

**《Web后端技术》创新实践训练项目**

**开题及调研方案报告**

**基于Web的内容管理系统**

**姓 名：**

**段斌、游卓儒**

**学 号：**

**2017213355、2017213358**

**指导教师： 兰名荥**

**北京邮电大学数字媒体与设计艺术学院**

**2020年03月16日**

**基于Web的内容管理系统**

**摘 要**

我们的项目是集简易信息聚合(RSS)和内容整理于一体、对内容统计并进行数据可视化、界面简约以及交互动态化的基于Web的内容管理系统(CMS)。

前端的部分在之前已经大致完成了，本次的目标是将其进一步完善，然后做好网站的后端服务器与接口，设计开发数据库，实现对于管理内容和订阅聚合信息的存储。

我们采用Django作为开发框架，Python为编程语言，Mysql为数据库。

**关键词：**Web后端开发，简易信息聚合（RSS），知识管理（KM） ，数据库 ，Python， Djan

**Web-Based Content Management System**

**Abstract**

Our project is a web-based content management system (CMS) that integrates Really Simple Syndication (RSS) and content organization, providing statistics of content and data visualization, simple interface, and interactive dynamics.

The front-end part of our project has been roughly completed before. The goal this time is to further improve it, then do a good job of the back-end server and interface of the website, design and develop a database, and realize the storage of management content and subscription aggregate information.

We use Django as the development framework, Python as the programming language, and Mysql as the database. .

**Keyword:** Web backend development, Really Simple Syndication (RSS), Knowledge Management (KM), Database, Python, Django

**目录**

[一 引言 5](#_Toc35430981)

[1.1 项目简介 5](#_Toc35430982)

[1.2 背景需求 5](#_Toc35430983)

[1.2.1 阅读的目的 5](#_Toc35430984)

[1.2.2 碎片化阅读 5](#_Toc35430985)

[1.2.3 网络阅读普及 6](#_Toc35430986)

[1.2.4 内容阅读平台增多 6](#_Toc35430987)

[1.3 市场同类产品分析 7](#_Toc35430988)

[1.3.1 内容聚合产品特点和不足 7](#_Toc35430989)

[1.3.2 内容聚合产品特点和不足 9](#_Toc35430990)

[1.4 项目可能的创新点 10](#_Toc35430991)

[1.4.1 模式创新 10](#_Toc35430992)

[1.4.2 内容创新 10](#_Toc35430993)

[二 方案设计 11](#_Toc35430994)

[2.1 总方案设计 11](#_Toc35430995)

[2.1.1 架构模式 11](#_Toc35430996)

[2.1.2 功能设计 12](#_Toc35430997)

[2.1.3 数据库设计 14](#_Toc35430998)

[三 进度安排 17](#_Toc35430999)

[3.1 研发计划 17](#_Toc35431000)

[四 项目团队 18](#_Toc35431001)

[4.1 团队成员简介 18](#_Toc35431002)

[4.1.1 组长 18](#_Toc35431003)

[五 项目难点 19](#_Toc35431004)

[5.1 项目难点 19](#_Toc35431005)

[5.2 可能遇到的问题 19](#_Toc35431006)

[六 参考文献 20](#_Toc35431007)

# 引言

## 项目简介

随着互联网的发展和普及，丰富的网络资源在给人们带来便利的同时，也带来了信息碎片化和信息过载的困扰。合理选择所需要的内容和及时管理内容，有利于提高我们的信息素养。

内容管理(Knowledge Management)由两部分组成，内容强调对象，管理强调方法。内容也即数字内容，可以是文本、图像、音频和视频等数字媒体。管理则为施加在内容对象上的处理，包括收集(聚合)、整理、搜索、存储等。

通过信息聚合RSS(Really Simple Syndication),用户可以订阅指定站点内容，当站点内容更新，RSS会对其内容进行及时推送更新，不用担心原文删除，同时也减少了在不同阅读平台之间切换的时间成本。

碎片化阅读的弊端则是思考不能深入，但如今生活碎片化时间不可避免，鉴于此，我们可以利用信息聚合来实现利用碎片化时间，再通过内容管理来进行深入思考。而我们的基于Web的内容管理系统，即将RSS或者API订阅内容进行聚合阅读，并可对其进行阅读标记和编辑形成内容结构，以达到个人内容管理的目的。从而一体化地高效率、高质量，高集中地完成了对大量碎片化信息的收集、存储，阅读，批注，管理。同时我们随时可以进行对内容进行搜索与归纳，实现对碎片化信息的“深入”思考。

## 背景需求

### 阅读的目的

阅读是人们最重要的认知方式,也是人们认识世界、传承文化的途径。阅读的载体虽然经历了从纸质印刷媒体到数字新媒体的变革,信息以更加丰富多元的方式呈现出来,但阅读是“读者接收材料并提取意义和情感信息的过程,是个体心理对信息加工内化的过程,是物质过程与精神过程的统一”这一本质始终没有改变[1]。

### 碎片化阅读

如今生活环境下，碎片化阅读逐渐成为人们生活中的主要阅读方式。不可否认，碎片化阅读充分利用了人们的零散时间，可以很好得充实人们的日常生活。但其带来的弊端则是不利于人们阅读能力的提高，容易造成人们思维的浅显性[2]。

面对快节奏的生活环境，正确认识碎片化阅读以及合理利用碎片化时间尤为关键。

### 网络阅读普及

网络的发展改变了人们的方方面面，手机、PC和ipad等各种便携设备让人们随时随地的进行沟通、娱乐等活动。同时，网络阅读也逐渐取代纸质阅读，成为现代人的主要阅读方式。

可见，对网络阅读相关的研究和开发不容忽视。

### 内容阅读平台增多

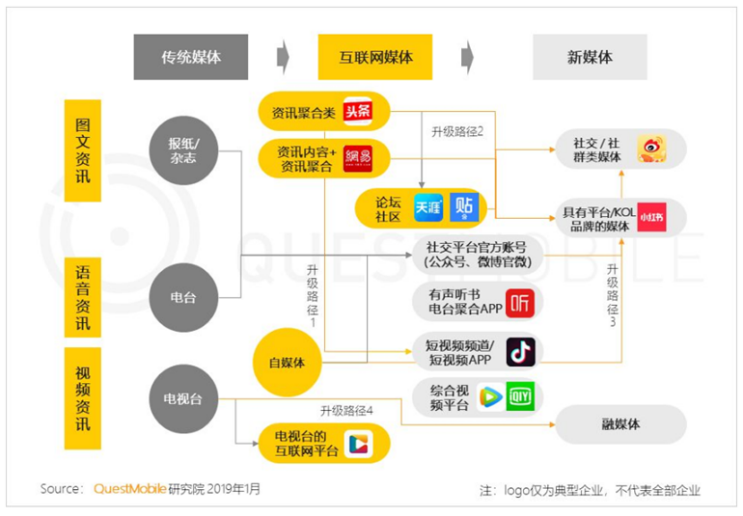


图 1‑ 2-5新媒介带来的价值转移

从图1-2-5可以看出，内容渠道从单一形态到多形态，并且资讯阅读类平台多资讯形态强势发展。身边的公众号，微博，今日头条等各种内容聚合平台层出不穷。面对眼花缭乱的内容聚合平台，如何对其进行选择令人困惑。

