



程序设计综合实验开题研讨

# 学生文档助手

中道崩组

信息与通信工程学院-北京邮电大学

## 目录

<b>1. 项目背景与需求分析</b>	<b>1</b>
1.1. 背景及用例	1
1.1.1. 第一部分：5W1H <sup>1</sup> 分析需求背景	1
1.1.2. 第二部分：8C 分析需求细节	1
1.2. 关键需求	3
1.2.1. 核心功能需求（Must Have）	3
1.2.2. 扩展性需求（Could Have）	3
1.2.3. 差异化需求（Value Added）	3
<b>2. 系统总体设计</b>	<b>3</b>
2.1. 总体架构	3
2.2. 分模块及动静关系	4
<b>3. 现有成果基础分析</b>	<b>5</b>
3.1. 已完成组件	5
3.2. 待改进项	5
<b>4. 预期成果与价值</b>	<b>5</b>
4.1. 增量目标	5
4.1.1. 核心版本（V1.0）	5
4.1.2. 增强版本（V1.5）	5
4.1.3. 扩展版本（V2.0）	6
4.2. 功能指标	6
4.3. 性能指标	6
4.4. 价值追踪	6
4.4.1. 技术价值	6
4.4.2. 应用价值	7
4.4.3. 创新价值	7
4.4.4. 社会价值	7
<b>5. 团队分工与计划</b>	<b>7</b>
5.1. 人员分工	7

---

<sup>1</sup>程序设计课上提供 下面 8C 同

目录	ii
<hr/>	
5.2. 开发计划 .....	7
参考文献	9

## 1. 项目背景与需求分析

### 1.1. 背景及用例

#### 1.1.1. 第一部分：5W1H<sup>2</sup> 分析需求背景

##### 1. What（是什么）

学生文档助手是一个将 PDF 转换为结构化 Markdown 文本，并通过 Ollama 嵌入模型存入基于 CherryStudio [1]的 SQLite3 数据库的程序。它支持额外的功能，如翻译、总结、概念提取，并对外提供 Restful API。

##### 2. Why（为什么需要）

- 许多学生需要处理大量 PDF 资料，提取关键信息较为困难。
- 现有 PDF 处理工具通常不支持嵌入数据库，难以进行高效检索和分析。
- 额外的功能（翻译、总结等）可以帮助用户更快地理解内容，提高学习效率。

##### 3. Who（谁提出的需求，谁受影响）

- 主要用户：学生、研究人员、教育工作者，希望高效整理和检索 PDF 资料的人。
- 受影响方：
  - 第三方开发者：可能会使用 Restful API 进行集成。

##### 4. When（何时需要）

- 目前项目处于 Windows 平台开发阶段，未来计划移植到其他平台。
- 初版需要完成核心功能(PDF 转 Markdown 并嵌入数据库)，后续迭代增加翻译、总结等附加功能。

##### 5. Where（在哪些场景应用）

- 个人学习：学生整理课程资料，方便搜索和复习。
- 研究工作：研究人员存储论文，进行概念提取和知识关联。
- 教育机构：教师处理教学资料，提高课程准备效率。

##### 6. How（如何实现）

- PDF 解析 → 转换为 Markdown → 支持翻译、总结等处理 → 使用 Ollama[2] 进行嵌入 → 存入 SQLite3。
- 对外提供 Restful API，让开发者可以集成该功能到其他系统。

#### 1.1.2. 第二部分：8C 分析需求细节

##### 1. Customer（客户）

- 核心用户：学生、研究人员。

---

<sup>2</sup>程序设计课上提供 下面 8C 同

- 潜在用户：希望处理和存储 PDF 资料的开发者、公司、教育机构。

## 2. Competitor（竞争对手）

- 现有解决方案：
  - 传统 PDF 处理工具（如 Adobe® Acrobat），但缺少 Markdown 转换和数据库嵌入功能。
  - Notion、Obsidian 插件可以提取 PDF 片段，但不能自动反义嵌入数据库进行语义搜索，同时提取的 PDF 片段不准确。
  - ChatGPT 插件和其他 AI 工具支持 PDF 分析，但通常不提供可供开发者调用的 API。
- 竞争优势：
  - 结构化 Markdown 便于二次整理。
  - Ollama 嵌入模型增强了智能检索能力。
  - 提供 API 方便第三方集成。

## 3. Capability（能力）

- 当前能力：
  - 已支持 PDF → Markdown 转换[3]。
  - 使用 SQLite3 存储数据，基于 CherryStudio 进行数据库嵌入。
- 待提升能力：
  - 未来需要支持其他平台（Linux、Mac）。
  - 可能需要扩展支持更多联网嵌入模型，以适应不同用户需求。

## 4. Cost（成本）

- 开发成本：
  - 主要涉及 PDF 解析、Markdown 转换、API 设计和数据库嵌入。
- 运行成本：
  - 由于使用 SQLite3，本地存储成本较低，但随着数据增长，可能需要优化数据库结构。

