

# 基于个性化推荐的文章阅读 app

## 摘 要

随着网络资讯产生速度不断加快，新闻和自媒体文章的种类和数量快速增长，用户需要花费大量的时间才能找到自己想看的文章。这种浏览大量无关的信息和产品过程无疑会使淹没在信息过载问题中的消费者不断流失。为了解决这些问题，文章个性化推荐系统应运而生。本产品致力于提供为用户提供个性化的文章推荐，向用户推荐用户感兴趣的文章。想用户之所想。

## 目 录

摘 要	1
第一章 设计需求分析	2
1.1 设计背景	2
1.1.1 产品描述	2
第二章 特色与创新	3
2.1 个性化	3
2.1.1 包罗万象	3
第三章 功能规划	4
3.1 功能介绍	4
第四章 软件架构和开发环境	5
4.1 软件架构 和开发环境介绍	错误！未定义书签。
第五章 其他	6
5.1 作品运营模式与成本	6
5.2 参考文献及说明	6

## 第一章 设计需求分析

（本章节主要描述面向什么需求设计了本作品，本作品主要解决了什么问题，或对什么现有产品或应用作出了改进等）

## 一.1 设计背景

随着网络资讯产生速度不断加快，新闻和自媒体文章的种类和数量快速增长，用户需要花费大量的时间才能找到自己想看的文章。这种浏览大量无关的信息和产品过程无疑会使淹没在信息过载问题中的消费者不断流失。为了解决这些问题，文章个性化推荐系统应运而生。本产品致力于提供为用户提供个性化的文章推荐，向用户推荐用户感兴趣的文章。想用户之所想。

## 一.2 需求分析

- 1.本系统主要分为三大部分：推荐系统设计与实现，安卓 app 的设计与实现，文章爬虫的设计与实现
- 2.文章爬虫是本项目的数据库基础，爬虫进程以天为单位，爬取今日头条，微信公众号等文章存入数据库
- 3.推荐系统是实现用户个性化阅读的核心。推荐系统基于 spark 计算框架，由 spark 集群并行计算。整个推荐系由 spark 离线批处理和 spark streaming 实时计算两大模块组成。
- 4.安卓 app 是为用户提供服务的窗口，同时也为推荐系统提供用户兴趣数据

## 一.3 产品描述

这是一个提供个性化文章推荐的移动端 app，个性化推荐是本 app 的一大特色。通过爬取网络上的优质新闻和文章到数据库，做好分类。根据每个用户的喜好记录进行推荐。

1. 用户打开 app，可以选择以游客身份浏览新闻和文章。每浏览一篇文章，在当前文章下方显示与当前文章相关的其他文章。

2.本 app 提供注册登录功能。用户可以自定义用户名注册，并绑定邮箱进行激活。

下次登录用户账号时，网站会在首页进行个性化推荐。

3. 主页面的内容均按照内容标签进行分类，分为时政，体育，游戏，技术，科技等。

[今日小站](#)
[关于本站](#)
[推荐](#)
[国际](#)
[时政](#)
[科技](#)
[社会](#)
[数码](#)
[军事](#)
[娱乐](#)
[游戏](#)
[体育](#)
[汽车](#)
[财经](#)

[登陆](#)
[注册](#)

新闻内容

不要忘了! 你的手机还有这个功能——NF C

read more

太疯狂! C罗点球命中后意大利解说直播中离衣庆祝?

read more

盘点泰山春晚上的四次失误, 几个细节体现出专业演员的素养

read more

【关注】他们建议恢复五一长假, 你怎么看?

read more

普通的人都能像孙悟空经纪人一样“洗澡有去

read more

[今日小站](#)
[关于本站](#)

[登陆](#)
[注册](#)

## 霸气回击! C罗掏裆庆祝复仇西蒙尼, 4万尤文球迷疯狂膜拜

奇迹, 这就是奇迹! 欧冠1/4决赛次回合, 尤文图斯主场3-0完胜马竞, 两回合3-2淘汰毕尔巴鄂, 闯入欧冠8强。C罗是当之无愧的主角, 两次头球破门, 关键时刻顶住压力打入点球。C罗用帽子戏法, 证明自己超级巨星的作用。他撞开C罗, 这样的C罗, 谁能不服?



