浙江大学

Techhub 后端知识库搜索引擎系统

软件需求规格说明书

Group 08

黄亦非、傅诤哲、胡瀚丹、陈鑫

2019/07/07

修订表：

编号	生成版本	修订人	修订章节与内容	修订日期
1	1.0	G08 小组所有成员	所有章节	2019/07/07
2	2.0	G08 小组所有成员	产品范围、产品描述、优先级表、系统特性	2019/07/08
3				
4				
5				

审批记录

版本	审批人	审批意见	审批日期
1.0	黄亦非	通过	2019/07/07
2.0	黄亦非	通过	2019/07/08

目录

1 引言

1.1 编写目的

本文档的主要目的是定义《Techhub 后端知识库搜索引擎系统 1.0 版》及其所有子系统的业务需求。包括了对文档本身的介绍、系统描述和特性、功能性与非功能性的需求。同时也对系统与文档所必须遵循的标准做出了规定，提高开发效率，也为日后软件的升级迭代提供参考。

1.2 文档约定

在阅读该文档时，你会注意到某些字词或者需求条款使用了不同的字体、大小和粗细。这种突出显示是有规律的并且各自有着不同的含义；用同一种风格来显示不同字词以表明它们属于同一种类型。

本文档中出现的“Techhub 后端知识库搜索引擎系统”、“Techhub 系统”等都是指代的都是本系统。

1.3 预期读者和阅读建议

本说明书针对了各种不同的预期读者，包括：

- 用户：建议阅读本文档的第 2、3、5 章，以了解软件的总体信息、具体功能、运行环境、配置以及运行时的要求约束等。
- 开发者：建议阅读全文，以充分了解文档结构、软件各方面需求、根据需求了解系统结构，从而完成系统分析与设计，高效率地进行系统开发。
- 测试人员：建议阅读本文档的第 4、5、6、8、9、10 章，以在测试时确定软件是否完成所有预期功能以及现有功能是否存在漏洞或者故障。
- 项目经理：建议阅读本文档的第 2、3、5、6 章，以了解系统的总体信息、大体功能、存在的问题以及需求工程中涉及的分析模型。
- 营销人员：建议阅读本文档的第 2 章，以了解系统的总体状况、功能、特性和运行环境等，从而更好地向用户介绍产品。

1.4 产品的范围

本次开发工程在之前完成了项目的《项目计划书》文档，对整个项目的大致计划进行了相关的规划，具体的计划和范围可参考该文档，本文档只提供关于该系统的简短描述。

用户范围包括了可以通过网站的首页输入查询语句或者相关的关键字进行搜索查询、同时能够支持用户通过不同的网站来源、搜索范围来进行过滤，得到用户最想要获取的相关信息。

该项目的功能范围及分类如下：

- 用户查询：用户输入关键字或者查询语句进行相关的搜索和查询。
- 结果筛选：用户通过关键字或者搜索范围进行查询结果过滤。
- 数据爬取与处理：建立项目的知识数据库，对网页信息进行大规模爬取，并且进行去重，结构化组织，数据筛选，正文的碎片化数据和相关的整合。
- 搜索引擎：能够自动进行分词和索引，根据计算算法进行排序，返回用户查询最相关的文章链接和内容。

1.5 参考资料

- 《课程项目描述》（课程资料）

  提供者：课程教学小组

- 《软件需求（第 2 版）》（课本）

  作者：Karl E. Wiegers（美）

  译者：刘伟琴　刘洪涛

  出版社：清华大学出版社

- 《IT 项目管理（第 6 版）》（课本）

  作者：Kathy Schwalbe（美）

  译者：杨坤

  出版社：机械工业出版社

- 《软件需求说明书国家标准规范（GB856T——88）》

  提供者：课程教学小组

2 综合描述

本章概述了《Techhub 后端知识库搜索引擎系统》软件的前景、用户类、运行的环境以及设计上的约束和限制。主要目的是给读者对该系统做一个总体的介绍。

2.1 产品描述

Techhub 后端知识库搜索引擎系统主要是面向在校大学生用户，致力于在互联网上学习后端的相关技术，能够帮助用户快速的搜索相关领域的技术文档，包括该技术的描述、使用手册、教学视频、使用当中的相关问题等方面。对于用户而言，可以根据输入的问题或者关键字，准确的定位用户的意图，并且返回最准确的结果给用户。同时用户也可以根据提供的几个过滤条件对返回的结果进行对应的过滤，实现更加精确的结果定位。

同时系统对于数据的来源和处理也非常的重视，致力于提供最准确、最完整、最精确的知识库架构。对于数据我们会进行相关的过滤、去重、结构化信息的提取等等数据处理的动作，保证系统数据的稳定和准确。

2.2 用户类及特征

用户类	特征与说明
网站用户	1. 主要用户 2. 同时使用该网站的用户数目可能较多，主要服务的对象为想要找到特定领域的专业技术文档或者是相关问题解答。 3. 要求能够返回足够精确和足够深度的结果 4. 能够根据用户的喜好进行返回数据的过滤 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首页。
系统管理员	1. 次要用户 2. 但是权限比较高，具有数据库的更新和审核等管理权限 3. 面向的用户数目比较多，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。

2.3 运行环境

- **终端环境：**

1. 使用终端主要是网站用户的智能手机和个人电脑。移动端操作系统主要集中于安卓或者 iOS，我们的网站能够做到多个移动端的自动适配。PC 端操作系统集中于 Windows、MacOS、Linux，浏览器为主流的浏览器，比如说 Firefox、Chrome、Safari 等主流的现代浏览器。

- **网络环境：**

2. 用户地理位置基本是在线的大学生在校或者是在家学习相关技术使用。需要保证至少 300 名用户同时使用我们的搜索引擎的要求。包括数据的存储能力，服务器响应能力，网络吞吐能力和数据安全等相关特性等方面。

2.4 设计与实现的限制

- LI-1：网站用户的范围主要是限制在需要学习相关技术的在校学生或者相关的技术人员。
- LI-2：网站的知识库主要聚焦在后端的特定领域，在后续网站进入维护阶段可以考虑新增其他技术的相关技术文档的支持。
- LI-3：用户的过滤条件有所限制，后期考虑可以加入更多的自定义的过滤条件，满足用户的需求。

2.5 假设与依赖

要成功开发该 Techhub 后端知识库搜索引擎系统，我们首先要了解甲方的需求，获得相应支持与认可；同时，还需要从不同的专业网站、数据库、论文知识库当中爬取相关的网页，经过数据处理之后形成我们的后端知识库，同时搭配相关的搜索引擎辅助，完成后端知识库搜索引擎的功能；还需要配备 Visual Studio Code, Eclipse, Xcode, IntelliJ IDEA，PyCharm 等专业开发软件和可联网的 Windows 端与 MacOS 端电脑。其次，我们团队要具有较好的合作精神、工作能力和空余时间。

假设如下：

AS-1：收集期望，按照使用者需求进行修改

AS-2：可以与管理员联系进行一些调整

AS-3：用户规模和信息存储空间具有最大上限

AS-4：网站用户能够根据实际的使用体验提出相关的建议

依赖如下：

DE-1：系统使用时，必须与服务器连接良好，正常工作

DE-2：用户输入查询的信息之后，搜索引擎能够很快速的响应

DE-3：用户能够根据需要进行返回结果的过滤

DE-4：后期考虑扩大后端知识库的内容和范围，提供更加精、深、专的专业技术搜索服务

3.系统优先级排序表

这是本次进行需求开发的后端知识库搜索引擎网站的需求优先级分析，主要是基于四个指标来进行相关的衡量，实现需求给客户带来的收益（Benefit），不实现需求给用户带来的损害（penalty），实现需求需要耗费的成本（Cost），实现需求的风险（Risk）。在评价该优先级的时候，我们使用的是质量功能部署法（QFD法），它是一种能够综合客户的意见和软件实际进行需求开发和设计实现的综合的评估方法。

本优先级计算出来最终优先级排序较高的功能将作为必须需要实现的功能，排序相对靠后的需求将作为可选的功能进行实现。

但是需要注意的是，本方法计算出的优先级仅仅作为后续活动当中的参考，准确性会受到包括收益、损失、成本、风险估计等能力的限制。因此本次的优先级表只能作为一种参考性的表，并且在后续的需求活动当中将会对该需求优先级表进行不断的改进和持续的跟进。

