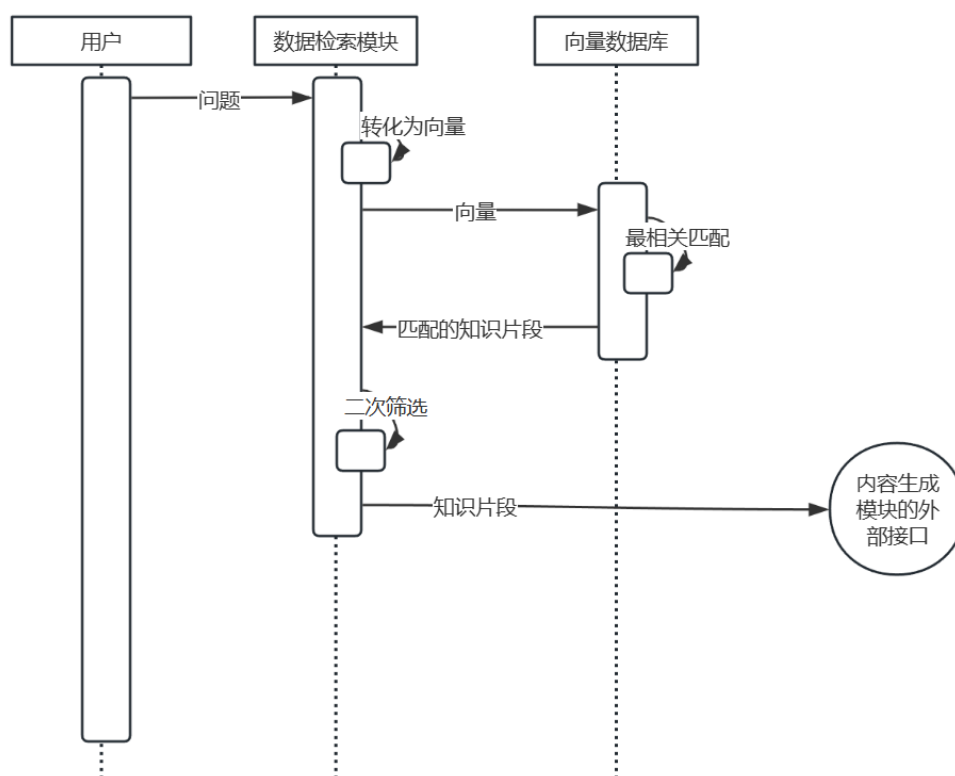


图 6 数据检索模块顺序图

2.3 内容生成模块

内容生成模块是项目的核心模块之一。该模块使用 LangChain 调用大语言模型阅读给定

的问题和知识片段，然后给出相应的回答。

2.3.1 功能描述

以下是内容生成模块的功能描述：

- **Prompt 模板设计：**将用户问题与检索片段按照预设模板组合，生成输入 Prompt。
- **生成式回答：**调用大语言模型，生成自然语言回答，确保回答逻辑清晰，符合用户提问语境。
- **回答后处理：**简化冗长回答，删除无关信息。同时校正潜在错误，确保回答准确性和专业性。

2.3.2 顺序图

内容生成模块描述了整个模块流程中不同功能之间的交互。

以下是内容生成模块的时序图，为了更加直观地体现交互，这里省略了中间的控制层。

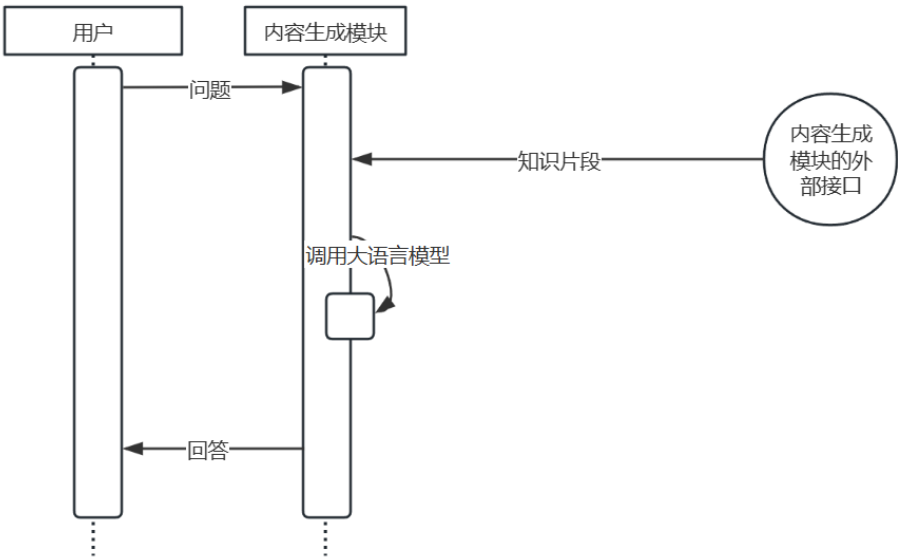


图 7 内容生成模块顺序图

2.4UI 界面

UI 界面从严格意义上来说并不是一个模块，这里将其单独作为一个模块进行设计，目的是为用户提供一个简洁、友好和便于交互的界面。

2.4.1 功能描述

UI 界面的功能描述如下：其中对话管理和知识库更新布局在左侧菜单栏中，对话窗口则放置在右边的子窗口。

- **对话管理：**支持用户创建新对话，记录每轮问答。对话记录按照时间排序，并显示摘要供用户快速查找。
- **知识库更新：**提供文件上传功能，支持用户更新本地知识库内容。
- **对话窗口：**用户可实时输入问题并查看回答。