截止到2019年初，中国移动网民手机中人均安装app的总量为51款。用户每天查看浏览的app数目并不多，但在app之间切换的时间成本和精力不同忽视。

鉴于以上四点，我提出一种以合理利用碎片化时间，减少切换成本，阅读即时加工内化为目标的新型阅读方式，并在此基础上开发出基于Web的内容管理系统。

## 市场同类产品分析

目前在市场上并没有找到完全同类型的产品，但是找到和分支功能相似的产品，主要就是内容编辑和内容聚合。

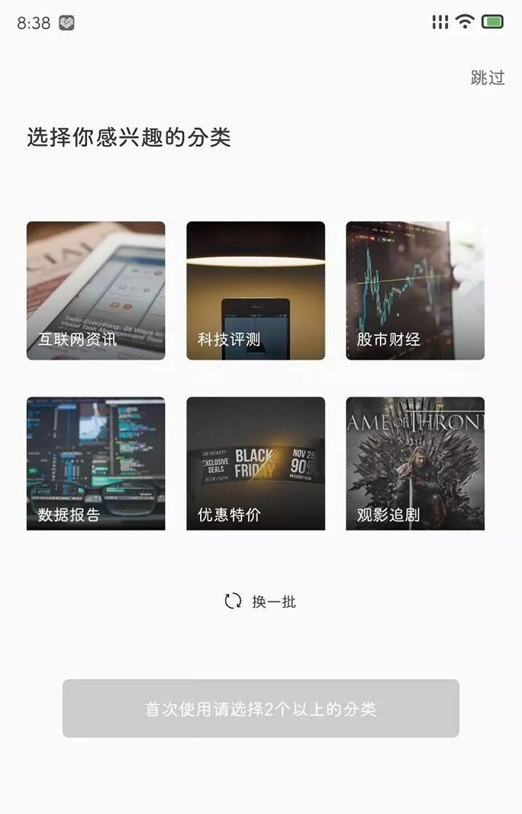
目前市场上只有以Feedly为首的只将内容进行聚合和以有道云笔记为首的只将内容进行编辑和管理的产品，将两者进行结合的产品还未发现。所以，在进行竞品分析时也主要是对这两类产品进行分析。

### 内容聚合产品特点和不足



图 1‑ 3 feedly内容聚合界面

这种内容聚合网页，可以添加很多的订阅源，比如RSS、微博、公众号或者知乎。对于用户来说主要的时间花费在了订阅源的选择上，很多时候订阅源都需要个人去寻找，也有一些人将其用过比较好的订阅分分享出来。有的产品比较好的地方就是可以根据平台订阅人数的多少或者订阅源的相关的程度来对订阅源进行一些推荐。比如快知APP，第一次登入时会要求你选择5个感兴趣的主题，选择之后会按照订阅热度自动给你推荐相关的订阅源，可以自主选择：



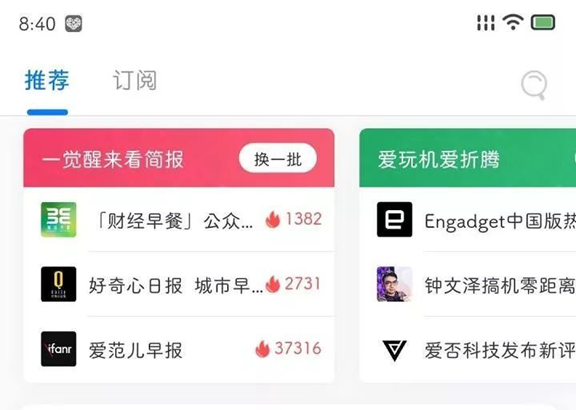


图 1-3-2，1-3-3 快知APP刚登入时的界面

和今日头条的直接推荐不同，这种推荐机制充分考虑了用户的选择权。

我认为这类产品的不足就是只提供了内容聚合阅读功能，并不能够对内容进行编辑和用户个人整理输出。个人在使用一些内容编辑软件过程中发现，阅读时对文章进行一些标记并且阅读后添加一些关键词或者简短摘要以便进行快速复盘很有必要。

### 内容聚合产品特点和不足



图1-3-2 有道云笔记内容编辑界面

内容管理和编辑类产品具备的功能就是可以编辑一些内容，并且组织内容之间形成一定的结构。但是，我认为的不足就是这些平台并不能够添加一些订阅源。我们也可以在手机或者其他设备上安装云笔记的对应版本，当看到比较好的内容时，可以选择发送到云笔记，但是这样我们仍然需要在不同的内容聚合平台之间进行切换。

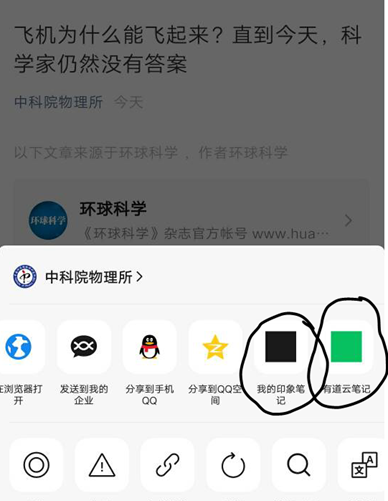


图1-3-5、1-3-6 内容管理和编辑类产品将内容发送到平台的方法

## 项目可能的创新点

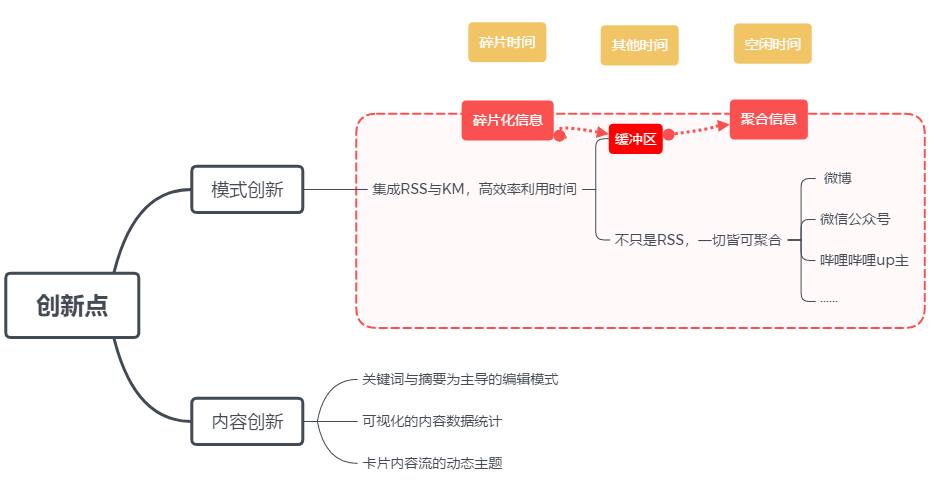


图1-4 创新点

### 模式创新

1. 常见的内容聚合平台和内容管理编辑平台将内容聚合阅读和内容管理编辑两个功能分开，用户不能在阅读后进行即时加工内化，本项目则将两者集成在一个平台。

2. 在聚合RSS内容的基础上，可以扩大聚合内容的范围，比如其他平台的API。

3. 借鉴缓冲区的思想，在碎片化的时间，对内容进行初步筛选，借助已有的云笔记等软件进行缓冲，最后使用该项目进行内容管理。

### 内容创新

1. 用户在阅读后通过编辑关键词和摘要的形式，对内容进行整体的归纳和概括。
2. 对内容管理系统进行搜索时，通过内容统计数据的可视化，对搜索起到一定的提示作用。
3. 在系统的内容流展示的地方，对于每个内容卡片，阅读后的卡片式的动态销毁给人一定的成就感。

