

行业知识问答助手 PRD（产品需求文档）

1. 产品概述

行业知识问答助手是一款基于 FastAPI 后端与 Streamlit 前端实现的文档知识库问答系统。用户可上传行业相关文档，系统会自动解析并向量化入库，随后通过对知识库的检索，实现基于上下文的专业问答。

本系统支持多轮会话、文件入库、知识库检索、智能问答等核心功能。

2. 产品目标

- 提供一个轻量、可扩展的行业知识库问答解决方案
- 支持 PDF / DOCX / TXT 等多格式文档入库
- 支持基于向量检索的精准知识问答
- 支持多轮上下文对话，使回答更连贯
- 提供简洁易用的 Web 可视化界面

3. 用户角色 & 使用场景

● 角色

- 行业工程师 / 操作员**：需要快速查询行业专业知识
- 文档维护人员**：上传行业规范、作说明
- AI 应用开发者**：需要快速搭建知识问答 Demo

● 场景

- 查询行业规范、标准、流程
- 查询项目文档、技术说明书
- 自动化 QA 客服测试环境
- 企业内部知识库问答入口

4.1 文档上传与入库

由 FastAPI 提供服务。`/upload`

功能说明

- 支持 PDF、DOCX、TXT 3 种文档格式
- 自动存储上传文件
- 自动解析文档
- 按块切分 (chunk_size = 800, 重叠=100)
- 使用 DashScope Embedding 进行向量化
- 使用 FAISS 本地向量库存储

上传成功返回

- 文件名
- 切分块数量

4.2 知识问答 (QA)

由 FastAPI 提供服务。`/qa`

功能说明

- 基于 FAISS 检索最相似的文档片段
- 拼接上下文 prompt 进行回答
- 多轮对话 (按 session_id 维护历史)
- 避免幻觉 → 回答严格来源于参考资料
- 若资料未包含 → 明确回答“资料中未提及”

4.3 文件列表

提供 获取当前已上传文件列表。`/files`

Streamlit 前端会缓存文件列表并美观展示。

4.4 Web 可视化界面

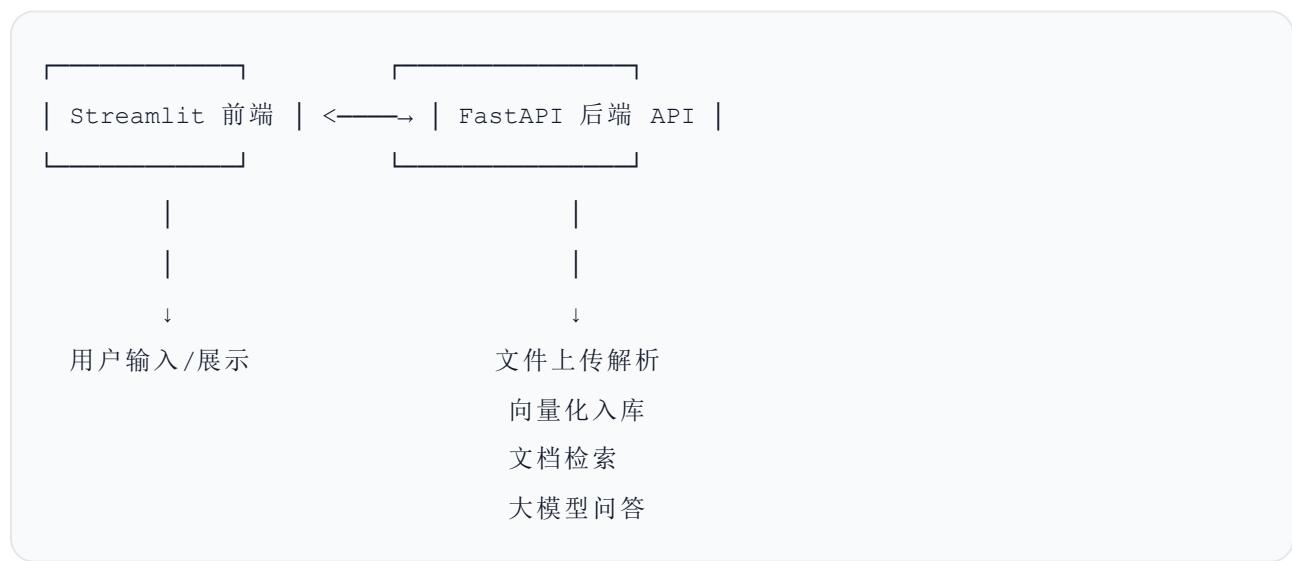
功能模块

- 文档上传区域
- 折叠面板展示已入库文件
- 聊天问答区域（左右气泡样式）
- 自动维护 session_id 与对话历史
- 按钮式界面作

设计亮点

- 气泡式对话框
- 简洁配色与清晰布局
- 自动刷新文件列表
- 使用 UUID 自动维护会话

5. 系统架构



6. API 设计

● 1) 上传文件

```
POST /upload  
Files: List[UploadFile]
```

● 2) 问答接口

```
POST /qa  
query: str  
session_id: str
```

● 3) 文件列表

```
GET /files
```

● 4) 系统状态

```
GET /
```

7. 技术栈

后端

- FastAPI
- 费斯
- LangChain
- DashScope 嵌入
- Qwen 大模型

前端

- Streamlit
- 请求

8. 性能指标

指标	目标
文档解析速度	100页 PDF < 5 秒
检索延迟	< 200 毫秒
单批上传文件数	≥ 10 个
多轮会话上下文	≥ 20 轮

9. 后续规划

- 增加 PDF 图片 OCR 能力
- 增加 图表解析、Excel 文档解析
- 支持在线知识库管理（删除、重建）
- UI 深度优化，增加主题色