

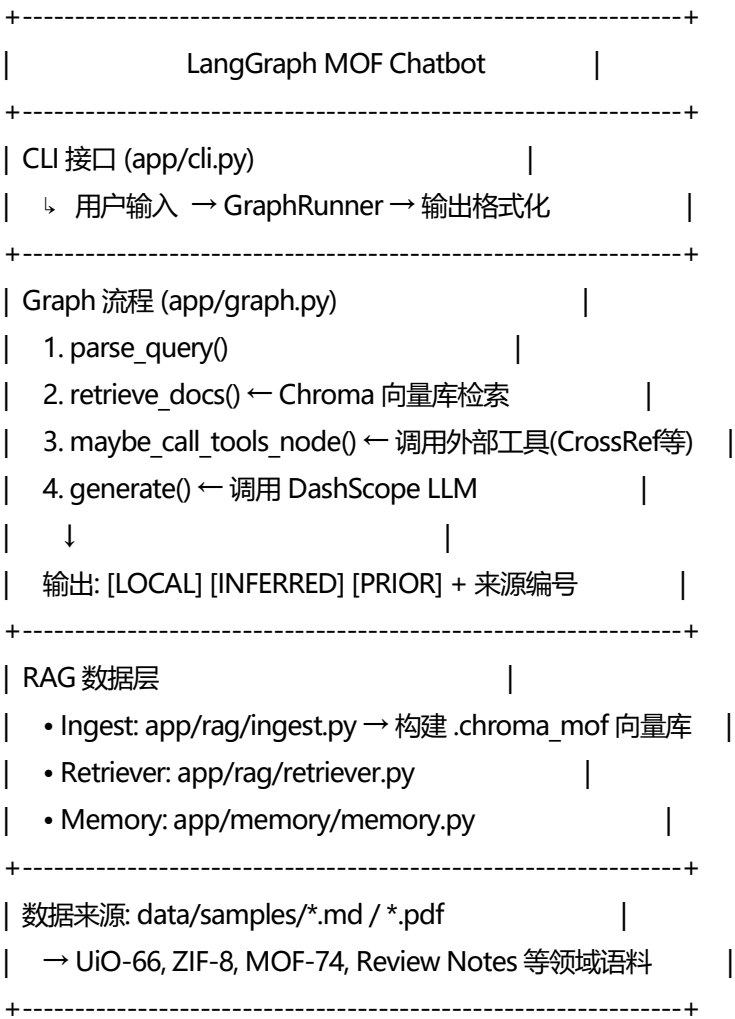
LangGraph MOF Chatbot 项目报告

1. 设计思路与总体架构

本项目旨在构建一个能够基于文献知识库进行问答的 MOF (Metal -Organic Frameworks) 领域 Chatbot，实现对科研资料（如 UiO-66、ZIF-8、MOF-74 等）的智能检索、阅读与解释。

核心思路：结合 LangGraph + LangChain + Chroma + DashScope (通义千问) 的 RAG (Retrieval-Augmented Generation) 流程，让模型在“读取知识库后再回答”。

系统总体架构：



2. 使用的技术与实现方法

模块	技术/库	功能说明
核心框架	LangGraph	构建可视化的对话状态流，定义节点与边，实现任务管线化
语言模型	ChatOpenAI (base_url 指向 DashScope)	使用通义千问 qwen-turbo 模型，支持中文科研对话
嵌入模型	DashScopeEmbeddings("text-embedding-v1")	向量化语料，支持中文与科技文本
向量数据库	Chroma	存储与检索嵌入；支持 persist_directory=.chroma_mof
文本切分	RecursiveCharacterTextSplitter	分块构建检索单元 (chunk size=600, overlap=120)
内存系统	Memory 类	记录上下文问答轮次，用于多轮对话
CLI 交互	typer + rich.console	提供命令行问答界面，彩色输出
系统提示词	多段模板 [LOCAL]/[INFERRED]/[PRIOR]	强制模型输出可追溯三段结构

3. 已实现的功能与流程介绍

功能	描述
语料导入	通过 <code>python -m app.rag.ingest</code> 自动扫描 <code>data/samples</code> 目录的 Markdown/PDF 文件并构建向量库
问答检索	用户输入问题后自动检索最相关文档段落
三段式回答	输出 [LOCAL]（来自文献）、[INFERRED]（逻辑推导）、[PRIOR]（模型常识）三段
来源编号映射	每条事实标注 [L1][L2] 并在末尾列出路径
严格模式	<code>--strict</code> 启动时禁用 PRIOR，仅允许本地信息
工具扩展	<code>maybe_tool_call</code> 支持后期接入 CrossRef、arXiv、ADS API 等文献外部搜索
调试输出	控制台打印检索命中 (hits) 和来源文件，便于验证

典型运行流程

- 语料导入:`python -m app.rag.ingest --input-dir ./data/samples --persist-dir ~/.chroma_mof`
- 启动 Chatbot:`python -m app.cli --persist-dir "$(pwd)/.chroma_mof"`
- 示例问答:

You > UiO-66 药物负载的典型范围和影响因素?

Bot >

[LOCAL]

- DOX (多柔比星) 负载常见 10–30 wt% ...

[INFERRED]

- 孔径约 6–12 Å 适合装载相近尺寸药物 [依据: L1]

[PRIOR]

- 药物分子间相互作用及 pH 效应属于药理学层面 (不在本地语料)

来源 (编号→路径):

[L1] UiO-66_drug_loading_notes.md

[L2] MOF_drug_delivery_review.md

4. 项目代码与资源

项目内容	链接
GitHub Repository	https://github.com/junzheyaojiezhao/mofbot
主要目录结构	app/, data/, .chroma_mof/, requirements.txt
可执行命令	python -m app.cli、python -m app.rag.ingest
环境要求	Python ≥ 3.10 , langchain, langgraph, dashscope, chromadb, typer, pypdf

演示内容建议（2-3分钟）： 1) 开场：介绍项目目的与RAG架构。 2)

展示：CLI问答交互、结果输出格式。 3) 总结：未来可扩展CrossRef实时检索。

6. 总结

LangGraph MOF Chatbot 实现了一个可追溯、结构化的科研问答系统。

系统基于RAG范式，结合了知识检索与生成能力，能够回答金属有机框架（MOF）相关问题，并保持可验证的信息来源输出，为科研类 Chatbot 架构提供了一个高质量的参考实现。