# 方案设计

## 总方案设计

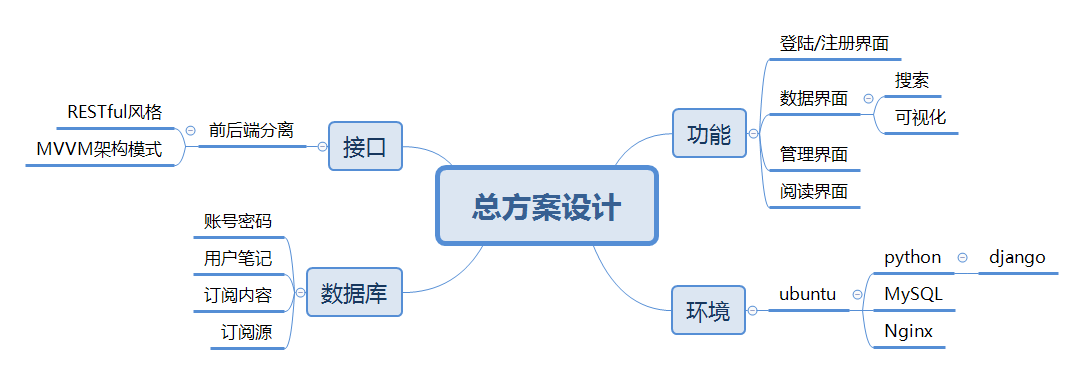


图 2‑1 总方案设计架构图

### 架构模式

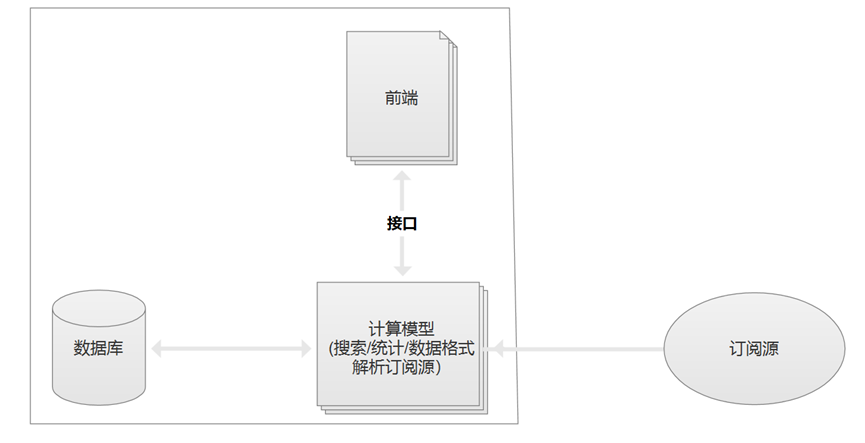


图 2-1-1 MVP架构模式图

项目采用经典的MVVM架构模式。前端通过接口和后端进行交互。典型账号密码验证、注册，搜索，数据库内容统计，笔记编辑和展示、内容阅读和订阅源参加等功能，通过计算模型处理以便更好的交互。

