

人工智能实训项目开题报告

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

一、选题方向

在当今数字化时代，信息的爆炸式增长带来了前所未有的挑战和机遇。如何从海量数据中快速、准确地检索和提取所需信息，已成为提高工作效率、增强决策能力的关键。本项目基于“云端 AI 应用开发”的选题方向，开发一个基于 LangChain 的本地知识库问答系统，以实现快速、准确的信息检索与自动回答，广泛适合各类应用场景，助力各类组织在特定领域的知识管理上的高效运作。

该问答机器人将充分利大语言模型（LLM, Large Language Model）的最新技术，结合本地知识库的特点，提供个性化、上下文相关的回答。通过更新本地知识库，问答系统能学习最新的当前领域的知识，以确保问答系统的回答的准确性和时效性。同时，通过与用户的交互，系统能够不断学习和适应，提高其回答质量和准确性。

二、项目背景与意义

2.1 项目背景

随着互联网和数字技术的迅速发展，信息的获取变得更加便捷。然而，海量信息的涌现也带来了知识的碎片化与管理困难。传统的信息检索方如搜索引擎+数据库的方式在一定程

度上缓解了一个问题，但是用户需要耗费相当多的时间和相当大的精力才能从已筛选的信息中获得自己正确需要的信息。

随着生成式大语言模型的发展，在一定程度上可以说完全解决了上述问题。大语言模型能在海量数据中筛选数据，然后整合相关信息，向用户提供个性化、上下文相关的回答。然而，生成式大语言模型的独立性也带来了信息准确性的问题，尤其在涉及特定领域的专业知识时，生成模型可能会产生“幻觉”——即生成与事实不符的内容。此外，大语言模型只能从网络存在的数据进行筛选，对于私有数据，大语言模型无法进行准确的回答。

为解决这一问题，RAG 技术应运而生。RAG 将生成模型与本地知识库检索相结合，使生成模型在回答前先检索相关知识库信息，将检索到的信息作为提示输入。通过这种方式，系统能更准确地回答问题，并支持更复杂的知识查找任务。

2.2 项目意义

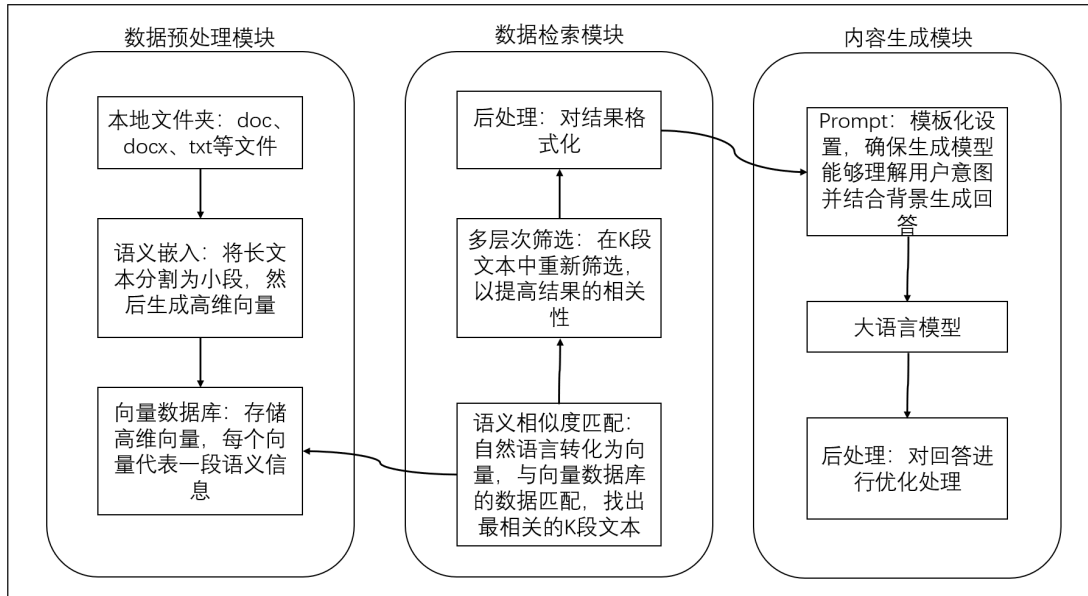
本项目的开展将有助于解决上述问题。通过构建一个基于 LangChain 的本地知识库问答系统，能够实现以下几点重要意义：

