

# 中国软件杯 大学生软件设计大赛



视障人士友好的资讯辅助软件

详细设计文档

## 悦读

更好的资讯辅助

编号	48011553
队名	好果汁你让我疯狂

# 目录

1 引言 .....	3
1.1 编写目的 .....	3
1.2 背景 .....	3
1.3 定义 .....	3
1.4 参考资料 .....	4
2 程序系统的结构 .....	5
3 Web 服务端设计说明 .....	5
3.1 程序描述 .....	5
3.2 功能 .....	6
3.3 接口 .....	6
3.4 软件流程与界面设计 .....	8
4 小程序客户端设计说明 .....	18
4.1 程序描述 .....	18
4.2 程序功能 .....	18
4.3 软件流程及界面设计 .....	19
4.5 接口 .....	26
5 推荐算法模型设计说明 .....	26
5.1 程序描述 .....	26
5.2 功能 .....	27
5.3 性能 .....	27
5.4 输入输出项 .....	27
5.5 算法构成 .....	28
5.6 流程逻辑 .....	28
5.7 接口 .....	33
5.8 限制条件 .....	33
6 后端设计说明 .....	34
6.1 程序描述 .....	34
6.2 功能 .....	34
6.3 流程逻辑 .....	35
6.4 存储分配 .....	43

# 1 引言

## 1.1 编写目的

本文档主要描述的是本团队对于 2022 年第十一届中国软件杯大赛 A3 赛题制作的软件的详细设计，对于系统内所有部分及所包含的模块的具体实现细节进行了详实地说明，同时也包括了对内外部接口、运行模块设计、数据结构设计以及出错处理设计的讲解。在基于需求规格说明书以及概要设计文档的基础上，明确需求规格并细化内容完善深度，从而编写了此文档，作为详细了解本软件开发内容的参考。

本文主要阅读对象为相关技术人员和评委老师。

## 1.2 背景

随着移动互联网以及信息技术的发展，大多数人可以通过手机进行出行、学习、娱乐、交流等生活活动。APP/小程序应用功能从根本上改变了生活娱乐的方式，带来了信息技术下带来的生活水平的提高。然而在考虑到大部分人群的同时，很多软件却忽视了视障人士、老年人等弱势群体，他们因为年龄或者疾病的原因，并不能很好的使用大多数的 APP/小程序。技术应当能够覆盖更多需要帮助的人群，让他们享受到科技进步的成果。

本团队开发的软件系统名为“悦读”，是一类视障人士友好的资讯辅助软件。该赛题的出题企业是中移（苏州）软件技术有限公司，开发者是本团队的全部成员，目标用户是存在视力障碍且无法通过佩戴眼镜等方式进行矫正的视障人群，我们的软件主要运行在 Web 端和微信小程序。

## 1.3 定义

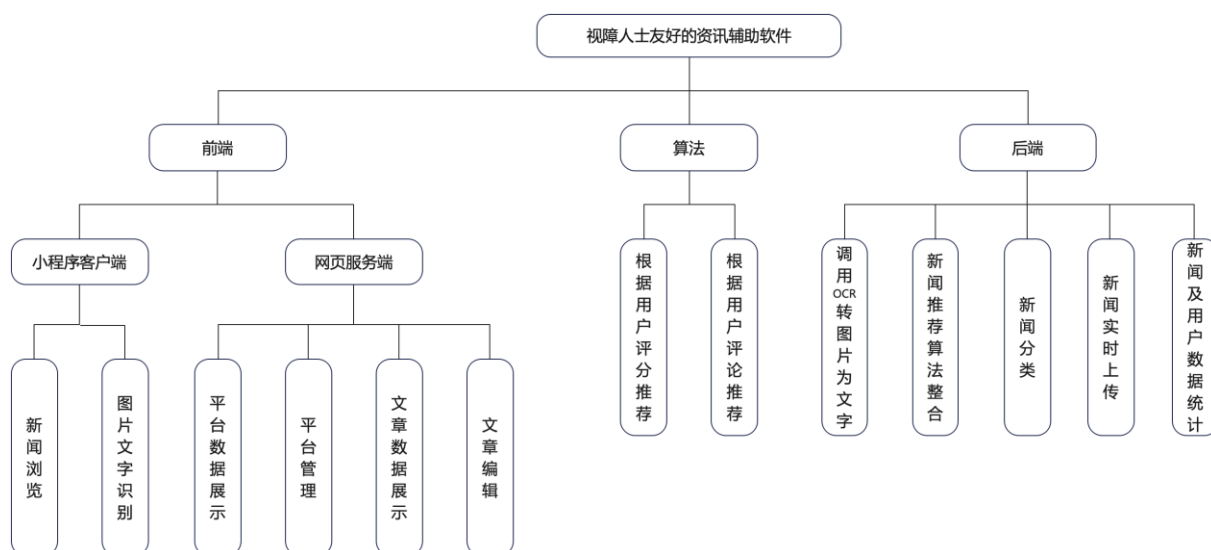
缩写、术语	解释
管理平台	指 Web 服务端设计的用于管理员和新闻上传者使用的后台管理 Web 平台
文章、新闻	本资讯平台内的内容皆为新闻形式的文章，故在下文两者含义相同，都是指供用户获取的资讯
token	用以验证用户的登录状态以及用户信息，存放在 http 的 header 中，除部分接口外用户发送请求时必须携带
editor	代表新闻上传者、编辑，属于服务端用户，为阅读者提供内容

reader	代表阅读者，即客户端用户
admin	代表平台管理员，属于服务端用户，可管理另外两类用户
词嵌入	词嵌入是自然语言处理（NLP）中语言模型与表征学习技术的统称。概念上而言，它是指把一个维数为所有词的数量的高维空间嵌入到一个维数低得多的连续向量空间中，每个单词或词组被映射为实数域上的向量。
文本分词	将文本拆分成独立的字词以便于选择复制，支持将文本中网址、邮箱、手机号码、单词等分割。
余弦相似度	余弦相似度，又称为余弦相似性，是通过计算两个向量的夹角余弦值来评估他们的相似度。余弦相似度将向量根据坐标值，绘制到向量空间中，如最常见的二维空间。

## 1.4 参考资料

- [1] 工业和信息化部办公厅. (2021, Apr 6). *工业和信息化部办公厅关于进一步抓好互联网应用适老化及无障碍改造专项行动实施工作的通知*, from 工业和信息化部网站: [http://www.gov.cn/zhengce/zhengceku/2021-04/13/content\\_5599225.htm](http://www.gov.cn/zhengce/zhengceku/2021-04/13/content_5599225.htm)
- [2] PanJiaChen. *Vue Element Admin 官方文档*, from: <https://panjiachen.github.io/vue-element-admin-site/zh/>
- [3] ElemeFE. *Element UI 官方文档*, from: <https://element.eleme.io/#/zh-CN>
- [4] 微信小程序开放文档, from: <https://developers.weixin.qq.com/miniprogram/dev/framework/>
- [5] Spring 官方文档: <https://spring.io/>

## 2 程序系统的结构



本软件可划分为前端、后端、算法三个部分，其中前端根据面向的使用者的不同，分为了小程序客户端和网页服务端。每个部分都包括了不同的功能模块，小程序客户端借助推荐算法，从后端获取到推荐给用户的新闻，网页服务端则通过交互界面控制改变后端数据库中的数据，从而对软件内容生态进行把控与创建。

## 3 Web 服务端设计说明

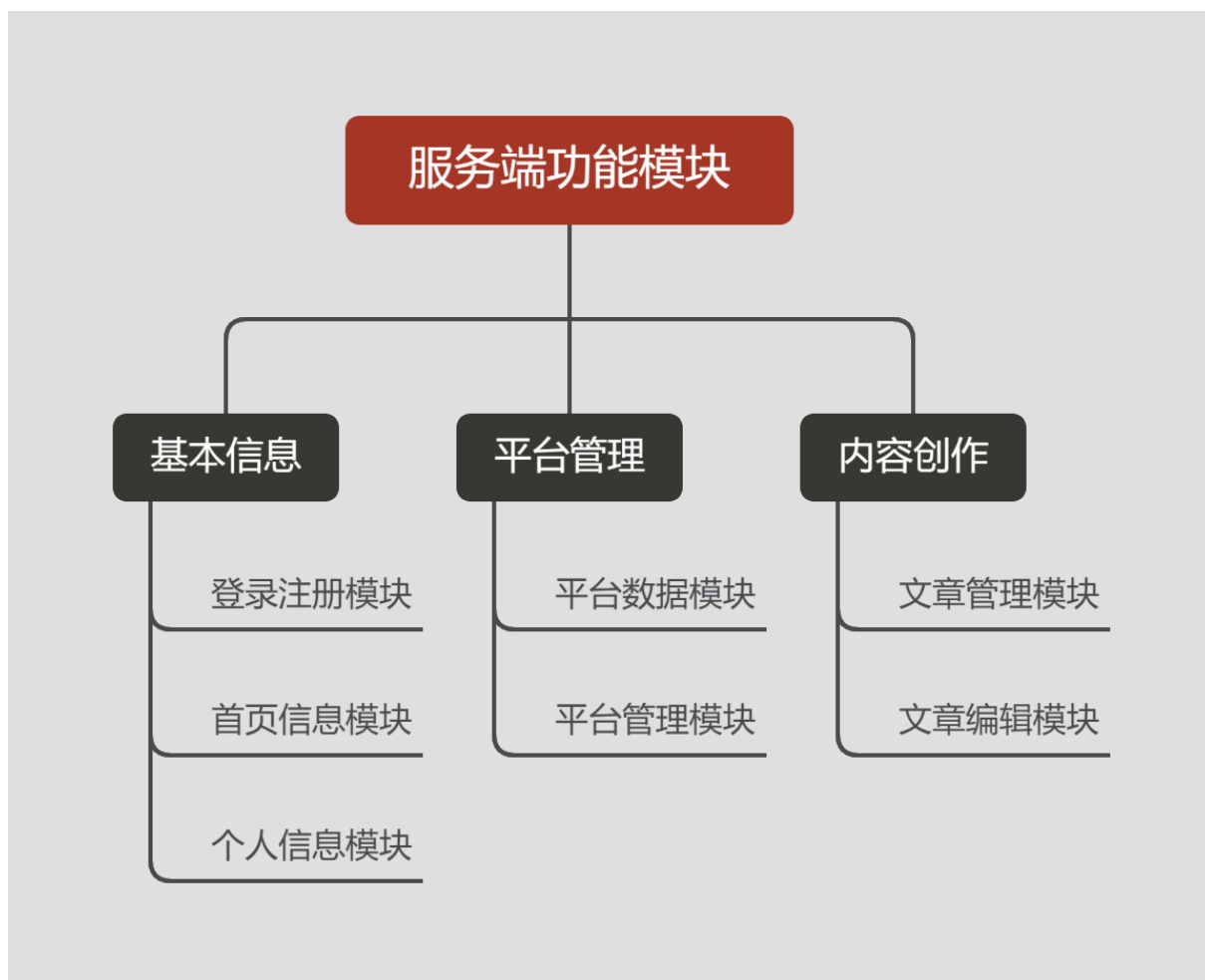
### 3.1 程序描述

该程序面向 2 类用户：平台管理员和新闻上传者。

平台管理员即 Admin，能够方便地查看平台上用户和文章的变化情况，同时借助用户和文章列表进行方便的管理，对于平台内生态进行管控。

内容编辑员即 Editor，能够便捷地进行文章管理与编辑，方便的查看文章的阅读量，赞数，评分等情况。

## 3.2 功能



服务端部分最终展出来的是一个 Web 端的管理平台，而根据功能可分为三个部分，其中又各自划分总共有七个功能模块。基本信息部分是管理员和新闻上传者所共有的，平台管理模块面向管理员，内容创作模块面向新闻上传者。

## 3.3 接口

服务端涉及的内部接口如下表：

系统元素	接口	HTTP 请求方法	接口名称
访问状态	/server/login	POST	登录
	/index/register	POST	注册
	/server/logout	POST	登出
	/token/refreshToken	GET	刷新 Token
用户信息	/index/admin	POST	Admin 首页信息

	/index/editor	POST	Editor 首页信息
	/index/avatar	POST	头像上传
	/index/profile	POST	个人信息修改
平台数据	/admin/userData	GET	平台用户数据
	/admin/articleData	GET	平台文章数据
用户管理（管理员）	/admin/userList/{page}	GET	平台用户列表
	/admin/userStatusChange	POST	改变用户状态
	/admin/userDelete	DEL	删除用户
新闻管理（管理员）	/admin/articleList/{page}	GET	平台新闻列表
	/admin/articleContent	GET	新闻详细内容
	/admin/articleStatusChange	PUT	改变新闻状态
	/admin/articleTypeChange	PUT	修改新闻类型
	/admin/articleDelete	DEL	删除新闻
新闻管理（上传者）	/editor/articleData/{page}	GET	新闻数据列表
	/editor/articleDelete	DEL	删除新闻
	/editor/articleContent	GET	获取新闻详情
新闻编辑（上传者）	/editor/uploadArticle	POST	上传新闻内容
	/editor/uploadArticleFile	POST	上传新闻文件

对每一个请求都会验证权限，所以这里针对业务封装了一下请求。首先通过 request 拦截器在每个请求头里面塞入 token，好让后端对请求进行权限验证。并创建一个 response 拦截器，当服务端返回特殊的状态码，统一做处理，如没权限或者 token 失效等操作。

服务端中，一个完整的前端 UI 交互到服务端处理流程是这样的：

#### 1. UI 组件交互操作；

2. 调用统一管理的 api service 请求函数;
3. 使用封装的 request.js 发送请求;
4. 获取服务端返回;
5. 更新 data;

服务端所有请求都是基于 axios 的封装, 便于统一处理 POST, GET 等请求参数, 请求头, 以及错误提示信息等, 封装了全局 request 拦截器、response 拦截器、统一的错误处理、统一做了超时处理、baseUrl 设置等。