### 功能设计

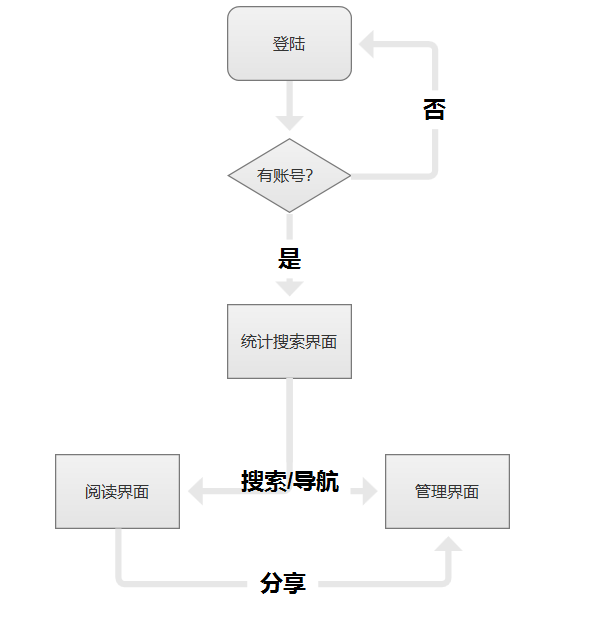


图 2-1-1-1 功能交互图

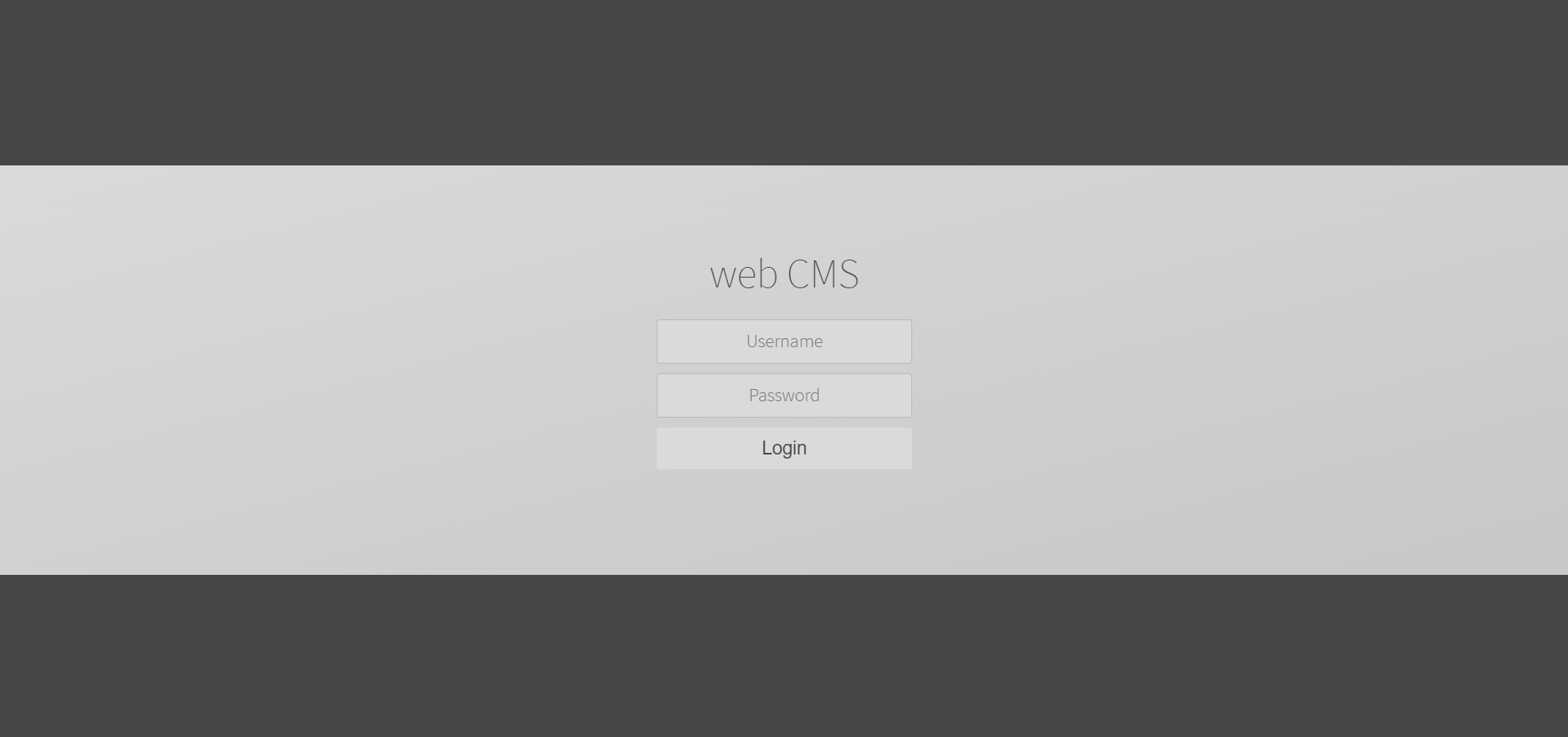


图 2-1-1-2 登陆注册界面前端图

我这里通过对项目的使用流程来对功能设计进行介绍。用户在使用系统时，首先进入项目的登陆/注册界面，这个界面如果用户有注册好的账号和密码，可以直接登陆，否则先进行注册，然后进行登陆。

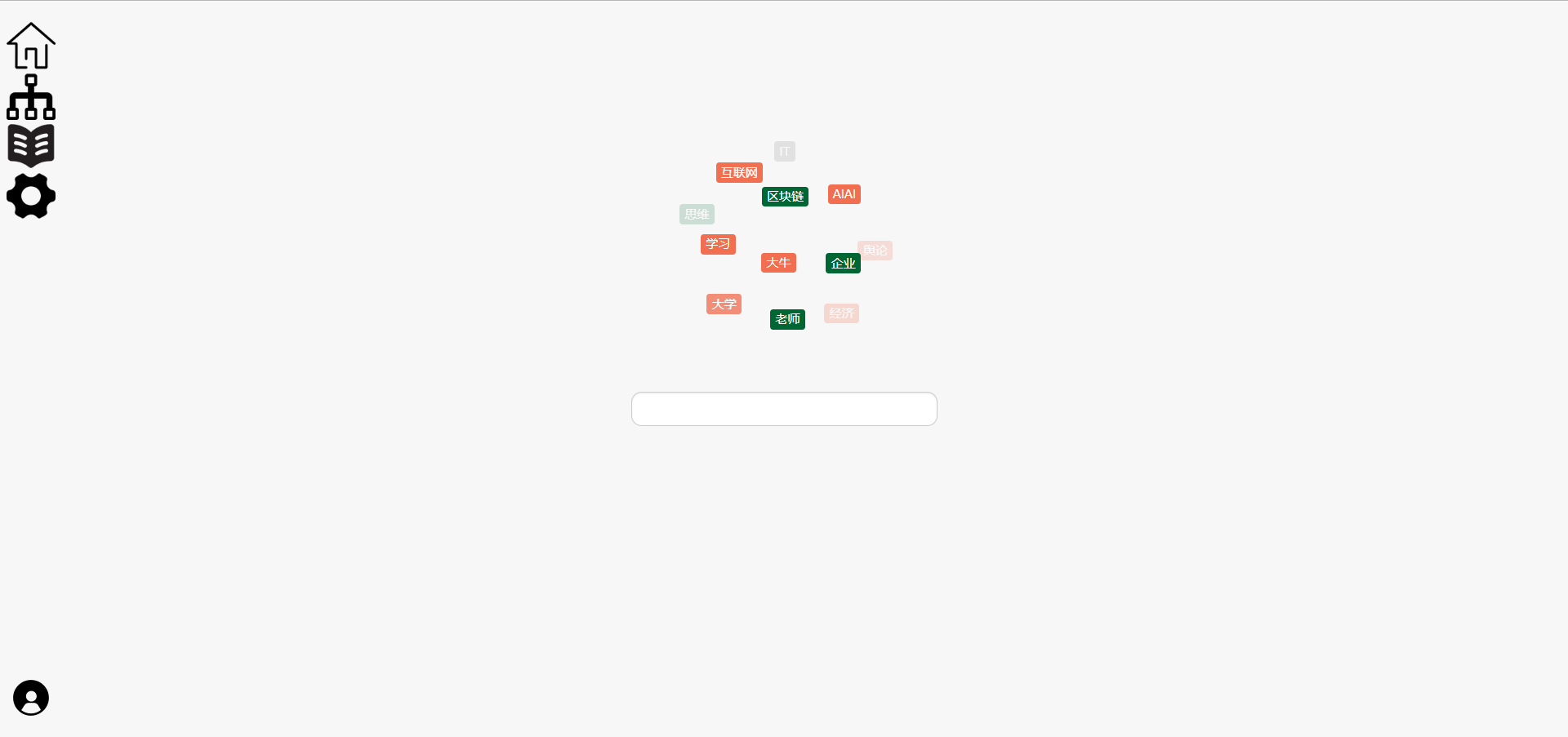


图 2-1-1-3 统计搜索界面前端图

用户登陆成功后，进入到统计搜索界面，相当于项目的主页。在这个界面，用户可以直接根据自己的需求关键字对已有的笔记或者订阅内容进行搜索，也可以根据搜索框上的系统统计信息进行辅助搜索。点击搜索后的结果链接可以跳转到对应的笔记以及内容阅读界面。并且，用户也可以通过侧边导航栏跳转到管理或者阅读界面。同时，侧边导航兰的设计功能也包含。。。

