

大模型面经——Langchain总结（下）

原创 喜欢瓦力的卷卷 瓦力算法学研所 2024年09月05日 17:57 广东

◇◇ 面试总结专栏 ◇◇

本篇将继续介绍Langchain相关面试题。

本次将继续关于Langchain的面试题回答，主要是对后三个问题的补充。

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

介绍一下LangChain Agent?

- LangChain Agent 是 LangChain 框架中的一个关键组件，专门用于构建和管理对话代理。这些代理负责根据当前的对话情况来确定接下来的操作步骤。LangChain 提供了一系列创建代理的方式，例如 OpenAI Function Calling、Plan-and-execute Agent、Baby AGI 和 Auto GPT 等。每种方式都拥有不同级别的自定义能力和功能，以便于开发者根据需求构建相应的代理。
- 代理能够借助工具包来执行特定的任务或动作。工具包实际上是一组供代理使用的工具集合，旨在执行特定的功能，比如语言处理、数据处理以及外部 API 的集成。这些工具既可以是定制开发的，也可以是预先定义好的，覆盖了多种功能。以下为代码中集成的工具包：

```
"""Core toolkit implementations."""

from langchain.tools.base import BaseTool, StructuredTool, Tool, tool
from langchain.tools.bing_search.tool import BingSearchResults, BingSearchRun
from langchain.tools.ddg_search.tool import DuckDuckGoSearchResults, DuckDuckGoSearchRun
from langchain.tools.file_management.copy import CopyFileTool
from langchain.tools.file_management.delete import DeleteFileTool
from langchain.tools.file_management.file_search import FileSearchTool
from langchain.tools.file_management.list_dir import ListDirectoryTool
from langchain.tools.file_management.move import MoveFileTool
from langchain.tools.file_management.read import ReadFileTool
from langchain.tools.file_management.write import WriteFileTool
from langchain.tools.gmail import (
    GmailCreateDraft,
    GmailGetMessage,
    GmailGetThread,
    GmailSearch,
    GmailSendMessage,
)
from langchain.tools.google_places.tool import GooglePlacesTool
from langchain.tools.google_search.tool import GoogleSearchResults, GoogleSearchRun
from langchain.tools.google_serper.tool import GoogleSerperResults, GoogleSerperRun
from langchain.tools.human.tool import HumanInputRun
from langchain.tools.ifttt import IFTTTWebhook
from langchain.tools.metaphor_search import MetaphorSearchResults
from langchain.tools.openapi.utils.api_models import APIOperation
from langchain.tools.openapi.utils.openapi_utils import OpenAPISpec
from langchain.tools.openweathermap.tool import OpenWeatherMapQueryRun
from langchain.tools.playwright import (
    ClickTool,
    CurrentWebPageTool,
    ExtractHyperlinksTool,
    ExtractTextTool,
    GetElementsTool,
    NavigateBackTool,
    NavigateTool,
```

LangChain 如何Embedding & vector store?

LangChain 框架提供了一种名为 "Embedding & vector store" 的方法，用于将文本数据转换为向量表示形式，并将其存储在向量数据库中。这种方法可以有效地处理自然语言处理任务，如文本分类、情感分析、命名实体识别等。以下是 LangChain 中 Embedding & vector store 的基本步骤：

- 文本预处理**：首先，对输入的文本进行预处理，包括去除停用词、标点符号、数字等，并进行分词或标记化操作。
- 嵌入模型选择**：选择合适的嵌入模型，例如 Word2Vec、GloVe、BERT 等。这些模型可以将单词或短语转换为固定长度的向量表示。
- 嵌入计算**：使用选定的嵌入模型，将预处理后的文本转换为向量表示。每个单词或短语都会被映射到一个向量空间中的点。
- 向量存储**：将生成的向量表示存储在一个向量数据库中，例如 Elasticsearch、Milvus 等。这样可以方便地查询和检索相似性较高的文本片段。
- 相似性搜索**：当需要根据给定的查询文本找到相似的文本时，可以使用向量数据库提供的相似性搜索功能。通过计算查询文本与数据库中已有向量的相似度，可以找到最相关的文本。

LangChain 存在哪些问题及方法方案

1. **LangChain 缺乏标准的可互操作数据类型问题。** LangChain为开发人员提供了一个标准化接口，以便利用大型语言模型（LLM）来执行自然语言处理任务。尽管LangChain能够处理较复杂的数据结构，它当前并不支持标准可互操作的数据类型。因此，在使用LangChain处理数据时，可能需要执行一些额外的处理和转换步骤。
2. **LangChain 太多概念容易混淆，过多的“辅助”函数问题。** LangChain包含了许多不同的模块和组件，每个模块都有其特定的功能。例如，有一些模块专门用于文本预处理、词嵌入、序列标注等任务。这些模块之间可能存在重叠的功能。LangChain提供了许多辅助函数，这些函数可以帮助用户更方便地实现各种功能。然而，过多的辅助函数可能会导致混淆，因为有些函数可能具有相似的名称或功能，但在不同的上下文中使用方式不同。此外，一些辅助函数可能需要额外的参数或配置，这可能会增加学习曲线。
3. **复杂性。** 从实际使用体验来看，LangChain框架并不完美。例如，LangChain中的提示词模板基本上只是对字符串的封装，但该框架提供了多种类型的提示词模板，它们之间的差异并不明显，这带来了一定的冗余。此外，这些模板缺乏安全性控制，并且存在较多的冗余。有些提示词默认是预先设定的，若要进行修改，用户需要查阅源代码以确定修改位置，这增加了使用的复杂性。

想要获取技术资料的同学欢迎关注公众号，进群一起交流~



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



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

面试 2 模型 13

面试 · 目录

上一篇 · 大模型面经——Langchain总结