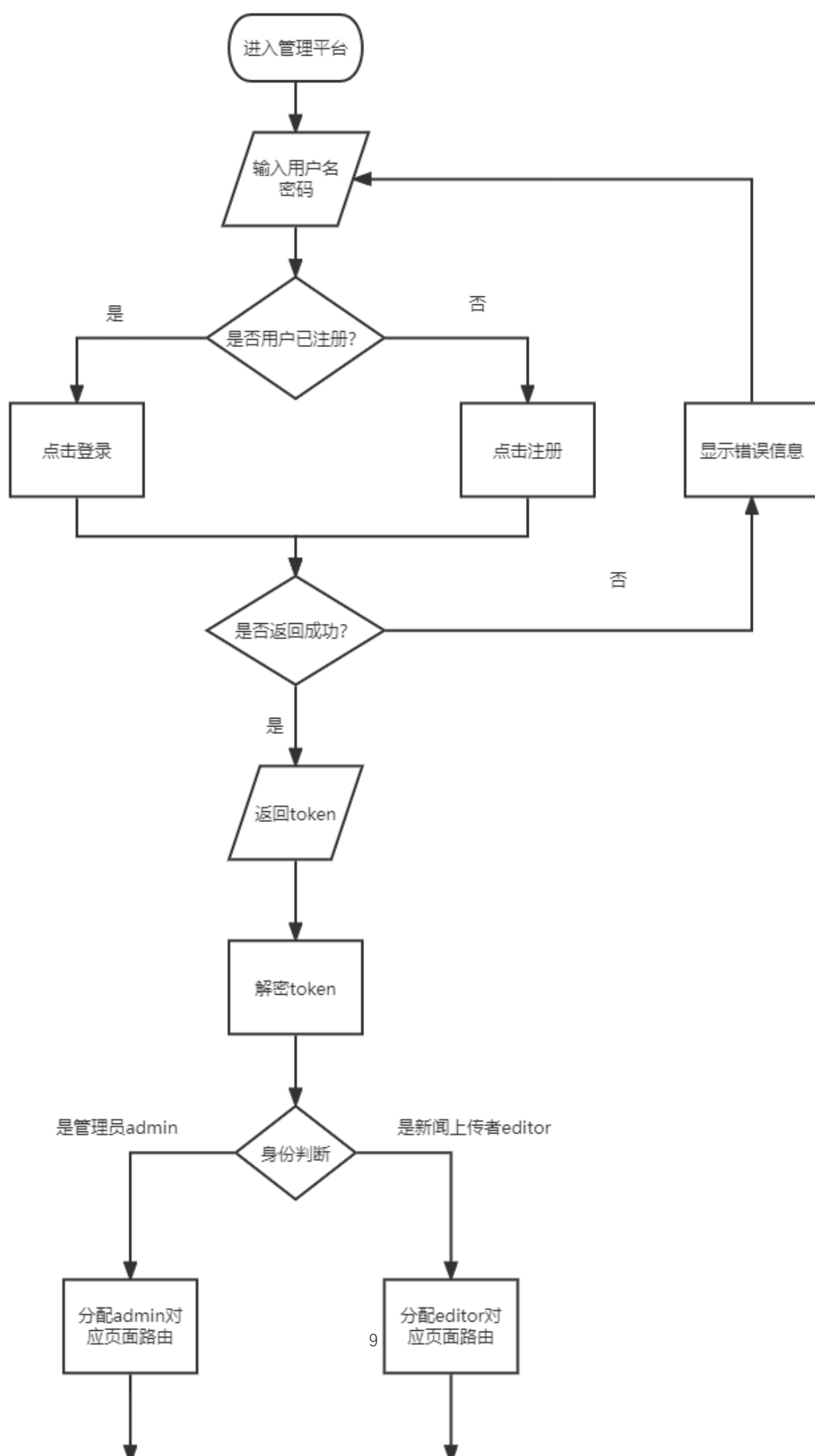
## 3.4 软件流程与界面设计

### 3.4.1 登录注册模块

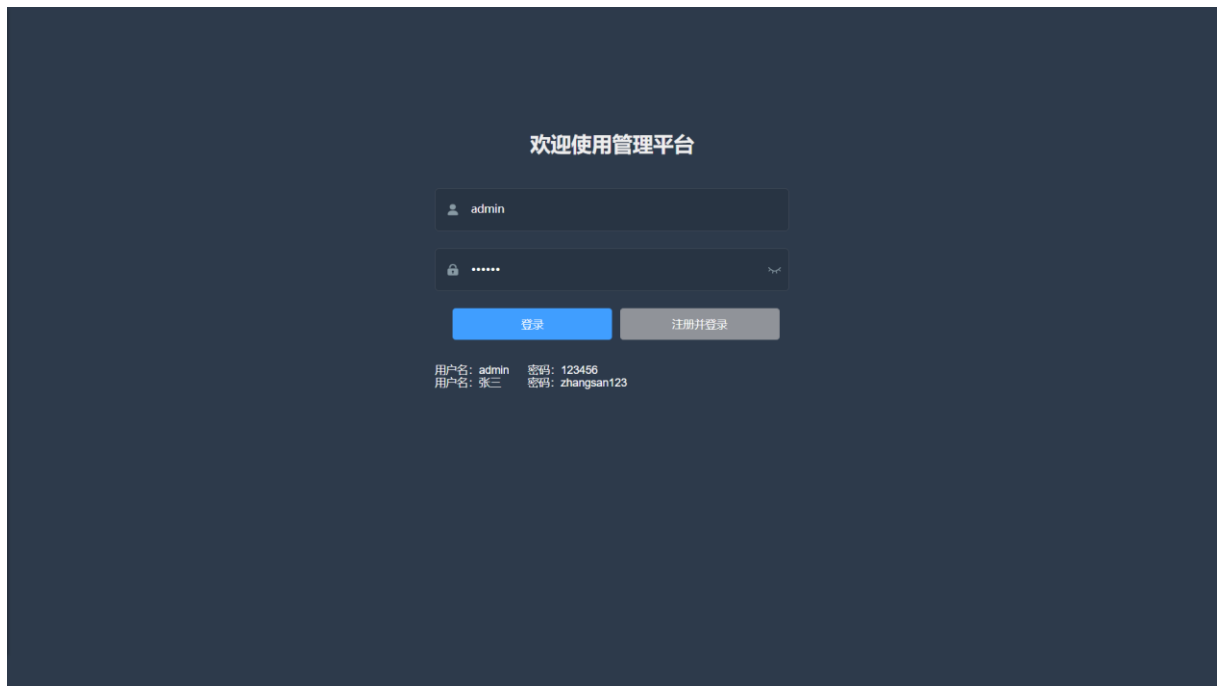
进入平台后首先是登录注册界面, 这里的流程是:

当用户填写完账号和密码后, 根据自己是否拥有账号, 选择登录或注册, 随之向服务端验证是否正确, 此时根据后端接口返回的信息, 若成功则根据返回的 token, 进行 jwt 的解密, 获取到身份组信息和对应有权限的路由, 动态挂载这些路由, 继而请求对应身份组数据信息的接口, 获取到接口的返回信息后, 跳转进入, 否则会进一步返回根据返回错误信息对使用者进行弹窗提示, 等待用户重新登录。

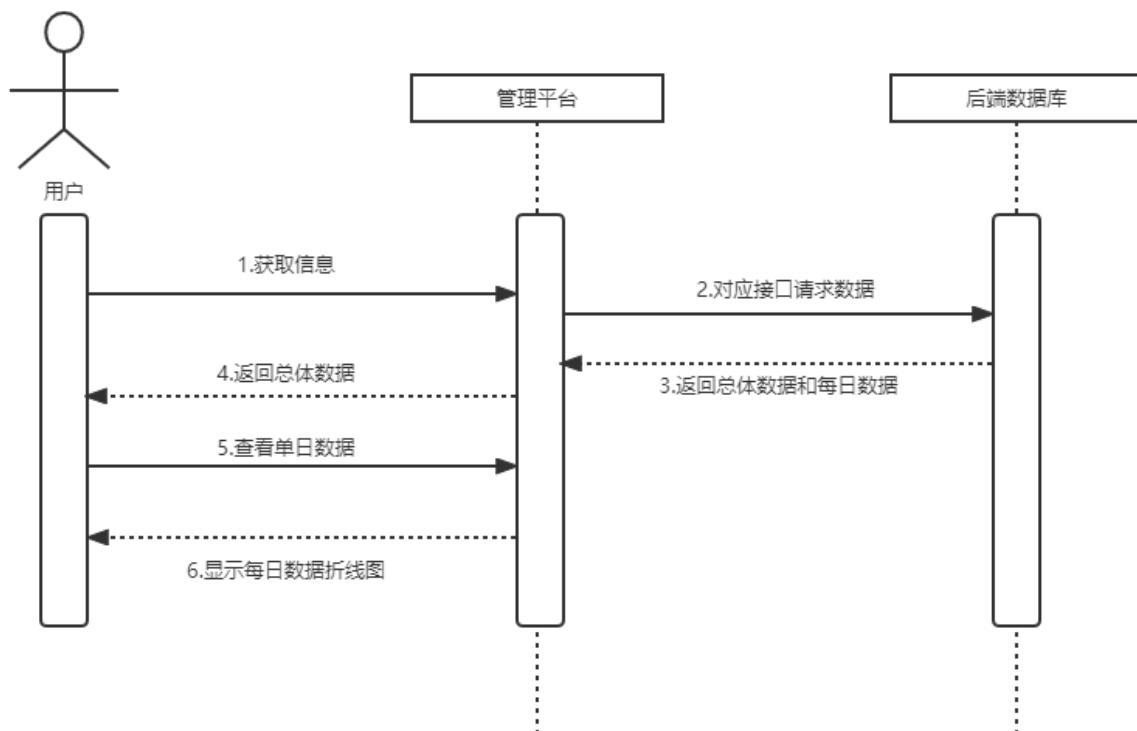




此处的界面设计简约明了，用户可以很快的关注到中心的信息表单，便于评委老师使用，在底下标注了不同身份组的测试账号。



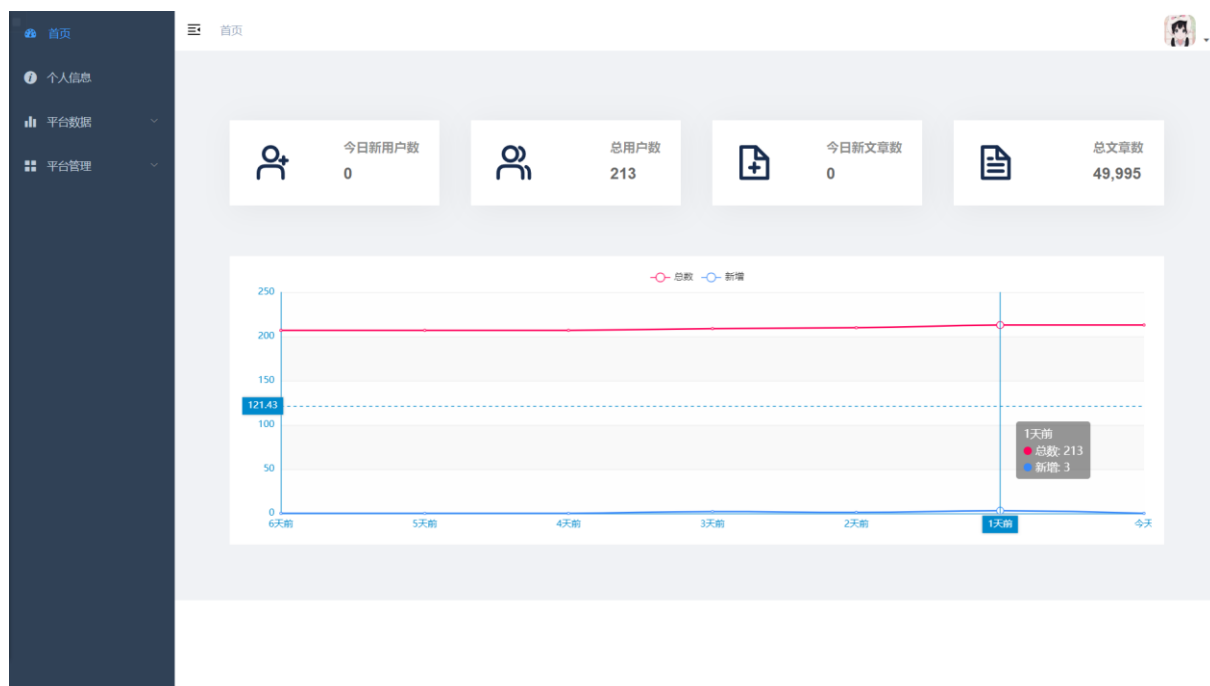
### 3.4.2 首页信息模块



成功进入管理平台后，会自动请求获取首页显示的信息，不同身份组的用户显示的信息不

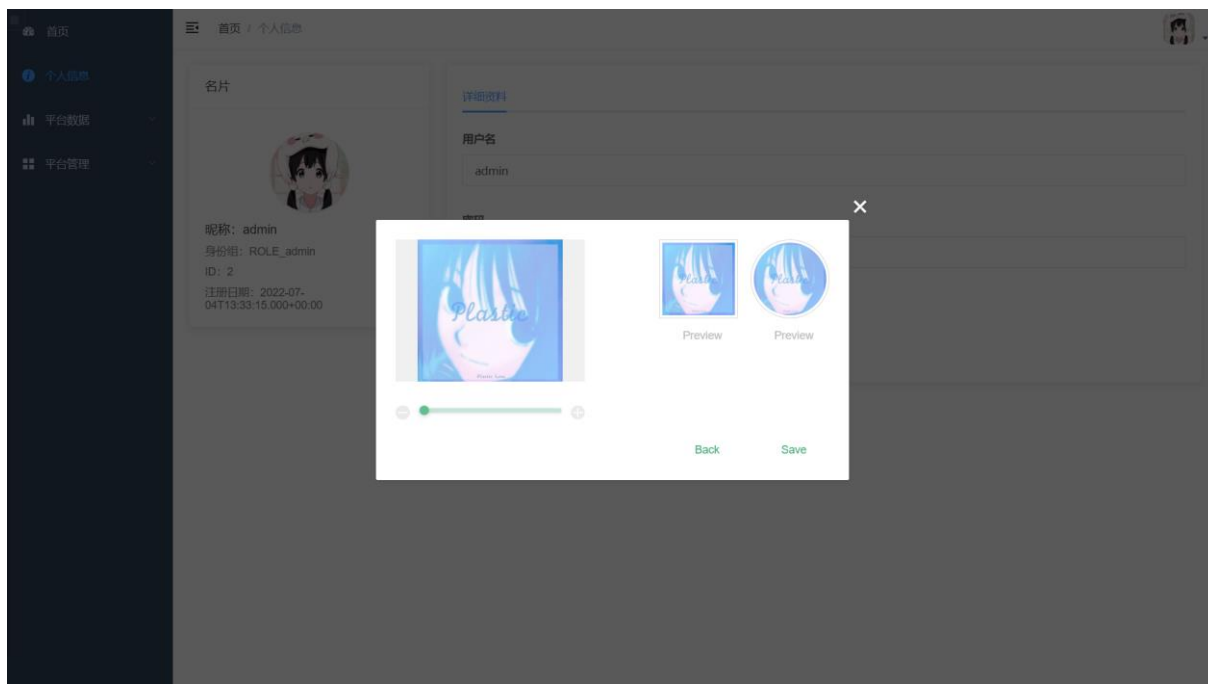
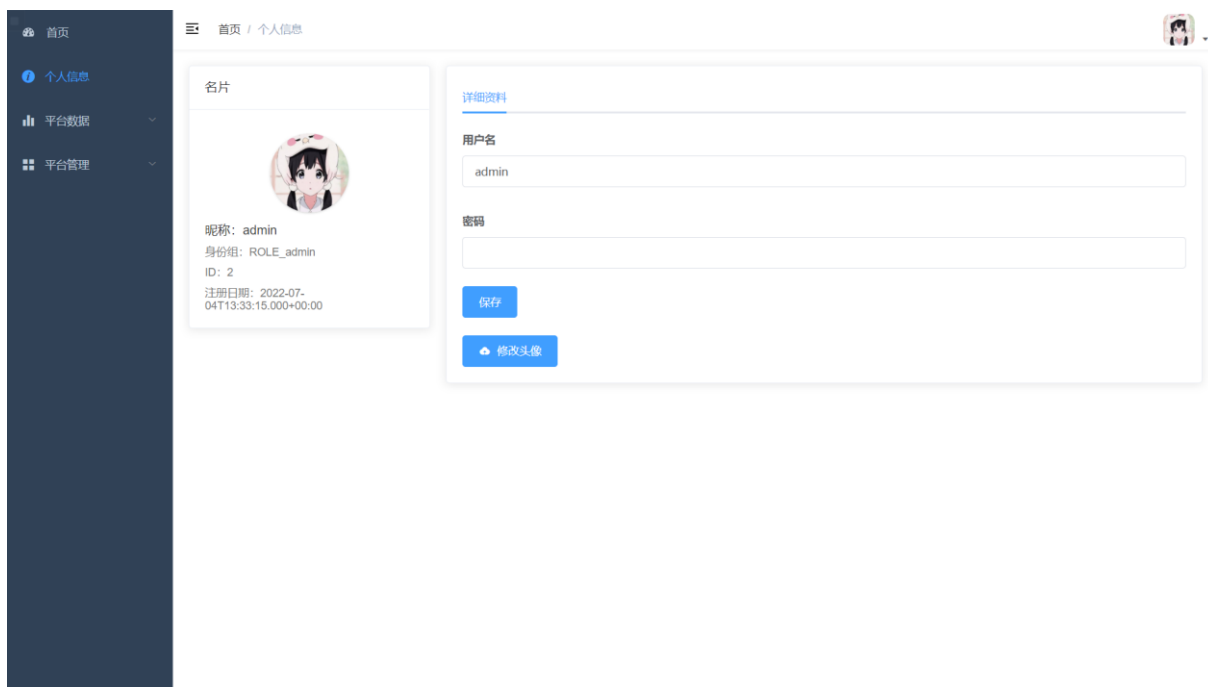
同。总体的数据会以单个卡片的形式进行显示，点击卡片，会以折线图的形式展示当前数据类的每日数据，方便查看变化。