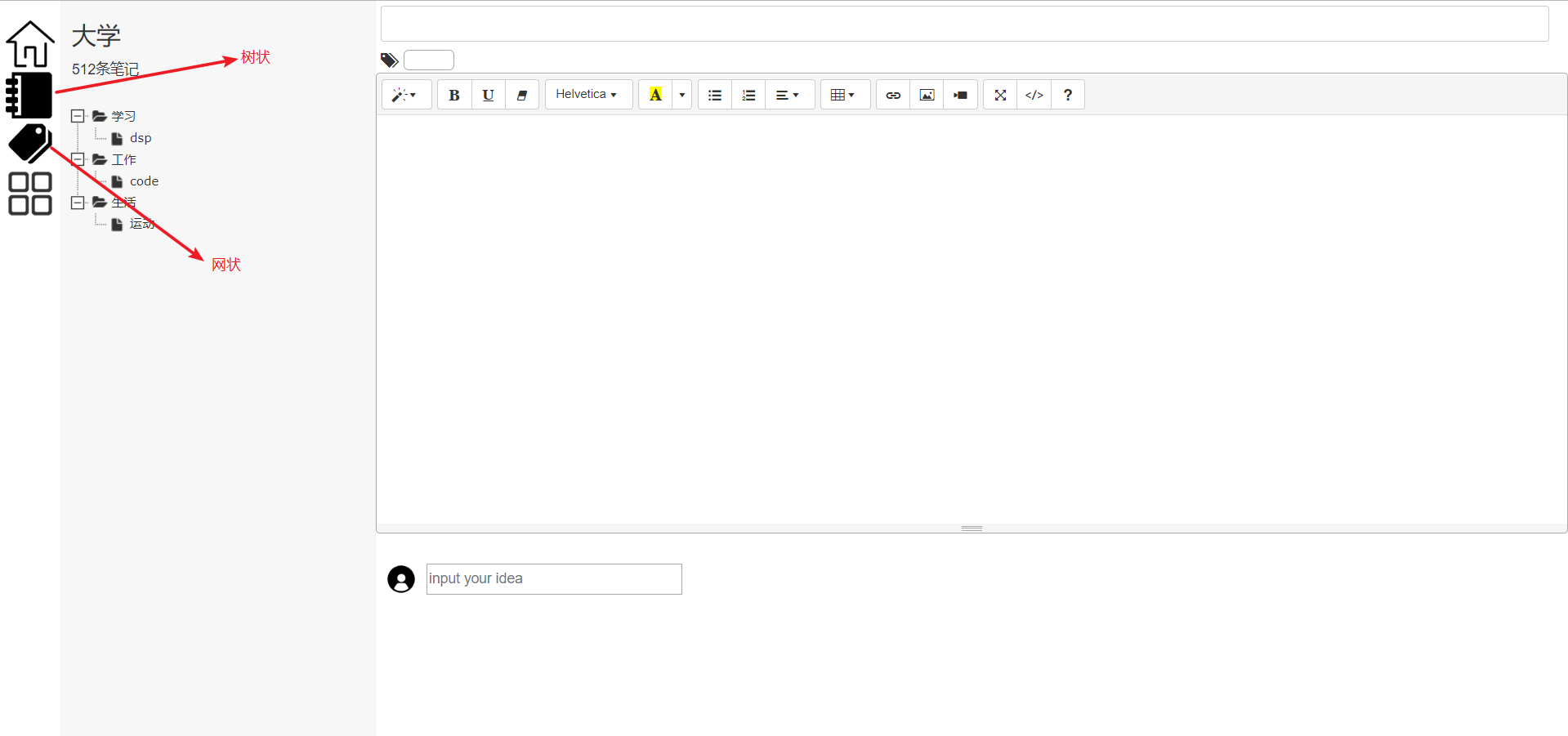


图 2-1-1-4 统计搜索界面前端图

当用户从统计搜索界面跳转到内容管理界面后，用户可以选择以树状或者网状结构形式对笔记内容进行呈现。同时在右方的编辑区域内可以对内容的标题、标签、文章内容、文章摘要进行修改和批注。

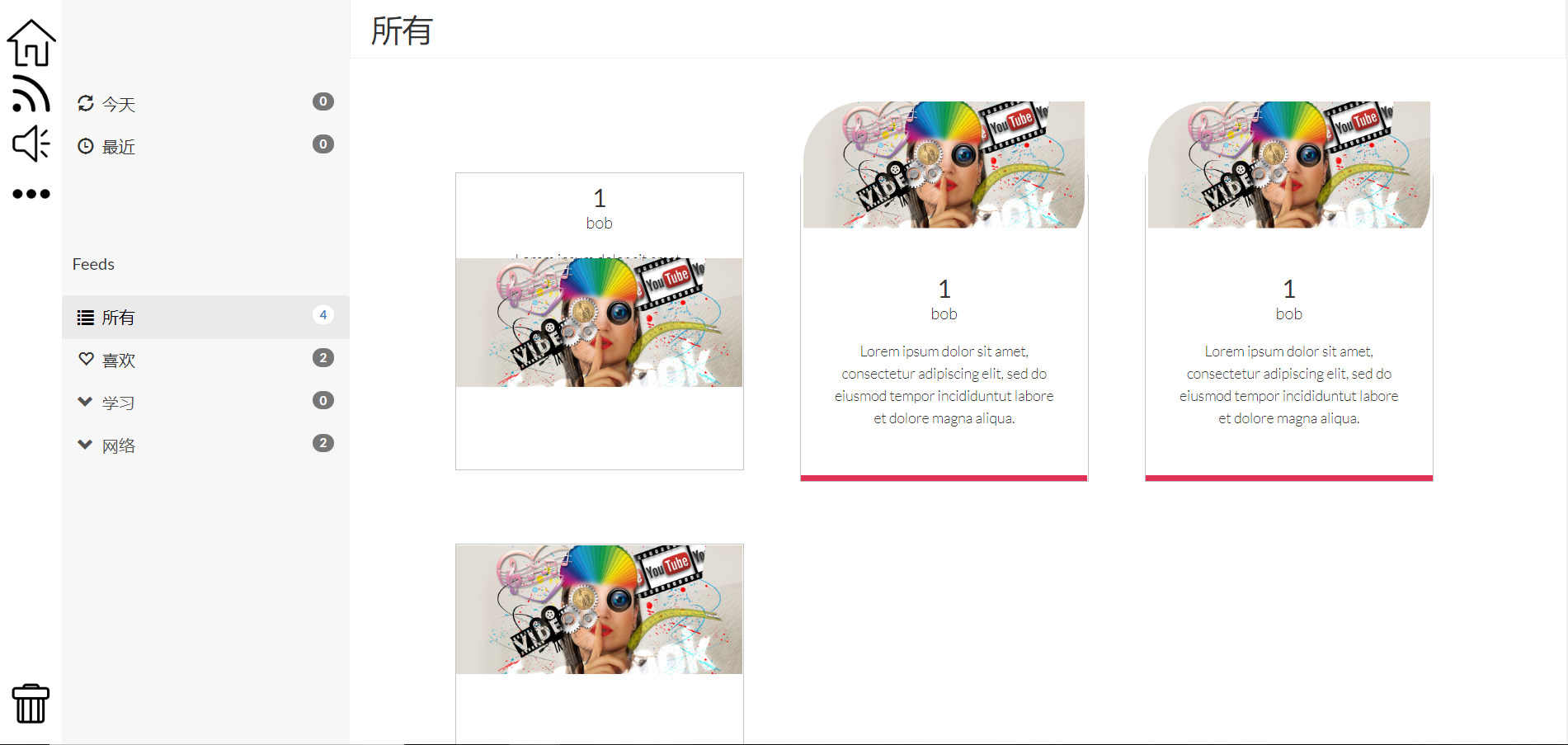


图 2-1-1-4 统计搜索界面前端图

在内容阅读界面，用户可以添加一些订阅源，并设置对订阅源设置简单的类别。用户可以通过文章标题以及文章摘要对内容进行初步筛选，并可通过阅读内容详情做进一步筛选，最后可以将自己喜欢的文章发送到管理界面。