2.4.2 顺序图

UI 界面描述了各个按钮的交互作用。
以下是 UI 界面的时序图，为了更加直观地体现交互，这里省略了中间的控制层。

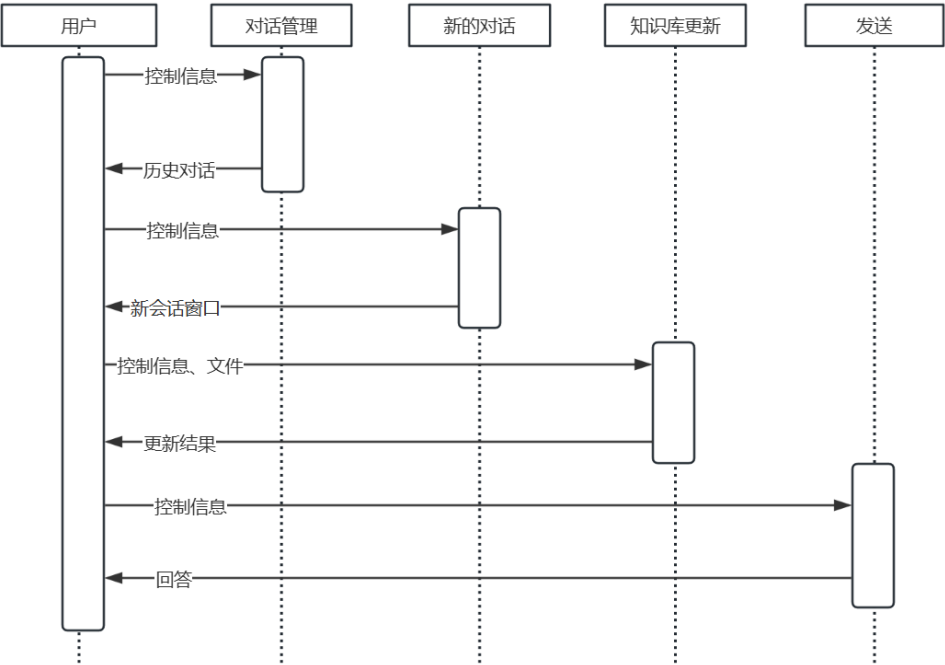


图 8UI 界面顺序图

3 努力方向

- 以下是尚未完成或已完成但需要进一步优化的功能：
- 1) **优化内容生成模块：**深化 Prompt 模板的设计，结合用户意图和上下文需求，进一步提升生成回答的精准性与实用性。
 - 2) **强化多轮对话能力：**引入对话历史的动态管理功能，通过滑窗机制控制上下文长度，避免信息冗余的同时保证语义连贯性。
 - 3) **提升用户界面体验：**优化 UI 布局和功能模块的交互设计，确保界面直观简洁。同时增强界面响应的流畅性，完善界面细节。
- 为了顺利完成上述任务，我制定了以下时间规划。

时间阶段	任务内容	预期目标
第 1-2 周	未完成部分的基本内容	完成项目的基本任务
第 2 周	优化内容生成模块	生成回答有效率达到 90%以上
第 3 周	强化多轮对话能力	多轮对话在长时间交互中能够保持语义连贯性，避免上下文信息丢失或冗余
第 4 周	优化页面布局	提供一个直观且易操作的用户界面，减少用户的学习成本

图 9 时间规划表

4 问题与解决思路

我将在完成项目的过程中遇到的问题总结成如下表格。我将表格分为问题描述和已使用或已计划地解决方案两部分，旨在清晰地描述问题与解决思路。

问题	已使用或已计划地解决方案
环境配置时 transformers 和 huggingface-hub 同时依赖于 sentence-Transformers 的不同版本，并且无法找到兼容的版本	在 LangChain 的 GitHub 仓库的 issues 中找到了解决办法：pip 虽然报错版本不兼容，但是实际上有部分版本可以兼容
生成式模型推理时间过长	通过模型压缩技术（如量化、蒸馏）减少模型计算量，或者引入多级缓存机制来加速重复问题的回答
生成模块回答中英混杂	使用自定义 Prompt 模板，结合规则校正生成结果

图 10 问题与解决方案

5 自我评价

与开题报告中的预期相比，项目进展总体符合计划。目前已完成 **70%** 的核心功能开发，整体进度良好。结合现阶段的进展，给予自己 **7/10 分** 的评价。后续将继续集中精力开发系统，同时尽力优化系统的性能和提高用户的体验，争取达到预期目标。