此处的界面主要分为三个部分，侧边栏，顶栏和主界面。侧边栏是使用 vue router 进行路由分配之后动态生成的，顶栏会根据所在界面显示面包屑导航，右上角的用户头像提示登录的账户，主界面是根据不同身份组，会有不同的内容显示，例如此时管理员的首页，会显示今日和总共的用户数、新闻文章数，点击对应的卡片会显示对应的逐日数据的折线图。



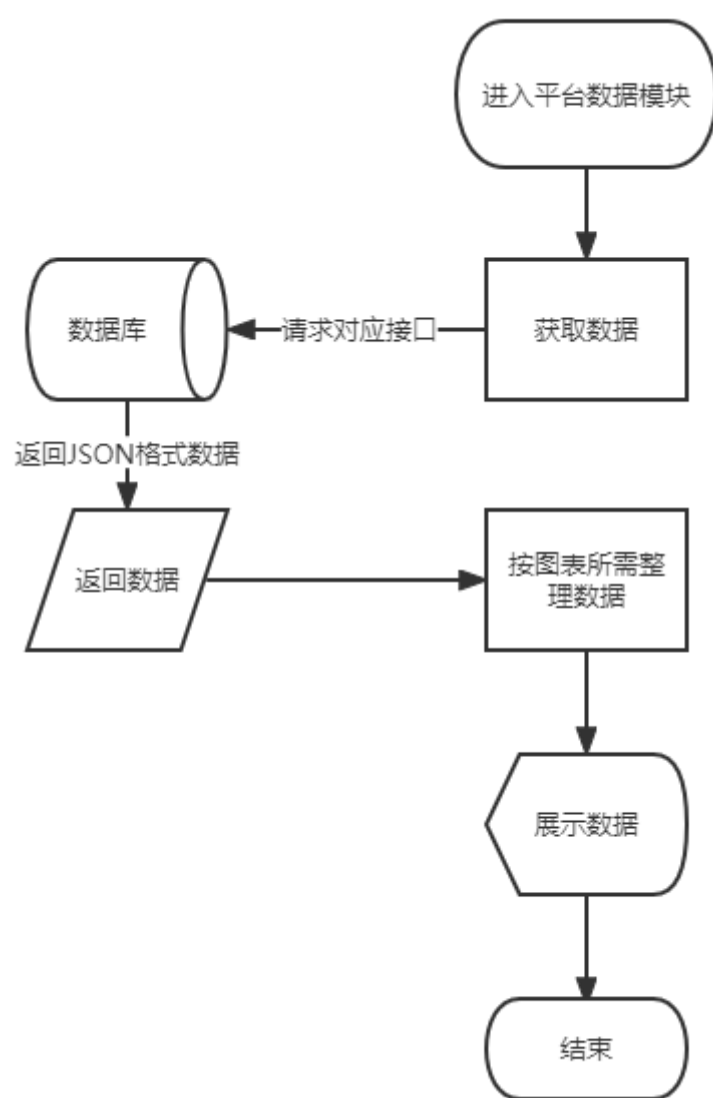
### 3.4.3 个人信息模块

此处显示和修改个人信息，也可以上传头像进行修改，设计了头像的预览界面。

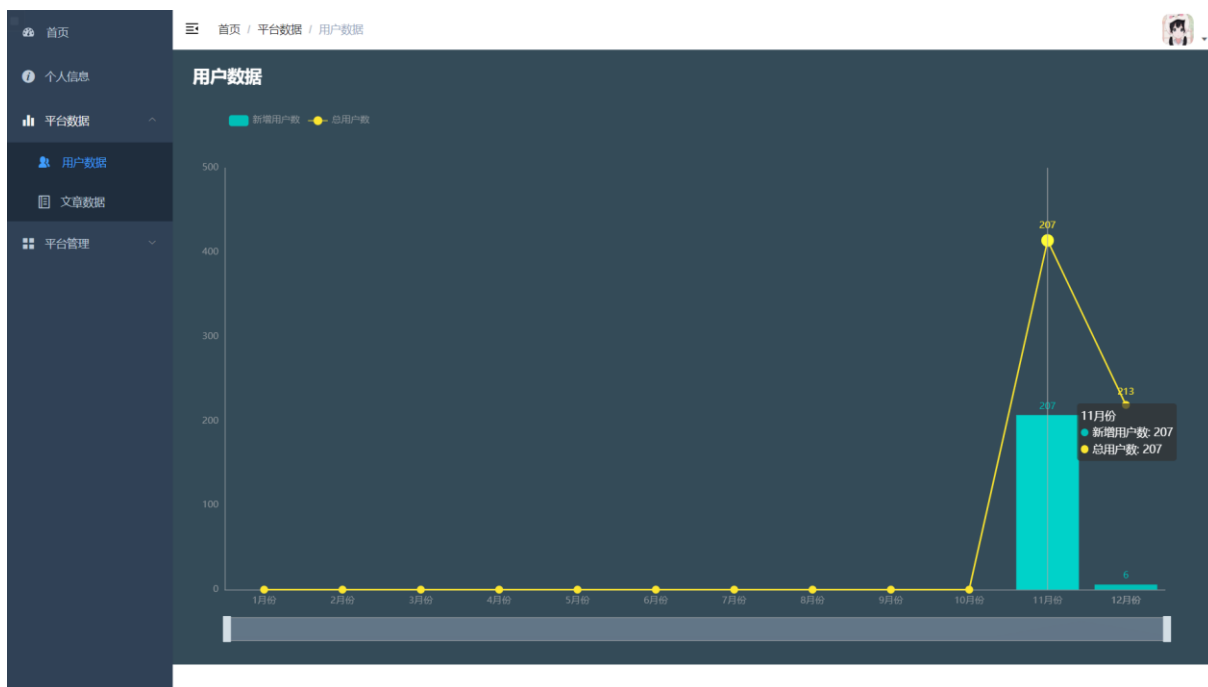


### 3.4.4 平台数据模块

此为管理员特有的功能模块。



此处引用 echarts 组件，使用了折线图、柱状图、饼图的形式分别展示所获取到的平台中的用户信息。



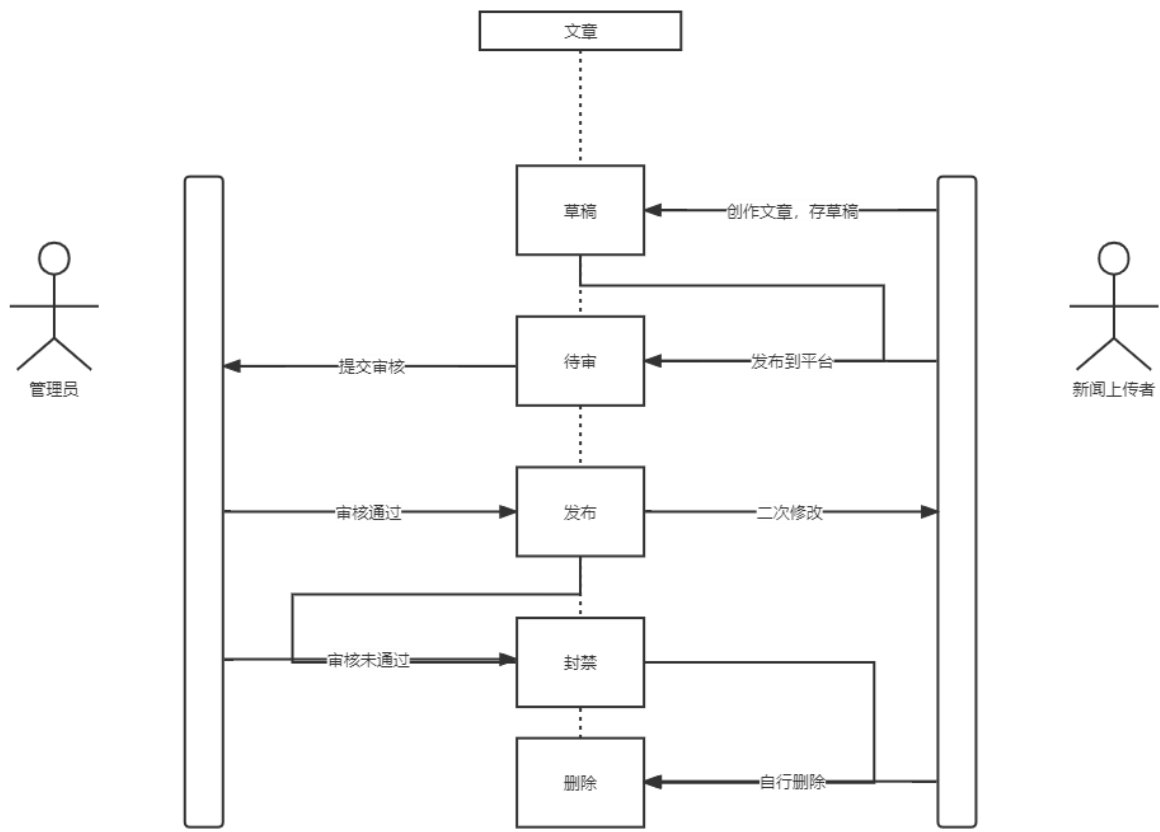
### 3.4.5 平台管理模块

平台管理是管理员特有的功能模块，主要分为用户管理和新闻文章管理两部分。两者的列表界面均可进行筛选、搜索、排序等操作，同时每页设置了最大行数为 20，引入 Pagination 分页组件进行分页显示。

用户管理部分，显示了用户 id，注册时间，用户名等信息，另外还涉及用户的封禁、解封、删除三种状态，仅有管理员有修改的权限。

文章管理部分，显示了新闻的 id，上传时间，标题，类型，作者，阅读量。操作中的详情

按钮点击后，会弹窗显示文章类型、标题、详情，管理员可以查看，同时可以修改文章的类型。另外涉及管理员和新闻上传者对于一篇新闻存储的 5 种状态，也即：草稿、待审、发布、封禁、删除。详细的不同身份的修改权限及新闻状态的改变如下图所示：



此处的界面设计如下：

首页

个人信息

平台数据

平台管理

用户列表

文章列表

首页 / 平台管理 / 用户列表

姓名

角色组

序号排序

搜索

编号	id	注册时间	用户名	状态	操作
1	1	2022-07-04 21:46	张三	可用	封禁 解封 删除
2	3	2022-06-29 22:49	李四	可用	封禁 解封 删除
3	5	2022-07-04 20:13	孤独气球	可用	封禁 解封 删除
4	7	2022-06-29 22:49	钟离宝贝	可用	封禁 解封 删除
5	8	2022-07-04 21:33	多情的诺言	封禁	封禁 解封 删除
6	9	2022-06-29 22:49	传统的蜜蜂	可用	封禁 解封 删除
7	10	2022-07-04 21:33	拉长的鸡翅	封禁	封禁 解封 删除
8	11	2022-06-29 22:49	余俱歌曲	可用	封禁 解封 删除
9	12	2022-06-29 22:49	矮小的毛衣	可用	封禁 解封 删除
10	13	2022-06-29 22:49	心灵美的微风	可用	封禁 解封 删除
11	14	2022-06-29 22:49	年菱白开水	可用	封禁 解封 删除
12	15	2022-06-29 22:49	无聊的芝麻	可用	封禁 解封 删除



### 3.4.6 文章管理模块

此为新闻上传者特有的功能模块。

列表内显示了文章的 id，上传时间，标题，类型，阅读量，分数，另外还涉及了状态的显示和修改，平台管理模块已有介绍不再赘述。此处操作种的编辑按钮，点击后会跳转到新闻文章的详情编辑页中。

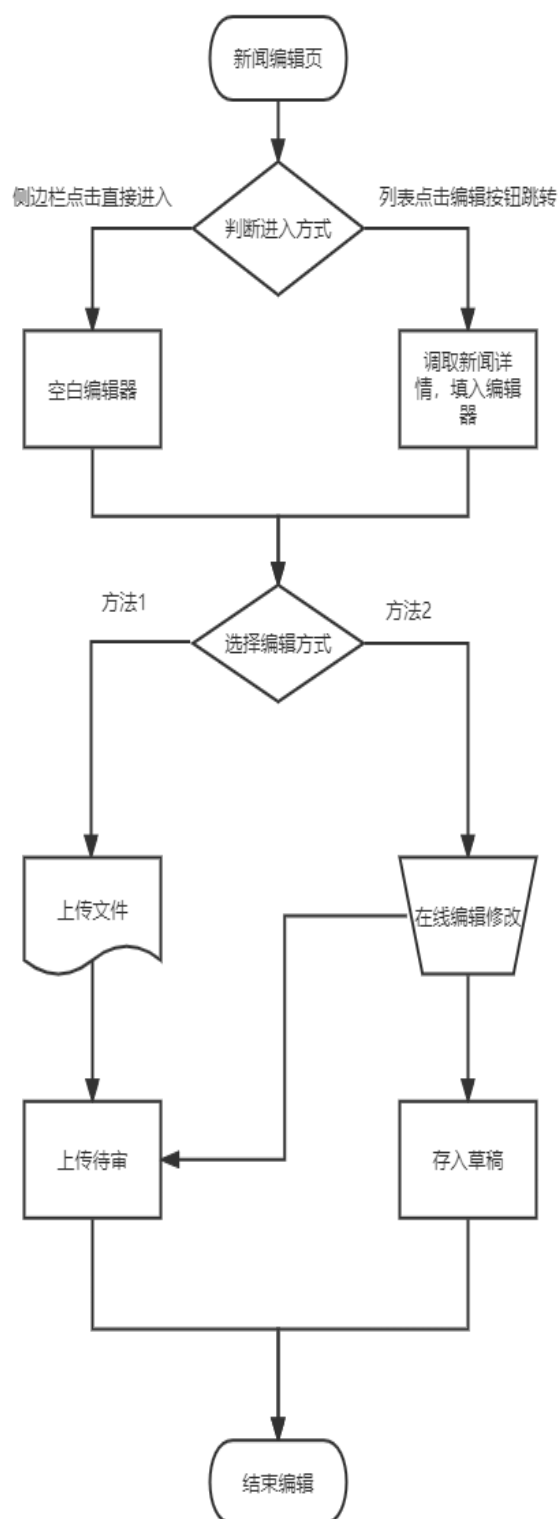
此处的界面设计如下所示:



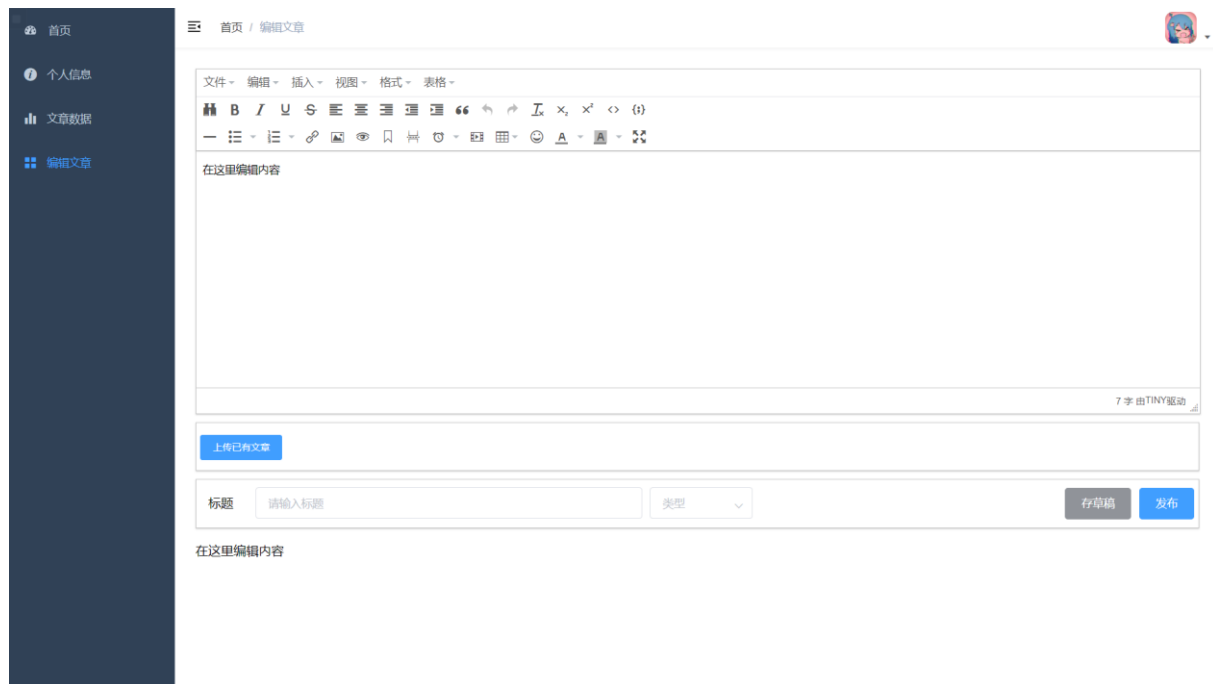


### 3.4.7 文章编辑模块

该模块是实现新闻上传者最关键的功能。使用开源富文本编辑器 TinyMCE 实现在线编辑的功能。同时提供文件上传的接口，允许多种文章创作功能。同时考虑到新闻文章可能不能一次性写完，提供了存草稿的功能。以下是该模块的流程逻辑图：