### API设计

从功能出发，尝试对前后端交互接口设计。但由于接口设计需要考虑前端项目的遗留问题，因为下面仅对已确定的接口进行设计，如下：

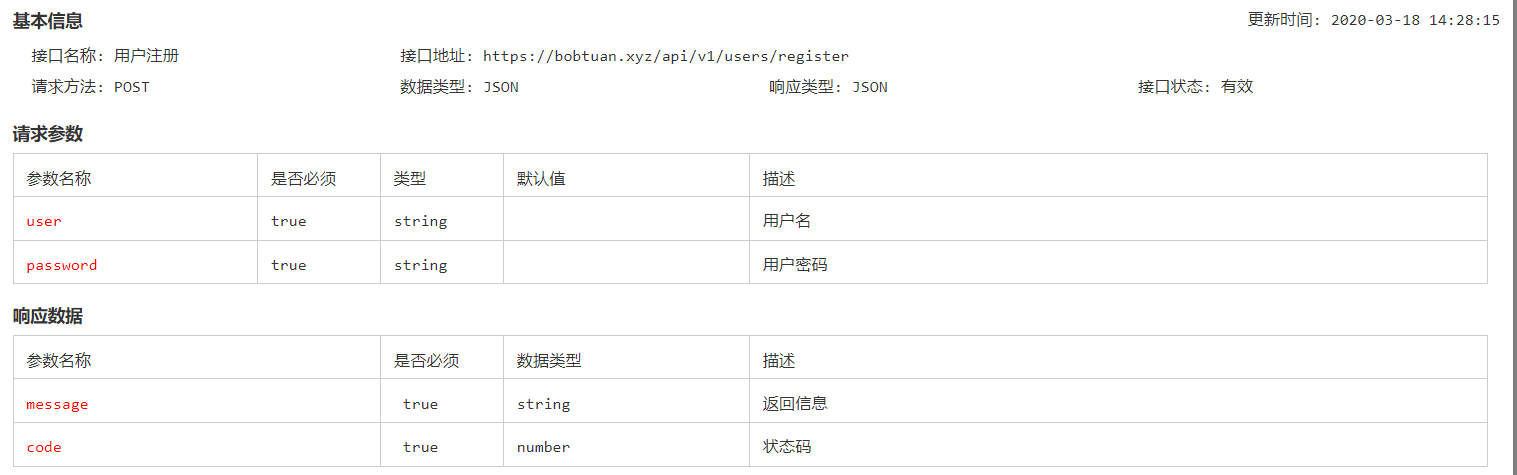


图 2-1-3-1用户注册接口

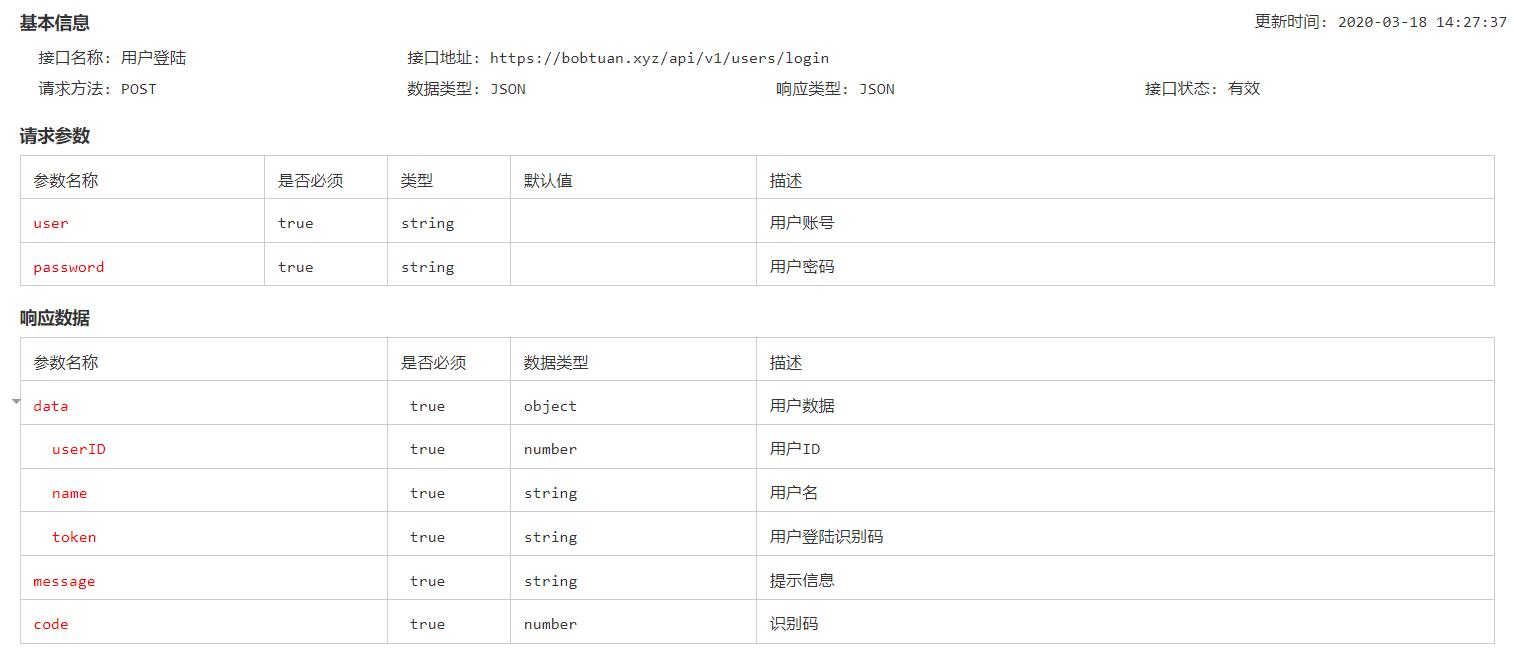


图 2-1-3-2用户登陆接口



图 2-1-3-3获取笔记接口



图 2-1-3-4添加订阅源接口

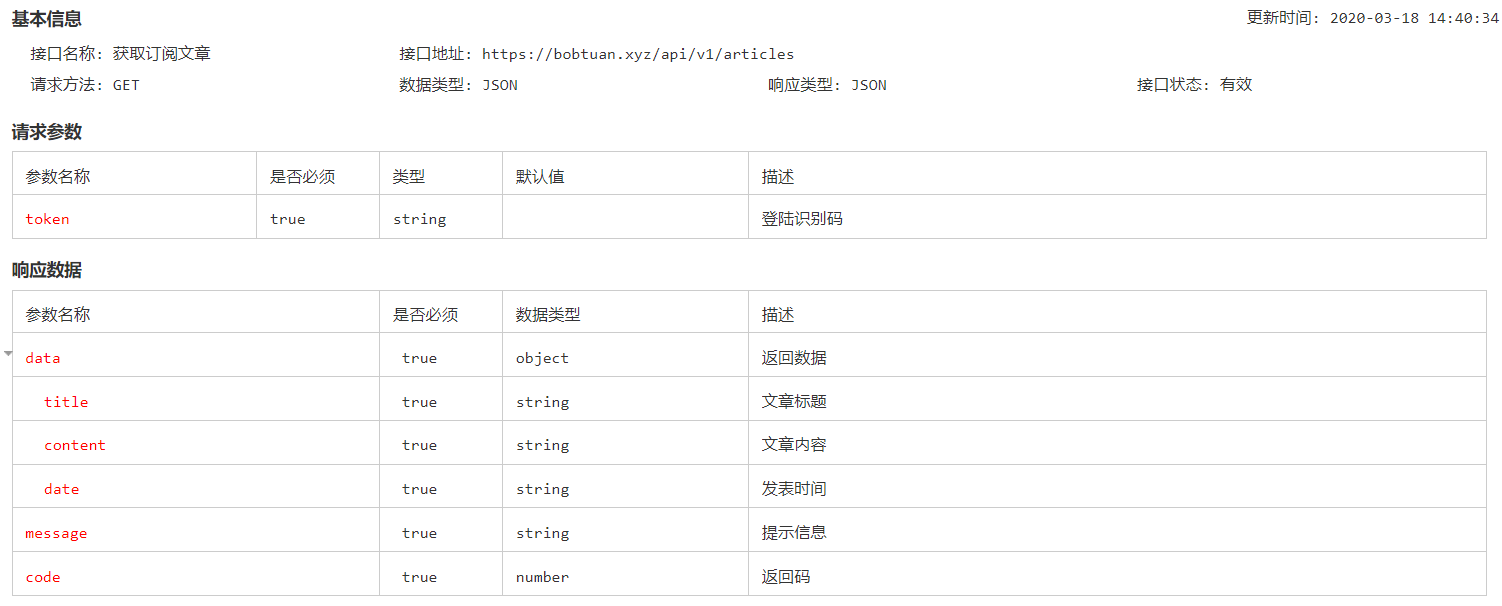


图 2-1-3-5获取订阅文章接口

### 后端设计

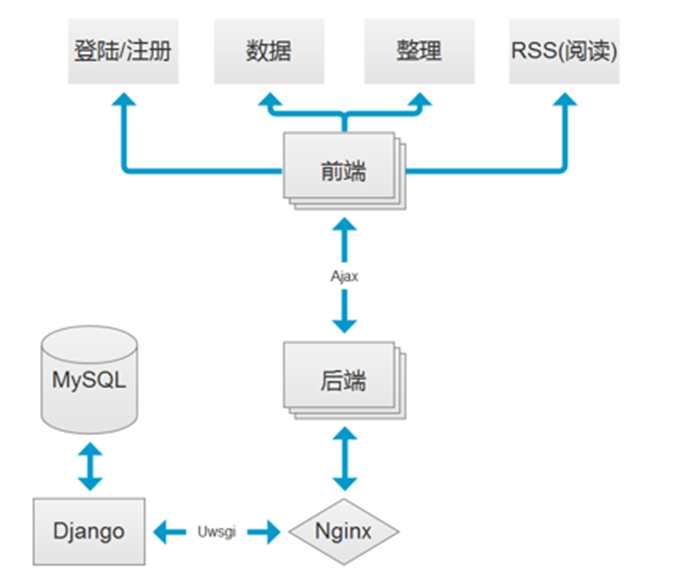


图 2-1-5 前后端架构图

从图中可以看到，后端主要用到了Django框架、Nginx服务器和MySQL数据库。使用Nginx这个轻量级的服务器主要将静态以及动态请求分开处理以提高访问的速度，同时Nginx将真正的后台-Django和前端分割开来，一定程度上有利于后台的安全。