**武汉大学**

**计算机学院实训**

**万卷书项目需求说明书**

**项目成员：赵一波、雷明康、李梓嘉、韩欣星**

**撰 写 人：雷明康**  2025 年 6 月24日

**评 审 人：**  年 月 日

武汉凡诺软件技术有限公司

2025年

**修订记录**

| 日期 | 修订版本 | 修改章节 | 修改描述 | 修订人 |
| --- | --- | --- | --- | --- |
| 2025.6.25 | 1.2 | 3 | 用例图修改 | 雷明康 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**目 录**

[万卷书项目需求说明书 1](#_Toc25827)

[1. 简介 1](#_Toc21947)

[1.1 编写目的 1](#_Toc18535)

[1.2 范围 1](#_Toc6556)

[2. 总体概述 1](#_Toc4674)

[2.1 软件概述 1](#_Toc27571)

[2.2 项目介绍 2](#_Toc15526)

[2.3 产品环境介绍 2](#_Toc31658)

[2.4 软件功能 2](#_Toc3064)

[3. 功能需求 2](#_Toc2602)

[3.1 用例图 2](#_Toc31921)

[3.2 系统模块 3](#_Toc11895)

[4. 性能需求 4](#_Toc19204)

[5. 接口需求 5](#_Toc13807)

[6. 用户接口需求 6](#_Toc28907)

[7. 总体设计约束 6](#_Toc21614)

[8. 其他需求 6](#_Toc5491)

[9. 需求分级 6](#_Toc21734)

[10. 待确定问题 6](#_Toc32632)

关键词：

摘 要：

缩略语清单：

| 缩略语 | 英文全名 | 中文解释 |
| --- | --- | --- |
| 万卷书 | Academic Paper Assistant Tools | 论文辅助工具 |
|  |  |  |
|  |  |  |
|  |  |  |

# 

# 简介

## 1.1 编写目的

本文档旨在介绍万卷书这一论文辅助工具的项目起源，功能，使用环境，以及性能需求，并附带展示软件实际开发中的设计用例图，系统模块，开发过程接口需求以及设计约束，并就需求进行分级，总体目的一是帮助用户了解我们软件怎样使用，如何使用，二是帮助开发人员明晰开发根本需求，紧抓用户体验，把握设计约束，协助进行更好的团队管理和项目进度管理，使得开发更加顺利。

本文档作为所有参与此项目人员工作的共同基础，对《论文辅助工具》（以下称名“万卷书”）的需求描述。通过此文，让业务人员了解项目组对《万卷书》需求理解是否准确无误。

1. 本文档作为《万卷书》供设计人员使用，作为系统设计的依据。
2. 本文档作为项目验收标准之一。
3. 本文档作为软件维护的参考资料

## 1.2 范围

该文档包括软件概述，功能，使用环境，开发中一系列用户图，系统模块，使用性能需求，用户接口以及设计约束和需求分级。

项目目标：

1. 界面简洁友好，符合业务需要，适合学生、学者群体不同的需求。
2. 采用web形式的基于Internet的网络应用系统。
3. 通用、高扩展性、简单、方便的论文辅助工具

# 总体概述

## 2.1 软件概述

论文辅助工具是一款基于自然语言处理（NLP）的智能化系统，旨在为学术研究人员、学生和教师提供高效的论文分析与优化服务。该系统通过N-Gram、NMT（神经机器翻译）模型等技术，实现对论文文本的错别字检测、主题提取、相似度分析等功能，并支持后台管理，满足不同用户角色的需求。

## 2.2 项目介绍

本项目是一个全新的独立系统，旨在解决学术论文写作与分析中的常见问题，如错别字纠正、主题提取、文本相似度计算等。相较于现有工具，本系统整合了NLP 技术与 AI 大模型，提供更智能化的论文优化建议，并支持后台管理，便于管理员监控用户行为与系统运行情况。

## 2.3 产品环境介绍

1.硬件环境：

* 应用服务器：阿里云 ECS 计算节点（8 核 Intel Xeon Platinum 8369CPU、32GB DDR4 内存、1TB NVMe SSD），用于部署 Flask 后端服务及 NLP 模型推理引擎。
* GPU 服务器：NVIDIA A10 GPU（24GB 显存）×2 台，集群部署以支持 N-Gram/NMT 模型的并行计算（错别字检测、相似度分析等任务）。