该部分的界面设计如图：



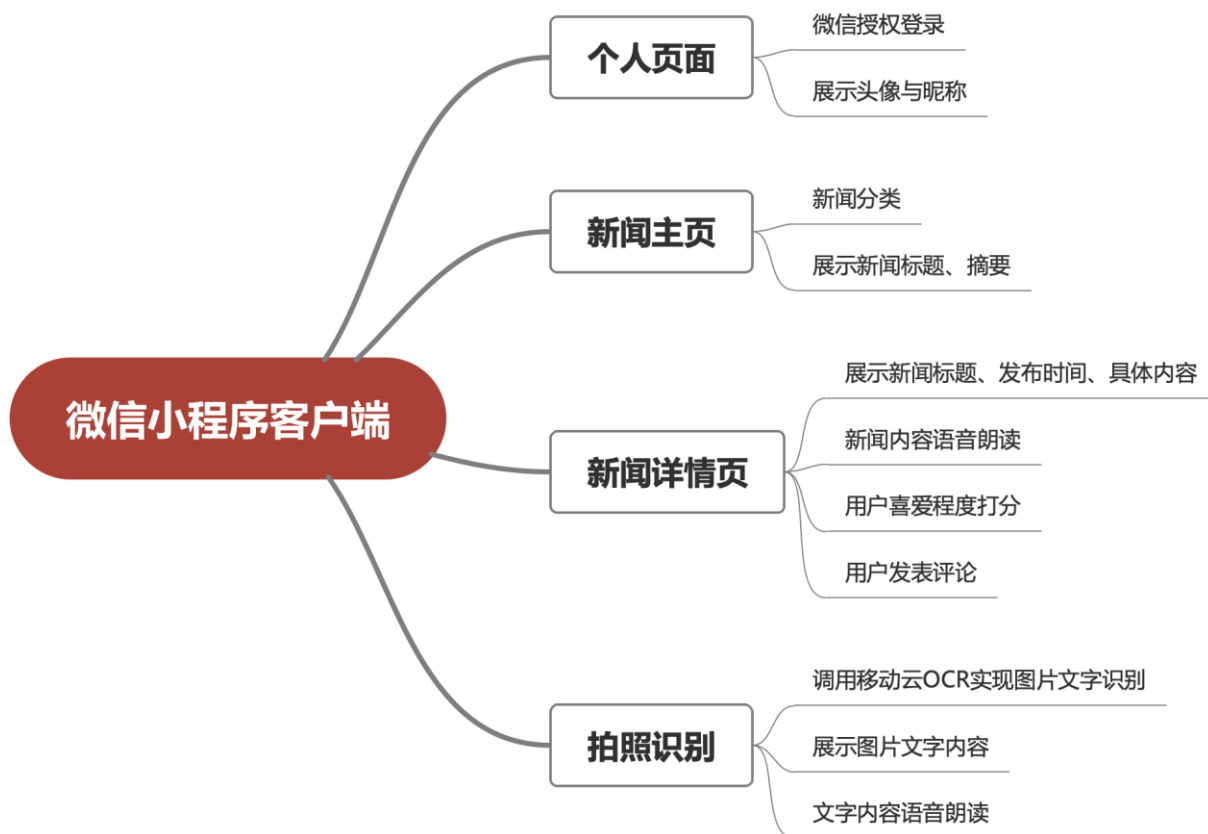
## 4 小程序客户端设计说明

### 4.1 程序描述

为了方便视障人士方便实用，本部分采用微信开放平台小程序作为软件客户端。小程序是一种新的开放能力，可以在微信内被便捷地获取和传播，同时具有出色的使用体验。

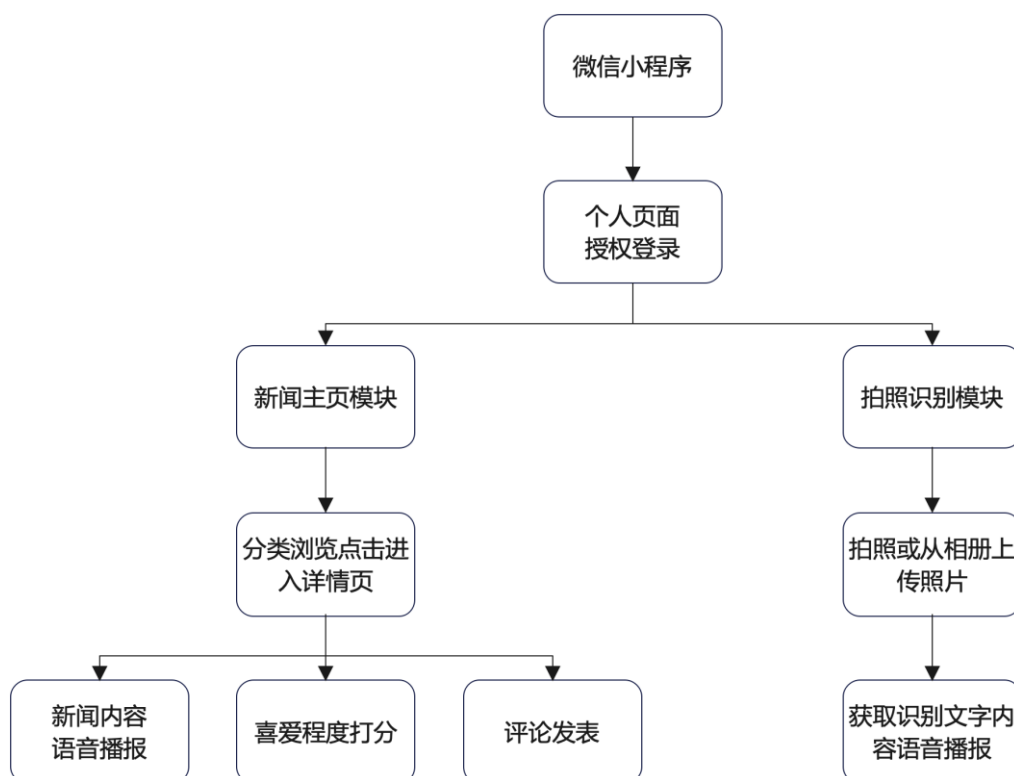
### 4.2 程序功能

小程序客户端主要分为个人页面，新闻主页，新闻详情页以及拍照识别模块。



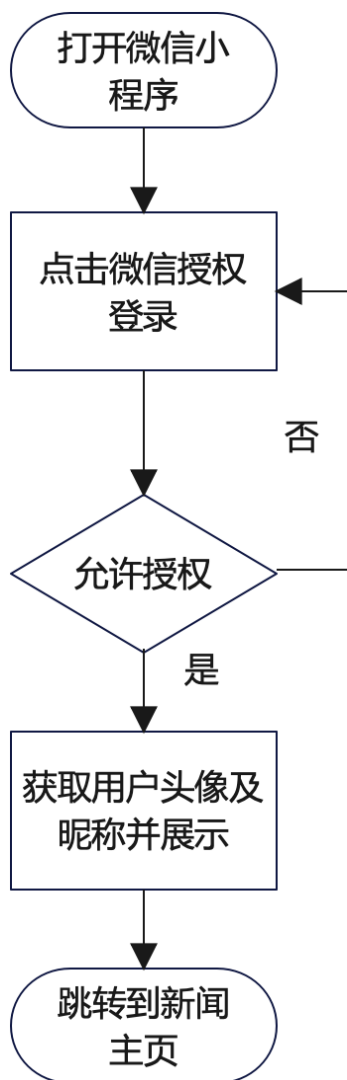
## 4.3 软件流程及界面设计

小程序主要功能流程如下图所示：



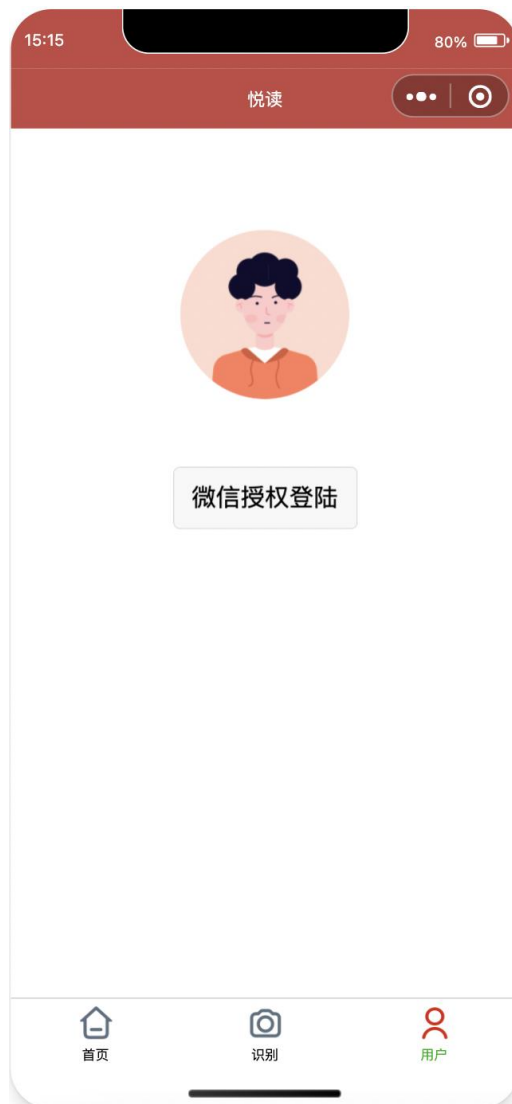
### 4.3.1 个人页面

个人页面为初始页面，打开微信小程序，点击微信授权登录，允许授权登录成功后跳转到新闻主页；若不允许，则返回初始页面。



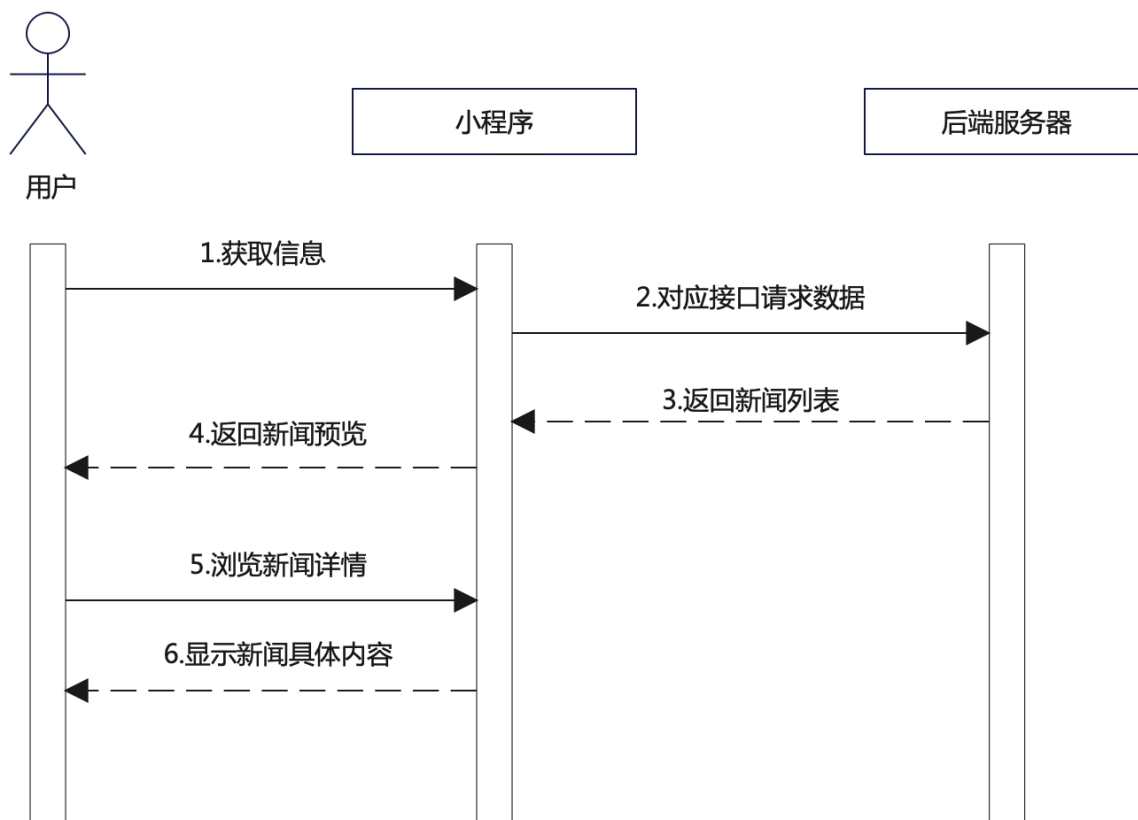
打开小程序，用户看到的是一个简易的登录界面，界面上部布局有初始头像，在其下方为“微信授权登录”按钮。用户点击按钮后将展现一个弹窗来向用户授权获取微信昵称、头像。用户可以点击允许，使用微信昵称及头像来当作此小程序用户基本信息。或点击“使用其他头像和昵称”，自定义新的头像和昵称作为此小程序的用户基本信息。授权登录获取到的信息将传递给后端服务器，后端接收到后返回信息作为用户标识，以便系统记录用户行为信息。

下图为个人页面界面设计：



### 4.3.2 新闻主页

用户登录成功后跳转到新闻主页获取新闻信息，小程序向后端发送请求，后端返回新闻列表给小程序。新闻列表内容包括新闻序号、新闻标题、新闻内容、新闻摘要、新闻发布时间以及新闻所属分类。当用户需要阅读新闻的具体内容时，点击新闻卡片进入新闻详情页。小程序将展现该条新闻具体内容。



新闻主页顶部展示的是新闻分类，点击某一新闻类型，页面将以卡片形式展现该分类下的一条新闻。上下滑动新闻卡片实现该分类下新闻的切换，左右滑动实现不同分类下新闻的切换。

新闻主页页面设计如下图所示：



点击任意一个新闻卡片则进入新闻详情页面。新闻详情页面包括新闻标题、新闻发布时间、新闻具体内容、用户喜爱度打分以及用户评论。在详情页面，用户借助手机自带阅读辅助激活文本朗读功能。例如苹果用户，在设置中打开旁白功能即可实现文本选择阅读，打开朗读屏幕功能，通过双指从屏幕顶部向下滑动激活文本朗读。

新闻阅读完毕后，用户可以进行新闻喜爱程度打分，同时可以进行评论，发表自己对该条新闻的看法。系统将记录下用户的行为信息以便后续分析为用户推荐自己感兴趣的新闻内容。

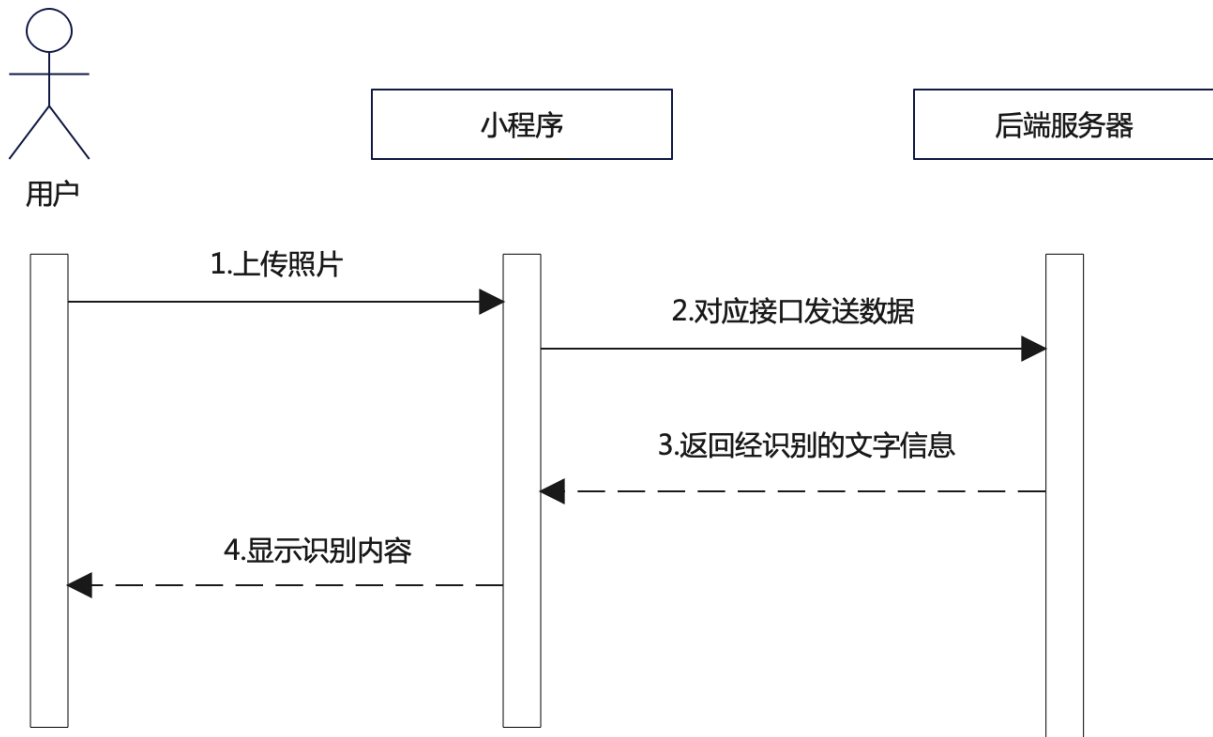
新闻详情页面如下图所示：



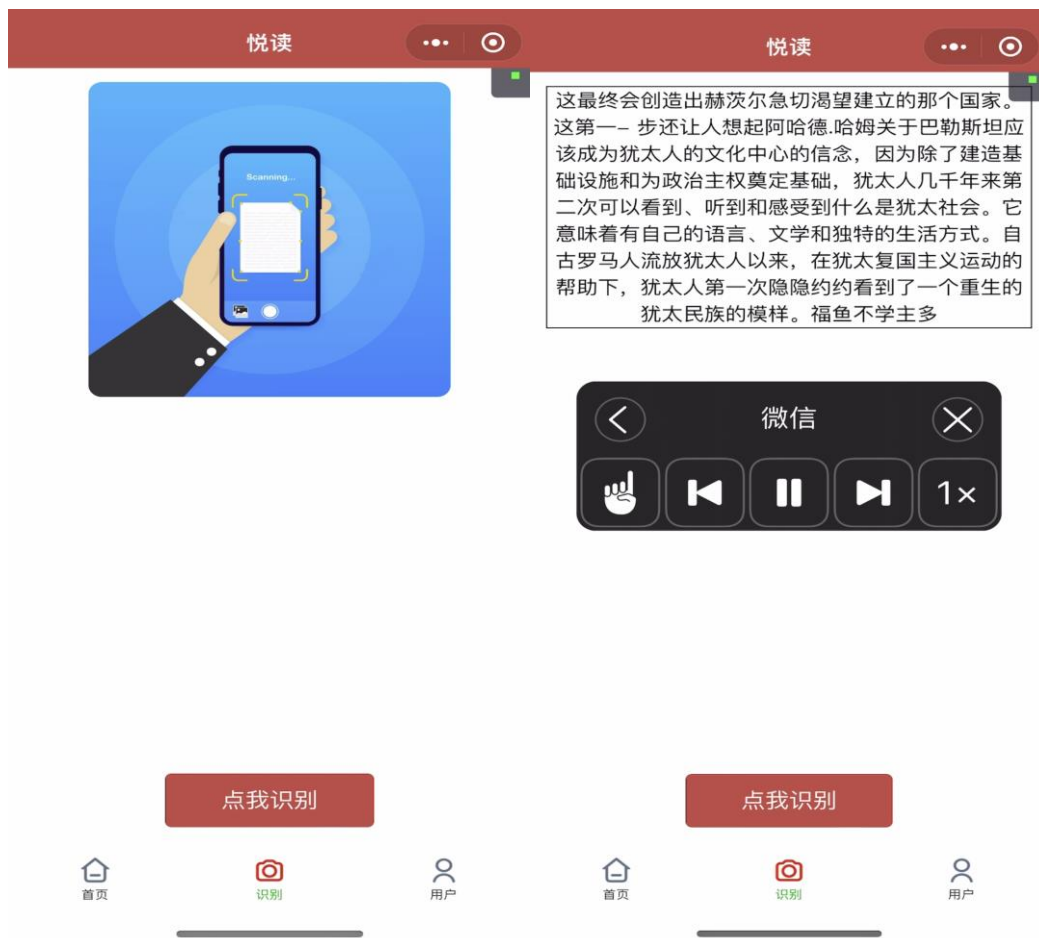
### 4.3.3 拍照识别

点击小程序下方相机按钮进入拍照识别模块。用户点击页面“点我识别”就可以进行拍照上传或者从相册中选择图片上传。上传至后端服务器的照片经识别处理返回识别后的文字信息，展示在方框内。用户可以使用新闻详情页文本朗读的方式进行识别内容的朗读。





拍照识别模块页面设计如下图所示：



## 4.5 接口

微信小程序客户端接口如下：

系统元素	接口	HTTP 请求方法	接口名称
用户信息	/reader/login	POST	授权登录
新闻主页	/index/reader	GET	获取新闻列表
新闻详情页	/reader/details /reader/like /reader/comment	GET POST POST	获取新闻详情 喜爱程度打分 发表评论
拍照识别	/reader/photo	POST	上传图片

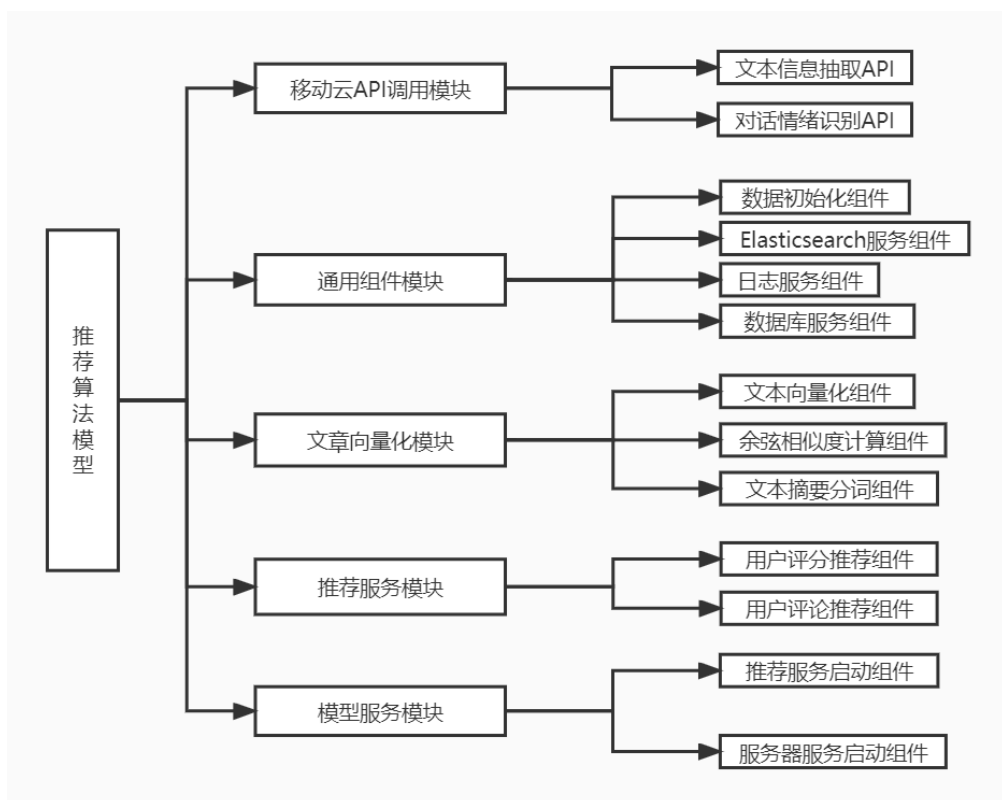
## 5 推荐算法模型设计说明

### 5.1 程序描述

本模块为新闻推荐模块，以下内容主要介绍新闻推荐模块的相关内容。

本模块采用词嵌入+余弦相似度算法思想构造模型。根据每一位用户对新闻的评价情况以及对于每条新闻的评论情况，记录并预测用户对于不同类型新闻的喜恶情况，从而对其进行个性化的新闻推荐。

下图为整个模型的组成架构：



## 5.2 功能

根据后端传入的用户 ID 以及推荐类型 (likes 或 comment)，向后端数据库查找相关数据并将参数返回模型进行计算。计算得出推荐度最高的一定数量（此处取值为 10）的新闻，将新闻 ID 及本次请求状态值返回给后端。

## 5.3 性能

由于官方没有提供相应的验证集，所以无法像一般推荐模型一样从准确率，召回率等维度进行衡量。故性能测试将从模型运行速度方面进行测试衡量。由于本模型查询文章采用 Elasticsearch 服务，速度极快，故测试新闻数量从五位数增加到六位数（万数量级到十万数量级）并不会对本模型运行速度造成影响。本新闻推荐服务分为据评分推荐服务和据评论推荐服务，经过反复测试得出：

1. 根据用户新闻评分的推荐服务：平均每一位用户的服务响应时间为：0.468818 秒
2. 根据用户新闻评论的推荐服务：由于该服务需要调用移动云对话情绪识别 API，故响应速度比前者稍慢一些。平均每一位用户的服务响应时间为：1.134472 秒