	Relative Benefit	Relative Penalty	Total Value	Value	Relative Cost	Cost	Relative Risk	Risk	Priority
查询搜索结果	9	9	27	0.203007519	5	0.147058824	5	0.142857143	0.700228833
手动导入数据	5	6	16	0.120300752	4	0.117647059	7	0.2	0.378724589
用户条件查询数据	8	9	25	0.187969925	6	0.176470588	6	0.171428571	0.540300025
信息爬取	7	8	22	0.165413534	5	0.147058824	6	0.171428571	0.519372309
自动滤重	5	6	16	0.120300752	4	0.117647059	3	0.085714286	0.591561548
自动分类	4	3	11	0.082706767	6	0.176470588	5	0.142857143	0.25900277
自动爬取	6	4	16	0.120300752	4	0.117647059	3	0.085714286	0.591561548
Total	44	45	133	100%	34	100%	35	100%	

4.系统特性

4.1 Web 前端子系统

4.1.1 描述

该子系统用于提供搜索引擎的用户界面与交互操作。用户可以输入想要搜索的内容，点击搜索按钮，前端发送请求给后端，返回符合的搜索结果并呈现给用户，也可以通过筛选功能进一步限定搜索条件。

4.1.2 刺激/响应序列

1. 查询搜索内容

刺激	响应
点击搜索栏	搜索栏聚焦
输入搜索内容	显示输入的语句
点击搜索按钮	将搜索内容发送给后端，进行查询与处理

2. 展示搜索结果

刺激	响应
后端返回搜索结果	前端接收数据
跳转页面	显示所有搜索结果条目
点击条目	打开原网站链接
点击下一页（分页）	向后端发送请求获取更多数据

3. 筛选搜索条件

刺激	响应
选择搜索来源	前端发送请求给后端，只搜索特定网站来源的数据
选择搜索范围（使用手册、相关问题等）	前端发送请求给后端，只搜索特定范围的数据

4.1.3 功能性需求

1. 查询搜索内容

SearchBar.input	输入搜索内容
SearchBar.search	点击搜索
SearchBar.clear	清空搜索内容

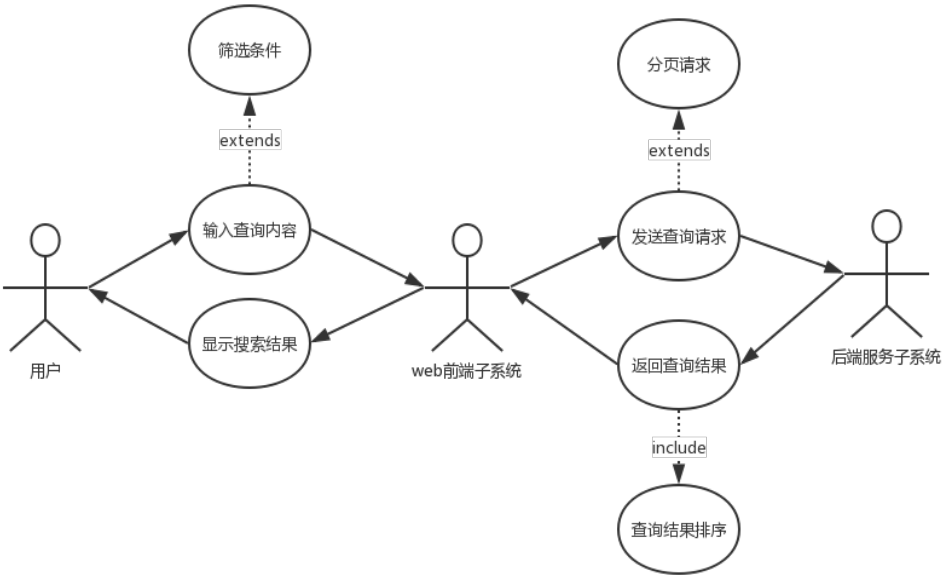
2. 展示搜索结果

Result.showAll	显示所有搜索结果
Result.sort	搜索结果排序
Result.reset	重置搜索结果
Result.showDetail	显示某个条目的具体内容
Result.nextPage	显示下一页搜索结果

3. 筛选搜索条件

Condition.select	选择搜索条件
Condition.apply	应用搜索条件
Condition.reset	重置搜索条件

4.1.4 用例图


筛选条件
extends
输入查询内容
显示搜索结果
用户
web前端子系统
分页请求
extends
发送查询请求
返回查询结果
include
查询结果排序
后端服务子系统

4.1.5 用例描述

ID 和名称	UC-1 查询搜索结果		
创建人	胡瀚丹	创建日期	2019 年 7 月 6 日星期六
首要角色	网站用户	次要角色	爬虫子系统、后端服务子系统
描述	用户可以在搜索栏中输入想要搜索的内容，点击搜索按钮，前端发送请求给后端，返回符合的搜索结果并呈现给用户，也可以通过筛选功能进一步限定搜索条件。		
触发条件	用户试图搜索信息		
前置条件	1. 系统加载好搜索界面 2. 信息数据库在线		
后置条件	1. 后端接收到前端的请求 2. 后端从数据库中查询内容 3. 返回显示搜索结果界面		
正常流程	1. 用户进入搜索主页面 2. 用户输入搜索内容 3. 点击搜索按钮 4. 前端发送请求给后端		

	5. 后端返回查询结果 6. 跳转页面，显示搜索结果
可选流程	1. 筛选网站来源 2. 筛选搜索范围 3. 搜索结果排序（按时间、点击量等）
异常	1. 搜索内容为空 2. 搜索结果不存在
优先级	高
使用频率	高，一天使用多次
商业规则	符合互联网相关的法律法规
其他信息	用户输入字符都需要进行转义，以防止攻击
假设	用户输入的内容属于知识库范围

4.2 搜索服务器（Solr）

4.2.1 描述

搜索服务器用于数据导入功能、数据建立索引、分词功能、排序功能、分页功能、搜索功能。后端 Web 服务器可以实时调用搜索服务器数据导入、数据搜索的接口。

4.2.2 刺激/响应序列

1. 手动导入数据

刺激	响应
手动导入数据	数据导入进 Solr

2. 条件查询记录

刺激	响应
条件查询记录	显示记录

4.2.3 功能需求

1. 手动导入数据

solrClient.dataImport	手动导入数据

2. 条件查询记录

solrClient.query	条件查询记录

4.2.4 用例图

4.2.5 用例描述

ID 和名称	UC-2 手动导入数据		
创建人	陈鑫	创建日期	2019 年 7 月 6 日星期六
首要角色	管理员	次要角色	管理员
描述	管理员可以进入 Solr 的管理界面，选择 Core 和 Entity 进行后动 data import。管理员可以利用后端 Web 服务器提供的 Http 接口进行 data import。		
触发条件	管理员试图去手动导入数据		
前置条件	1. 管理员加载好 Solr 管理界面		
后置条件	2. 管理员进入 Solr 管理界面 3. 管理员选择 Core 和 Entity		
正常流程	1.0 直接浏览器输入 Solr 管理界面的地址 1. 选择 Core 2. 选择 Entity 3. 点击 data import 按钮		
可选流程	1.1 管理员通过搜索界面的管理员入口进入 Solr 服务器		

	1. 管理员点击搜索界面的管理员入口 2. 管理员进入 Solr 管理界面 3. 选择 Core 4. 选择 Entity 5. 点击 data import 按钮
异常	1.0 E1 选择错误的 Entity 和 Core 3. 选错 Entity 4. 选错 Core
优先级	高
使用频率	较高，约 1 天一次
商业规则	符合相关的法律法规
其他信息	需要进行管理员登录
假设	管理员身份是真实的

4.3 后端 Web 服务器

4.3.1 描述

后端 Web 服务器提供搜索记录的接口，并且定时调用 Solr 提供的 data import 服务进行数据更新。

4.3.2 刺激/响应序列

1. 条件搜索

刺激	响应
用户选择分类	无
用户输入想要搜索的内容	验证是否为空
点击搜索按钮	调用 Solr 服务器搜索接口，返回给前端

2. 定时 data import

刺激	响应
每 10 分钟导入一次	调用 Solr 服务器数据导入接口

4.3.3 功能需求

1. 条件搜索

search	条件搜索

2. 定时 data import

dataImport	定时 data import

4.3.4 用例图

4.3.5 用例描述

ID 和名称	UC-3 用户条件查询数据		
创建人	陈鑫	创建日期	2019 年 7 月 6 日星期六
首要角色	用户	次要角色	用户
描述	用户可以在搜索界面先选择搜索的方向分类，然后输入想搜索的内容进行搜索。如果数据库有相关信息则直接返回，否则就帮用户在线查找相关内容。		
触发条件	用户试图搜索		
前置条件	1. 用户加载好搜索页面		
后置条件	1. 用户选择分类 2. 用户输入想要搜索的内容		
正常流程	1.0 用户条件搜索内容 1. 用户进入搜索界面 2. 用户选择分类 3. 用户输入想要搜索的内容 4. 用户点击搜索按钮		
可选流程	无		

