

2025 商网智能问答 项目开发文档

项目介绍

基于各地市政府已有的政府文件、faq 问答等数据，利用 LLM 技术完成智能问答系统的升级迭代（需要支持多轮问答），具体可参照黄山市的智能问答（这个是科大讯飞做的）



<https://www.huangshan.gov.cn/>

数据源：政府公文数据（doc、pdf等非结构化），部分faq数据，办事服务指南（半结构化数据，类似表格），部分新闻类性质数据。

数据总量：非结构化数据总量在10000条附近，数据量不多，但是覆盖面较杂

遇到过的问题：

- 1、数据源较为杂的情况下，去年用的召回方法效果不够稳定。用的 ES 文本 + 向量召回，向量基于 bge-base，rerank 是基于业务思维写的，没有上 bge 去做。目前资源够的情况下看可以整体优化
- 2、部分文件的结构较为复杂，包含表格图片等数据，简单的解析方法无法处理部分数据。这里可以用 deepdoc 这种方法去拓展解析的能力
- 3、政府数据存在说新老交替的情况，比如 23 年发的一个文 A，24 年会少量修改后再发一次。目前的数据库里面没有这类的关系，去年给的解决方案是需要建立一个这样的关系（因为这种文的标题不一定就一样，从标题判断很容易遗漏），但是这个方案推行起来客户那边有困难，不知可有其他解决方案
- 4、多轮问答，目前是基于问题改写做的；
- 5、敏感话题屏蔽，比如各种涉敏，涉政的东西，目前是从 prompt 里面处理的；

数据样例

<https://xc.ahzwfw.gov.cn/>

<https://www.xuancheng.gov.cn/Xzgfwjk/show/2293113.html>

<https://www.xuancheng.gov.cn/Service/show/4028e6a6901628f801902e718e2b6586/fd1f98f343dd4342818f1c16303950a1.html>

项目时间规划

项目计划表						
序号	名称	我方事项	责任人	商网事项	责任人	时间计划
1	第一周（第一阶段）	Embedding模型部署及API功能测试		2卡4090云服务到位，用于模型部署测试（价格太贵可以1卡，但是效果跟后续情况差距较大，不是很建议） 最晚周三到位		
		Reranker模型部署及API功能测试				
		Deepseek部署及速度测试，推理加速功能测试				
		完成开发用LLM模型部署及相应问答调用功能接口调试，具体包含如下： 1、文档总结能力API，用于在预处理的过程中提供相应支持，提升后续召回的准确度 2、问答生成API		一套ES服务，用于存储预处理后的数据，本地或者云服务都可以，如果是云服务的话就部署在GPU机器上		
		数据处理功能部署，ES增删改查功能 完成数据预处理功能开发测试，同工程侧联调		数据处理接口对接		
		办事服务数据处理入库 12345数据处理入库				
2	第二周（第二阶段）	完成层次召回功能开发，召回效果测试		问答接口对接、联调，可以支持初步问答测试		
		prompt构建、初步测试问答功能及效果		完成前端界面，可用于问答测试使用		
		政策文件处理优化入库，尽量支持表格等复杂数据处理（入库80%）		问答交互中，不同数据类型问答情况下的业务交互逻辑分析，可参照贵山		
		问答的合规性测试及规范性调整				
3	第三周（第三阶段）	政策文件相关问答功能测试及整体效果测试、数据处理方案优化		功能测试，对于效果差的一些case提出反馈，同算法端一起判断case是否合理以及期望方向等一系列问题。		
		多轮问答功能实现及测试				
		召回效果及情况分析、相关召回算法优化				
4	第四周（第四阶段）	prompt调试及问答效果优化		协助完成最终测试，功能调整对接		
		市长之窗数据接入（尽量完成）				
		bad case原因分析及算法调整				
		政策文件全量入库（剩余20%），模拟测试数据更新情况下的效果及问答召回整体稳定性				

项目后端模块设计

模块名称	子模块名称	接口名称	功能
大模型	本地 DeepSeek 大模型调用	/v1/algo/llm	
	云端大模型调用		
算法模块	文本编码	/v1/algo/embedding	
	文本重排序	/v1/algo/rerank	
	关键词提取	/v1/algo/keyword	
	文本分类	/v1/algo/classify	
	内容提取	/v1/algo/ocr	
数据模块	插入文档	/v1/doc	
	更新文档 / 删除文档 / 获取文档	/v1/data/doc/{doc_id}	
	插入 faq	/v1/data/faq	
	更新 faq / 删除 faq / 获取 faq	/v1/data/doc/{faq_id}	
问答模块	智能问答	/v1/biz/chat	
	删除对话 / 获取对话	/v1/biz/chat/{chat_id}	
	提问联想	/v1/biz/similar-question	
	常见问题	/v1/biz/common-question	
	历史对话记录	/v1/biz/history-chat	