

文脉医典——

# 医学文献智能识别与检索系统

小组：第七小组

汇报人：昌宇



目

录

1

项目介绍

2

过程管理与架构设计

3

项目展示

4

总结展望



第一部分  
项目介绍

文脉医典是一款集医学文献检索、智能推荐、深度阅读与知识分享于一体的综合性在线平台。它不仅为医学专业人士（如医生、护士、科研人员、医学生等）提供了便捷、高效的文献获取途径，同时也向广大健康关注者开放，满足他们对医学知识的探索与学习需求。



# 人员分工

## 昌宇

文献管理和组织、数据分析、数据采集、知识图谱、模型调整、系统测试

## 翟菴坤

文献管理和组织、智能文献检索、数据分析、数据采集、知识图谱、模型调整、系统测试



## 谭雪瑶

文献管理和组织、用户管理、数据分析、知识图谱、模型调整、系统测试

## 饶艺

文献展示与阅读、文献管理和组织、智能文献检索、数据分析、知识图谱、模型调整、系统测试

# 项目优势

## 高效精准的文献检索

提供智能推荐和过滤功能，帮助用户快速筛选出最相关的研究成果

## 用户管理功能完善

方便用户管理自己的账户和个人信息，保障用户账户的安全性和使用的便利性。

## 文献展示与阅读体验

能够显示文献的基本信息，如标题、作者、发布日期等，让用户快速了解文献的概况。

## 数据分析深入

涵盖了用户日活报表、功能页面分析、用户点击率分析等多个维度，可以深入了解用户的使用行为和习惯。

## 数据采集与清洗全面

数据爬取和数据清洗为后续的检索和分析提供高质量的数据基础。

## 技术栈成熟稳定

主题模型采用 LDA 和 LSI，帮助用户更好地发现相关文献，提升检索的准确性和相关性。



第二部分  
过程管理与架构设计

## 一、项目启动阶段

明确要为特定期刊建立文献检索网站的目标，划定功能与技术范围，召集成员分配职责，制定包含时间与资源规划的项目计划，开启项目进程。

## 二、需求分析阶段

多渠道收集各方对文献检索网站的需求，梳理分类后确定优先级，组织评审完善需求内容，形成详尽准确、指导开发的需求规格说明书。

## 三、系统设计阶段

规划网站前端、后端及数据库架构，设计用户界面交互流程，完成数据库结构搭建，经严格评审确定最终设计文档，指引开发工作。

## 四、开发阶段

前端开发人员运用相关技术构建页面实现交互，后端人员用 Python 和 Flask 编写业务逻辑及数据库操作代码，边开发边集成并进行单元测试。

## 五、测试阶段

依据需求和设计制定全面测试计划与用例，严格执行各类测试，详实记录缺陷问题，编写分析报告，依此评估系统质量与可靠性。

## 六、运行维护

部署系统并完成配置与初始化，充分测试验证，确保运行无误

## 七、项目完成

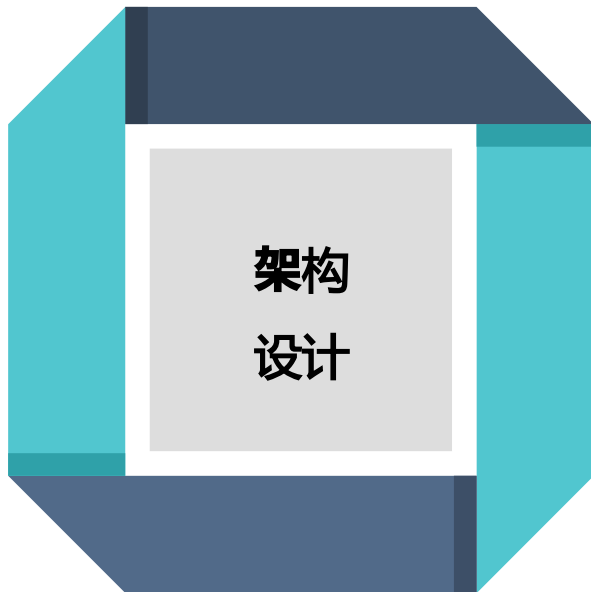
全面总结项目经验，涵盖技术应用、团队协作、需求管理等方面；整理项目成果，包括代码、文档、报告等资料；评估项目成效，分析目标达成度与用户满意度，为后续项目提供宝贵借鉴，正式宣告项目圆满结束。