异常	1.0 E1 搜索内容为空 1. 搜素内容
优先级	高
使用频率	较高，约 1 分钟一次
商业规则	符合相关的法律法规
其他信息	用户输入字符都需要进行转义，以防止攻击
假设	用户输入的内容是可被识别的

4.4 爬虫子系统

4.4.1 描述

该爬虫子系统提供系统后台知识库的数据来源，会对不同的网站上的多元异构的数据进行爬取，然后存放到后台数据库当中。系统会对收集到的数据进行清洗、加工、处理，然后形成系统后台知识库的数据库存储，为搜索引擎提供数据支持。

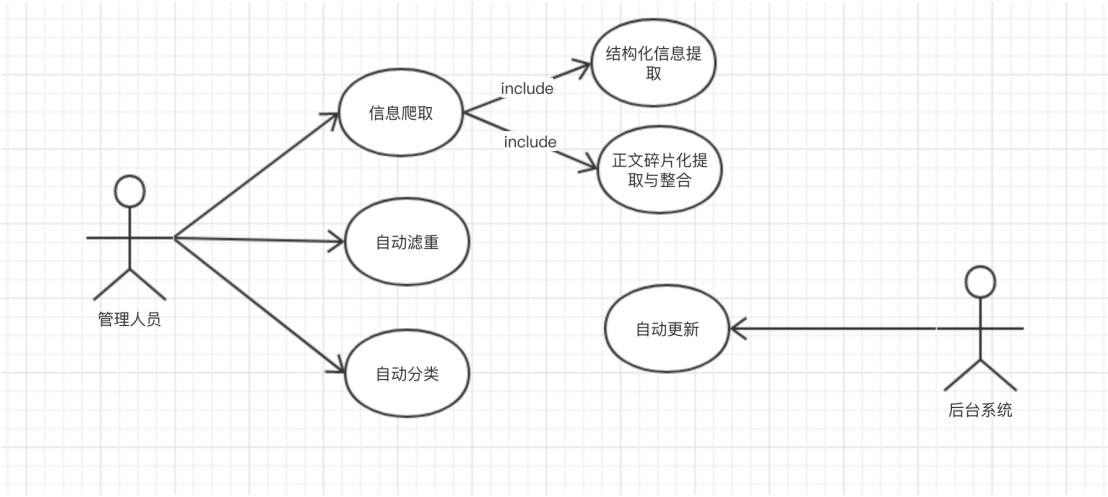
4.4.2 刺激/响应序列

刺激	响应
管理人员运行信息爬取程序	信息爬取
管理人员在本机上运行自动滤重程序	自动滤重
管理人员在本机上运行自动分类程序	自动分类
后台要求动态爬取	自动爬取

4.4.3 功能性需求

Scrapy.get	信息爬取
Scrapy. removeDuplicates	自动滤重
Scrapy. setclass	自动分类
Scrapy. dynamicget	自动爬取

4.4.4 用例图



4.4.5 用例描述

ID 和名称	UC-4 信息爬取		
创建人	傅诤哲	创建日期	2019 年 7 月 6 日星期六
首要角色	管理人员	次要角色	
描述	系统在服务用户的过程中一般情况下不会实时爬取来自网络的数据，而是通过读取数据库中经过提炼和整合的数据来更好得服务用户。为了保证系统的信息具有一定的实时性，管理员必须隔一段时间爬取一次来自网络的各种信息。		
触发条件	管理员调用接口爬取数据		
前置条件	1. 系统所在机器有空余计算能力		
后置条件	1. 系统在指定位置保存爬取数据		
正常流程	0 信息爬取 1. 管理人员在本机上运行程序 2. 系统遍历资源池中的地址，开启多个线程递归式爬取网站内容，并提取与整合正文碎片化并提取结构化信息，将资源保存在本地指定目录。 3. 指定目录按网站名建立文件夹储存网站资源		
可选流程			

异常	E1 网站反爬或布局更新，爬虫无效
优先级	高
使用频率	约每周一次
商业规则	
其他信息	
假设	技术性博客更新迭代具有一定迟缓性

ID 和名称	UC-5 自动滤重		
创建人	傅诤哲	创建日期	2019 年 7 月 6 日星期六
首要角色	管理人员	次要角色	
描述	在管理人员爬取大量网络文本资源后，需要对一些相似性很高的文本进行过滤重复或者是合并，由于网络上的资源有很多重复的内容，给用户提供多个重复内容的情况在各类搜索引擎中并不少见，因此对文本数据做滤重十分重要。		
触发条件	管理员调用接口对指定文件夹中的文本自动滤重		
前置条件	1. 系统所在机器有空余计算能力 2. 管理员已经爬取了大量网络资源并保存在指定位置		
后置条件	1. 系统在指定位置保存滤重后的数据		
正常流程	0 自动滤重 1. 管理人员在本机上运行信息爬取程序 2. 管理人员在本机上运行自动滤重程序 3. 系统遍历资源文件夹中的文件，计算文件两两之间的相似度，对相似度高于阈值的数据进行合并或过滤。 4. 指定目录按网站储存过滤后的资源		

可选流程	新增数据中需要对新增数据去重
异常	E1 阈值过低或者算法不准导致遗失一些数据
优先级	高
使用频率	约每周一次
商业规则	
其他信息	
假设	技术性博客有大量重复资源

ID 和名称	UC-6 自动分类		
创建人	傅诤哲	创建日期	2019 年 7 月 6 日星期六
首要角色	管理人员	次要角色	内容数据库
描述	技术博客一般都具有一些明显的标签来表面他是属于什么技术的、以及介绍了技术什么反面的知识，因此，如果对他们做一定的分类，可以有效得提高搜索的效率。		
触发条件	管理员调用接口对指定文件夹中的文本自动分类		
前置条件	1. 系统所在机器有空余计算能力 2. 管理员已经爬取了大量网络资源并保存在指定位置		
后置条件	1. 数据库中导入了分类完的数据		
正常流程	0 自动分类 1. 管理人员在本机上运行信息爬取程序 2. 管理人员在本机上运行自动滤重程序 3. 管理人员在本机上运行自动分类程序 4. 系统遍历资源文件夹中的文件，选择博客合适的标签 5. 系统将带有标签的数据保存到数据库中		
可选流程	新增数据中需要对新增数据分类		

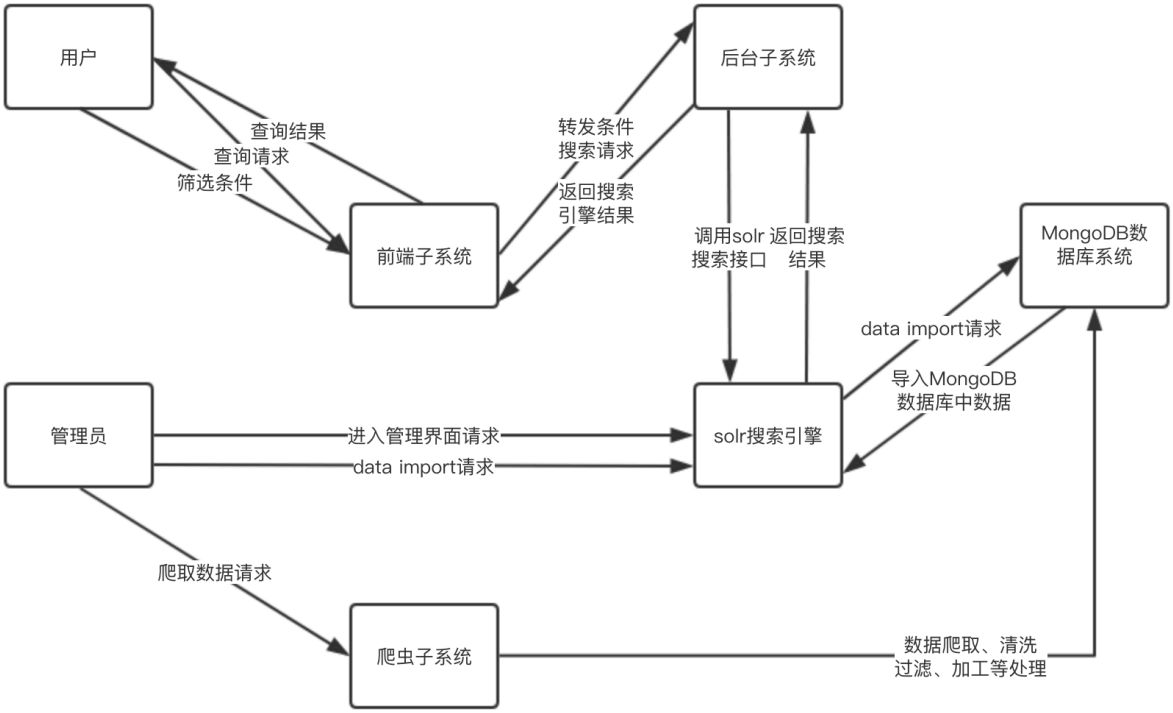
异常	E1 分类算法未能很好的表现
优先级	高
使用频率	约每周一次
商业规则	
其他信息	
假设	技术性博客有一定的标签可以用来分类