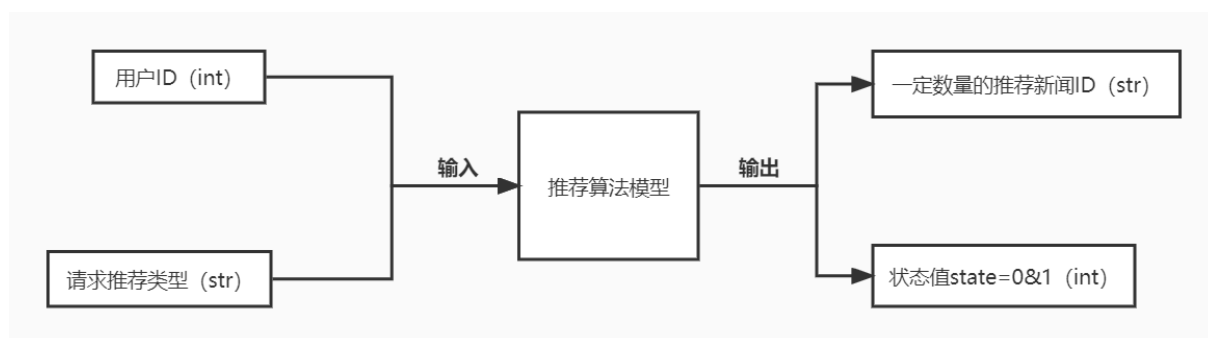
## 5.4 输入输出项

输入项：用户 ID (int)

请求推荐类型 (str)

输出项：一定数量的推荐新闻 ID (str)

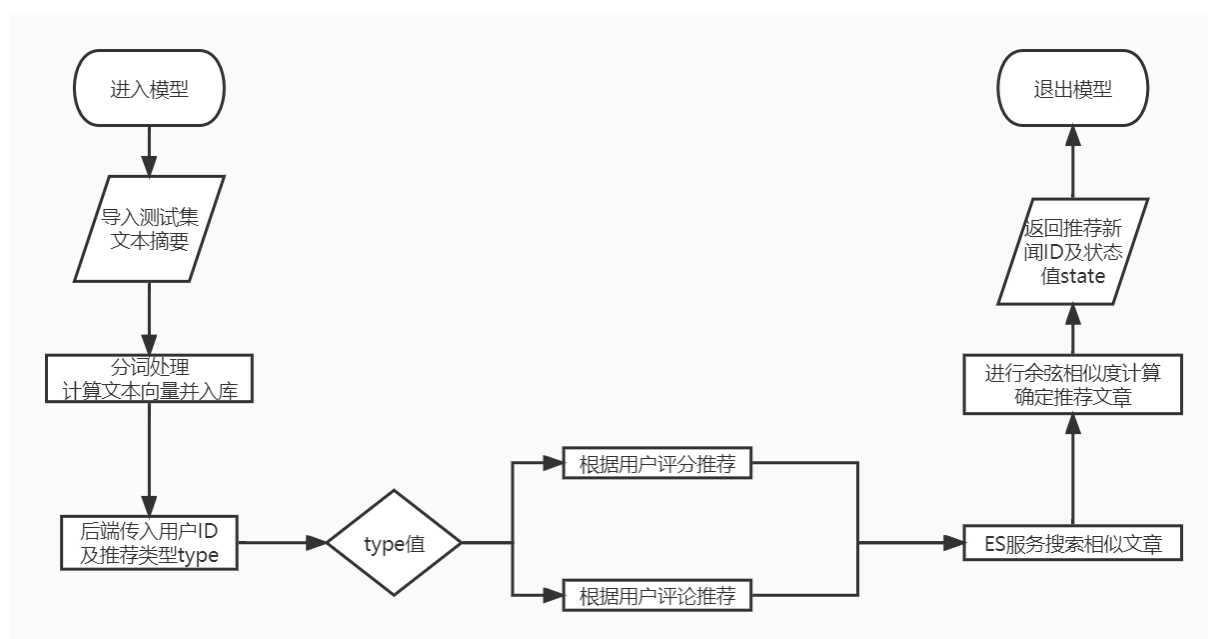
状态值 state=0&1 (int)



## 5.5 算法构成

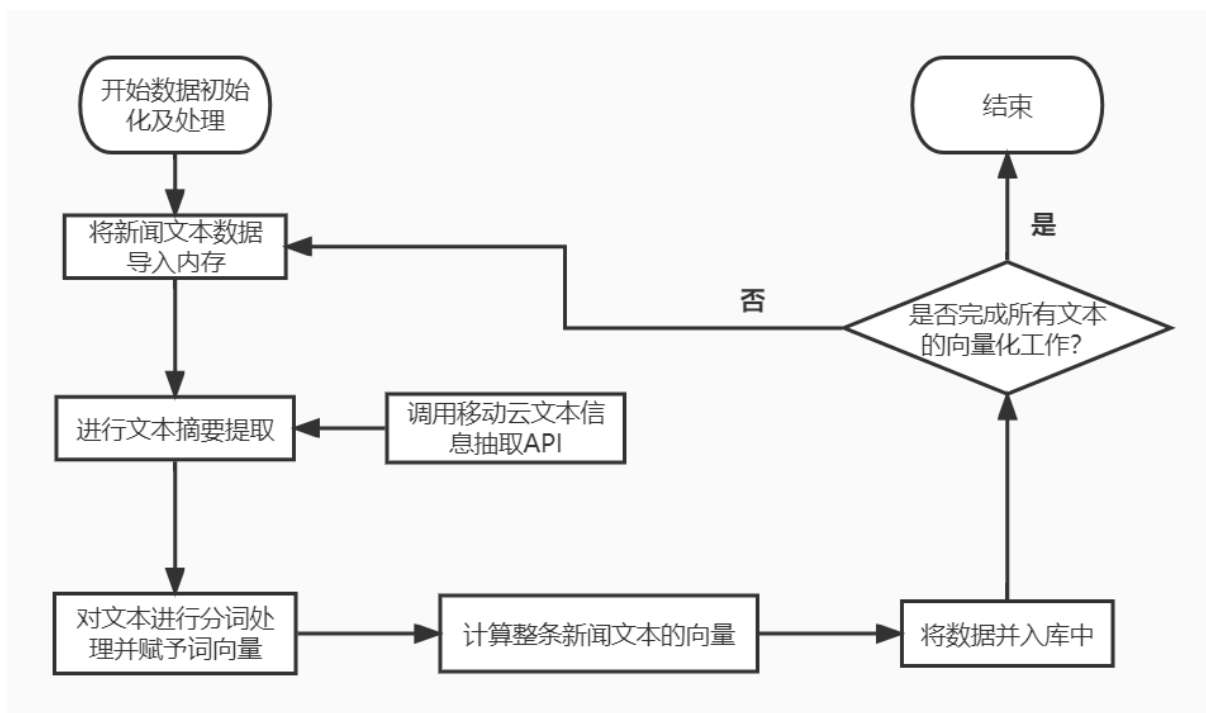
## 5.6 流程逻辑

本节介绍推荐模型的工作流程逻辑。本模型的工作流程逻辑包括：数据初始化及处理，推荐逻辑，搜索相似文章，返回结果等内容。下图为推荐算法模型的工作流程图：



### 5.6.1 数据初始化及处理

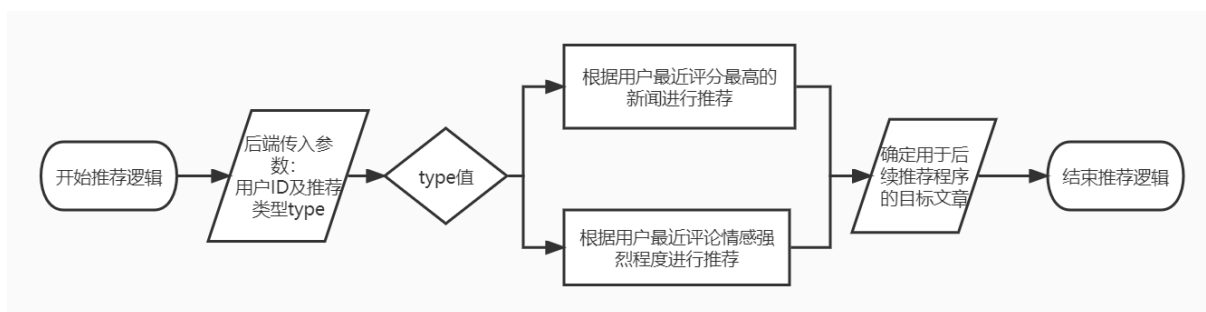
该部分为数据初始化及处理部分，该部分工作流程图如下：



1. 首先将测试数据集中的新闻文本导入内存。调用移动云自然语言处理基础服务：文本信息抽取 API，对新闻文本进行文本摘要。
2. 提取出文本摘要之后对其进行分词处理，并根据词嵌入向量文件“2000000-small.txt”将分词转换为词向量。  
注：词汇参照向量文件“2000000-small.txt”为腾讯开源的词嵌入向量集，其中收录了汉语中的两百万个词汇，并赋予每个词汇以一个 200 维的向量，以使每一个词语在向量域中具有唯一的特殊标识，方便后续文章间的余弦相似度计算。
3. 将该新闻的所有分词向量进行叠加运算并取其平均值，得出的结果即作为该条新闻文本的总向量。
4. 以上为对单条新闻所作的操作。对于测试集中的每条新闻重复如上操作，进行文章向量化工作并将每篇文章的总向量入库。

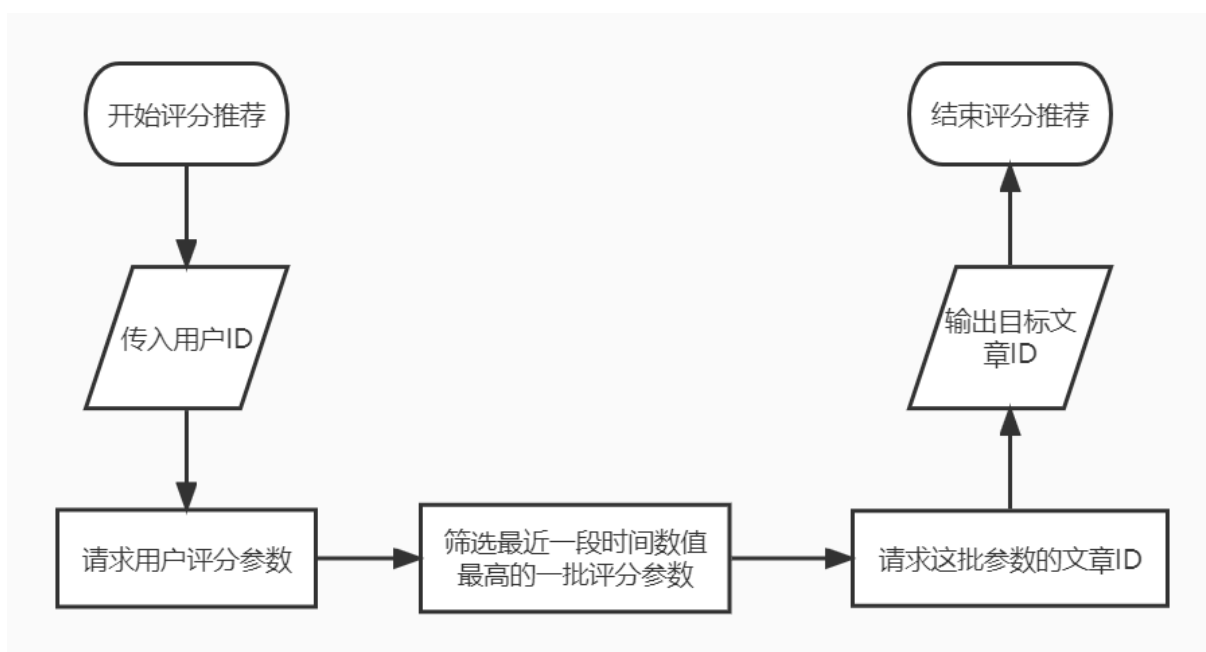
## 5.6.2 推荐逻辑

完成上述数据初始化及处理流程之后，现开始推荐进程。本推荐服务共构建两条推荐逻辑：根据用户评分进行推荐，根据用户评论进行推荐。现有后端传来请求，传输参数：用户 ID (int), 推荐类型 type (str)。现根据参数推荐类型 type 为 likes(评分)或 comment (评论) 启动推荐服务。该部分工作流程图如下：



### 5.6.2.1 根据用户评分进行推荐

该部分为根据用户评分进行推荐部分，该部分工作流程图如下：



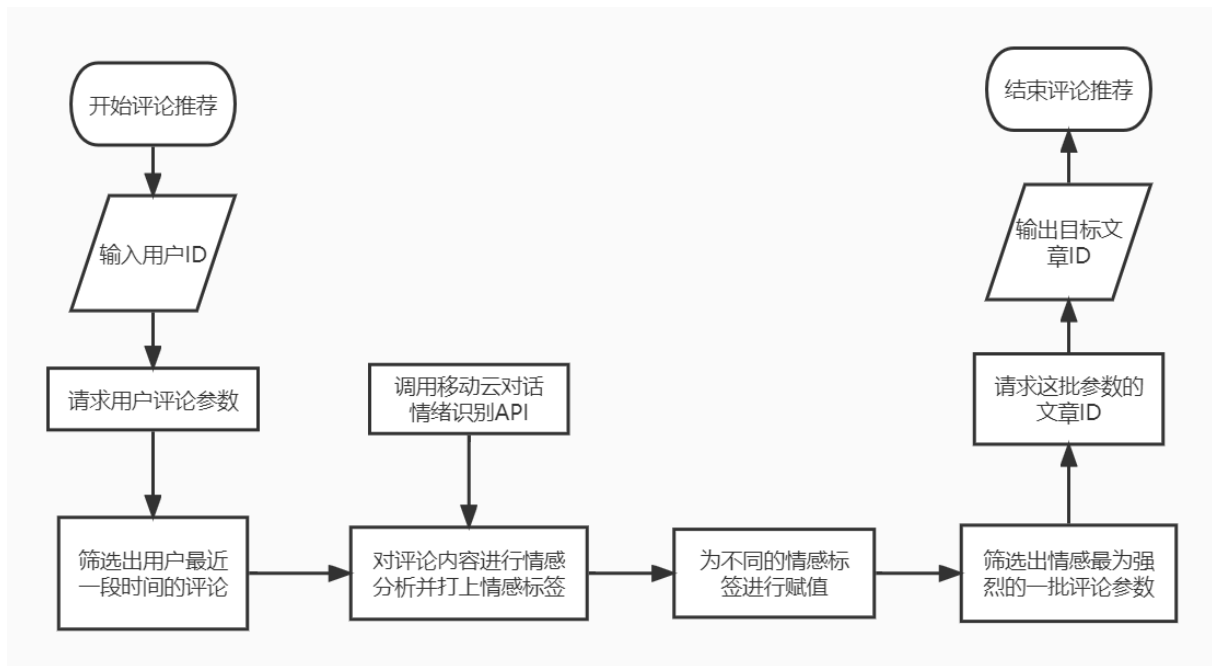
1. 根据传入的用户 ID 访问后端数据库，请求该用户的评分数据参数。
2. 将各个评分数据按照打分时间进行倒序排序，筛选出该用户最近时间段评分最高的一批新闻。

