

周莎俊

现居北京 | 18858790598 | 每周可实习 4-5 天 | 可现场办公 | 可连续实习 6 个月 | 最快一周内到岗

教育背景

2025.09-2028.06

北京城市学院

软件工程（硕士研究生，2028 届毕业）

个人优势

软件工程硕士在读（2028 届），具备企业级 RAG、知识图谱及 AI 应用后端项目经验，参与过文档解析、检索重排、Query Planner、知识抽取与接口开发，熟悉 Python、FastAPI、LlamaIndex、Milvus、Spring Boot，具备独立开发与问题排查能力。

专业技能

大模型应用： Python、FastAPI、LlamaIndex；了解 LangChain、LangGraph、Tool Calling、任务规划与多轮上下文管理。

RAG / 检索： 文档解析、语义分块、Embedding、Milvus、Top-K 召回、BGE Reranker、来源追溯。

知识图谱： Schema 设计、实体/关系抽取、实体对齐及图谱增强检索。

工程开发： Java、Spring Boot、MySQL、Git、Docker、Linux，具备接口调试、日志排查经验。

项目经历

2026.01-2026.08 | 医疗器械企业级知识服务中台

项目描述：面向企业内部文档问答场景，参与建设基于 LlamaIndex 的 RAG 知识服务中台，覆盖文档解析、语义分块、索引构建、向量检索、重排筛选、答案生成、多轮对话上下文管理和接口化调用，为企业知识查询与业务问答提供可检索、可追溯能力。

RAG 端开发

- 基于 Python、FastAPI、LlamaIndex 与 Milvus 搭建企业知识库 RAG 服务，支持 DOCX/PDF/XLSX 多格式文档解析、增量建库、检索重排、LLM 生成与 SSE 流式问答，并返回文档及章节级来源信息。

- 针对 Word 标题、段落和表格设计结构感知解析，构建原文、关键词、摘要及文档/章节分层节点，通过稳定 parent_id 与 documentId 关联多视图数据，支持单文件增量建库和来源追溯。

- 设计并实现 Query Planner，定义 9 类意图、5 种查询范围和 4 种检索策略，支持复杂问题拆解、指代补全及模型失败后的规则回退，并按细节问答、章节概述和全文总结路由由向量优先、摘要优先或混合检索。

- 优化检索与数据一致性：组合 Milvus 多路召回、多视图加权聚合、父块回填、Reranker 精排与阈值过滤；增加文档级权限过滤，并通过带锁原子 upsert 及并发、异常测试避免侧存跨文档覆盖

Java 业务后端

技术栈：Java 17、Spring Boot 3、MyBatis-Plus、MySQL、WebClient、ApachePOI

- 基于 Spring Boot 分层架构开发 AI 文档业务后端，完成检测报告绑定、异步解析、生成记录持久化等 REST 接口，并通过 WebClient 对接 Python RAG 服务。

- 将业务要求转化为数据校验规则：仅汇总存在服务端生成快照且四部分正文完整的测试项，使用来源哈希、数量与生成时间维护过期状态，阻止失效内容继续预览或导出。

- 基于 Apache POI 实现复杂 Word 模板导出，按照前端 rows + cellMeta 数据还原多级表头、合并单元格及标题/正文样式，并通过单元测试、Maven 构建、日志和接口响应完成验证。

2025.09-2025.12 | 知识图谱构建

北科院技术情报监测平台

- 围绕方法模型、理论原理、研究问题等概念参与领域本体与 Schema 设计，定义实体类型、语义关系、字段约束及知识抽取规则。

- 基于 LLM 与 Prompt Engineering 完成实体识别、关系抽取及结构化结果整理，并针对字段缺失、类型不一致、关系方向错误等异常进行规则校验。

- 针对技术名词同义词、缩写及中英文异名，通过语义向量召回候选实体并结合名称、类型规则完成实体对齐，同时完成图数据入库、索引配置及查询验证。