2.软件环境：

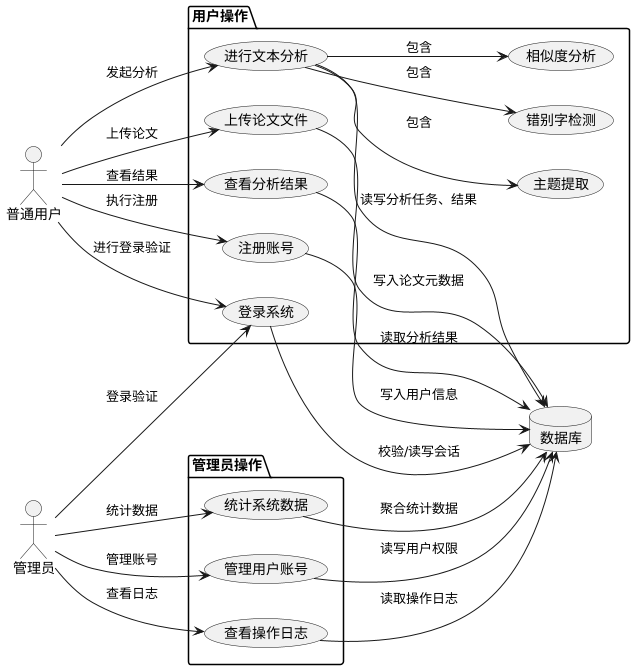
* 前端开发环境：Visual Studio Code，Vue v3.5.17，Typescript
* 后端开发环境：Visual Studio Code,Python+Flask,Mysql

## 2.4 软件功能

|  |  |
| --- | --- |
| **模块** | **功能** |
| 用户界面 | 登录、注册、修改密码 |
| 文本处理 | 错别字检测（N-Gram/NMT）、主题提取、相似度分析 |
| 用户管理 | 查看所有用户、获取最后登录时间、用户权限管理、删除用户 |
| 日志管理 | 查看所有操作以及对应用户、下载操作文件、查看操作时间 |
| 统计 | 活跃用户、文件数量、存储空间、系统状态 |

# 功能需求

## 3.1 用例图



**Use Case 1 用户登录系统**

**Goal in Context**

用户通过账号密码认证进入系统，获取使用权限。

**Preconditions**

1. 系统已启动并正常运行。
2. 用户已注册账号（未注册则自动注册账号）。

**End Condition**

* Success End Condition
* 用户成功登录，系统跳转至用户工作台，生成 JWT 令牌用于权限控制。  
  Failed End Condition