注：将评分数据按照打分时间进行倒序排序，目的为筛选掉评分较高但评价时间较为久远的一批数据，让筛选出的目标新闻具有时效性，从而使得推荐结果更加准确和个性化。

3. 将这批新闻设置为目标新闻，后续程序将以目标新闻为样本搜寻相似新闻进行推荐。

### 5.6.2.2 根据用户评论进行推荐

该部分为根据用户评论进行推荐部分，该部分工作流程图如下：



1. 根据传入的用户 ID 访问后端数据库，请求该用户的评论数据参数。
2. 将各个评论数据按照评论时间进行倒序排序，筛选出该用户最近时间内的一部分评论。  
注：将评论数据按照评论时间进行倒序排序，目的为筛选掉评论时间较为久远的一批数据，让筛选出的评论具有时效性，从而使得推荐结果更加准确和个性化。
3. 调用云服务接口：移动云自然语言处理基础服务：对话情绪识别 API，对评论内容进行分析并打上情感标签。

情感标签说明：

根据对话情绪识别 API，该接口将会返回 5 个值，分别对应 5 种文本的情感标签：正面情绪：感谢(0)、愉快(1)；中性情绪(2)；负面情绪：抱怨(3)、愤怒(4)。

其中将：感谢(0)和愤怒(4)标记为情感强烈，愉快(1)和抱怨(3)标记为情感较强，中性情绪(2) 标记为情感平淡。

情感值	情感分类
感谢(0)和愤怒(4)	情感强烈
愉快(1)和抱怨(3)	情感较强
中性情绪 (2)	情感平淡

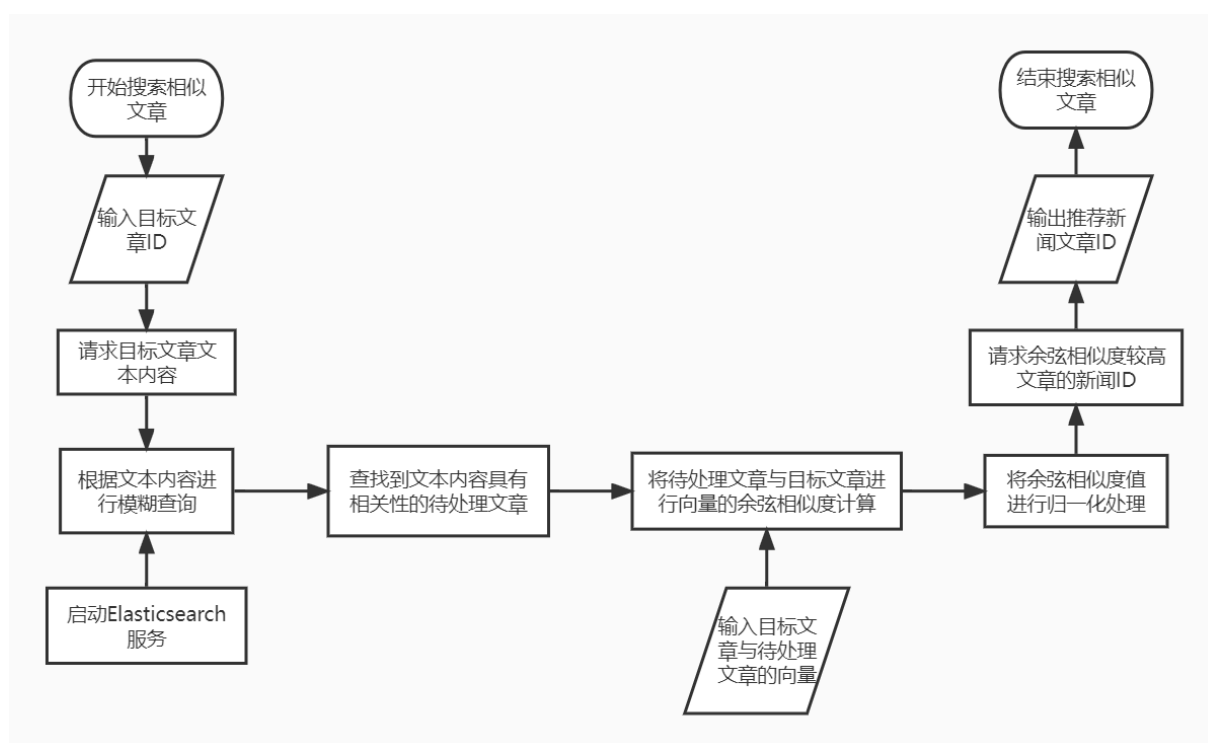
4. 根据每条新闻情感标签上的标记分别赋值，筛选出评论情感最为强烈的一批新闻作为目标新闻，后续程序将以目标新闻为样本搜寻相似新闻进行推荐。

注：此处将愤怒(4)和抱怨(3)归类为情感强烈并将此作为目标新闻进行推荐，此处的考量为：虽然该用户面对此类新闻表达了自己较强的负面情绪，但正因为该用户对其比较

关注，反应较大才会做出情感强烈的负面反馈。举例来说，该用户对社会上出现的暴力违法犯罪等新闻（如近期的唐山烧烤店打人事件）发表了自己的负面情绪评论，其大概率还是会迫切关注该事件的后续进展或其相似事件。

### 5.6.3 搜索相似文章

该部分为搜索相似文章部分，该部分工作流程图如下：



1. 确定目标新闻之后，访问获取目标新闻文章的文本内容。
2. 根据目标新闻文本内容进行模糊查询，寻找文本内容较为相似的一批待处理新闻文章。
3. 将目标新闻的向量与待处理新闻文章的向量进行余弦相似度的计算，并进行归一化处理，使相似度值归一到区间[0,1]。
4. 将相似度值排名前一定数量（本项目暂定取 10）的新闻文章提取其文章 ID，标记为推荐新闻文章。

注：步骤 2 中用到了 elasticsearch-7，在此对其进行一个简单说明：

Elasticsearch 是一个实时的分布式存储、搜索、分析的引擎。相对于数据库或其他查询工具，Elasticsearch 的强大之处就是可以模糊查询：

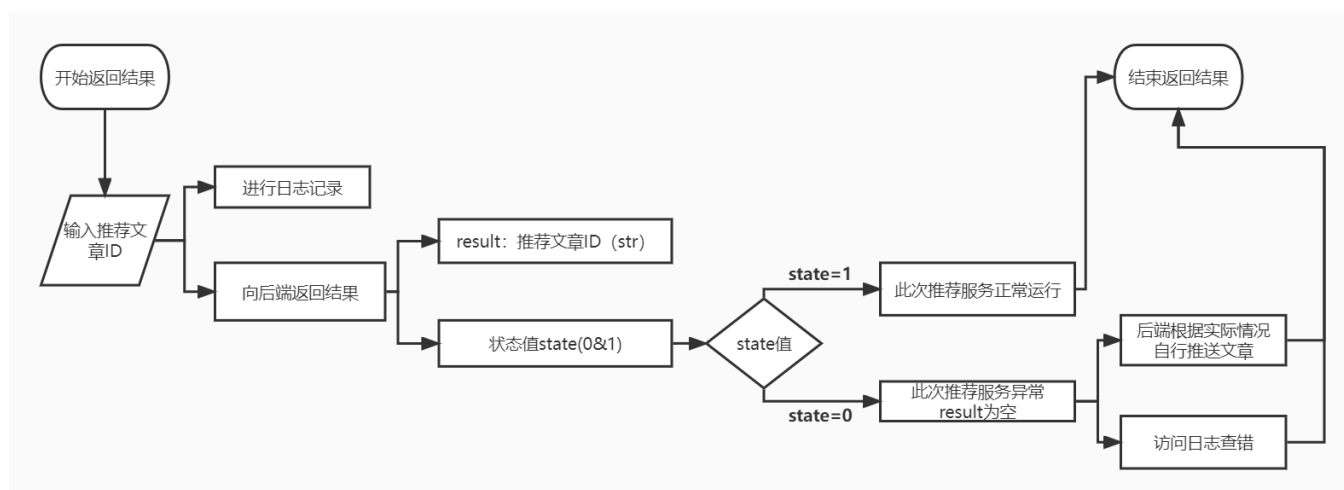
- (1) Elasticsearch 对模糊搜索非常擅长（搜索速度很快）。
- (2) 从 Elasticsearch 搜索到的数据可以根据评分过滤掉大部分的，只要返回评分高的给用户即可（原生就支持排序）。
- (3) 没有那么准确的关键字也能搜出相关的结果（能匹配有相关性的记录）。



正因为使用了 Elasticsearch，可以实现较快匹配相关新闻数据。虽然测试数据集里只有近五万条新闻，但该模型完全可以搭载本赛题开源的新闻集“THUCNews.zip”里全部的 83 万条新闻数据，响应速度和初始化所需时间只与所搭建的服务器内存有关。

#### 5.6.4 返回结果

该部分为返回结果部分，该部分工作流程图如下：



确定所推荐的新闻文章 ID 之后，向后端进行结果返回，并进行日志记录。

返回值为：{"result": ['256822', ..., '287006'], "state": 1}

result 为推荐文章 ID (str)，共十个。

state 为状态值 (int)：值为 1 表示推荐服务正常运行：值为 0 表示推荐服务异常，需后端根据情况向用户自行推荐新闻。

#### 5.7 接口

接口名称	接口信息
推荐服务请求接口	http://47.103.7.30:8001/recom
Elasticsearch 服务接口	http://47.103.7.30:9200
数据库接口	http://39.99.60.47:3306/cnsoftbei
移动云 API 接口	https://ecloud.10086.cn

#### 5.8 限制条件

1. 服务器部署问题：若想要搭载开源新闻集中的全部 83 万条新闻，模型响应速度方面没有问

题，但全部新闻的向量化进程受到服务器内存的极大限制。目前只能开通内存 4G 的服务器，该服务器完成 83w 条新闻的向量化进程保守估计至少需要 10 小时，希望官方在复赛提供服务器时能够好好考虑一下。

2. Elasticsearch 服务若在服务器上部署的话，在面对上述情况时处理速度会极大地依赖于服务器内存大小。移动云上也有相应的 Elasticsearch 服务，但可惜不能免费试用，所以排除 API 调用 Elasticsearch 服务。
3. 移动云自然语言处理基础服务：文本信息抽取 API：若要搭载全部 83 万条新闻，该服务的免费限用额度较少，所以不得不减少测试数据集大小或采用本地文本摘要服务。

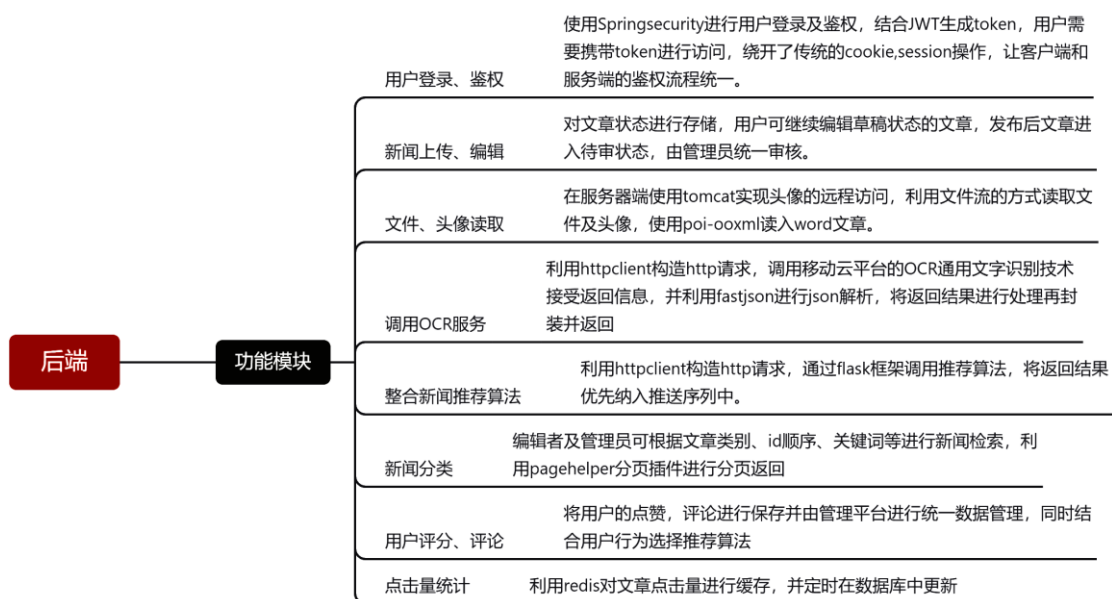
## 6 后端设计说明

### 6.1 程序描述

该程序负责进行数据库交互，整合各个模块进行数据处理，并按接口约定返回数据至微信端或小程序端。

### 6.2 功能

- (1) 用户登录、鉴权；
- (2) 新闻上传、编辑；
- (3) 文件、头像读取；
- (4) 调用 OCR 服务；
- (5) 整合新闻推荐算法；
- (6) 新闻分类；
- (7) 用户评分、评论；
- (8) 点击量统计。



## 6.3 流程逻辑

本部分将对代码中的重点功能进行详细介绍。

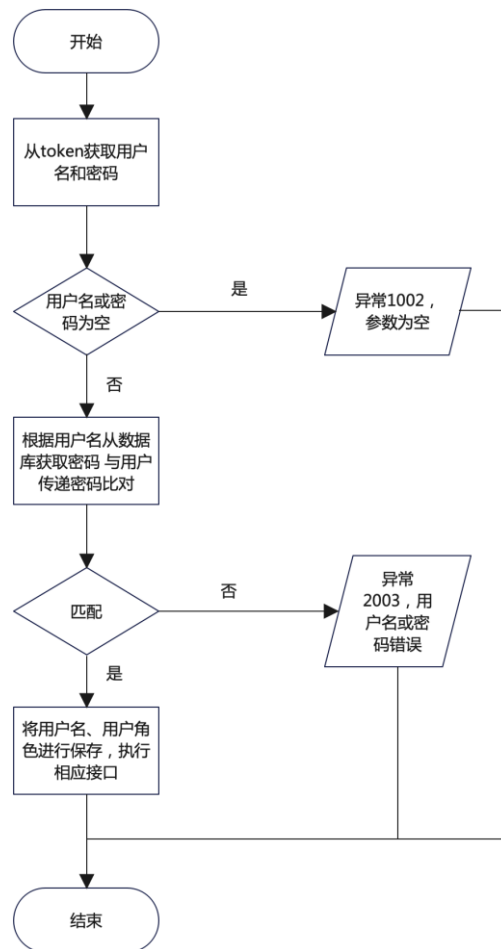
### 6.3.1 editor、admin 登录

后端主要采用 Springsecurity 进行用户登录，鉴权。登录模块的主要功能是为用户提供 token 以便用户能访问其他页面。

后端接收用户输入的用户名和密码，当用户登录时首先校验用户名或密码是否为空，若为空则抛出异常，否则根据用户的用户名在数据库中进行查询，若未查询到或密码不匹配则抛出异常，否则根据用户的用户名，角色代码生成 token 并返回。

