

大模型面经——Langchain总结

原创 喜欢瓦力的卷卷 瓦力算法学研所 2024年08月25日 10:00 广东

◇◇ 面试总结专栏 ◇◇

本篇介绍Langchain相关面试题。

本次将会分为上下两个部分，本篇章将会介绍前三个问题，下一次在将后三个问题补充完毕。

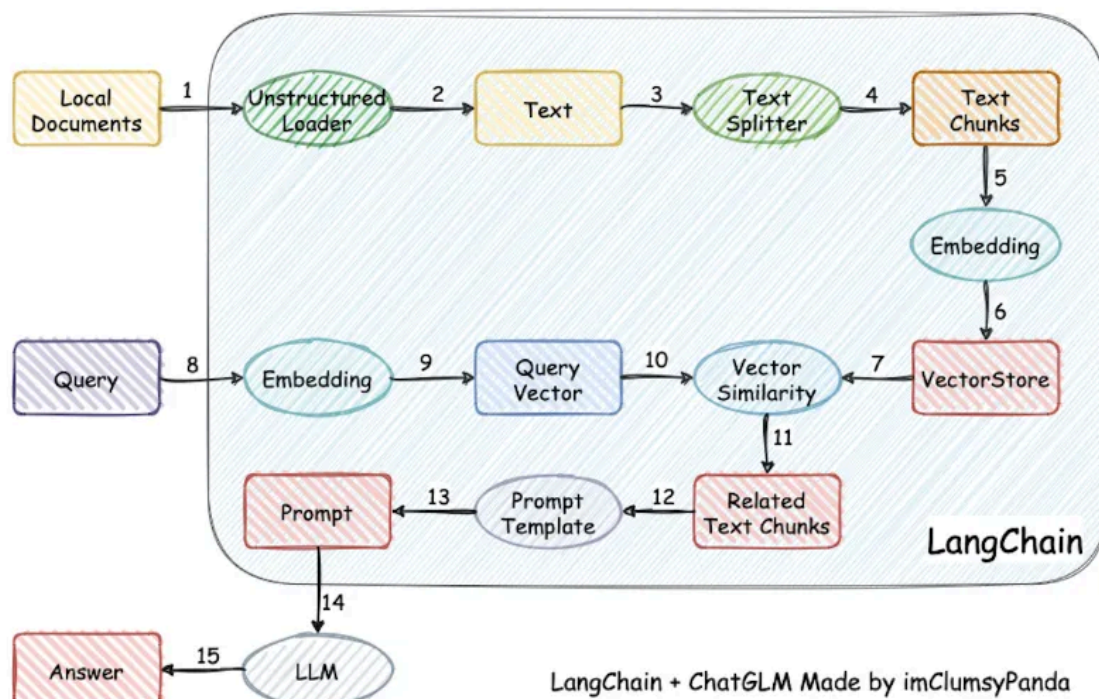
以下是一个快捷目录：

1. 什么是 LangChain?
2. LangChain 包含哪些部分?
3. LangChain 中 Chat Message History 是什么?
4. 介绍一下LangChain Agent?
5. LangChain 如何Embedding & vector store?
6. LangChain 存在哪些问题及方法方案

什么是 LangChain?

LangChain 是一个基于语言模型的框架，用于构建聊天机器人、生成式问答（GQA）、摘要等功能。它的核心思想是将不同的组件“链”在一起，以创建更高级的语言模型应用。LangChain 的起源可以追溯到 2022 年 10 月，由创造者 Harrison Chase 在那时提交了第一个版本。

LangChain 包含哪些部分?



为了能够帮助大家理解，附上LangChain的流程图，如上所示。从上图可知，主要包括以下部分：

- **模型 (Models)**：这指的是各种不同的语言模型以及它们的集成版本，例如GPT-4等大型语言模型。LangChain对这些来自不同公司的高级模型进行了概括，并封装了通用的API接口。利用这些API，用户能够方便地调用和控制各个公司的大模型。
- **提示 (Prompts)**：涉及到提示的管理、优化和序列化过程。在大语言模型的应用中，提示词发挥着至关重要的作用，无论是构建聊天机器人还是进行AI绘画，有效的提示词都是不可或缺的元素。
- **记忆 (Memory)**：它负责存储与模型交互时的上下文状态信息。由于模型本身不具备保存上下文的能力，因此在与模型交流时，传递相关的对话内容上下文变得十分必要。
- **索引 (Indexes)**：用于将文档结构化，以便更加高效地与模型进行互动。
- **链 (Chains)**：代表了对各种组件进行的一系列调用操作。
- **代理 (Agents)**：确定模型应当执行哪些操作，它们会管理执行流程并监控整个过程直至任务完成。

LangChain 中 Chat Message History 是什么？

Chat Message History 是 Langchain 框架中的一个组件，用于存储和管理聊天消息的历史记录。它可以跟踪和保存用户与AI之间的对话，以便在需要进行检索和分析。Langchain 提供了不同的Chat Message History 实现方式，包括以下：

- **StreamlitChatMessageHistory**：用于在 Streamlit 应用程序中存储和使用聊天消息历史记录。它使用 Streamlit 会话状态来存储消息，并可以与 ConversationBufferMemory 以及链或代理一起使用。

- **CassandraChatMessageHistory**: 使用 Apache Cassandra 数据库来存储聊天消息历史记录。Cassandra 是一种高度可扩展和高可用的 NoSQL 数据库，适用于存储大量数据。
- **MongoDBChatMessageHistory**: 使用 MongoDB 数据库来存储聊天消息历史记录。MongoDB 是一种面向文档的 NoSQL 数据库，使用类似 JSON 的文档进行存储。

想要获取技术资料的同学欢迎关注公众号，进群一起交流~



喜欢卷卷的瓦力

扫一扫上面的二维码图案，加我为朋友。

添加瓦力微信

算法交流群 · 面试群

大咖分享 · 学习打卡

🗨️ 公众号 · 瓦力算法学研所



瓦力算法学研所

我们是一个致力于分享人工智能、机器学习和数据科学方面理论与应用知识的公众号。我...
117篇原创内容

公众号

面试 2 模型 13

面试 · 目录

下一篇 · 大模型面经——Langchain总结（下）