系统设计报告

---BookWise 书籍垂直搜索引擎

《软件工程管理》课程 G13 小组

组长：王伟杰

组员：安俊霖、包越、陈华杰、黄乐诚、刘逸杰

2023.11.25

修改历史

修订日期	版本号	修改人	修改内容	审核人
2023-11-17	Version1.0	全体组员	初稿	王伟杰
2023-12-1	Version2.0	全体组员	终稿	王伟杰

目录

1 引言

1.1 编写目的

本系统设计报告以《项目计划书》和《需求规格说明书》为基础，具体讲述系统的总体架构，各个功能的实现方式以及数据库设计方法，明确了各个模块的外部接口、内部接口以及用户接口，为软件系统的开发提供指导，为软件系统的维护提供参照。

- 预期读者：
- 项目经理
- 系统分析人员
- 系统设计人员
- 系统开发人员
- 系统测试人员
- 系统质量分析员
- 系统维护人员

1.2 软件项目背景

本项目是浙江大学 2023 学年《软件工程管理》课程的课程项目，目标是实现一个书籍垂直搜索引擎。

随着互联网的发展，人们获取信息的方式发生了巨大的变化。特别是在图书领域，读者需要更加高效和便捷地获取他们感兴趣的书籍信息。传统的搜索引擎虽然可以提供大量的信息，但是在特定领域的

深度搜索和精准推荐方面仍然存在不足。因此，开发一款专注于书籍领域的垂直搜索引擎成为了当下的需求。

该书籍垂直搜索引擎项目旨在为读者提供一个专注于书籍的搜索平台，通过整合各种图书信息资源，包括线上线下书店、图书馆、电子书平台等，为用户提供更加全面、深入的书籍搜索和推荐服务。这将有助于读者更快速、更准确地找到他们感兴趣的书籍，推动图书行业的数字化和信息化发展。

项目团队将致力于开发一款功能强大、用户体验友好的书籍垂直搜索引擎，通过技术手段提高图书信息的获取效率，为用户提供更加便捷的阅读体验。同时，该搜索引擎还将为图书出版商、书店和图书馆等机构提供更广泛的推广和宣传渠道，促进图书产业的发展和繁荣。

综上所述，书籍垂直搜索引擎项目的开发将填补当前书籍领域搜索服务的空白，满足用户对于高效获取图书信息的需求，促进图书产业的数字化和信息化发展。

1.3 定义

1.3.1 HTML

HTML（Hyper Text Markup Language）即超文本标记语言。 HTML 是由 Web 的发明者 Tim Berners-Lee 和同事 Daniel W. Connolly 于 1990 年创立的一种标记语言，它是标准通用化标记语言 SGML 的应用。用 HTML 编写的超文本文档称为 HTML 文档，它能独立于各种

操作系统平台(如 UNIX， Windows 等)。使用 HTML，将所需要表达的信息按某种规则写成 HTML 文 件，通过专用的浏览器来识别，并将这些 HTML 文件“翻译”成可以识别的信息,即现在所见到的网页

1.3.2 CSS

层叠样式表（Cascading Style Sheets），是一种用来表现 HTML 等文件样式的 计算机语言，在网络中能够对网页中元素位置的排版进行像素级精确控制

1.3.3 JavaScript

JavaScript（简称“JS”） 是一种具有函数优先的轻量级，解释型或即时编译型的编程语言。虽然它是作为开发 Web 页面的脚本语言而出名，但是它也被用 到了很多非浏览器环境中，JavaScript 基于原型编程、多范式的动态脚本语 言，并且支持面向对象、命令式、声明式、函数式编程范式

1.3.4 Vue

Vue 是一个用于创建用户界面的开源框架,也是一个创建单页应用的 Web 应用框架，一套用于构建用户界面的渐进式框架

1.3.5 Spring Boot

Spring Boot 是由 Pivotal 团队提供的全新框架，其设计目的

是用来简化新 Spring 应用的初始搭建以及开发过程。该框架使用了特定的方式来进行配置， 从而使开发人员不再需要定义样板化的配置。通过这种方式，Spring Boot 致力 于在蓬勃发展的快速应用开发领域(rapid application development)成为领导者

1.3.6 MySQL

一个小型关系型数据库管理系统

1.4 参考文献与资料

- 《软件工程开发国家标准》
- 《软件工程项目开发文档范例》
- 《软件需求（第三版）》，Karl Wiegers Joy Beatty，清华大学出版社
- 《软件工程 实践者的研究方法》，Roger S.Pressman，机械工业出版社
- [G13]“书籍垂直搜索引擎”项目计划书
- [G13]“书籍垂直搜索引擎”需求规格说明书

2 需求规定

2.1 用户需求规定

2.1.1 搜索词条

用户可以输入关键字进行搜索，关键字支持与或非等逻辑运算。

系统支持用户进行高级搜索，用户可以根据搜索符号规则组织搜索输入语句，进行搜索范围限定，获得更加精准的搜索结果。例如：关键词过滤、关键词并集、关键词交集等。

用户输入时系统可以进行智能补全并识别用户输入的拼音搜索。

用户输入时，搜索框默认显示用户最近的两次搜索内容，在自动补全时，优先显示用户历史搜索内容

2.1.2 查看搜索结果

用户在输入关键词或者查询语句，点击搜索按钮后，系统可以将搜索结果呈现给用户。搜索结果可以以卡片列表的形式展示出书籍的基本信息，例如书名、 图片、作者等。

用户可以通过点击具体的搜索条目进入查看书籍详情，书籍详情页展示书籍更多元的信息，例如书籍的主要内容信息、角色信息（小说）等，同时还会智能化地猜测你喜欢的书籍并给出推荐。

搜索结果默认按照搜索的相关度、时间排序。同时，用户可以通过自主选择排序规则和筛选条件，例如仅搜索书名或仅搜索作者名等，

对搜索结果进行二次排序和筛选

用户可以在页面右侧查看到系统针对用户输入进行的个性化推荐内容，即一些同类词条，用户可以通过点击跳转进入相应页面

2.1.3 详情内容展示

用户可以通过点击搜索结果，进入搜索引擎对该搜索结果进行聚合分析后的页面。该页面包含有书籍的基本信息和其他相关数据。同时该页面还会展示一些分析得出的数据，以图形化和列表等多样化的形式展现，例如书籍相关的信息，相关书籍推荐等

2.2 其它需求规定

2.2.1 性能需求

- 系统应保证运行稳定，避免出现崩溃
- 主流浏览器均能正常访问本系统
- 系统应能保证至少 100 人的并发访问
- 当用户登录以及进行任何操作时，系统应该能及时进行反应，反应的时间在 1s 以内
- 系统应该能及时检测出各种非正常情况，如与设备的通信中断断开，无法连接数据库服务器等情况，避免用户长时间等待每个页面。一般情况下应在 1s 内加载完毕，高峰期应在 7s 内加载完毕；
- 系统保证在半个月内不超过一次维护与重启

2.2.2 输入需求

- 用户输入数据时，应对数据输入进行数据有效性和安全性检查，同时过滤掉一些错误的、非法的输入，同时要防止 SQL 注入攻击
- 用户搜索书籍时，应对数据的有效性和合法性进行检查；
- 此外，系统应通过程序控制出错几率，减少系统因用户人为的错误引起的破坏，开发者应当尽量周全地考虑到各种可能发生的问题，使出错的可能降至最小。

2.2.3 数据传输与并发需求

- 用户输入账号密码点击登录后，对登录的响应时间不能超过 1 秒，在此时间内将登录结果显示在屏幕上
- 系统能支持 10 名用户同时搜索书籍以及查看详细信息
- 系统能支持 20 名用户同时进行书籍信息的条件查询
- 系统应支持 100 名用户并发使用，并保证性能不受影响
- 在网页中，系统生成的所有 Web 页面可以在不超过 5 秒的时间内可以全部下载下来