系统显示 "账号或密码错误" 提示，用户仍处于未登录状态  
**Actors**

普通用户

**Trigger**

用户在登录界面输入用户名 / 邮箱和密码，点击 "登录" 按钮  
**Description**

1.用户打开系统登录页面

2.输入注册时使用的用户名 / 邮箱和密码

3.点击 "登录" 按钮触发登录请求

4.系统接收用户输入信息，进行后端加密比对

5.系统验证用户信息有效性

6.根据验证结果返回相应状态

**Step**

1.用户访问系统登录页面

2.在登录表单中输入用户名 / 邮箱和密码

3.点击 "登录" 按钮

4.系统将用户输入的密码进行加密处理

5.系统查询数据库验证用户信息

6. 若验证通过，生成 JWT 令牌，跳转至用户工作台；若验证失败，显示错误提示

**Use Case 2 上传论文文件**

**Goal in Context**

用户将论文文件上传至系统，以便进行文本分析。

**Preconditions**

1. 用户已成功登录系统。
2. 系统文件处理服务正常运行。

**End Condition**

* Success End Condition
* 论文文件上传成功，系统生成分析任务编码，显示上传成功提示。  
  Failed End Condition

系统显示 "文件格式错误"、"文件大小超出限制" 等提示，文件未上传。  
**Actors**

普通用户

**Trigger**

用户在系统中选择论文文件，点击 "上传" 按钮。  
**Description**

1.用户登录后进入论文上传页面。

2.选择本地计算机中的论文文件（支持 PDF/DOCX 格式，≤10MB）。

3.点击 "上传" 按钮。

4.系统接收文件并进行格式和大小验证。

5.验证通过后，系统解析文件并提取文本内容。

6.返回上传结果。

**Step**

1.用户登录系统，进入论文上传模块。

2.点击 "选择文件" 按钮，从本地选择论文文件。

3.系统自动检查文件格式和大小。

4.若符合要求，显示上传进度条；若不符合，显示错误提示。

5.文件上传完成后，系统生成分析任务编码。

6.显示上传成功提示，用户可选择分析类型。

**Use Case 3 进行文本分析**

**Goal in Context**

系统对上传的论文进行错别字检测、相似度分析或主题提取，提供分析结果。。

**Preconditions**

1. 用户已成功上传论文文件。
2. 系统 NLP 服务正常运行。

**End Condition**

* Success End Condition
* 完成文本分析，生成分析报告，用户可查看分析结果。  
  Failed End Condition

系统显示 "分析失败" 等提示，未生成分析结果。  
**Actors**

普通用户

**Trigger**

用户选择分析类型（错别字 / 相似度 / 主题），点击 "开始分析" 按钮。  
**Description**

1.用户在论文上传成功后，选择需要进行的分析类型。

2.点击 "开始分析" 按钮。

3.系统接收分析请求和分析任务编码。

4.系统调用 NLP 模型进行相应的文本分析处理。

5.分析完成后，将结果存储到数据库。

6.返回分析结果页面。

**Step**

1.用户在上传成功页面选择分析类型（错别字检测、相似度分析或主题提取）。

2.点击 "开始分析" 按钮。

3.系统显示 "分析中" 状态，并提示预计等待时间。

4.系统根据选择的分析类型，调用相应的 NLP 模型（N-Gram、NMT 等）进行文本分析。

5.分析完成后，从数据库获取分析结果。

6.跳转到分析结果页面，展示分析内容。

## 3.2 系统模块

* 前台功能

|  |  |  |
| --- | --- | --- |
| 功能名称 | 子功能 | 描述 |
| 用户端 | 用户登录/注册模块 | 用户登录注册 |
| 论文上传 | 上传论文处理 |
| 查看结果 | 获得处理结果 |

* 后台功能

|  |  |  |
| --- | --- | --- |
| * 功能名称 | * 子功能 | * 描述 |
| * 首页管理 | * 通知管理 | * 管理通知 |
| * 内容管理 | * 内容CURD |
| * 统计分析 | * 用户分析 | * 用户操作分析 |
| * 功能分析 | * 功能使用分析 |
| * 性能分析 | * 性能使用率分析 |
| * 日志记录 | * 管理所有日志记录 |
| * 设置 | * 系统设置 | * 系统属性设置 |
| * 功能设置 | * 功能CURD |
| * 用户管理 | * 系统用户管理 | * 系统用户CURD |
| * 普通用户管理 | * 普通用户CURD |
| * 论文分析 | * 文本查错模块 | * 显示短文本的错字 |
| * 文本相似度分析模块 | * 显示两个短文本的相似度 |
| * 文本主题提取模块 | * 提取短文本的主题 |