- **解决信息查询的效率问题：**用户可以通过自然语言输入问题，系统将自动识别并返回相关回答。该系统将显著减少用户在繁杂信息中查找所需知识的时间成本。特别是对于知识密集型的用户，这一系统提供了一种便捷的知识访问方式。
- **增强生成式模型的回答准确性：**通过引入 RAG 结构，本项目将生成模型与本地知识库的语义检索技术相结合，解决了生成模型在处理专业知识或私有数据时容易生成不准确信息的问题。这种结构使回答具有较高的准确性和可信度，适合应用于需要严格知识支持的领域，如医疗、法律等。
- **提升知识管理的个性化体验：**本项目的私人知识库问答机器人支持用户创建、更新和维护个性化的知识库。不同用户可以根据自己的需求添加不同的知识，系统将根据用户的私人知识库进行回答，使得问答体验更加个性化。
- **实现知识库的动态更新与多轮对话：**知识库问答系统具备动态更新功能，支持知识库内容的定期更新或即时补充。系统还结合大语言模型的多轮对话能力，使用户能够与系统进行连续对话，获得多步的指导信息。此功能进一步提高了系统的实用性。

三、项目内容与特色

3.1 项目内容

本项目的具体实现分成三个模块，分别为**数据预处理模块**、**数据检索模块**、**内容生成模块**。系统架构图如下图所示。



系统架构图

3.1.1 数据预处理模块

首先系统使用 LangChain 对整理好的数据（doc 文件、docx 文件等）进行清洗和分割，将长文本拆分成小段，以便于生成语义嵌入。每个片段应涵盖完整的语义，便于之后检索和生成时使用。然后使用语义嵌入生成工具，将每个片段生成高维向量，这些向量代表文本片段的语义信息。最后，每个向量会被存储在向量数据库中，系统可在检索阶段对向量进行快速匹配。

此外，系统会在该模块提供数据管理接口，用于支持增量式更新本地知识库。

3.1.2 数据检索模块

该模块通过语义相似度匹配、多层次筛选和检索后处理，在向量数据库中实现快速的语义检索。

1. **语义相似度匹配：**在检索过程中，系统将用户的问题转化为向量形式，并在向量数据库中寻找最相似的片段。通过语义相似度匹配技术，系统能精准找到与问题最相关的 K 个知识片段。
2. **多层次筛选：**为保证检索的准确性，系统设置多层次筛选机制。在初次筛选除 K 个最相关的知识片段后，系统会在这 K 个知识片段中再次筛选，以提升检索结果的相关性。
3. **检索后处理：**检索到的片段会在生成阶段作为提示输入，为生成模型提供背景信息。在检索后处理阶段，系统会对检索结果进行格式化，确保生成模块能无缝整合这些信息。

3.1.3 内容生成模块

该模块可以进一步细分为三个部分：**Prompt 模板设计**、**生成式大语言模型**和**内容生成后处理**。

1. **Prompt 模板设计：**Prompt 模板是生成模块的核心设计。系统将用户问题和检索到的相关片段组合成一个完整的 Prompt，通过模板化设置，确保生成模型能够理解用户意图并结合背景生成回答。
2. **生成式大语言模型：**生成式大语言模型我们可以根据自己的需要选择 OpenAI 的 GPT，也可以选择 LLaMA 等本地模型。生成模型的作用为根据 Prompt 生成答案，系统在得到回答后将其返回给用户。
3. **内容生成后处理：**生成回答后，系统会对回答进行优化处理。例如，对过于冗长的回答进行简化、对不准确的回答进行修正等。

3.2 项目特色

本项目具有以下几个特色：

- **高效的检索增强生成（RAG）架构：**通过 RAG 结构，系统的回答精准性显著提升，尤其在知识密集型领域中，系统可以提供具有较高可信度的答案。
- **本地知识库集成：**不同用户可根据个人需求建立私人知识库，回答将基于用户个性化知识库，提供专属的问答体验。
- **动态的知识库管理：**系统支持知识库内容的动态更新，使知识库内容始终保持时效性，并能根据用户的需求灵活调整。
- **基于 LangChain 的灵活性：**利用 LangChain 的强大功能，实现多种数据源的整合和复杂问答逻辑的处理，提升系统的智能化水平。
- **多轮对话支持：**系统具备多轮对话能力，使得用户能够在一段连续的对话中获得系统性的知识解答，满足复杂问题的连续咨询需求。

四、预期目标

本项目开发一个能够处理自然语言查询并提供准确回答的问答系统。在性能方面，系统回答准确率超过 90%，并且响应时间不超过 5s。在体验方面，开发一个直观友好的用户界面，使用户能方便地提出问题和浏览回答。此外，系统能够实现本地知识库的自动更新和管理机制，确保系统中知识的时效性和准确性；并且系统能够提供知识分析和统计功能，帮助用户了解知识使用情况，为后续决策提供依据。

4.1 系统功能

- **智能问答功能：**用户能够通过自然语言提问，系统能提供准确的回答。
- **个性化知识库管理：**用户可以自由管理本地知识库，随时更新内容。
- **多轮对话支持：**系统能够利用大模型处理多轮对话，保持上下文连贯性。
- **直观的用户界面：**设计简洁、美观的 UI，提升用户体验。