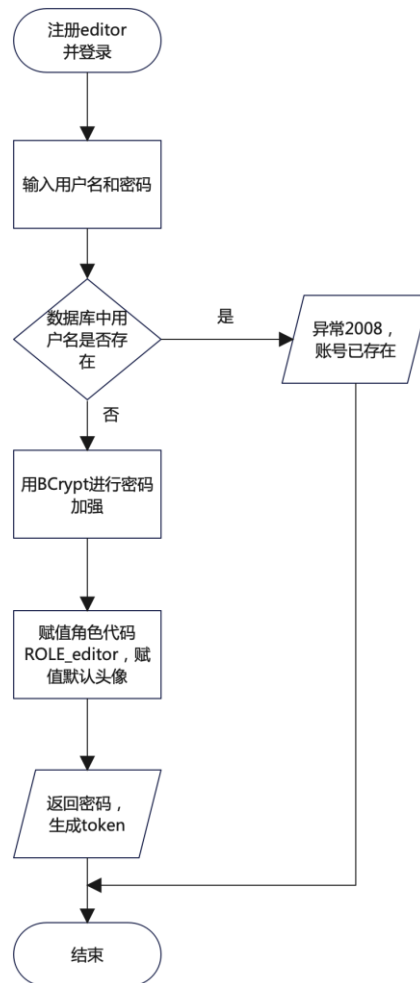
异常由 Controller 层统一接收，并封装进统一的返回结果中，若无异常则将返回信息置入统一的返回结果，后文不再赘述。

数据库中的密码均采用 BCrypt 进行加密，并进行加盐操作，安全性高。流程图如下：



### 6.3.2 editor 注册

当 editor 进行注册时，后端接收用户输入的用户名和密码，首先校验数据库中是否存在该账号，若存在则抛出异常，否则对用户密码进行加密存储并返回 token，让用户处于登录态。流程图如下：

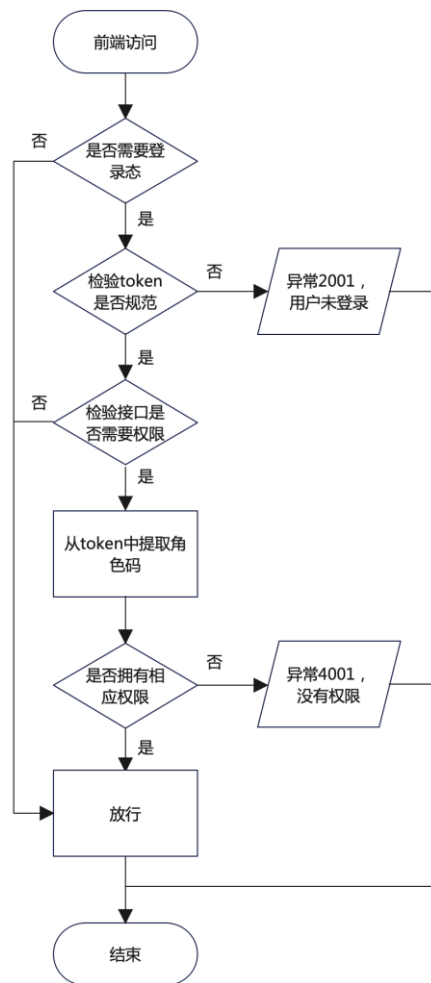


### 6.3.3 用户鉴权

后端多数接口采用 Springsecurity 提供的注解配置注明了用户登录需要的权限，当用户访问这些需要鉴权的接口时，Springsecurity 就会对请求进行拦截。

查看请求头中带有的 token 是否符合规范，若不符合抛出异常，否则根据 token 获取用户权限判断与接口所需权限是否相符。

若没有相应权限则抛出异常，否则进行相应的访问。流程图如下：

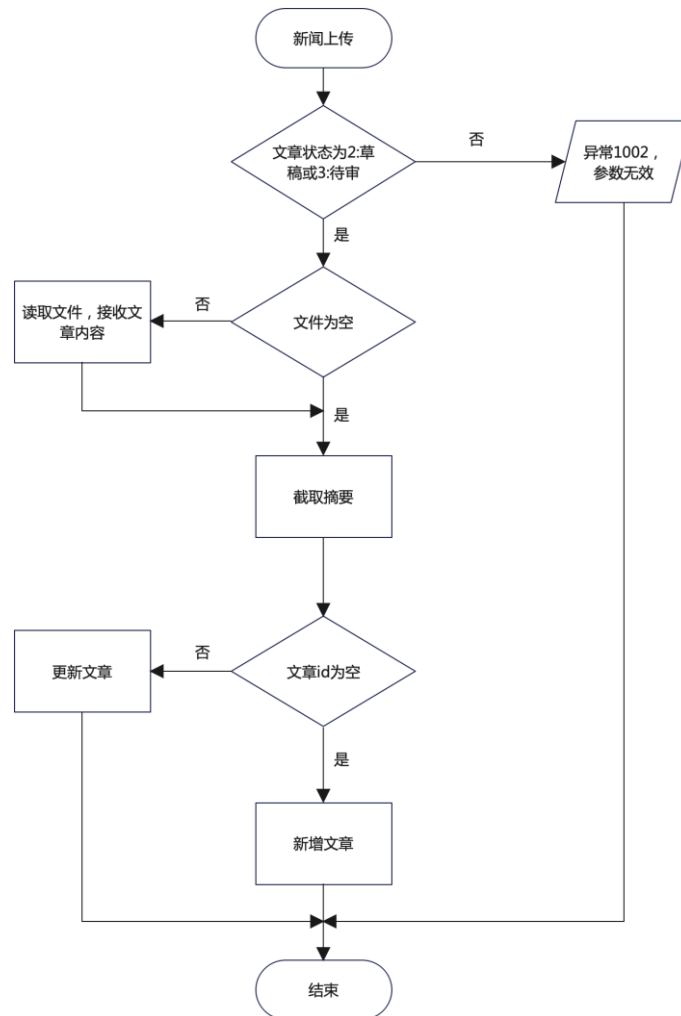


### 6.3.4 新闻上传

当编辑者上传文章时，将携带文章标题 title, 文章类型 IDtypeId, 文章状态 status, 文章 ID id, 文章内容 content/文章文件 file，其中文章状态必须携带。

若文章状态不为草稿或待审则抛出异常，因为其他状态用户无权进行操作。之后判断用户是否以文件形式上传文章，若是则调用相关工具类读取内容，否则将 content 作为文章内容。

之后判断文章 ID 是否为空，若 ID 为空则说明是 editor 新增的文章，需要在数据库中进行添加操作，否则说明是 editor 在编辑之前保存为草稿态的文章，需要在数据库总进行修改操作。流程图如下：

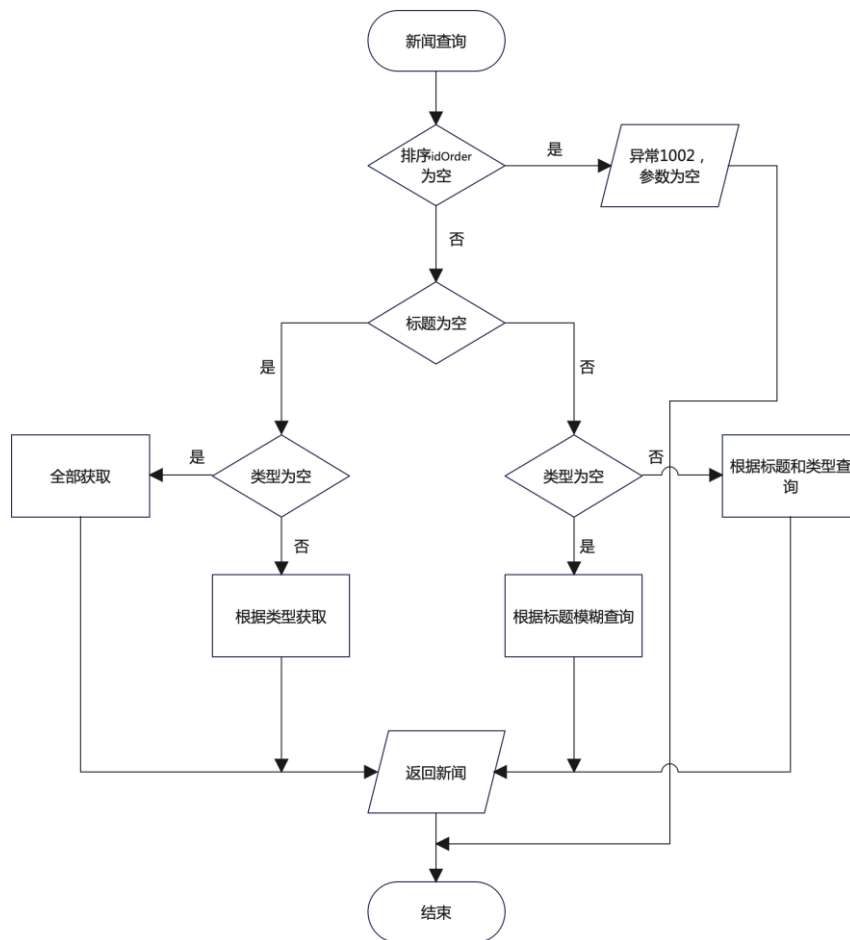


### 6.3.5 文章分类

当 editor 或 admin 查询文章时 (editor 只能查询自己发表的文章, admin 能查询到所有文章) 后端会接收文章类型 typeId, 文章 ID 排序 idOrder, 文章标题 title, 文章页数 page (默认每页 20 条数据)。

其中 idOrder 为必须携带的参数, 若为空则抛出异常。

首先判断 title 是否为空, 进而判断 typeId 是否为空, 根据这两个值是否为空产生的四种情况分别做为对应的查询条件进行查询, 最后由 pagehelper 分页插件根据 page 进行分页, 若 page 不符合规范则默认 page 为 1, 最后将结果返回流程图如下:



### 6.3.6 调用新闻推荐算法

此模块调用了两个包，一是 apache 公司提供的 httpclient，用以构建 http 请求，二是阿里巴巴公司开源的 fastjson，用以 Json 对象和 java 对象的相互转换。

新闻推荐算法为 Post 请求，接收用户的 Id 以及根据评分或根据评论两种推荐方式,若成功返回 10 条文章 Id。其接收的参数为 json 格式，故需要先通过 fastjson 将对象转换为 json 字符串再进行传递，并且对返回值同样通过 fastjson 进行获取。

首先根据用户的评论以及评分数目决定推荐方式，若获取的 state 值为 0 则说明为新用户，暂无相关行为，则随机每类新闻推送 5 条，若 state 值为 1 说明调用成功，将得到的十条新闻的 ID 存入返回结果集，并随机每类再存入 4 条，将结果返回至首页。

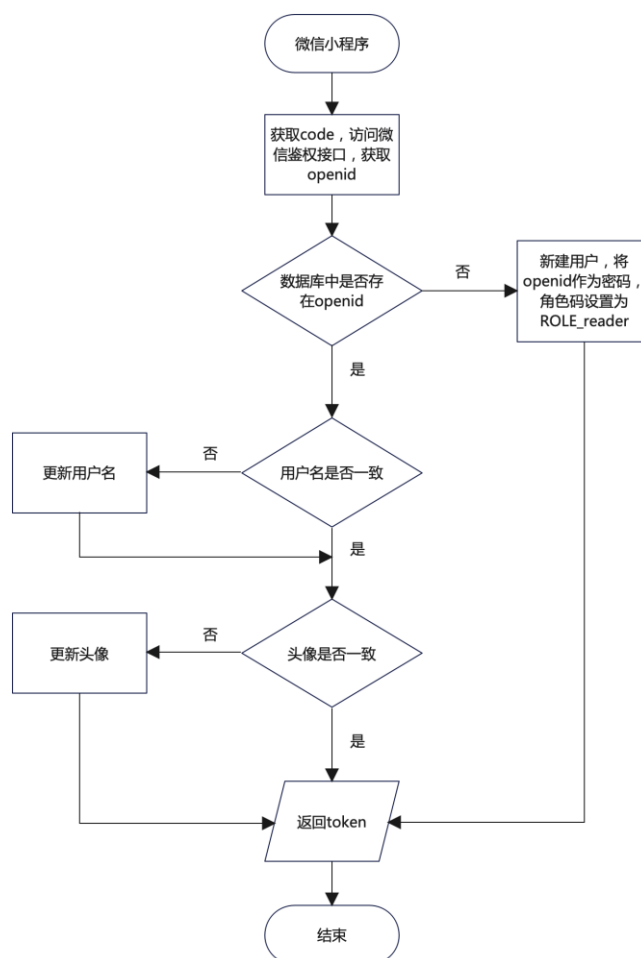
### 6.3.7 reader 注册、登录

当 reader 注册/登录时，接收用户名，code，头像地址三个参数。

首先需要根据 code 调用微信的鉴权接口获取用户的唯一标识，以便于数据库中的用户信息能和微信中的相关信息相统一。通过 httpclient 远程访问微信提供的接口，需携带 appId 以及 appSecret，这两个参数需要在微信平台获取并配置在配置文件中。



根据返回的用户唯一标识 openid 查询数据库中是否有与之匹配的用户，如果有则进一步判断用户的用户名，头像是否需要更新；如果没有则将 openid 作为密码，新增用户并进行存储。流程图如下：

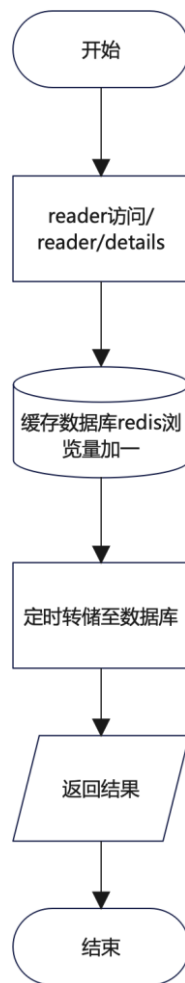


### 6.3.8 调用 OCR 服务

根据移动云官方提供的 sdk 包进行调用，并取返回结果中 items 数组的所有 itemstring 并将其连贯成段，以字符串的形式返回给前端。

### 6.3.9 点击量存储

当 reader 浏览文章详情时，会触发点击量的增长，这里为了避免频繁访问数据库，故采用了利用 redis 进行缓存，并将浏览量定时存入数据库的策略。redis 的读写效率远远高于 mysql 数据库，此做法提高了程序的运行效率，减轻了服务器端的压力。流程图如下：



### 6.3.10 其他模块整合

后端通过自定义状态码构建了统一的异常处理机制，并且构建了统一的返回类型，使得无论是客户端还是服务端都能接收到统一的返回信息，为开发工作提供了极大的便利。此外后端此次采取了 jwttoken 的登录、鉴权方式，避免了小程序端缺失 cookie session 机制引起的麻烦。统一返回结果如下：

```
{
  "success" (boolean, 是否成功) ,
  "code": (int, 状态码) ,
  "message": (String, 状态信息) ,
  "data" (Object, 返回数据)
}
```

## 6.4 存储分配

程序产生的相关数据保存在数据库中，故存储空间即为数据库软件所在的磁盘空间。对于用户上传的头像，程序做了存储工作，需要开发者配置存储路径以及访问路径，对于用户上传的头像设置了文件大小不超过 10MB，对于以文件形式上传文章则由系统开辟内存进行暂存，读取内容后立即释放，将占用一定的内存。