2.2.4 数据管理需求

系统既要与其他系统有接口，又必须保证本系统的独立性与完整性。即应防止未经授权的各类人员对本系统进行设置和修改或进行有关统计。

系统服务器软件必须提供可靠的数据备份和恢复手段，在服务器软件或硬件出现严重故障时，能够根据备份的数据和账户信息等必要

的配套信息，迅速彻底地恢复正常运行环境。

系统的用户信息管理相关模块，决定了其他众多系统的账户安全性，必须保证统计数据准确、安全，用户信息应当提供完整的备份及恢复措施。

无论访问者账户信息还是管理者账户信息，都必须提供完备手段由用户自定义和备份保存，软件开发者不得在系统中预留任何特殊账户和密码。

除此之外，系统应具备加密登录、数据加密传输等安全方面的保障，保证数据在不用系统间传输过程中的保密性与安全性。

以下为具体细则：

- 当系统崩溃后，系统应能在 24 小时内恢复运行
- 数据库可支持表的最大行数达到 600 行
- 本系统用于日志等记录的数据增长约为 20MB/月，具体增长速度由用户的使用频率及所发生业务的数据量决定
- 当出现重大事故造成数据丢失后，系统应能在 48 小时内恢复数据
- 系统管理员每两个月应至少维护备份一次数据
- 系统服务器应具备至少 20GB 的存储空间

2.2.5 权限与安全需求

在我们的系统中，对于安全与权限进行了如下设计：

- 所有涉及功能信息或个人信息的网络事务，都应进行加密操作

- 只有系统管理员有权查看及修改底层数据库与搜索引擎的数据，且行为应被系统日志记录，用户无法非法修改数据库
- 系统应该能够记录系统运行时所发生的所有错误，包括本机错误和网络错误，以便于查找错误的原因。系统日志应同时记录下用户的关键性操作信息
- 当流量过大时，优先限制游客流量防止恶意访问
- 系统对可能发生严重后果的操作要有补救措施，通过补救措施用户可以回到原来的正确状态。同时对可能造成等待时间较长的操作应该提供取消功能
- 对错误操作支持可逆性处理，如取消系列操作。在输入有效性字符之前应该阻止用户进行只有输入之后才可进行的操作

2.2.6 可视化需求

- 要将搜索结果进行良好的可视化呈现并进行一定的排序
- 要将书籍的信息使用美观简洁的表格、统计图等形式呈现
- 要将书籍版本信息使用图表形式等进行可视化呈现
- 要将书籍相关资讯用良好的可视化界面呈现

2.2.7 防护性需求

- 文件格式错误时，系统提出警告，保持数据库数据不变
- 数据库误删除时，可以使用撤销删除修复
- 系统应该保护未开放下载权限的资料不被下载
- 重复操作导致卡死时，系统提出警告

- 系统应该提供验证码防止恶意使用(如爬虫等)
- 系统应该及时信息备份防止病毒攻击
- 系统应该能检测到恶意操作，提出警告并在一段时间内不允许操作
- 访问无权限时，系统发出提示并禁止用户访问

2.2.8 软件质量属性

- 可用性：系统保证早上 6 点到晚上 12 点之间可用，但在发生紧急情况时允许停止运行一段时间
- 可维护性：系统运行时要保存运行日志，用来维护分析。每周一的凌晨 1 点到 5 点为维护时间，在此期间用户不能使用系统。此外，维护人员需要在系统正常运行时能保持联系
- 兼容性：系统需要保证在主流浏览器（Chrome、Edge）上可以正常浏览和使用，对于其他市场占有率超过 5%的浏览器，保证实现系统的主要功能
- 易用性：系统界面应该简洁明了、操作简单，功能按钮的位置符合用户的日常习惯。此外，系统应该要有导航和清晰简短的用户使用手册
- 可扩充性：系统在设计上考虑到了网站可能的后续发展，在后端设计和前端设计上尽可能地在满足所有需求的同时，增强了网站的可扩充性。一旦有扩展需要，客户可以联系系统维护人员，维护人员需要在 1-4 个工作日内完成客户的内容扩充需求,主要包括增加新的功能、增加新的模块、界面优化、系统性能提升等

2.2.9 其他需求

- 软件对用户的使用应该有较为清晰的引导
- 软件对用户的误操作或不合法操作进行检查，并给出提示信息

3 总体设计

3.1 功能设计

表 3-1-1 功能设计

功能模块	功能
前端子系统	查询搜索内容
	搜索结果排序
	搜索结果筛选
	搜索结果与详情展示
	搜索结果相关信息展示
后端子系统	数据导入
	前端搜索接口
	Elastic Search 接口
爬虫子系统	信息爬取
	自动分类
	自动滤重
	自动爬取

3.2 用户类型及用户特征

表 3-2-1 用户类型及用户特征

用户类	特征与说明
网站用户	1. 主要用户 2. 主要服务的对象为想要找到书籍的信息以及书籍相关问题解答的学习者 3. 要求能够返回足够精确和足够深度的结果 4. 能够根据用户搜索的结果进行相关的推荐 5. 支持根据用户的输入进入搜索引擎搜索，并且返回相关的内容到网站搜索结果列表页面
系统管理员	1. 次要用户 2. 权限比较高，具有数据库和搜索引擎数据库的更新和审核等管理权限 3. 面向的用户数目和书籍数据比较多，要求能够提供方便并且快速的管理员接口进行管理 4. 操作频率相对较低，但是每一次的操作对于系统造成的影响比较大

3.3 运行环境

本网站主要服务于需要书籍资源的网友，保证至少 100 名网友同时取得服务的需求，包括数据存储能力和网络吞吐能力，保证账户一定的安全性。

表 3-3-1 软件运行环境

项目	名称	版本
操作系统	Windows 7 及以上，Linux	
网站服务器	Nginx	1.15.8
数据库服务器	Linux Socket	
数据库服务器类型	MySQL，ElasticSearch	8.0

浏览器	Chrome，Edge	

表 3-3-2 硬件运行环境

项目	名称
操作系统	CPU:CORE i3 及以上 内存：2G 以上 硬盘：500G 以上
数据库服务器	内存：512M 及以上 硬盘：50G 及以上
通讯设备	网线

3.4 基本概念和处理流程

服务器

以 Nginx 为服务器，Java 语言编写后端代码，数据库采用 MySQL，搜索引擎使用 Elasticsearch。当用户通过浏览器使用网站系统时，浏览器接收用户的请求，并传送到服务器。通过搜索引擎 Elasticsearch 查询相关内容，并且从数据库接口函数向数据库发送 SQL 查询语句，数据库接收 SQL 查询语句后执行，返回查询结果，处理查询结果后返回给前端，并显示在网站页面上。

客户端

浏览器采用常用的 Chrome、Firefox、Edge 等。客户端在不频繁的操作页面时完成操作后断开与数据库的连接以减轻服务器负荷，在操作频繁时保持连接以增加访问速度。