## 前端

采用 **HTML5**、**CSS**、**JavaScript** 结合 **XHR** 技术，确保页面布局合理、样式美观且交互流畅，为用户提供友好的操作界面。

## 后端

选用 Python 语言搭配 **Flask** 框架，Python 的简洁性与丰富库支持，加上 Flask 的轻量级和灵活性，能高效处理业务逻辑与数据请求。利用 `py-mysql` 实现与 MySQL 数据库的稳定连接与数据交互。



## 数据库

**MySQL** 数据库具备良好性能、可靠性和扩展性，可有效存储与管理大量文献数据，支持复杂查询与数据关联操作。

## 主题模型

运用 **LDA** 和 **LSI**，深入挖掘文献主题结构，提升检索准确性与相关性，助力用户精准定位所需文献。

## LDA-- 潜在狄利克雷分配

LDA 是一种基于概率图模型的主题模型。它假设文档是由多个主题混合而成，每个主题由一系列单词的概率分布表示。

它的工作原理是利用贝叶斯推理，根据文档中单词的共现情况，计算出每个文档属于各个主题的概率分布以及每个主题下单词的概率分布。这有助于用户在检索文献时，快速定位到与特定主题相关的内容，提高检索的准确性和相关性，让用户能更高效地获取所需的医学知识。

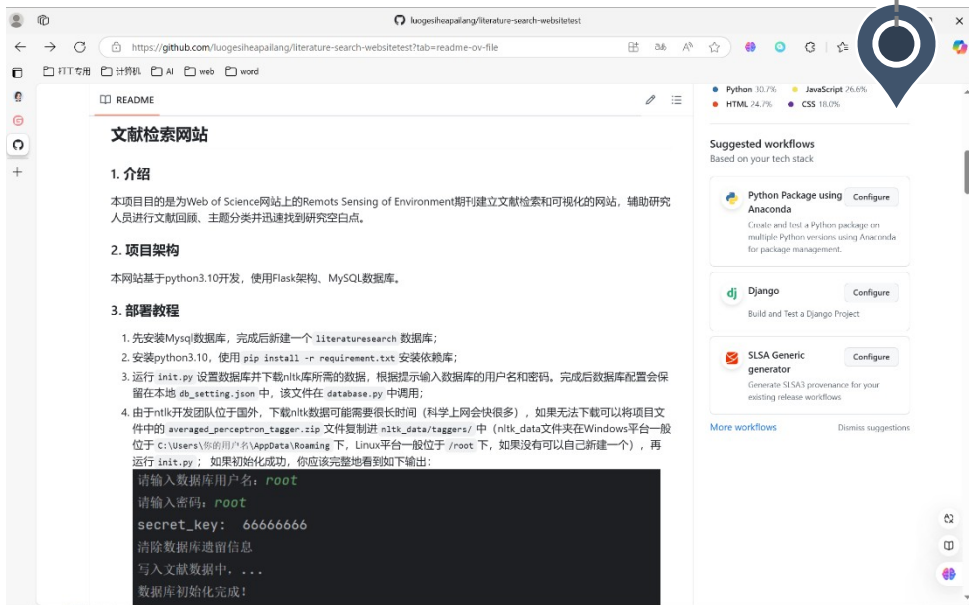


## LSI-- 潜在语义索引

LSI 则是通过对文本矩阵进行奇异值分解（SVD）来提取语义信息。它将文档和单词映射到一个低维的语义空间中，在这个空间里，语义相近的单词和文档会距离更近。

在医学文献处理中，LSI 能够发现词语之间的潜在语义关系，即使它们没有直接的词汇关联。这使得系统在检索时能够跨越词汇的表面差异，为用户提供更全面、更符合语义逻辑的检索结果，帮助用户发现更多有价值的文献资料，拓宽研究视野。