ID 和名称	UC-7 自动更新		
创建人	傅诤哲	创建日期	2019 年 7 月 6 日星期六
首要角色	后台系统	次要角色	内容数据库
描述	当查询得到的数据不够多不能满足用户的要求，后台会启动爬虫去一些著名搜索引擎爬去数据，并和己方数据库中的内容做比对来去重和分类，最后将数据插入数据库供以后查询所用。		
触发条件	后台搜索引擎搜索得到的数据从量和质量上来说低于阈值		
前置条件	1. 系统已成功部署 2. 后台调用接口启动爬虫自动更新		
后置条件	1. 内容数据库更新		
正常流程	0 自动更新 1. 后台搜索得到匹配数据，并发现数据条目较少，切匹配度不高 2. 后台启动爬虫自动更新 3. 系统爬取数据，去重，分类 4. 系统将数据保存在数据库		
可选流程			

异常	E1 分类算法未能很好的表现
优先级	高
使用频率	约每周一次
商业规则	
其他信息	
假设	一开始爬取的数据不能很好的满足用户需求

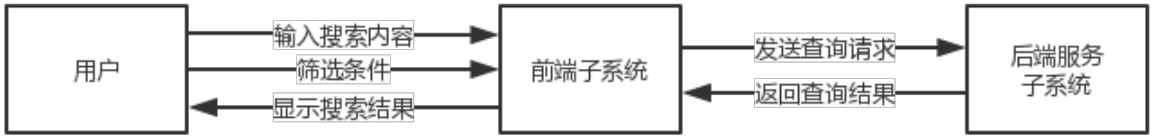
5. 数据流图

5.1 环境层数据流图