客户端动态页面：嵌入 Vue，动态网页以数据库技术为基础，能

降低网站维护的工作量。

Vue：页面的各种搜索框、筛选框与按键操作能够完成，同时能实现无刷新页面的一些动画效果，包括下拉菜单等。

3.5 结构

3.5.1 用户需求分析图

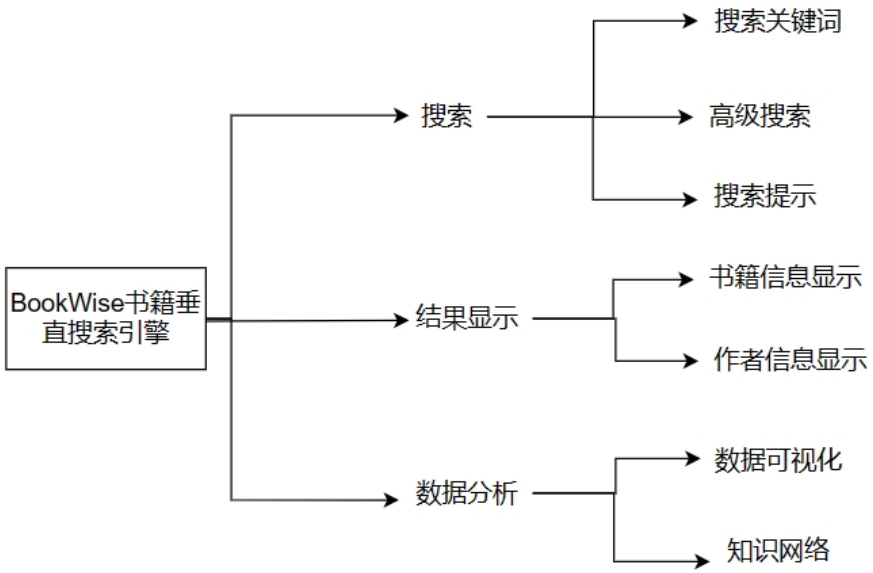


图 3-5-1 用户需求分析图

3.5.2 系统模块架构图

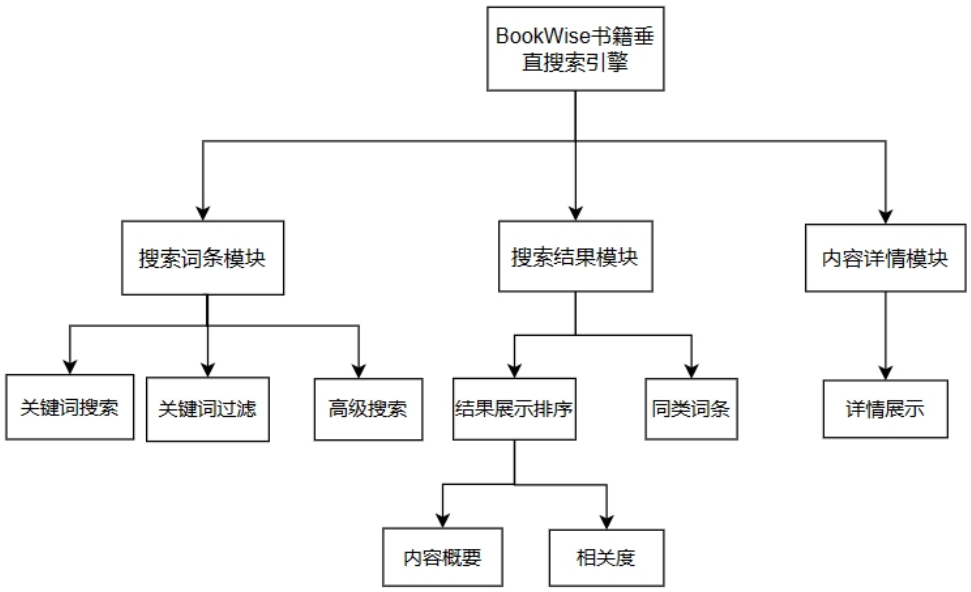


图 3-5-2 系统模块架构图

3.5.3 数据流图

3.5.3.1 搜索引擎子系统数据流图

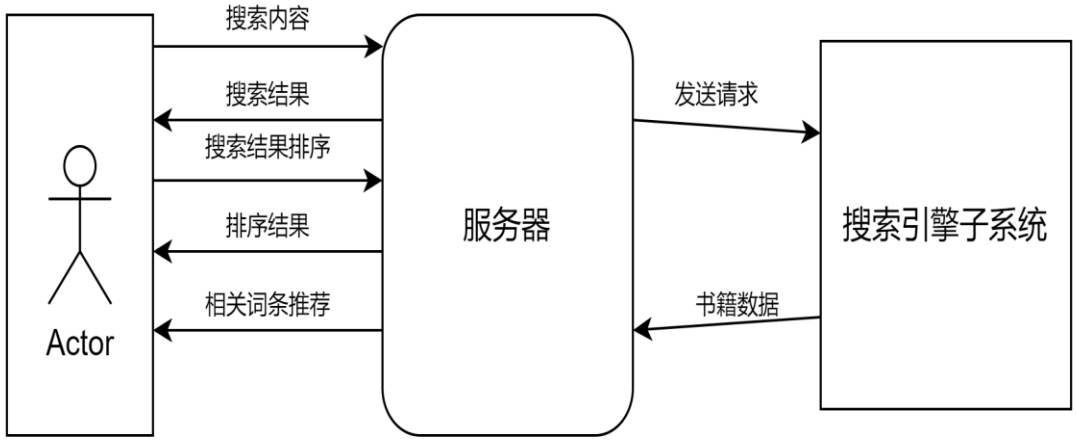


图 3-5-3 搜索引擎子系统数据流图

3.5.3.2 网站维护子系统数据流图

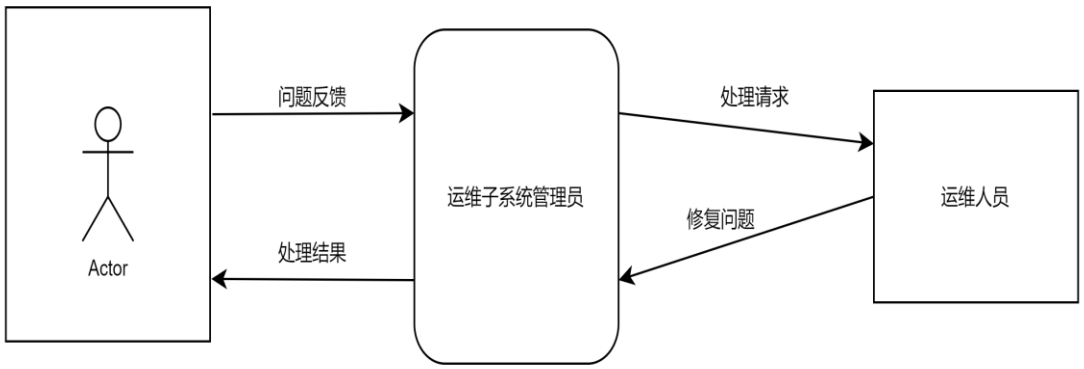


图 3-5-4 网站维护子系统数据流图

3.5.4 ER 图

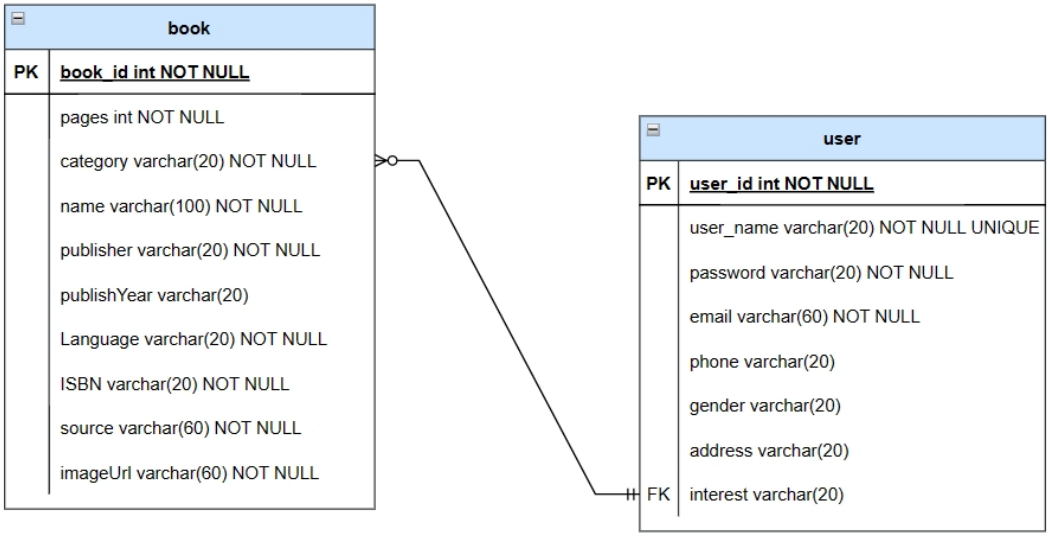
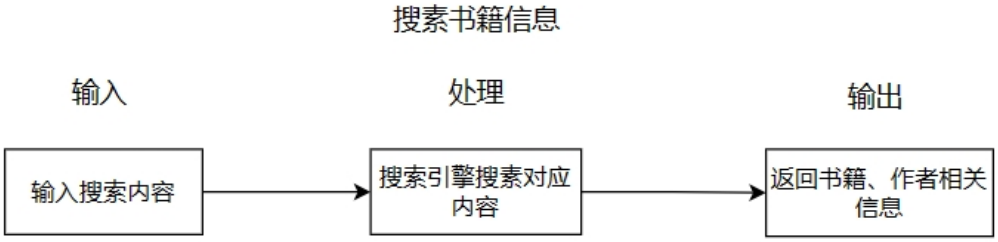


图 3-5-5 ER 图

3.5.5 关键 IPO 图



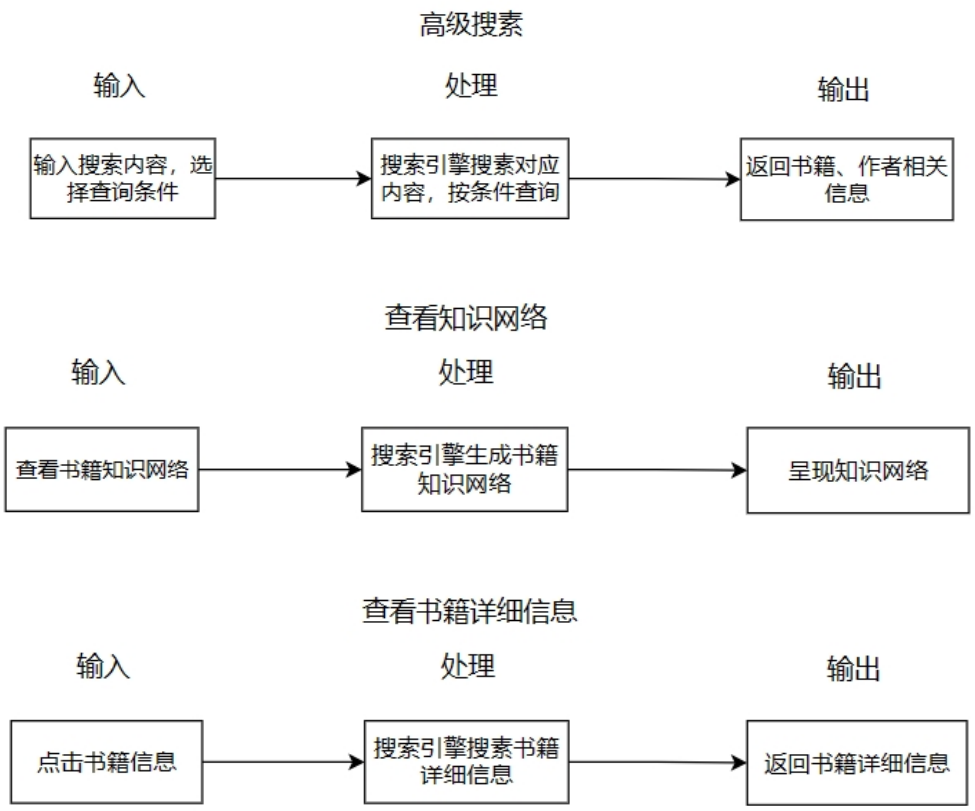


图 3-5-6 关键 IPO 图

3.5.6 数据字典

3.5.6.1 数据元素定义表

表 3-5-1 数据元素定义表

编号	数据元素名	类型	值域	说明
E1	查询字符串	字符	/	用户在搜索引擎输入框中输入的查询字符串
E2	书籍名称	字符	/	/
E3	书籍作者	字符	/	/
E4	书籍类型	字符	/	/

E5	书籍语言	字符	/	/
E6	书籍页数	整形	/	/
E7	书籍链接	字符	/	/
E8	出版时间	长整型	当日 0 点的时间戳格式	/
E9	出版社	字符	/	/
E10	ISBN	字符	/	书籍编号

3.5.6.2 数据精确表

表 3-5-2 数据精确表

编号	数据元素名	类型	精度要求	说明	示例
E1	查询字符串	字符	128 个字符以内	用户在搜索引擎输入框中输入的查询字符串	软件 计算机