Django则是考虑到学习过python同时，Django也是一个比较热门的后端框架。对于MySQL则是由于项目数据对象的结构化的特点决定。Nginx和Django之间的uwsgi中间件则是考虑到Django自带wsgi的性能不可靠。

### 数据库设计

基于项目功能设计以及交互逻辑，分析得到如下的ER图。

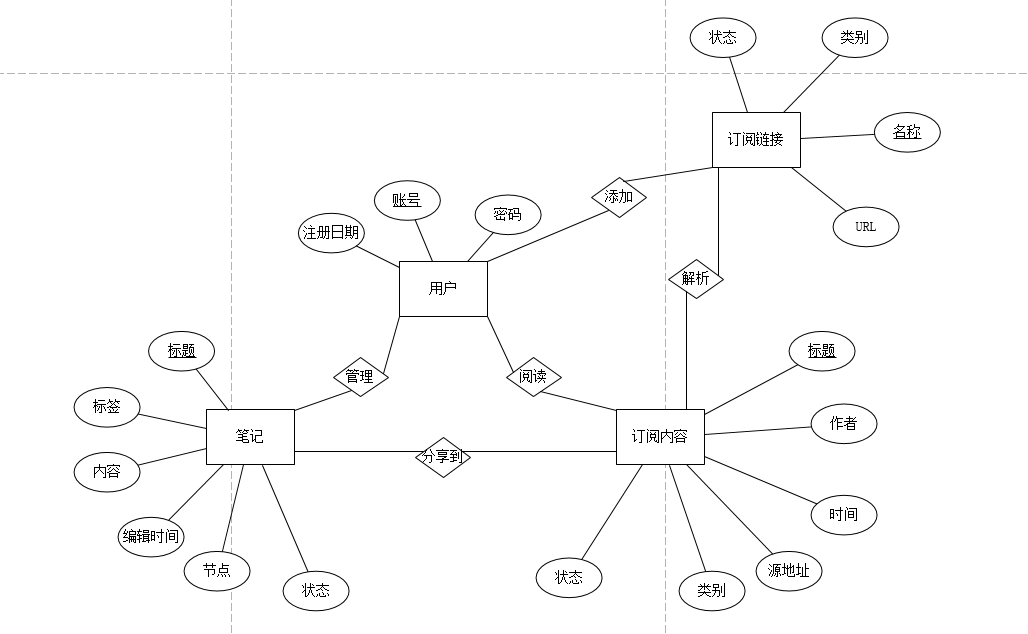


图 2-1-3 -1数据库ER图

从图2-1-3中可以看到，数据对象主要由4个实体5个关系组成。

进一步，将上图的ER图转化为初步的数据库表，如下：

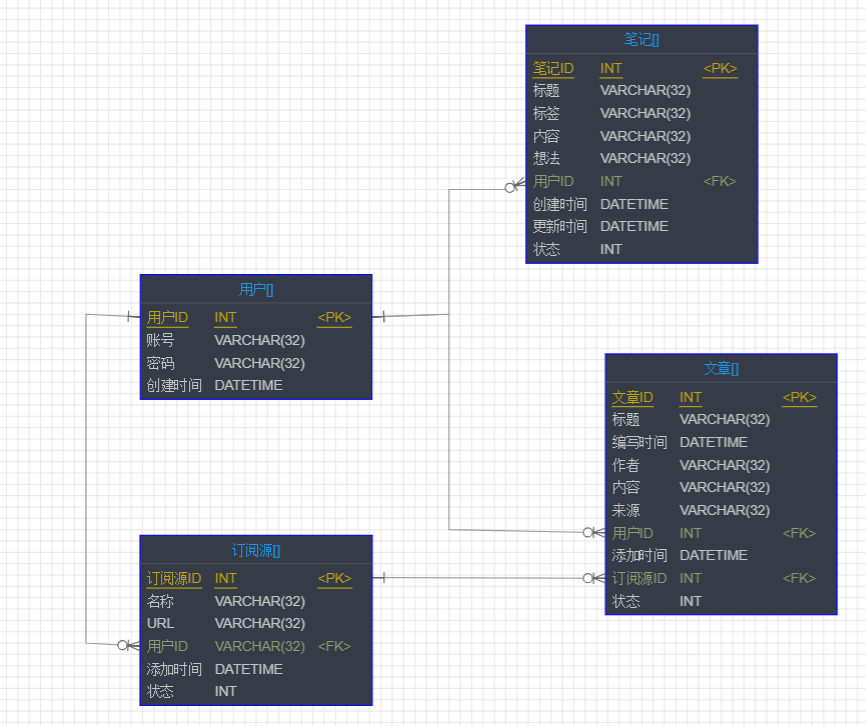


图 2-1-3-2 满足第二范式关系表

根据ER图中的实体和属性构建的关系表已经满足了第一范式的要求，同时关系表之间的关系通过主外键进行联系。同时，为了简化数据表的记录的标识，在每个数据表中增加ID字段作为该表的主键。