利用 Git 进行代码管理，团队成员提交代码到 Github 仓库的不同分支。每次提交都应遵循清晰的提交信息规范，便于追溯代码变更历史和理解修改意图。通过分支策略确保代码的有序整合与管理。



The screenshot shows a GitHub repository page for 'luogesheapalang/literature-search-websitetest'. The main content is the README file, which is titled '文献检索网站' (Literature Search Website). The README is divided into three sections: '1. 介绍' (Introduction), '2. 项目架构' (Project Architecture), and '3. 部署教程' (Deployment Tutorial). The '部署教程' section contains a list of steps for setting up the project, including installing MySQL, Python 3.10, and running the initialization script. A terminal output snippet is shown at the bottom of the README, displaying the results of the initialization process. On the right side of the page, there are 'Suggested workflows' for Python Package using Anaconda, Django, and SLSA Generic generator.

## 文献检索网站

### 1. 介绍

本项目目的是为Web of Science网站上的Remots Sensing of Environment期刊建立文献检索和可视化的网站，辅助研究人员进行文献回顾、主题分类并迅速找到研究空白点。

### 2. 项目架构

本网站基于python3.10开发，使用Flask架构、MySQL数据库。

### 3. 部署教程

1. 先安装Mysql数据库，完成后新建一个 literaturesearch 数据库；
2. 安装python3.10，使用 pip install -r requirement.txt 安装依赖库；
3. 运行 init.py 设置数据库并下载nlk库所需的数据，根据提示输入数据库的用户名和密码，完成后数据库配置会保留在本地 db\_setting.json 中，该文件在 database.py 中调用；
4. 由于nlk开发团队位于国外，下载nlk数据可能需要很长时间（科学上网会快很多），如果无法下载可以将项目文件中的 averaged\_perception\_tagger.zip 文件复制进 ntk\_data/taggers/ 中（nlk\_data文件夹在Windows平台一般位于 C:\Users\你的用户名\AppData\Roaming\下，Linux平台一般位于 /root 下，如果没有可以自己新建一个），再运行 init.py；如果初始化成功，你应该完整地看到如下输出：

```
请输入数据库用户名: root
请输入密码: root
secret_key: 66666666
清除数据库遗留信息
写入文献数据中, ...
数据库初始化完成!
```

### Suggested workflows

Based on your tech stack

- Python Package using Anaconda
- Django
- SLSA Generic generator



第三部分  
项目展示



第四部分  
总结展望



## 项目 前景

### 竞争优势

技术专业性强，可解决行业痛点，文献管理功能完善。注重用户体验，持续优化性能与功能，能在市场竞争中吸引和留住用户。

### 市场需求

科研与行业对专业文献检索需求大且增长快，本项目专注特定领域，能满足科研人员及相关企业需求，用户群体有望持续扩大。

### 技术发展趋势契合度

项目紧密结合人工智能技术在文献处理中的应用趋势，利用大数据分析挖掘用户行为与文献知识关联。

在本次项目中，技术能力上实现多技术融合但仍需紧跟前沿提升，如优化文献处理技术；团队协作分工明确但沟通和应急机制待完善；实践经验积累了开发与运维流程经验，但进度把控和运维策略需加强，整体为后续项目提供重要参考与改进方向。

## 技术能力

本项目成功融合 HTML5、CSS、JavaScript、Python、Flask、MySQL 及主题模型技术，构建起功能完备的文献检索系统。在开发中深刻体会到各技术间协同作用。但也认识到技术更新换代快，需持续学习新特性与最佳实践。

## 团队协作

明确分工促进高效开发。项目启动时清晰界定各成员职责，确保每个环节都有专人负责，避免职责不清导致的混乱与延误。团队成员间相互学习与支持是项目成功关键。通过定期技术交流与问题讨论，实现知识共享与经验传承。

## 实践经验

项目实践中积累丰富开发流程经验。从需求调研、系统设计、编码实现到测试上线，严格遵循软件工程规范，确保项目有序推进。

# 感谢您的聆听!

文脉医典 -- 医学文献智能识别与检索系统