E2	书籍名称	字符	64 个字符以内	/	软件工程
E3	书籍作者	字符	64 个字符以内	/	Patt
E4	书籍类型	字符	32 个字符以内	/	教育
E5	书籍语言	字符	32 个字符以内	/	英语
E6	书籍页数	整形	/	/	30000
E7	书籍链接	字符	128 个字符以内	/	www.baidu.com
E8	出版时间	长整型	当日 0 点的时间戳格式	/	1577808000000
E9	出版社	字符	128 个字符以内	/	机械工业出版社
E10	ISBN	字符	/	/	0201103311

3.6 人工处理过程

在本系统的运行过程中，可能会出现一些系统无法自动解决的问题，需要人工处理介入来解决，包括书籍记录的删改，相关书籍信息的爬取等。

3.7 尚未解决的问题

无

4 接口设计

4.1 用户接口

本系统作为垂直搜索引擎系统，用户所有行为均在网页页面上实现，用户通过鼠标点击或键盘输入完成与系统的交互。用户主要通过输入框、按钮、筛选框、下拉选择框等可视化元素与服务器后端进行交互。用户的主要接口有：

1. 搜索书籍信息
2. 查看搜索结果列表
3. 查看书籍知识网络等可视化信息
4. 查看书籍详细信息

4.2 外部接口

本系统的部分数据存储在服务器及数据库中，搜索引擎所需数据以文本形式存储。资源文件及不适宜数据库表项存储的超长文本存储在文件中。网页前端获取用户输入后，由网页后端完成与服务器及数

据库的交互。利用 Java、 Vue 与 MySQL、搜索引擎 Elastic Search 之间的接口完成网站外部接口设计。 本系统的初始数据依靠人工导入。

4.3 内部接口

本系统总体分为前端、后端、数据爬取三个模块。各模块之间耦合度较低，各模块之间使用 JSON 进行数据传输。

前后端接口：

- 关键词搜索
- 书籍内容筛选
- 搜索提示
- 书籍详情搜索

数据爬取与后端接口：

- 数据库存储

5 运行设计

5.1 运行模块的组合

本系统按照交互逻辑划分模块，每个模块不共享界面，相对独立。每个模块 按照流程划分为客户端界面，客户端脚本和后台服务器程序。 各个模块之间不会共享界面，但共享数据库数据和搜索引擎，