## 5. Convenience（便利性）

- 对用户：
  - 提供 Markdown 格式输出，便于编辑和检索。
  - 额外的翻译、总结功能减少手动处理时间。
- 对开发者：
  - 通过 Restful API 提供标准化接口，方便集成到其他应用。

## 6. Communication（沟通）

- 需要详细的 API 文档，确保外部开发者能够快速集成。

## 7. Channel（渠道）

- 提供 Windows 版本下载，后续支持 Linux 和 Mac。

8. Compliance（合规性）

- 提供本地处理模式，避免数据外泄。
- 确保 API 访问权限受控，防止未授权的数据泄露。

1.2. 关键需求

1.2.1. 核心功能需求（Must Have）

1. 精准文档转换
  - 支持 PDF→Markdown 转换准确率≥95%（不含表格/公式场景）
2. 数据持久化
  - 数据库写入成功率 100%（断电保护机制）

1.2.2. 扩展性需求（Could Have）

1. 插件体系支持：
  - 允许第三方开发转换模板（开放 Schema 定义）
  - 支持动态加载 AI 模型
2. 跨平台兼容：
  - Windows/Linux/macOS 平台功能一致性≥98%
  - 系统依赖库版本锁定机制

1.2.3. 差异化需求（Value Added）

2. 智能辅助
  - 辅助翻译 关键词提取解释 内容总结
  - 语义检索（基于 Ollama 模型）
  - 语义检索准确率≥95%（不含表格/公式场景）
  - 为网盘提供翻译预览