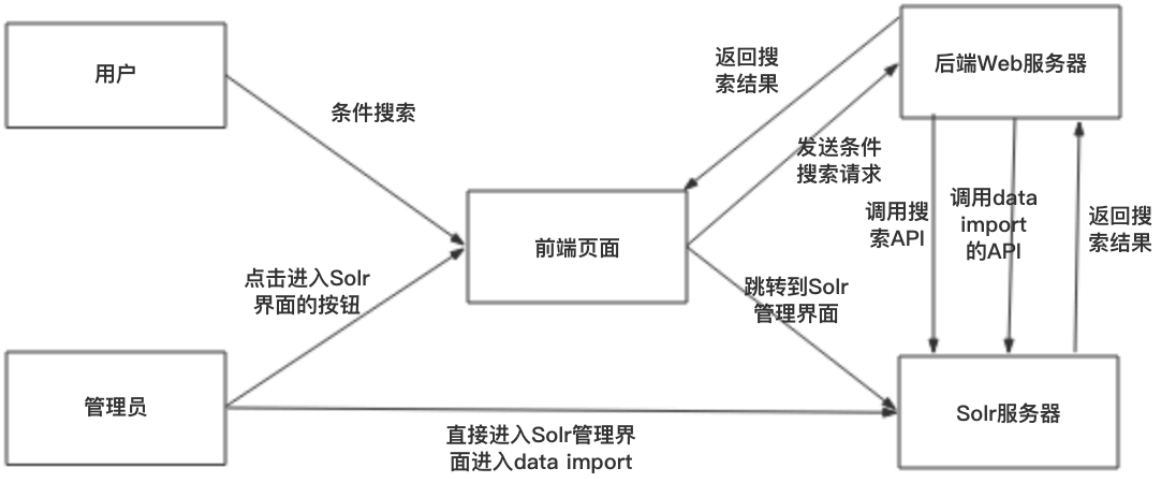


5.2 第二层数据流图

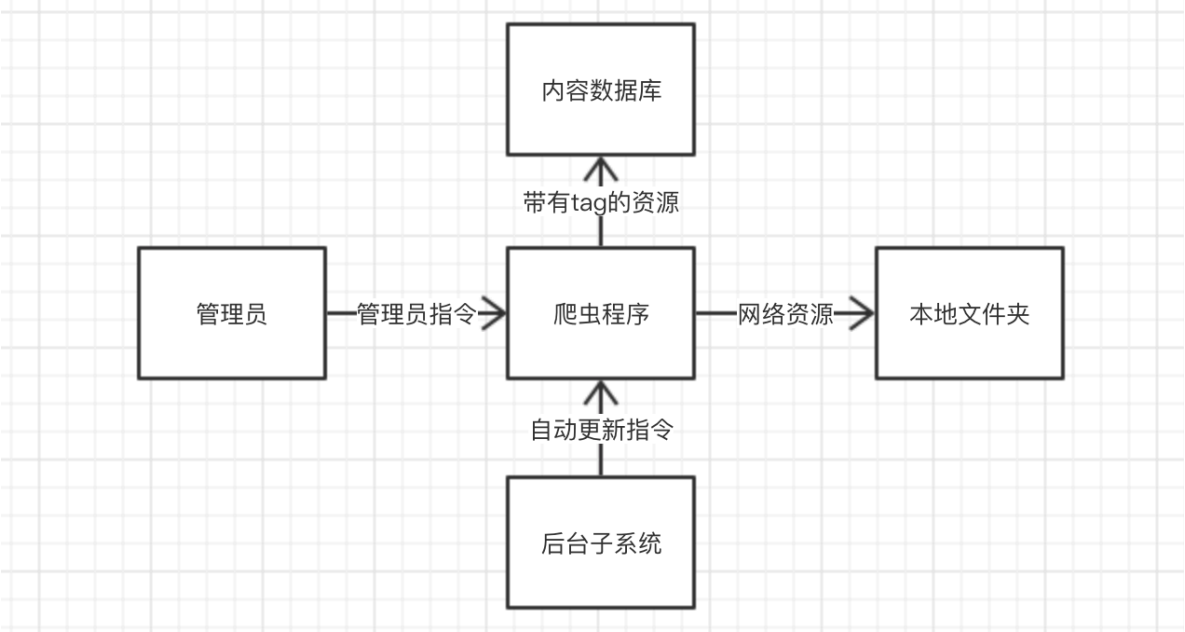
5.2.1 前端子系统数据流图



5.2.2 后端子系统数据流图

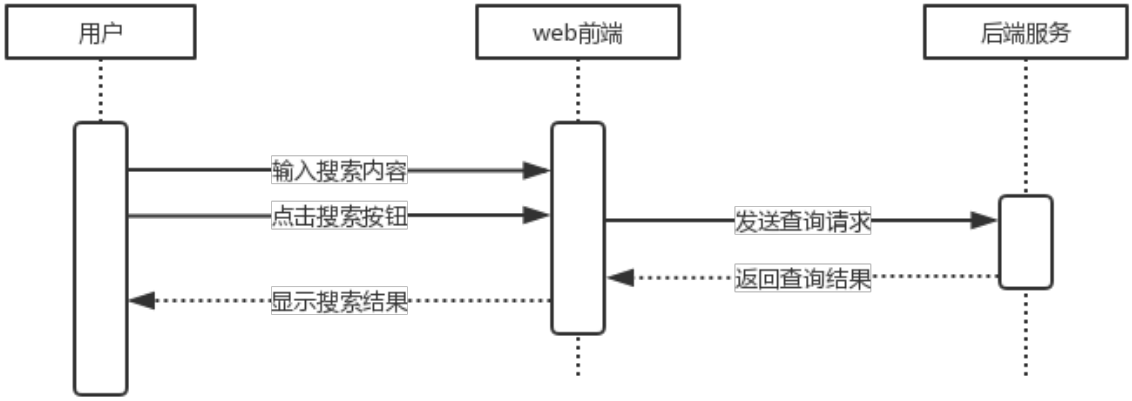


5.2.3 爬虫子系统数据流图



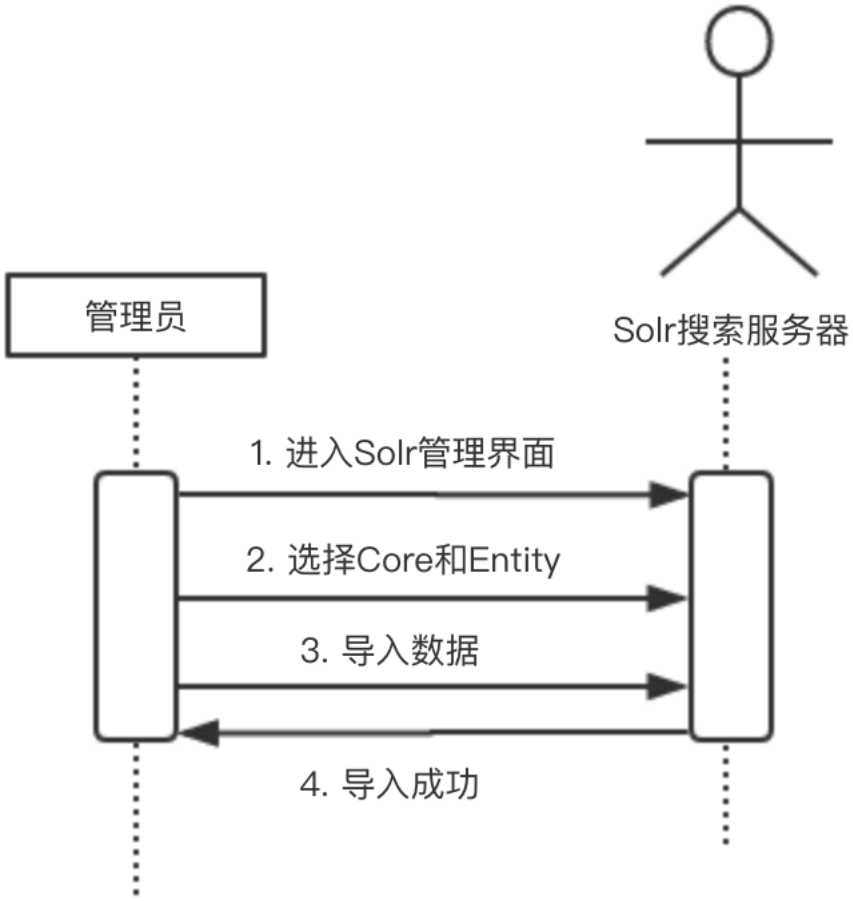
6. 时序图

6.1 前端子系统

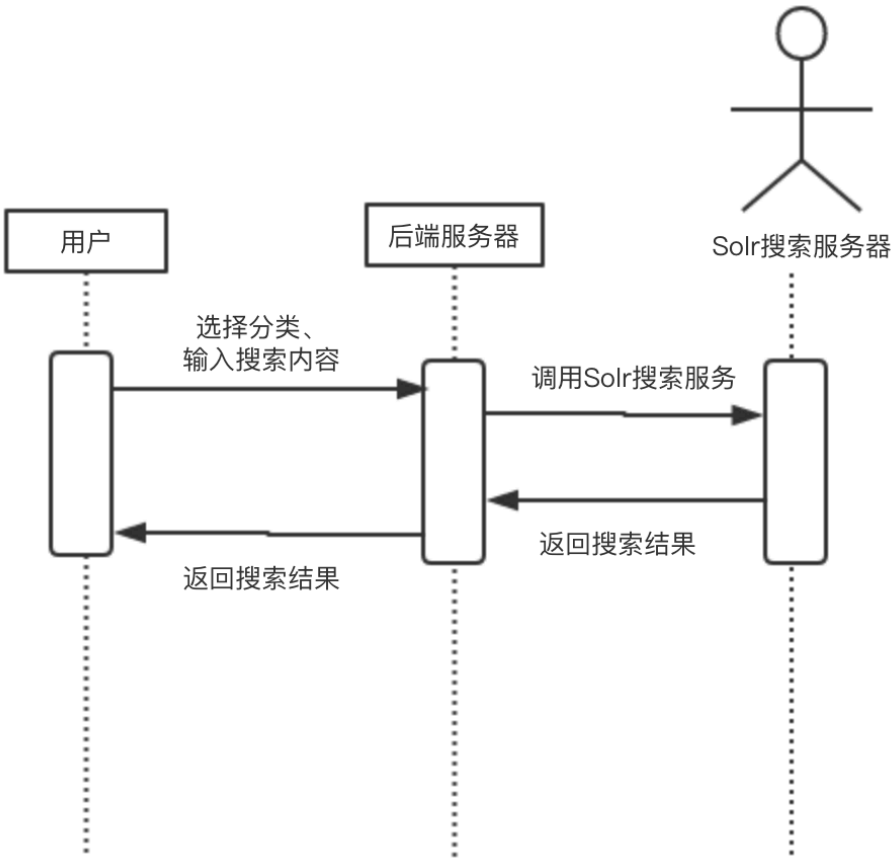


6.2 后端子系统

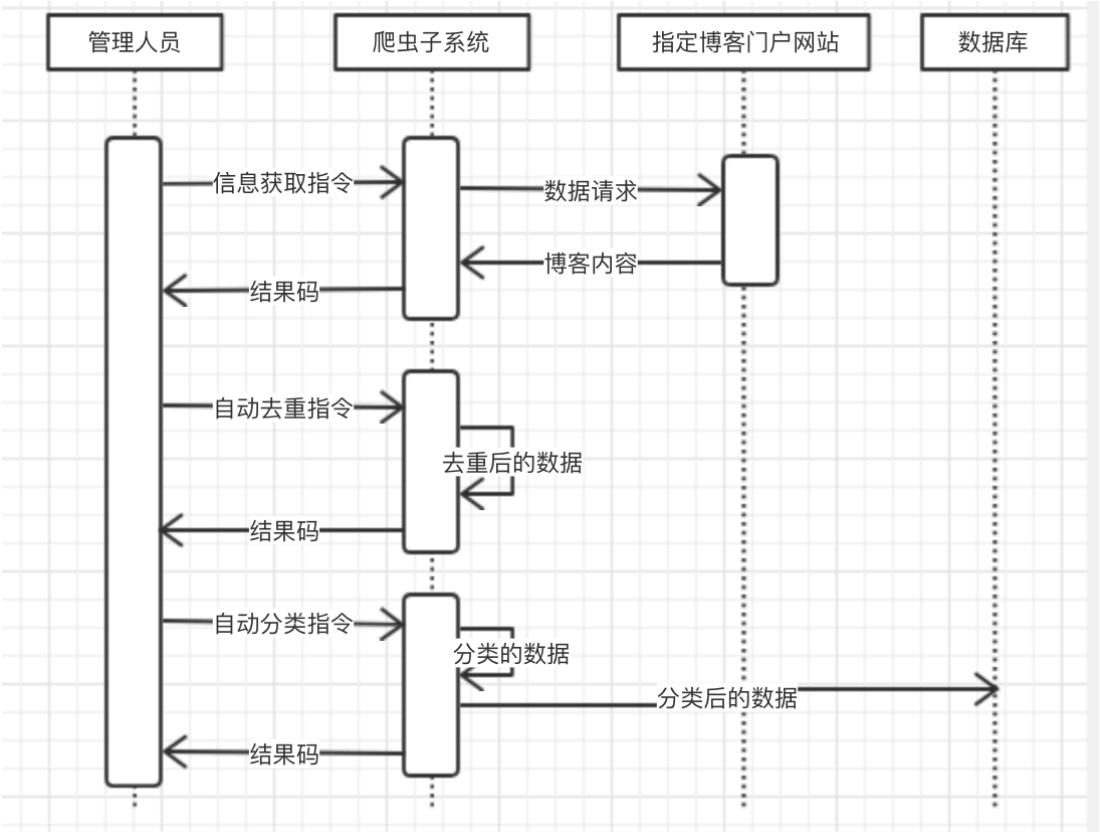
1. 管理员导入数据



2. 用户搜索

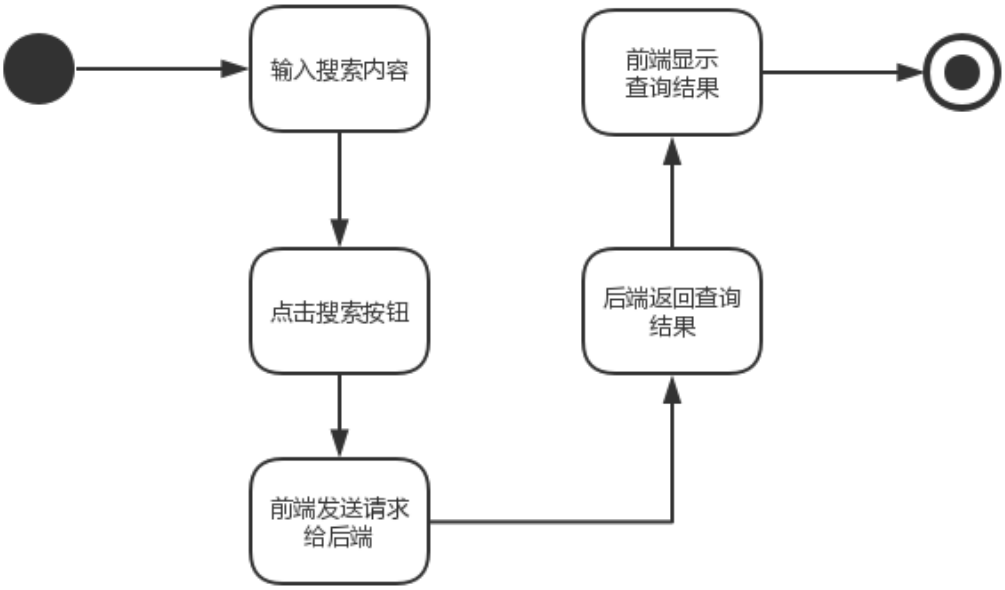


6.3 爬虫子系统


管理人员
爬虫子系统
指定博客门户网站
数据库
信息获取指令
数据请求
博客内容
结果码
自动去重指令
去重后的数据
结果码
自动分类指令
分类的数据
分类后的数据
结果码