后台程序只共享数据库连接和搜索引擎。

5.2 运行控制

5.2.1 界面

界面是用户直接与系统交互的部分，界面力求简洁而不简陋，能引导用户进行无碍操作。设计时，以在提供用户便捷操作的基础上增加美观度为基准。

5.2.2 运行控制的条件与限制

本项目的开发要求小组成员足够的参与度，能及时保质保量完成任务。且项目开发过程中可能会有技术上的难点和设备、服务器资源等方面的欠缺，需要开发小组合理利用现有设备和资源，积极查找资料解决问题，在完成项目开发的基础上，同时保证项目的可用性、安全性、可维护性等。

5.2.3 前台与后台的关系

前台主要展示搜索结果信息、电影详情内容等显示信息，后台主要负责业务 流程，控制前台显示信息，负责与搜索引擎和数据库交互。

5.3 运行时间

用户在做搜索时候，前端不断地向 ES 搜索引擎请求数据，会频繁与数据交互以获取信息，这会占用较多的数据库资源

6 总体数据设计

6.1 数据存储

项目使用标准 MySQL 数据库以及 ElasticSearch 搜索引擎，按照数据产生、转换和存储的策略，通过将数据导入数据库和搜索引擎的方式进行数据的存储操作

6.2 数据安全

将从完整性、保密性和可用性三个层面来保护用户的数据安全

完整性

要求数据未经授权不得进行修改，确保数据在传输和存储过程中不被篡改、 盗用和丢失。通过利用安全的框架（如 spring data），在加密的基础上，运用多种方案和技术实现。

保密性

要求对数据进行加密，只有授权者才能使用。这一特性要求加密技术必须自动、实时、精确、可靠。

可用性

要求做到避免因为系统数据泄露而使得合法使用者无法接触可用数据，通过对使用者身份的验证，为合法使用者提供更加安全便捷的使用。

6.3 逻辑结构设计要点

6.3.1 Elastic Search 索引设计

6.3.1.1 标识

在 Elastic Search 的搜索引擎中，一个索引的定义由字段名和类型组成。如果该字段的类型是一个对象（Object）或者嵌套对象（Nested），则对象中的各个属性的定义也同样是由字段名和类型组成。

在一定程度上，我们可以将 Elastic Search 理解为一个文档型数据库。索引（Index）定义了文档的逻辑存储和字段类型，每个索引可以包含多个文档类型，文档类型是文档的集合，以索引定义的逻辑存储模型。

- 索引：相当于数据库，用于定义文档类型的存储
- 文档类型：相当于关系表，用于描述文档中的各个字段的定义
- 文档：相当于关系表的数据行，存储数据的载体，包含一个或多个存有数据的字段

以下是定义一个索引基本模板：

```
{
    "mappings":{
        "properties":{
            "<字段名>":{
                "type": "<字段类型>",
                "index": "<是否索引>",
                "store": "<是否存储>",
                "analyzer": "<分析器>"
            }
        }
    }
}
```

图 6-3-1 ES 搜索引擎索引基本模板

其中 mapping 指明了这是索引的映射配置，properties 指明其中的内容配置的是索引的字段，type 指明类型，index 指明该字段是否为索引，store 指明 该字段是否为存储，analyzer 指明索引的分词器

本文中涉及的索引类型 type 类型如下：

- text：普通文本，分词器会将这一类文本进行分词，进行倒排索引
- keyword：关键词，分词器不会将这类文本进行分词
- date：日期，格式为 ”YYYY-MM-dd” 或毫秒数
- integer：普通整数
- long：长整数
- double：双精度浮点数
- object：普通对象，在 Elastic Search 的底层存储中，object 对象会被扁平化成数组存储

- nested：嵌套对象，在 Elastic Search 的底层存储中，nested 对象不会被扁平化，而是将嵌套子对象存储在单独的文档（doc）中

6.3.1.2 书籍索引设计

表 6-3-1 书籍索引设计

Properties	Type	Analyzer
book_id	keyword	
name	text	ik_pinyin_analyzer
categories	text	
authorname	text	ik_pinyin_analyzer
publishYear	interger	
bookeLanguage	text	ik_pinyin_analyzer
pages	interger	
source	keyword	
ISBN	keyword	
img.url	keyword	

6.3.2 MySQL 数据库设计

6.3.2.1 用户

表 6-3-2 用户 MySQL 数据库设计

Field	Type	Allow Null
user_id	int	No
username	varchar(20)	No
password	varchar(20)	No
email	varchar(60)	No
phone	varchar(20)	Yes
gender	varchar(20)	Yes

address	varchar(60)	Yes
interest	varchar(60)	Yes

6.3.2.2 书籍

表 6-3-3 书籍 MySQL 数据库设计

Field	Type	Allow Null
book_id	int	No
name	varchar(100)	No
category	string	No
pages	int	No
publisher	string	No
publishYear	string	Yes
bookLanguage	string	No
ISBN	string	No
source	string	No
imageUrl	string	No

6.4 物理结构设计要点

一些基本的可以用来充当索引的数据项存储在 Elastic Search 中，从而来提供索引功能。

全部的完整的数据存储在 MySQL 数据库中，包括一些比较详细的信息，通过 SQL 语句访问数据库获取。

文件资源存储在磁盘中，通过搜索引擎访问存储位置获取文件

7 系统出错设计

7.1 出错信息

表 7-1-1 出错信息

输出信息形式	含义	处理方式
数据库连接失败	由于并发操作的用户数量很大，导致 ES 访问读写率降低；或者 ES 的节点配置不对，导致 ES 连接失败	修改 ES 节点配置，尝试重连
磁盘损害	由于物理因素等，导致数据库中的数据丢失	定期对数据库中的数据进行备份
数据库读取乱码或汉字输出为乱码	客户端页面、数据库、搜索引擎读取过程编码不一致	统一各处的编码方式
搜索结果列表为空	搜索引擎无法获得书籍相关内容	手工检索

7.2 补救措施

系统备份

定期备份系统数据，当系统数据因不可抗拒力丢失时，可以启用备份数据。

分布式部署

将系统部署到不同计算机上，减小硬件损坏造成的数据丢失的影

响。

8 系统维护设计

8.1 概述

- 连接数据库时，需要在创建数据库连接、销毁数据库连接时使用 try catch 语句捕获异常，对不同的错误信息尽量区分输出。使用 MyBatis 中间件，有效防止 SQL 注入攻击,保障网站的安全。
- 管理员有权对整个网站的状况进行控以防系统出现不可预计的错误防止系统显示不合法信息
- 系统维护人员每次维护后需要留下完备可读的系统维护日志便于管理员和其他维护人员查看
- 做好代码版本管理和溯源的工作,发现异常和 bug 时及时排查并找到责任人进行修复

8.2 检测点设计

8.2.1 搜索词条

- 关键词普通搜索
- 关键词与运算搜索
- 关键词或运算搜索
- 关键词非运算搜索
- 关键词混合逻辑搜索

- 关键词过滤高级搜索
- 关键词条件搜索
- 搜索过程中显示历史记录
- 搜索过程中进行智能补全
- 拼音搜索

8.2.2 查看搜索结果

- 搜索结果页面显示
- 搜索结果类型筛选
- 书籍详情页面跳转
- 搜索结果排序
- 书籍搜索结果筛选
- 个性化推荐内容显示

8.2.3 详情内容展示

- 书籍基本信息显示
- 书籍知识网络可视化呈现
- 个性化词条推荐

8.3 相互维护设计

- 硬件资源维护：定期清理服务器硬盘垃圾，可根据网站实际需求选择升级服务器性能

- 数据库维护：定期备份数据库文件
- 系统功能升级：根据用户实际访问平台的需求，对于系统功能进行合理的更新

9 模块设计计划

9.1 项目架构设计

BookWise 电影垂直搜索引擎采用 B/S 结构，一共分为前端、搜索服务器、后端服务器、爬虫四个大模块。模块之间采用 JSON 进行数据传输。其中，前端使用 Vue 全家桶技术，后端使用 Spring Boot，数据持久化使用 Elastic Search、MySQL，项目管理工具使用 cornerstone，代码托管平台使用 GitHub。