4.2 系统 UI

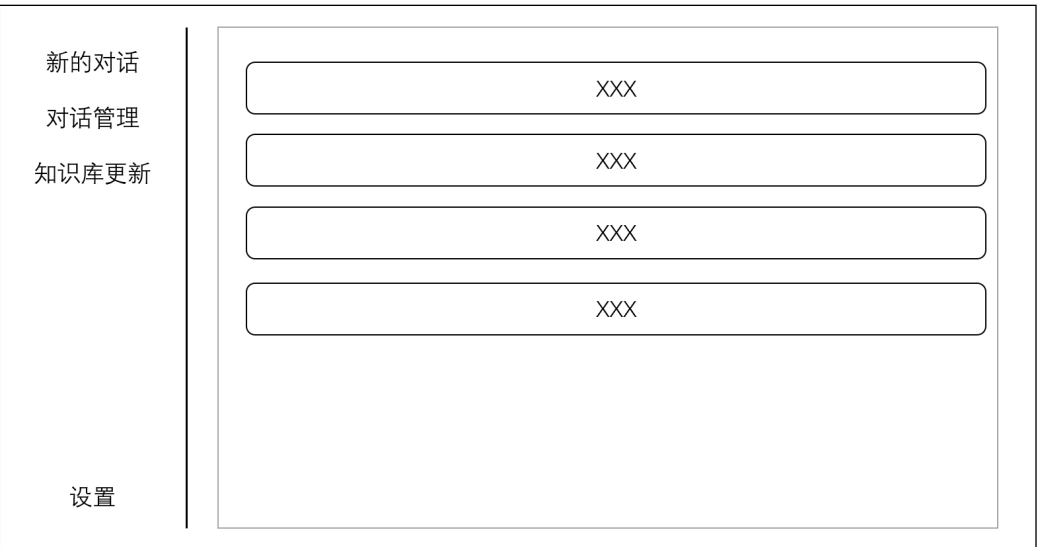
主界面分为左侧功能选择栏和右侧对话框（如下图所示）。功能选择栏的作用为允许用户手动更新本地知识库，查看用户与系统的对话记录和开启新一轮的对话。



系统 UI 主界面

系统 UI 主界面即默认界面就是新的对话窗口，方便用户进入应用后进行提问。

对话管理界面会按对话创建的时间顺序进行排序，最晚创建的对话在顶部，最早创建的对话在底部。同时，该界面会显示每个对话的摘要，默认取前 30 个字，方便用户寻找需要的对话，概念图如下图所示。



对话管理

知识库更新界面则是提供上传按钮，用户可以上传本地文件夹作为数据，更新本地知识库。概念图如下图所示。



知识库更新

五、相关工作

1. **问答系统的研究:** 近年来, 基于深度学习的问答系统得到了广泛的研究。各种预训练模型被广泛应用于问答任务中, 通过语言理解和生成技术提高了回答的准确性。RAG 作为一个新的架构, 在提高生成式问答系统准确性方面具有显著效果。
2. **用户体验的设计:** 近年来, 用户体验在产品开发中的重要性日益凸显。研究表明, 直观的界面和流畅的交互设计能够显著提升用户满意度。本项目将借鉴 ChatGPT, 优化问答系统的用户界面。

六、现有基础

在开展本项目之前, 我已经具备了一定的基础, 包括:

1. **技术能力:** 我熟悉 Python 编程, 有使用 Python 进行简单开发的经验。同时, 我选修过机器学习相关课程, 具有一定的学习经验。
2. **已有资源:** 我搜集到了一些关于本地问答系统的参考文档以及部署教程。

七、参考资料

- **本地知识库问答:** [基于 LangChain+LLM 的本地知识库问答: 从企业单文档问答到批量文档问答 langchain 本地知识库-CSDN 博客](#)

- **本地知识库问答:** [基于大语言模型的本地知识库问答（离线部署） 基于本地大语言模型的知识库-CSDN 博客](#)
- **本地知识库问答:** [【本地知识库】本地知识库+语言大模型=知域问答-CSDN 博客](#)
- **LangChain:** [构建检索增强生成（RAG）应用 | LangChain 中文](#)

八、自我评价

从可行性方面来说，该项目的架构清晰且简洁，并且项目实现的复杂度适中，既有一定的挑战性，又不会过于困难，无法完成预期的目标。

综合项目目前进展，我给自己的打分为 4/10 分。

九、诚信承诺

本人郑重承诺在项目实施的过程中不发生任何不诚信现象，一切不诚信所导致的后果均由本人承担。

签名：

黄亮铭