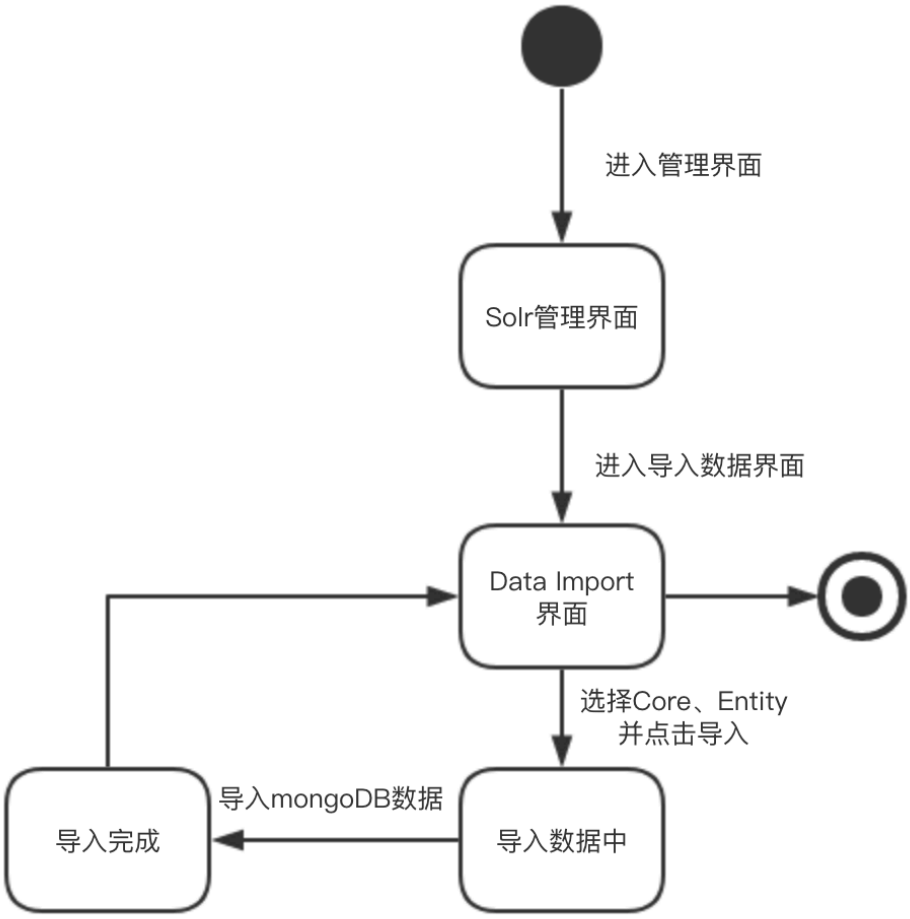
7. 状态图

7.1 前端子系统

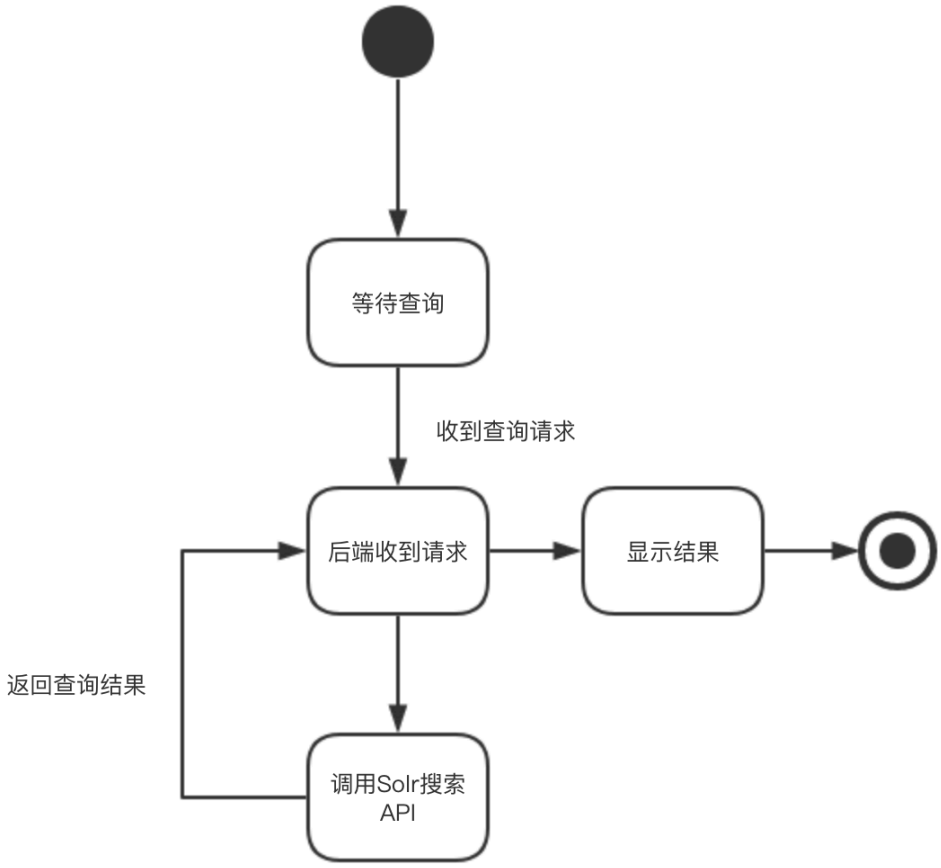


7.2 后端子系统

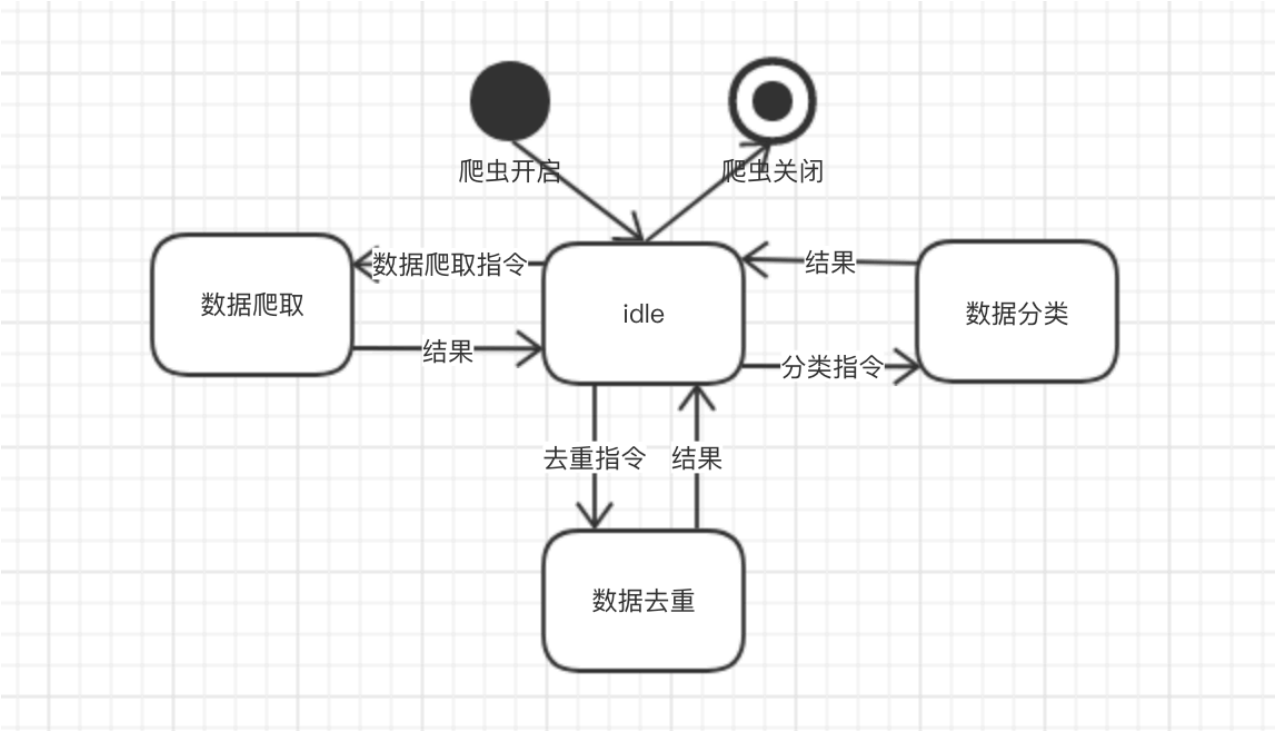
1. 管理员导入数据



2. 用户搜索



7.3 爬虫子系统



8. 数据需求

8.1 CRC 卡

8.1.1 web 前端子系统

类：web前端子系统	
职责	协作者
查询用户输入内容	
显示查询结果	后端服务系统
筛选搜索条件	

8.1.2 爬虫数据子系统

类：爬虫数据子系统	
职责	协作者
爬取数据信息	
自动滤重与分类	
自动更新	
提取结构化信息	
正文碎片化提取与整合	

8.1.3 后端服务子系统

类：爬虫数据子系统	
职责	协作者
搜索引擎data import	爬虫数据系统
建立索引	
数据排序	
调用solr服务	
查询分页	web前端系统
封装查询服务，返回搜索结果	Web前端系统

8.2 数据字典

8.2.1 数据元素定义表

编号	数据元素名	类型	值域	说明
E1	搜索内容	Varchar(32)	4{[1..9|a..z|A..Z|_]}32	
E2	文章标题	Varchar(32)	8{[1..9|a..z|A..Z|_]}32	
E3	文章链接	Varchar(64)	8{[1..9|a..z|A..Z|_]}32	
E4	文章标签	列表		
E5	筛选条件	表单		

8.2.2 数据流定义表

编号	数据	来源	去向	组成	说明
D1	用户搜索内容	Web前端	后端服务	E1	
D2	搜索结果	搜索引擎	后端服务、Web前端	E1+E2+E3+E4	
D3	条目详情	Web前端	后端服务	E2+E3+E4	

D4	筛选搜索条件	Web前端	搜索引擎、后端服务	E1+E5	

8.2.3 外部项定义表

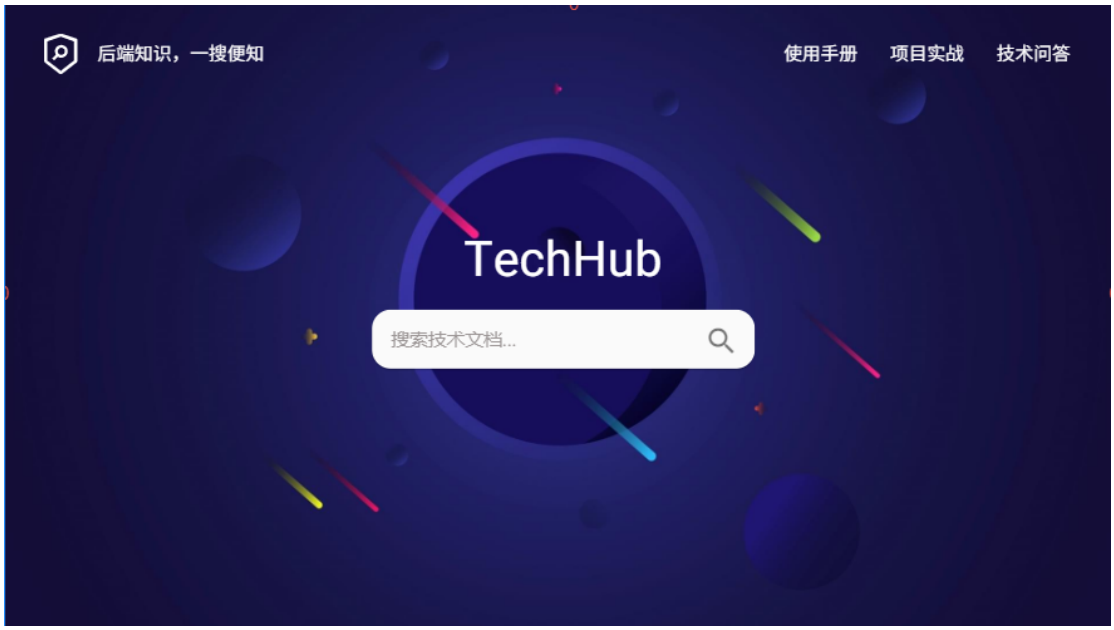
编号	外部项名称	输入数据流	输出数据流	说明
W1	用户	搜索内容	搜索结果	