前端提供搜索引擎的图形化交互界面，合理、美观地呈现搜索结果以及内容展示；后端接收前端的请求，配置资源，查询底层数据库与搜索引擎信息，返回结构化数据；数据爬取端收集多维度的信息，持久化到数据库中，同时进行数据分析工作。

9.2 项目任务分解

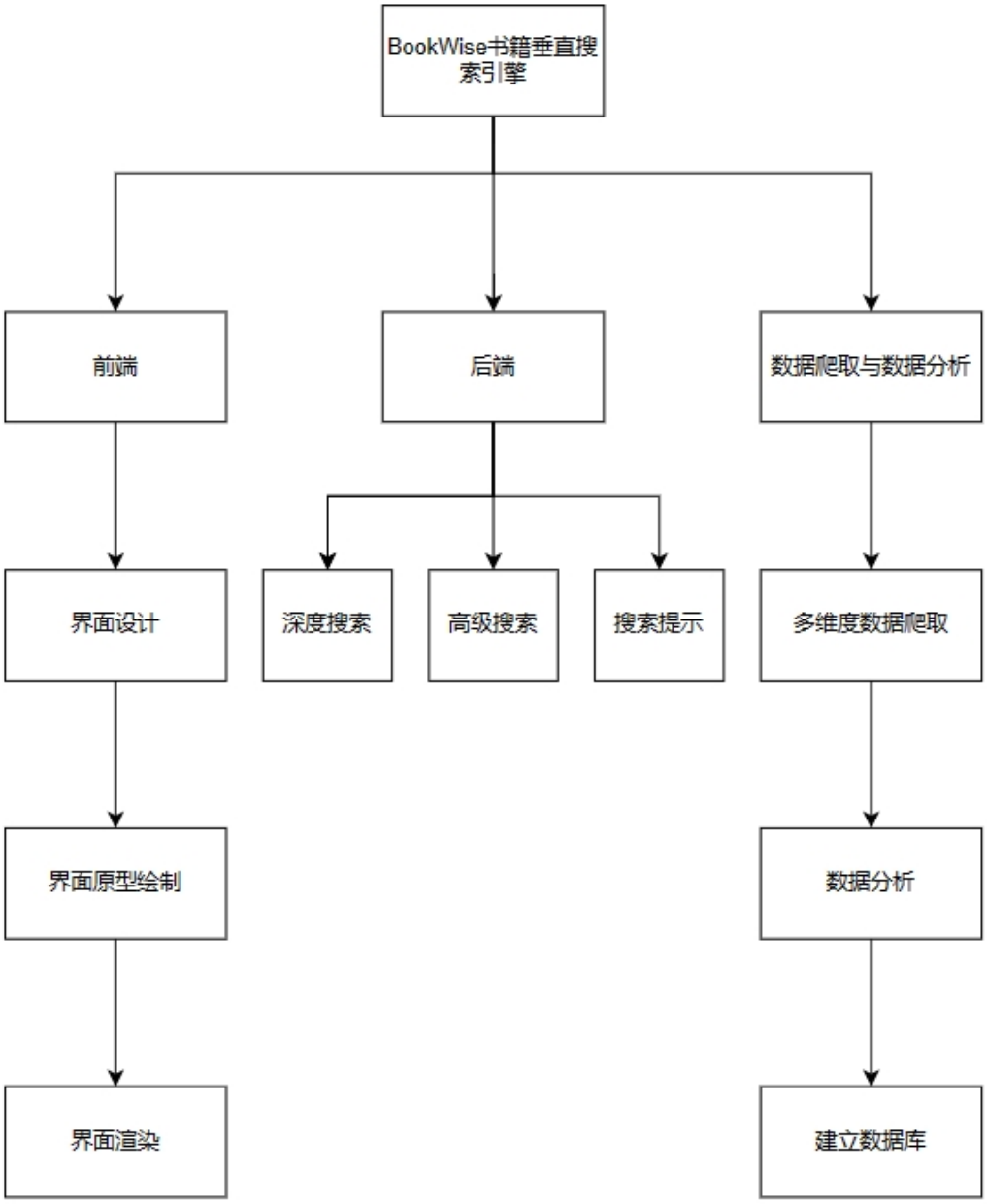
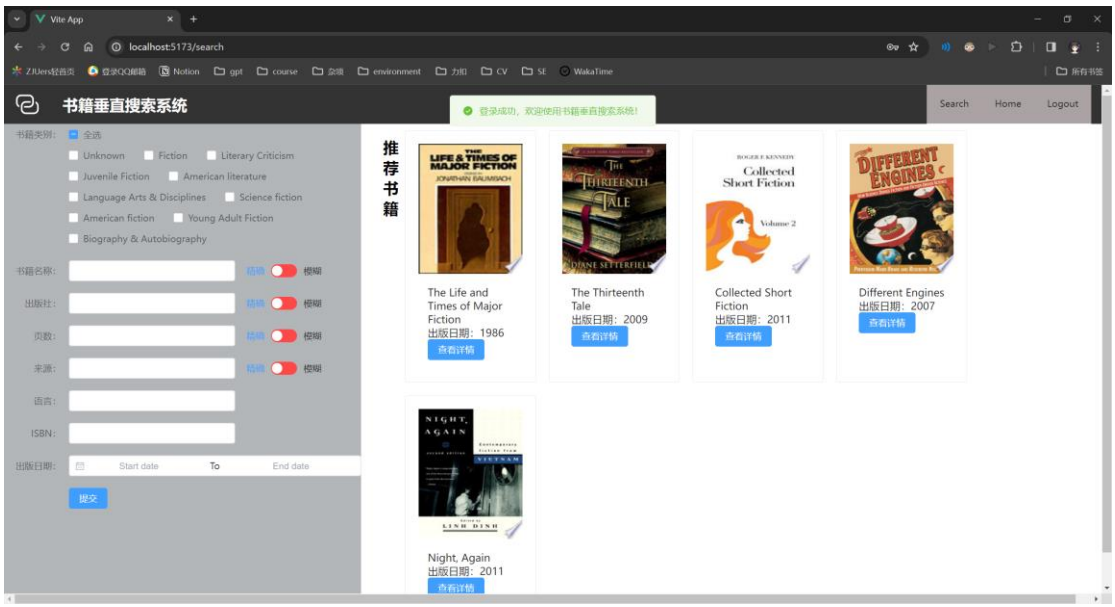