登录注册模块

a． 描述：用户通过账号密码认证登录系统，未注册用户可申请账号。

b． 输入：用户名/邮箱、密码

c． 加工：验证用户信息（后端加密比对）、生成 JWT 令牌（用于权限控制）

d． 输出：登录成功跳转至用户工作台；登录失败则提示“账号或密码错误”。

操作日志模块

a. 描述：记录用户的历史操作，支持按时间筛选查看。

b. 输入：用户ID（自动获取）、时间范围（可选）

c. 加工：数据库查询（MySQL操作日志表）

d. 输出：操作日志列表（操作类型、时间、文件名称）

论文上传模块

a. 描述：用户上传论文文件，系统进行文本分析处理。

b. 输入：论文文件（DOCX，≤10MB）

c. 加工：文件解析（提取文本内容）、NLP处理（错别字检查、相似度分析等）

d. 输出：分析任务编码、上传成功/失败提示

查看结果模块

a. 描述：查看论文分析结果，支持下载报告。

b. 输入：分析任务编码

c. 加工：从数据库读取分析结果

d. 输出：分析报告（可视化展示）、下载链接（PDF/JSON格式）

首页管理模块

a. 描述：管理系统首页内容和通知。

b. 输入：通知内容（富文本）、展示内容数据

c. 加工：内容存储至数据库

d. 输出：首页内容更新状态

统计分析模块

a. 描述：展示系统使用数据统计。

b. 输入：统计时间范围、统计指标类型

c. 加工：数据聚合分析

d. 输出：统计图表（折线图/柱状图）、数据报告

用户管理模块

a. 描述：管理系统用户账号和权限。

b. 输入：用户信息、操作指令

c. 加工：数据库增删改查操作

d. 输出：用户列表、操作结果提示

论文分析管理模块

a. 描述：查看和管理所有论文分析任务。

b. 输入：筛选条件

c. 加工：数据库查询

d. 输出：分析任务列表、任务详情

# 性能需求

系统的性能需求主要有：

|  |  |  |
| --- | --- | --- |
| 需求项 | 指标 | 说明 |
| 单篇论文大小限制 | ≤10MB | 支持PDF/DOCX格式，超出大小拒绝上传 |
| 文件存储空间 | ≥1TB | 云存储/本地存储需预留扩展空间 |
| 用户认证响应 | <2秒> | JWT令牌快速验证，Redis缓存用户权限 |
| 论文上传与解析 | <5秒> | 异步文件处理，进度回调（WebSocket/轮询） |
| 文本分析 | <10秒> | 预加载NLP模型，分布式任务队列（Celery/RabbitMQ） |
| 主题提取 | <15秒> | 算法优化（LDA并行计算） |
| 高并发场景 | 峰值：100并发用户 | 负载均衡（Nginx）、API限流（令牌桶算法） |
| 日均API请求：50,000次 | 数据库连接池优化，读写分离 |

（一）实时性要求：

用户操作日志记录延迟 ≤1秒（MySQL事务提交+Elasticsearch同步）

通知推送延迟 ≤3秒（WebSocket长连接）

（二）容灾与备份：

数据库每日全量备份，Binlog日志实时同步

论文文件异地冗余存储（如AWS S3跨区域复制）

（三）扩展性要求：

支持横向扩展NLP服务节点（Docker容器化部署）

动态扩容文件存储空间（无需停机）

# 接口需求

**一、软件接口需求**：

**前后端交互接口**：协议RESTful API（基于 HTTP/2）、数据格式JSON（Content-Type: application/json）、认证方式JWT 令牌（Authorization: Bearer {token}）

接口示例：POST /api/v1/auth/login：用户登录接口，接收用户名 / 密码，返回 JWT 令牌。

**NLP 模型服务接口**：调用方式RPC（远程过程调用），基于 gRPC 框架、协议HTTP/2 + Protobuf 序列化。

接口功能：1.SpellingCheckService.Check(text: string)：错别字检测，返回错误位置及建议词

2.SimilarityService.Compare(text1: string, text2: string)：文本相似度计算，返回百分比值

3.ThemeService.Extract(text: string, topN: int)：主题提取，返回关键词列表

**数据存储接口：**

关系型数据库（MySQL）：ORM 框架：SQLAlchemy（Python）

接口示例：# 用户查询接口（伪代码）

def get\_user\_by\_id(user\_id: str) -> User:

return db.session.query(User).filter(User.id == user\_id).first()

**第三方系统接口：NLP 算法工具包接口**

N-Gram 模型：框架NLTK（Python 自然语言工具包）、接口示例nltk.ngrams(text, n=3)（生成 3-gram 词组）。

NMT 神经机器翻译模型：框架Transformers（Hugging Face）、预训练模型BERT-base-chinese（用于语义理解）

1. **接口文档引用** 1.《NLP 服务接口规范》：描述 gRPC 接口的服务定义（.proto 文件）及调用示例。

2.《硬件设备管理手册》：包含 F5 负载均衡器、MinIO 存储集群的配置接口说明。

# 用户接口需求

* 登录/注册

|  |  |
| --- | --- |
| * 用户操作 | * 系统反馈 |
| * 输入用户名/邮箱和密码 |  |
| 点击"登录"按钮或"注册"链接 | * 登录成功：跳转至用户工作台 |
| * 登录失败：显示红色提示"账号或密码错误" |
| * 注册：若账户不存在则自动注册 |