2. 系统总体设计

2.1. 总体架构

程序总体架构如 图 1 所示：

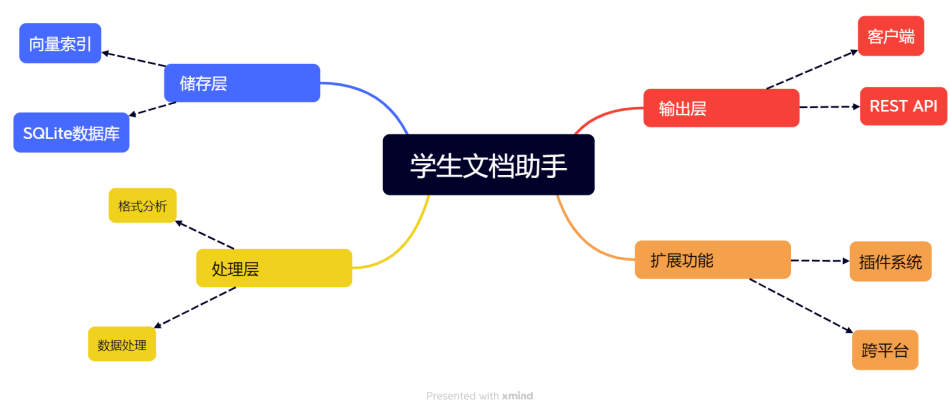


图 1 程序总体架构

2.2. 分模块及动静关系

系统的静态结构和动态行为如下图所示：

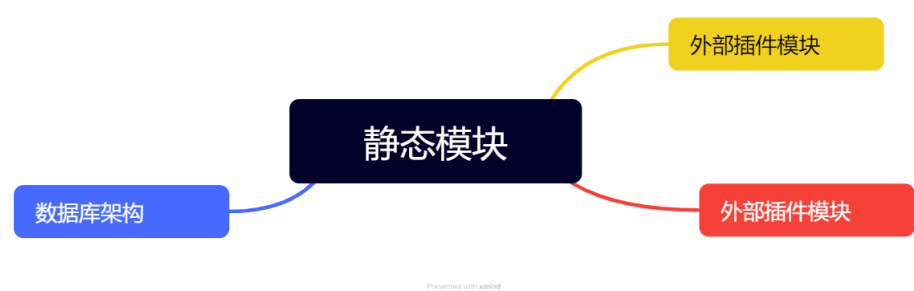


图 2 系统静态结构图

上述 图 2 展示了系统的模块组成，而 图 3 展示了系统运行时的交互流程。

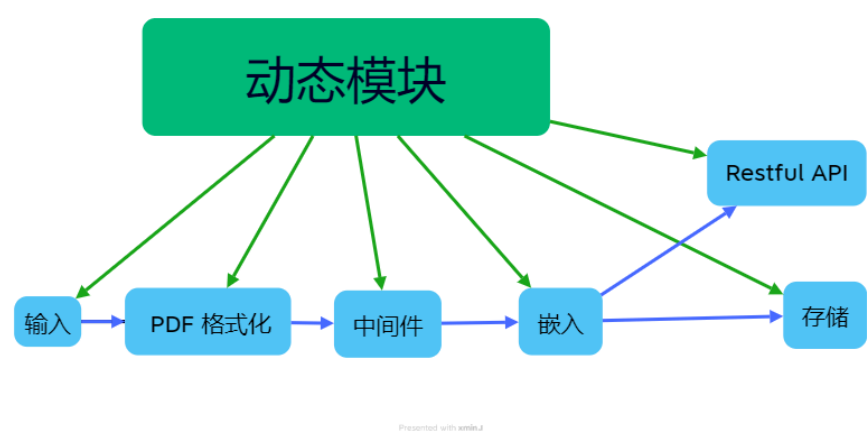


图 3 系统动态行为图

### 整个系统分成了以下几个模块

- PDF 解析模块：负责将 PDF 文档解析为 Markdown 格式。
- 数据存储模块：使用 SQLite3 存储解析后的数据，并提供 API 接口供其他模块调用。
- 插件模块：支持第三方插件的加载和使用，允许用户自定义转换模板。
- API 模块：提供 Restful API 接口，供外部系统调用。
- 嵌入模块：使用 Ollama 嵌入模型进行语义检索和智能辅助功能。

## 3. 现有成果基础分析

### 3.1. 已完成组件

PDF 文本提取模块：

- 使用 [3] 提取 PDF 文本，支持多种格式的 PDF 文档。

PDF 翻译模块：

- 使用翻译模块[4]，支持多种语言的翻译。

Restful API 模块：

- 使用 [5] 提供 Restful API 接口，支持多种请求方式。

### 3.2. 待改进项

- 现在的模块没有整合到一起
- 没有实现嵌入模型的功能

## 4. 预期成果与价值

### 4.1. 增量目标

#### 4.1.1. 核心版本（V1.0）

- 完成 PDF→MD 基础流程
- 实现 SQLite3 嵌入式存储

#### 4.1.2. 增强版本（V1.5）

- 添加翻译（中英互译）模块
- 部署总结生成功能



4.1.3. 扩展版本（V2.0）

- 多平台客户端（Windows/Linux/macOS）
- 可视化检索界面

4.2. 功能指标

功能项	指标要求
PDF 页面解析	支持 $\geq 10$ 页的 PDF 文档
嵌入向量维度	$\geq 384$ 维
数据库容量	支持 $\geq 100\text{MB}$ 文本数据
数据持久化	支持断电保护

4.3. 性能指标

功能项	指标要求
并发写入	支持 $\geq 2$ 个并发写入
检索响应时间	$\leq 3$ 秒/次
翻译响应时间	$\leq 5$ 秒/次
总结响应时间	$\leq 5$ 秒/次
语义检索准确率	$\geq 95\%$
翻译准确率	$\geq 95\%$
总结准确率	$\geq 95\%$

4.4. 价值追踪

4.4.1. 技术价值

- 提供开源的 PDF 文档处理解决方案
- 实现本地化的文档语义检索能力
- 形成可复用的插件开发框架

4.4.2. 应用价值

- 提高学生文献阅读效率
  - 预计可减少 30% 文献整理时间
  - 降低 90% 重复查询工作量
- 优化文档管理流程
  - 支持智能分类与检索
  - 实现知识快速定位

4.4.3. 创新价值

- 结合本地大模型的文档处理方案
- 支持多维度的文档语义分析
- 提供开放的二次开发接口

4.4.4. 社会价值

- 降低文献学习门槛
- 促进知识高效流通
- 支持开源社区发展

5. 团队分工与计划

5.1. 人员分工

成员	主要职责	具体任务
A	技术开发	<ul style="list-style-type: none"><li>• PDF 解析模块开发</li><li>• API 接口实现</li><li>• 辅助文档编写</li><li>• 用户手册撰写</li></ul>
B	技术开发	<ul style="list-style-type: none"><li>• 需求分析文档</li><li>• 接口文档编写</li><li>• 辅助功能测试</li><li>• SQLite3 数据库对接</li></ul>

5.2. 开发计划

时间节点	计划任务	负责人
------	------	-----

3 月	<ul style="list-style-type: none"><li>• 需求分析完成</li><li>• 技术方案确定</li></ul>	A、B
4 月	<ul style="list-style-type: none"><li>• PDF 解析模块开发</li><li>• 数据库设计实现</li></ul>	A、B
5 月	<ul style="list-style-type: none"><li>• API 接口开发</li><li>• 文档撰写</li></ul>	B、A
6 月	<ul style="list-style-type: none"><li>• 成果展示</li></ul>	B

## 参考文献

- [1] 《GitHub - CherryHQ/cherry-studio: 🍒 Cherry Studio is a desktop client that supports for multiple LLM providers. Support deepseek-r1 — github.com》. 2025 年.
- [2] 《GitHub - ollama/ollama: Get up and running with Llama 3.3, DeepSeek-R1, Phi-4, Gemma 3, and other large language models. — github.com》. 2025 年.
- [3] B. Wang 等, 《MinerU: An Open-Source Solution for Precise Document Content Extraction》. [在线]. 载于: <https://arxiv.org/abs/2409.18839>
- [4] 《Helsinki-NLP/opus-mt-en-zh · Hugging Face — huggingface.co》. 2025 年.
- [5] 《FastAPI — fastapi.tiangolo.com》. 2025 年.