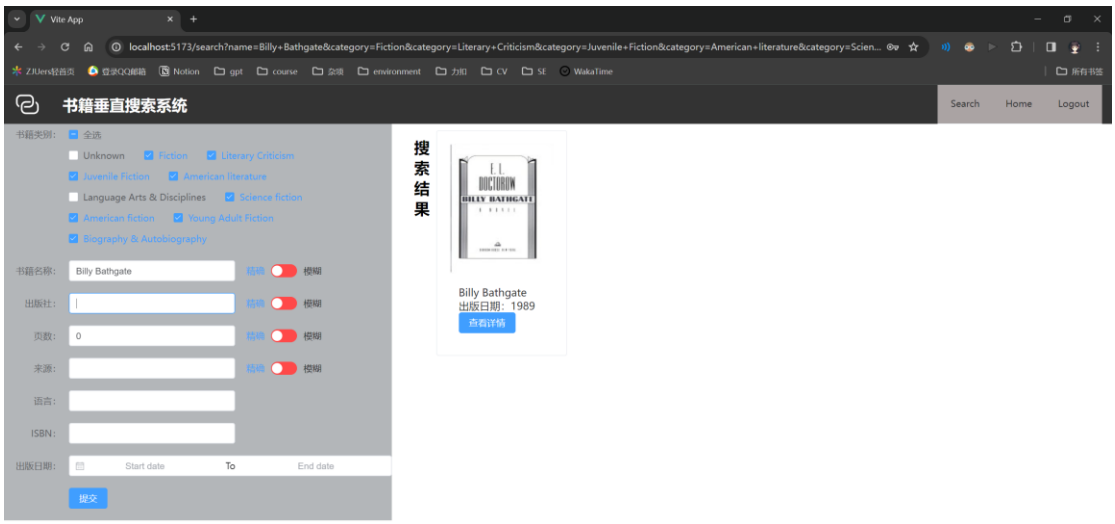
图 9-2-1 项目 WBS 划分图

9.3 前端

9.3.1 搜索首页模块

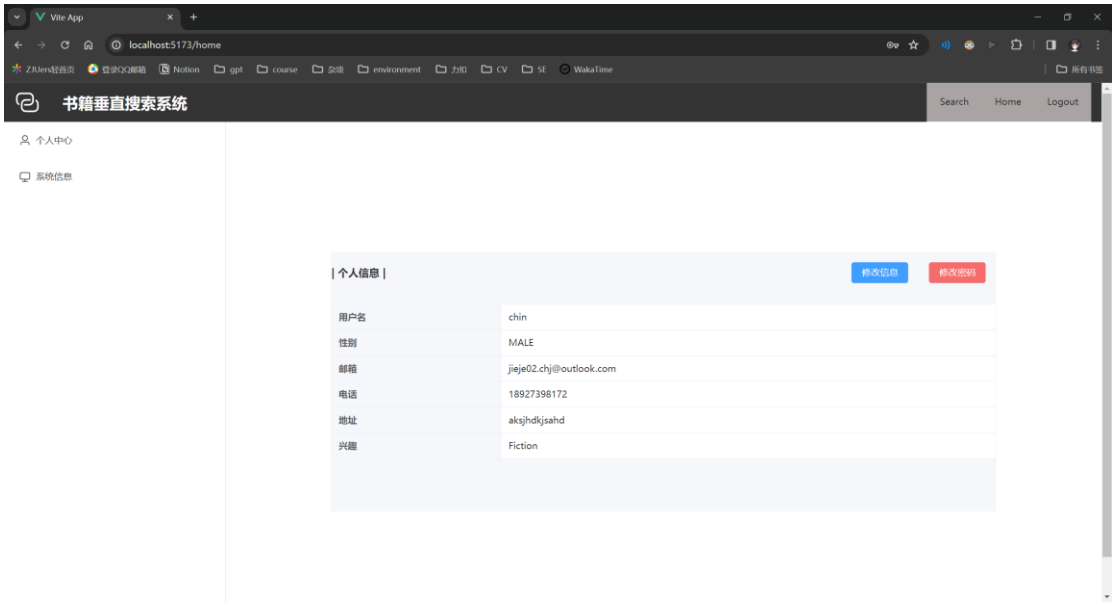


9.3.2 搜索结果展示模块

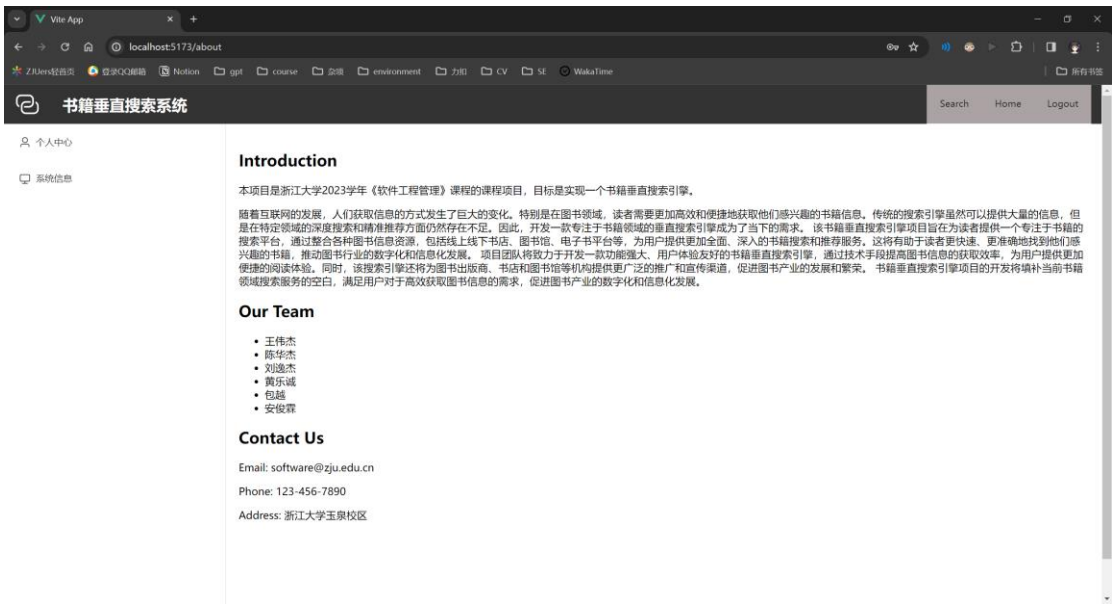


9.3.3 书籍信息展示模块

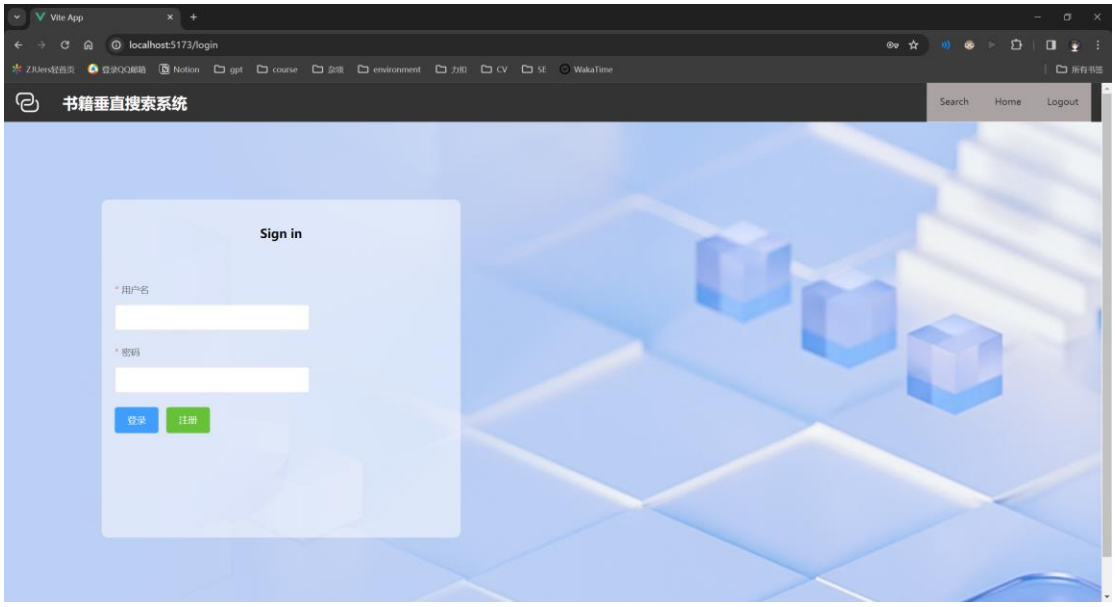
9.3.4 用户信息模块



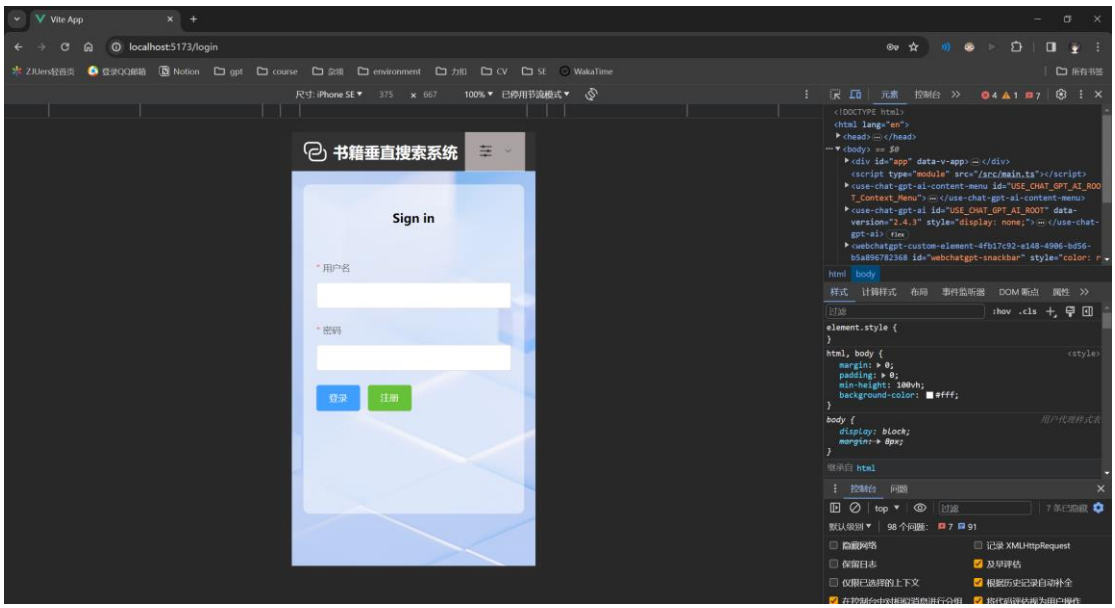
9.3.5 系统信息模块



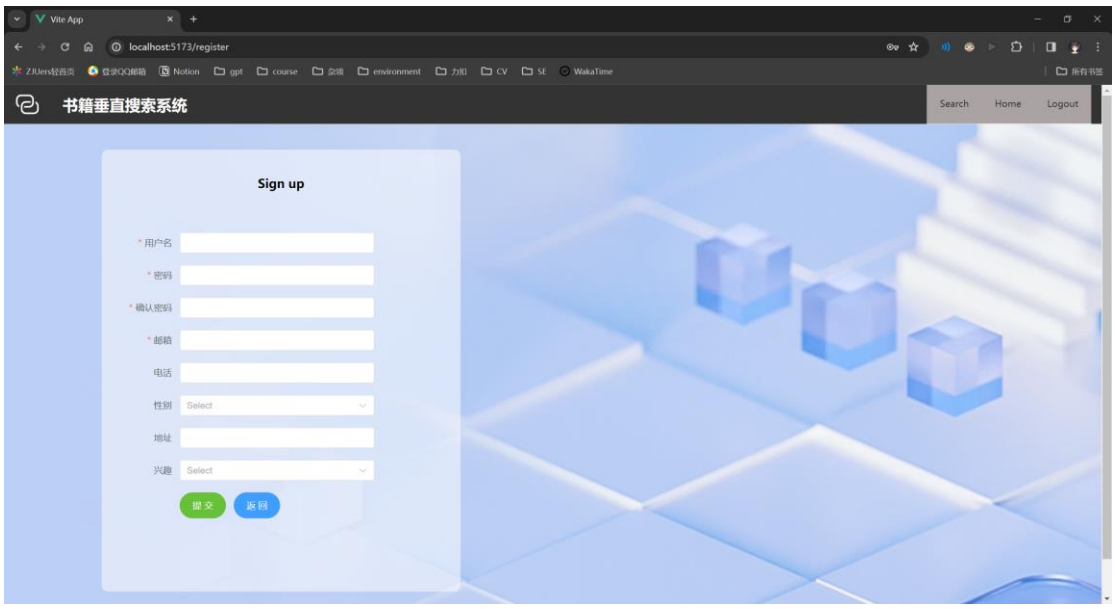
9.3.6 登录模块



移动端适配：



9.3.7 注册模块



9.4 搜索服务器

表 9-4-1 搜索服务器

功能编号	功能名称	详细描述	负责人	截止时间
ES-0001	数据导入	向后端服务器提供导入数据的 HTTP API 向管理员提供直接可执	王伟杰	2023.12.6

		行的脚本和 JSON 模板配置文件		
ES-0002	数据搜索	向后端服务器提供搜索的 HTTP API	王伟杰	2023.12.6
ES-0003	索引建立	向后端服务器提供导入数据的 HTTP API 向管理员提供直接可执行的脚本和 JSON 模板配置文件	黄乐诚	2023.12.6
ES-0004	数据更新	向后端服务器提供搜索的 HTTP API	包越	2023.12.6

9.5 后端服务器

表 9-5-1 后端服务器

功能编号	功能名称	详细描述	负责人	截止时间
BE-0001	数据爬取	定期爬取新的数据并更新到数据库和 Elastic Search 服务器中	王伟杰	2023.12.6