		筛选条件		
		排序方式		

9. 外部接口需求

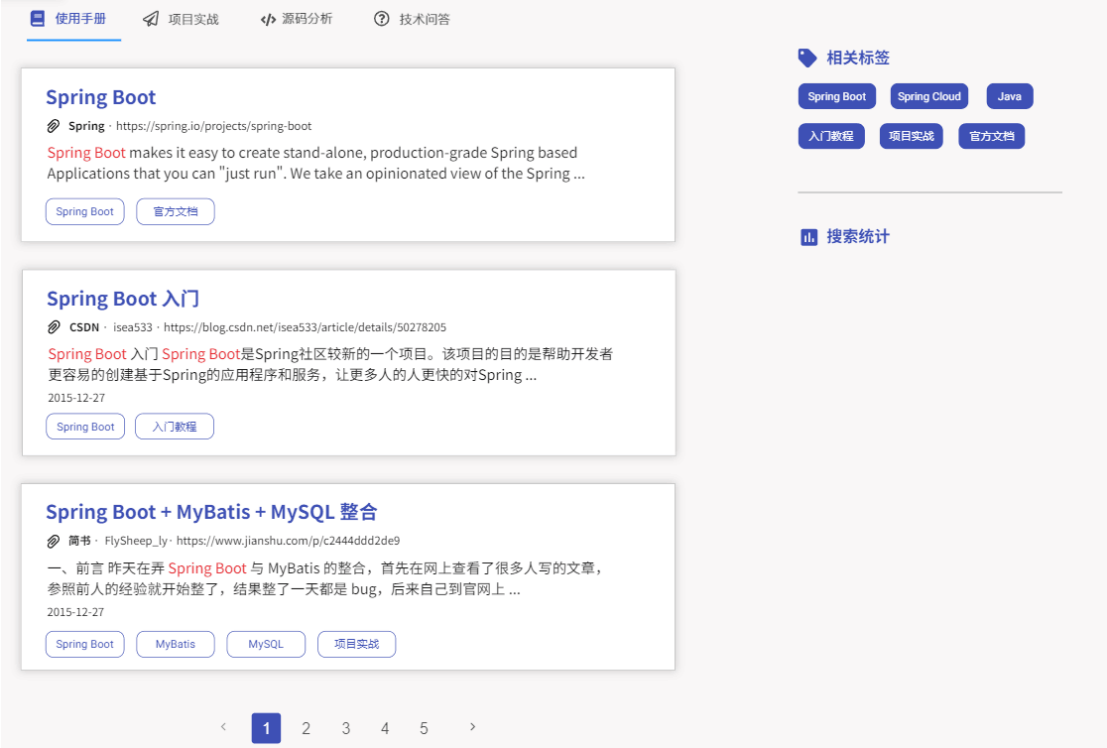
9.1 用户接口

- 所有界面采用 WEB 页面形式并适配移动端
- 简洁易用，方便用户的使用习惯
- 用户界面的具体细节将在软件概要设计说明书中描述

用户搜索界面原型：



搜索结果展示界面原型：


使用手册
项目实战
源码分析
技术问答
Spring Boot
Spring · https://spring.io/projects/spring-boot
Spring Boot makes it easy to create stand-alone, production-grade Spring based Applications that you can "just run". We take an opinionated view of the Spring ...
Spring Boot
官方文档
Spring Boot 入门
CSDN · isea533 · https://blog.csdn.net/isea533/article/details/50278205
Spring Boot 入门 Spring Boot是Spring社区较新的一个项目。该项目的目的是帮助开发者更容易的创建基于Spring的应用程序和服务，让更多人的人更快的对Spring ...
2015-12-27
Spring Boot
入门教程
Spring Boot + MyBatis + MySQL 整合
简书 · FlySheep_ly · https://www.jianshu.com/p/c2444ddd2de9
一、前言 昨天在弄 Spring Boot 与 MyBatis 的整合，首先在网上查看了很多人写的文章，参照前人的经验就开始整了，结果整了一天都是 bug，后来自己到官网上 ...
2015-12-27
Spring Boot
MyBatis
MySQL
项目实战
1
2
3
4
5
相关标签
Spring Boot
Spring Cloud
Java
入门教程
项目实战
官方文档
搜索统计