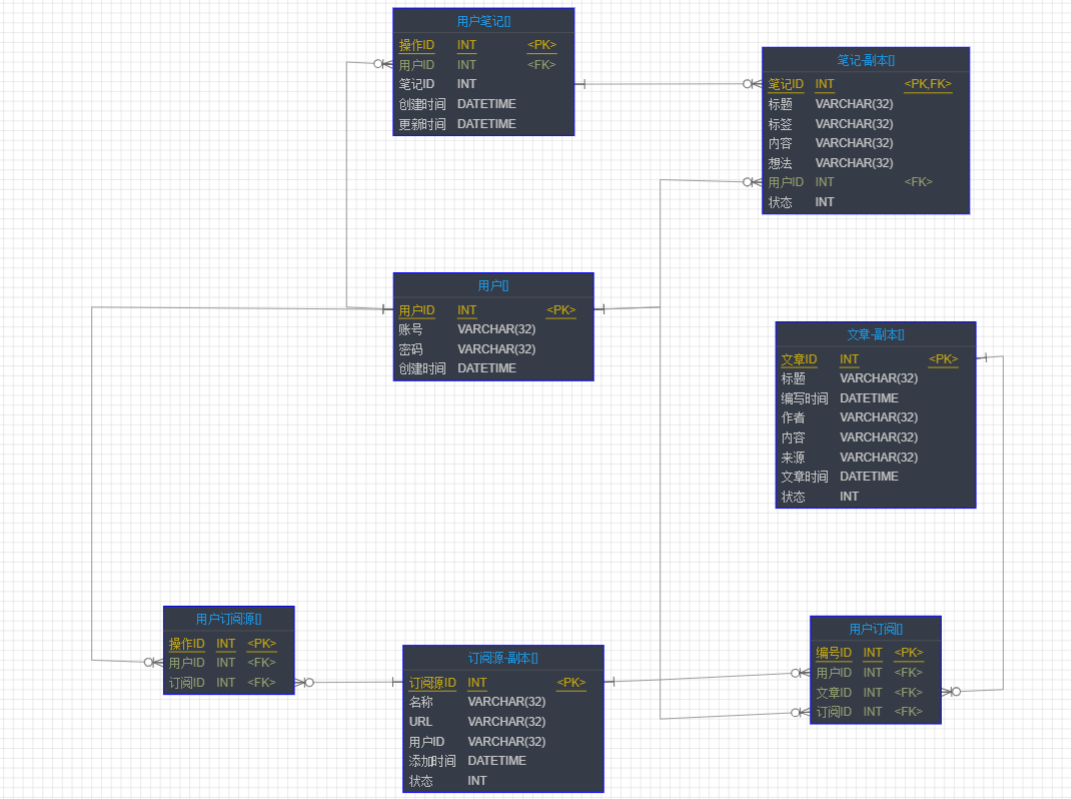


图 2-1-3-3 第三范式数据库关系模型

在满足第一、二范式关系表的基础上，通过去除关系表中的非主关键字以外的函数依赖，得到上图中的满足第三范式的关系表。

其中，对于满足第二范式关系表中的文章表，其中的文章的相关属性函数依赖订阅源ID和文章时间，所以选择将其抽离出来。同时，笔记表中的标题、标签等字段函数依赖用户ID、创建时间这些字段，所以也将其进行抽离。订阅源表中的URL字段也是函数依赖用户ID和添加时间属性，故将其抽离出来。

# 进度安排

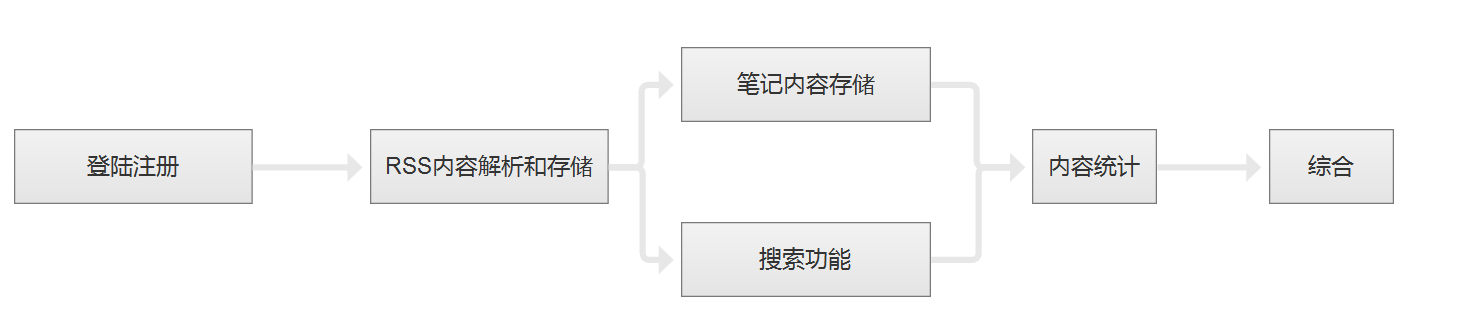


图 3 进度安排图

在执行项目时，我们将按照图3中的任务分解顺序执行。从现在到答辩还有12周，大概1周登陆注册，可以进行账号登陆和注册。2周RSS内容解析和存储，最重要找到RSS的检测更新算法。5周笔记内容存储和搜索功能并行处理，将前端遗留的笔记问题处理好，并且可以达到较好的搜索效果。1周实现内容统计，希望对数据库内容进行统计展示。最后3周进行综合优化。

# 项目团队

## 团队成员及分工

### 组长-段斌

负责部分后端和前端遗留问题。喜欢探索技术。由于从上学期一直做该项目，所以对一些问题的解决方案有所了解。尽量规范好接口，以便组员更好的面向接口编程。

### 组员-游卓儒

主要负责后端与数据库，参与一些前端的改进。比较擅长进行测试与发现问题。对于RSS与内容管理不是特别了解，但是对于python语言比较熟悉。使用Pycharm，主要使用Win10系统，也会使用Ubuntu18.04系统。

# 项目难点

## 项目难点

1. 搜索的实现方式，是使用数据库的模糊搜索还是通过0开始建立索引表然后进行搜索。
2. 搜索内容如何呈现，我们不想做循规蹈矩的类似于百度搜索的呈现方式。
3. RSS内容的检测更新。对于RSS链接包含了最近几天的内容。不同时间的获取内容可能存在交叉地方，如何识别？
4. 设置后台抓取RSS链接内容的时间间隔为多少？
5. 笔记的存储方法，备份，回收站的存储？

6. 笔记的导出

## 可能遇到的问题

1. 数据库的设计问题
2. 服务器容量、负荷问题
3. 如何对微信公众号、微博等信息进行聚合的问题

# 参考文献

1. 贺逸群. 当代大学生碎片化阅读研究[D].河南大学,2019.
2. 周天慧.碎片化阅读的弊端[J].传播力研究,2019,3(24):234.