BE-0002	数据清洗过滤	对爬取到的数据和信息进行合理的预处理和清洗过滤，去掉相似性过高的数据，对数据中的一些缺失内容进行补全	王伟杰	2023.12.6
BE-0003	数据搜索	由后端向前端提供调用 ES 搜索引擎的接口	黄乐诚	2023.12.6
BE-0004	数据查询	向管理员提供一系列的接口供管理员创建索	包越	2023.12.6

		引、导入数据		
BE-0005	数据管理	后端提供格式化查询的数据接口，可以帮助用户在掌握了基本信息之后从关系型数据库中查到更多的详细信息	刘逸杰	2023.12.6

9.6 数据模块设计

9.6.1 信息爬取

为了满足用户对于信息资料的需求，需要在网络上爬取满足该垂直领域的所有资源。为了保证系统的高响应性，采用提前爬取而非实时爬取的方式，将相关数据经过提炼和整合，根据其应用范围的不同存储入搜索引擎、数据库中。

9.6.2 定期爬取

为了保证搜索引擎的时效性，需要定期对数据资料进行更新、追加。对于时效性极高的搜索指数等数据，系统通过定时脚本以天为单位启动相应爬虫实现定时爬取的目的。对于时效性要求较低且信息更新速率低的电影基本信息等数据，可由系统管理员以周或月为单位手动操作，以减少对于网络的负担。

9.6.3 唯一标志

数据资料需要在系统中有唯一的标志来进行区分，系统采用数据

来源网站简写+资源在该网站中的 ID 予以标志，并将书籍 ID、作者 ID 等特定的数据存入数据库中，以指导下一步的资源爬取以及在定期爬取中去重。

9.6.4 数据分析

基于垂直搜索引擎中的深度分析需求，系统提供了书籍相关词云、同类词条推荐等功能。在爬取完数据分析所需的资料后将其存入数据库中，启动脚本对于采集到的数据进行再加工，将统计分析后的数据、生成的词云等存入数据库中，当用户访问时返回深度分析后的数据。