9.2 软件接口

- 数据库接口：实现数据库与 web 页面数据的通讯与存取
- 操作系统接口：实现爬虫爬取信息的整理和收集，以及进行相关的数据处理和过滤。

9.3 硬件接口

- 服务器端建议使用专用服务器

9.4 通信接口

将使用的通信功能的需求：

- Web 浏览器
- HTTP 协议
- TCP/IP 协议
- 数据传输速率至少达到 50kb/s

10. 其他非功能性需求

10.1性能需求

- PR-1：系统将容纳 500 以上的静态用户（注册用户）以及 200 以上的动态用户（在线用户），并发数将有 100 以上。
- PR-2：当单个用户在线时，系统响应用户动作的时间将小于 0.2s。
- PR-3:对相关搜索信息或者关键字进行搜索时,系统响应时间不超过5s。
- PR-4：系统将能够连续运行一周。

10.2防护性需求

- SA-1：当网站崩溃时，系统将把用户尚未保存的资料自动进行保存。
- SA-2：允许用户进行数据的备份和恢复，以弥补数据的破坏和丢失。
- SA-3：该系统将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。这些错误记录便于查找错误的原因，日志同时记录用户的关键性操作信息。

10.3安全性需求

- SE-1：每个用户的查询数据尽可能的进行保密
- SE-2：除管理员之外的用户不能进行后台知识库的手动添加
- SE-3：当系统发现用户对进行交换的数据进行无意或恶意的修改时，系统将防止此操作
- SE-4：系统对一些重要的数据按一定算法进行加密，如一些重要参数等

10.4 软件质量属性

10.4.1 对用户重要的属性

10.4.1.1 可用性

- AV-1：工作日期间，当地时间早上 6 点至午夜，系统的可用性至少达到 99.5%；晚上 6 点到 12 点，系统的可用性至少达到 99.95%。

10.4.1.2 有效性

- EF-1：在预计的峰值负载条件下，至少 25%的处理器能力和应用程序可用内存必须留出备用。
- EF-2：在过多用户同时在线时，性能降低程度不会超过百分之二十。

10.4.1.3 灵活性

- FL-1：一个至少具有 6 个月产品支持经验的程序维护人员，可以在大于一个小时的时间内为系统添加一个新的模块，包括代码修改和测试。

10.4.1.4 完整性

- IN-1：只有管理员权限的用户能对后台的知识数据库进行手动导入和相关修改，系统将组织其他任何用户对后台知识数据库的修改。
- IN-2：系统能撤回管理员权限用户的有误操作。

10.4.1.5 可靠性

- RE-1：由于软件故障引起搜索失败或数据更新失败的概率不超过 1%。

10.4.1.6 健壮性

- RO-1：如果在用户搜索关键字时候发生故障，那么下次同一用户启动程序时，搜索引擎能恢复在故障发生时的相关搜索结果。
- RO-2：系统具有一定的容错和抗干扰能力，在非硬件故障或非通讯故障时，系统能够保证正常运行，并有足够的提示信息帮助用户有效正确地完成任务。

10.4.1.7 易用性

- US-1：网站用户应该能在 5 秒之内，得到相关的搜索引擎返回的结果。
- US-2：系统管理员应该可以在 30s 内，完成对后台知识库的导入和修改。
- US-3：以前从未使用过该系统的网站用户，能在不超过 5 分钟的适应后，正确的使用该系统的搜索引擎和相关的过滤条件得到自己想要的搜索结果。
- US-4：该系统能够对数据的来源进行相关的过滤，聚焦在数据的精、深、和准确度等方面，完成垂直搜索引擎的相关特性和功能特点。

10.4.2 对开发者重要的属性

10.4.2.1 可维护性

- MA-1. 程序维护人员应该在 20 小时或更短时间内，对现有报告进行更改。
- MA-2. 函数调用的嵌套层次不能超过两层。
- MA-3. 每个软件模块中，注释与源代码语句的比例至少为 1:2.

10.4.2.2 可测试性

- TE-1. 一个模块的最大循环复杂度不超过 20。

10.5 用户文档

- 安装手册：Word 格式文件、PDF 格式文件
- 用户手册：Word 格式文件、PDF 格式文件
- 在线帮助

11. 运行环境规定

11.1 设备

服务器设备要求	
CPU	≥2.0GHz
内存	≥4.0GB
硬盘	≥100GB
硬盘转速	≥5400rpm
外设	
键盘	能用即可
鼠标	能用即可
显示器	能用即可
通讯设备	
网线	具有良好的数据传输能力
网卡	100M

PC 端设备要求
能上网的计算机即可，推荐使用 Chrome 浏览器

移动端设备要求
Android 或 iOS 操作系统、能上网的智能手机

11.2支持软件

操作系统：Windows10/Mac OS

前端开发框架：React

后端开发：Eclipse、Pycharm

数据库平台：Mongodb

Web 服务器：Apache

搜索引擎：Solr

开发工具：能支持网页开发的工具均可，如 JetBrain WebStorm、Visual Studio Code

测试工具：请测试人员自行选择，推荐使用 LoadRunner

建模工具：根据项目情况，请自行选择合适方便的建模工具，《软件工程-实践者的研究方法》一书中介绍了很多建模工具，可对比参考。

办公软件：Microsoft Office 2016 系列产品

浏览器：Chrome、Firefox、IE11、Safari

附录 A：词汇表

词汇	解释
权限	指用户职能的范围，即各种用户所登录的界面、所接触的数据、所进行的操作等范围。
角色	指系统中具有某些特定权限的用户。
用例	Use Case（用例）是一个 UML 中非常重要的概念，在使用 UML 的整软件开发过程中，Use Case 处于一个中心地位。用例是对一组动作序列的抽象描述，系统执行这些动作序列，产生相应的结果。这些结果要么反馈给参与者，要么作为其他用例的参数。
数据流图	数据流图（Data Flow Diagram）：简称 DFD，它从数据传递和加工角度，以图形方式来表达系统的逻辑功能、数据在系统内部的逻辑流向和逻辑变换过程，是结构化系统分析方法的主要表达工具及用于表示软件模型的一种图示方法。
时序图	时序图（Sequence Diagram），亦称为序列图或循序图或顺序图，是一种 UML 交互图。它通过描述对象之间发送消息的时间顺序显示多个对象之间的动态协作。它可以表示用例的行为顺序，当执行一个用例行为时，时序图中的每条消息对应了一个类操作或状态机中引起转换的触发事件。
状态图	状态图(State chart Diagram)是描述一个实体基于事件反应

	的动态行为，显示了该实体如何根据当前所处的状态对不同的事件做出反应的。通常我们创建 UML 状态图是为了以下的研究目的：研究类、角色、子系统、或组件的复杂行为。
CRC 卡	在 CRC 建模中，用户、设计者、开发人员都有参与，完成对整个面向对象工程的设计。 CRC 卡是一个标准索引卡集合，每一张卡片表示一个类。 类代表一系列对象的集合，这些对象是对系统设计的抽象建模，可以是一个人、一件物品等等，类名写在整个 CRC 卡的最上方。 职责包括这个类对自身信息的了解，以及这些信息将如何运用。诸如，一个人，他知道他的电话号码、地址、性别等属性，并且他知道他可以说话、行走的行为能力。这个部分在 CRC 卡的左边。 协作指代另一个类，我们通过这个类获取我们想要的信息或者相关操作。这个部分在 CRC 卡的右边。 CRC 卡片的背面往往记载着这个类的详细描述和在 CEC 设计中的一些注意事项。
数据词典	数据字典是指对数据的数据项、数据结构、数据流、数据存储、处理逻辑、外部实体等进行定义和描述，其目的是对数据流程图中的各个元素做出详细的说明。 数据字典（Data dictionary）是一种用户可以访问的记录数

	据库和应用程序元数据的目录。主动数据字典是指在对数据库进行修改时，内容可以由 DBMS 自动更新的数据字典。被动数据字典是修改时必须手工更新其内容的数据字典。
系统集成	将不同的系统，根据应用需要，有机地组合成一个一体化的、功能更加强大的新型系统的过程和方法。
管理员	负责该系统日常管理与维护，完成数据的导入与审核，是该系统的角色之一。
网站用户	使用该后端知识库的实际用户，能够通过搜索引擎准确的获取需要的资源。