* 论文上传

|  |  |
| --- | --- |
| * 用户操作 | * 系统反馈 |
| * 拖拽文件至上传区域/点击"选择文件" | * 上传成功：显示进度条（实时更新） |
| * 格式错误：提示"仅支持PDF/DOCX格式" |
| * 选择分析类型（错别字/相似度/主题） |  |
| * 点击"开始分析" | * 分析中：显示任务ID和预计等待时间 |
| * 分析结果界面 |

* 论文分析



|  |  |
| --- | --- |
| * 用户操作 | * 系统反馈 |
| * 查看高亮标注的错别字 | * 错别字：黄色高亮显示，悬停展示建议词 |
| * 点击相似论文条目查看详情 | * 相似度：百分比进度条+相似论文列表 |
| * 主题：词云图展示关键词权重 |
| * 使用"导出报告"按钮 | * 导出报告 |

* 后台管理

|  |  |
| --- | --- |
| * 管理员操作 | * 系统反馈 |
| * 删除用户 | 用户账号被删除 |
| * 用户权限控制 | * 操作成功：用户可使用权限被修改 |
| * 权限不足：弹出"需要管理员权限"提示 |
|  |  |

# 总体设计约束

**一、技术约束**

1.NLP模型依赖：错别字检测依赖 N-Gram 模型与 BERT 预训练模型，需兼容 Hugging Face Transformers 框架（版本≥4.25.1）；主题提取算法受限于 LDA 模型的并行计算效率，单篇分析耗时需控制在 15 秒内（参考《4. 性能需求》）。

2.技术栈限制：前端必须使用 Vue 3 + TypeScript，后端仅限 Python 3.8+ + Flask 框架；数据库连接池最大并发数限制为 200（基于 MySQL 默认配置）。

3.兼容性约束：支持 Chrome 80+、Firefox 78+、Edge 80 + 浏览器，不兼容 IE 11 及以下；移动端需适配 iOS 13+、Android 8.0 + 系统的 WebView。  
 **二、法律与合规约束**

1.数据隐私：用户论文数据需遵循《个人信息保护法》，存储时长不超过 6 个月，支持用户随时删除；文本分析过程中禁止留存用户敏感信息（如姓名、学号等）。

2.知识产权：NLP 模型使用开源框架（如 NLTK、Transformers），需在文档中声明开源协议（Apache 2.0）；禁止使用未授权的商业 API（如非开源的文本相似度算法）。

# 其他需求

**一、测试需求**

1.测试类型：单元测试：覆盖 80% 以上的代码逻辑（尤其是 NLP 分析模块）；集成测试：验证前后端接口、数据库交互的正确性；性能测试：模拟 100 并发用户场景，验证响应时间（参考《4. 性能需求》）。

2.测试用例：错别字检测：准备包含常见语病的测试文本（如 “的得地” 误用）；相似度分析：使用不同重复率的论文样本（10%/30%/50%）。

# 需求分级

|  |  |  |
| --- | --- | --- |
| 需求ID | 需求名称 | 需求分级 |
| RQ-001 | 用户登录 / 注册 | A |
| RQ-002 | 论文上传与格式验证 | A |
| RQ-003 | 错别字检测 | B |
| RQ-004 | 管理员用户管理 | B |
| RQ-005 | 主题提取 | C |
| RQ-006 | 分析结果可视化（词云图） | C |

# 待确定问题

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 需求ID | 问题描述 | 影响  (H/M/L) | 风险 | 责任人 | 解决日期 | 状态(Open/Close) |
| DP-001 | NLP 模型在长文本场景的准确率波动 | H | 分析结果可信度下降 | 雷明康 |  | Close |
| DP-002 | 分布式存储集群的成本优化方案 | M | 预算超支 | 李梓嘉 |  | Close |
| DP-003 | 移动端 WebView 的文件上传兼容性问题 | M | 部分用户无法使用 | 韩欣星 |  | Close |
| DP-004 | 高并发场景下的数据库死锁问题 | H | 系统服务中断 | 赵一波 |  | Close |