从禁 @体育足球世界

C罗带着愤怒, 愤怒的C罗无比可怕。马竞想守住晋级欧冠2-0的优势, 但C罗不答应。上半场, C罗力压胡安弗兰, 霸气头球破门; 下半场, C罗在马德里人山人海之中泰山压顶, 头球梅开二度; 第86分钟, C罗又打入制胜点球! 何为超级巨星? 何为力挽狂澜? 何为救世主? 看看C罗, 就是答案。

[今日小站](#)
[关于本站](#)

[登陆](#)
[注册](#)

相关推荐

C罗+齐达内! 他是皇马最缺的不死战魂 想看C罗英吗? 没门

read more

6年间5次被C罗羞辱淘汰! 马竞铁杆被彻底打服: C罗是世界最佳

read more

各路名嘴大吹C罗! 就连西蒙尼和格列兹曼都被C罗征服了!

read more

马竞还是逃不过一个叫C罗的男人! 尤文花1亿欧元买C罗太值了

read more

欧冠之王: 什么是领袖气质? C罗这个圈

read more

## 第二章 特色与创新

（本章节主要描述作品与现有其他技术或应用相比所产生的优势和创新点）

### 二.1 个性化

本 app 使用协同过滤等推荐算法为用户推荐用户最可能感兴趣的内容，减少了用户查找的时间。相比其他仅仅根据文章的热门程度和文章的新旧来规定出现次序的网站相比，本 app 更贴近用户，帮用户从海量的新增文章中查找出用户最有可能关心的资讯。

### 二.2 包罗万象

本 app 规划后续引入包括新闻，搞笑，教育，公众号文章这四个模块。用户可以在本网站获得尽可能多的所需的资讯。

本设计理念来源于 58 同城网站。在 58 同城上，我们可以体验到找工作，找房子，买卖二手车，买卖二手家电，搬家等服务。让用户的需求一应俱全。

本 app 在后续的更新规划中，也会不断引入新的模块，提高用户粘度。本 app 致力为用户提供新闻，娱乐，科普，教育为一体的文章在线阅读服务，让用户的体验更多元化。同时，在阅读模式上，我们紧扣用户的喜好，为用户推荐他最可能感兴趣的资讯，为用户过滤不感兴趣的文章。

## 第三章 功能规划

（本章节主要描述作品根据需求分析所规划设计的各种功能，着重体现这些功能的作用）

### 三.1 功能介绍

App 移动端主要分为 4 个网页-----主页，文章页，登录/注册，个人中心。

主页：主页在游客第一次访问的时候置顶的是最新的时事热点，往下拉到底的时候会动态加载出新

的文章内容。还有实现了导航栏的功能，点击不同的分类时，主页会自动切到相应的类别文章。登录之后，主页的文章会根据你的观看记录来给你个性化推荐。如果没有观看过文章时，就会自动推荐时事热点。

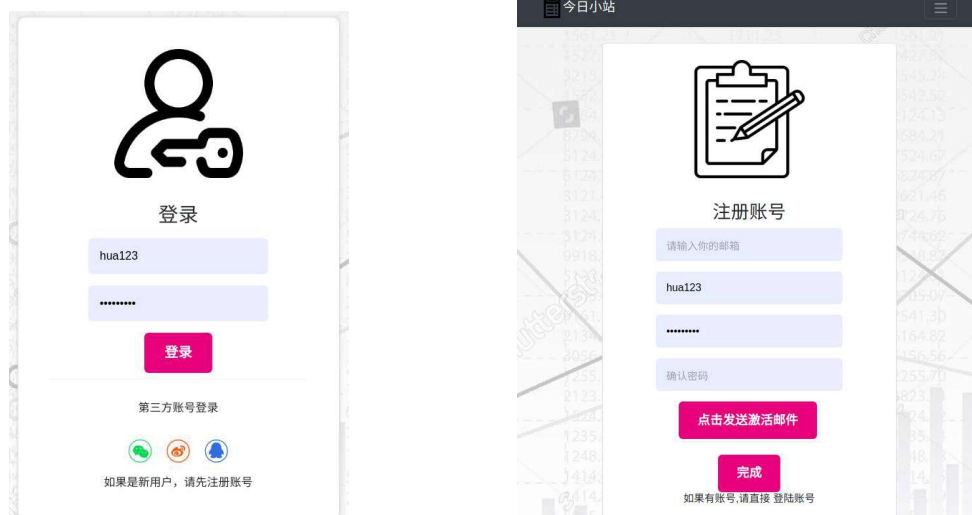
文章页：文章页就是文章的主体内容，底部的就是跟这篇文章的类型一致的 6 篇文章，也是推荐的一部分。

登录/注册的页面：都实现了前端的账号密码输入的检测功能（后端还会过滤），包括输入的长度，输入的字符的类型是否符合（正则表达式）等等。此外注册的功能还实现了邮箱注册的功能，可以使用各种邮箱来注册账号。

个人中心：里面有历史记录的查看，个人信息的完善，以及头像上传等功能。(待完善)

爬虫：每天爬取文章，负责每天更新文章。为 web 网站和推荐系统提供数据来源。

推荐系统：计算文章相似度，实时更新用户推荐列表，基于用户的协同过滤计算



## 第四章 软件架构和开发环境

（本章节主要描述软件架构设计，包括各功能模块子系统、服务端和客户端、数据库、硬件的嵌入

式开发、移动互联应用开发和云计算平台等的应用，并描述开发环境如.Net、Java、Android 等)

#### 四.1 软件架构和开发环境介绍

(爬虫和安卓 app)

##### 1. App 客户端端:

- 1) 主要开发语言:java
- 2) 开发工具: android studio

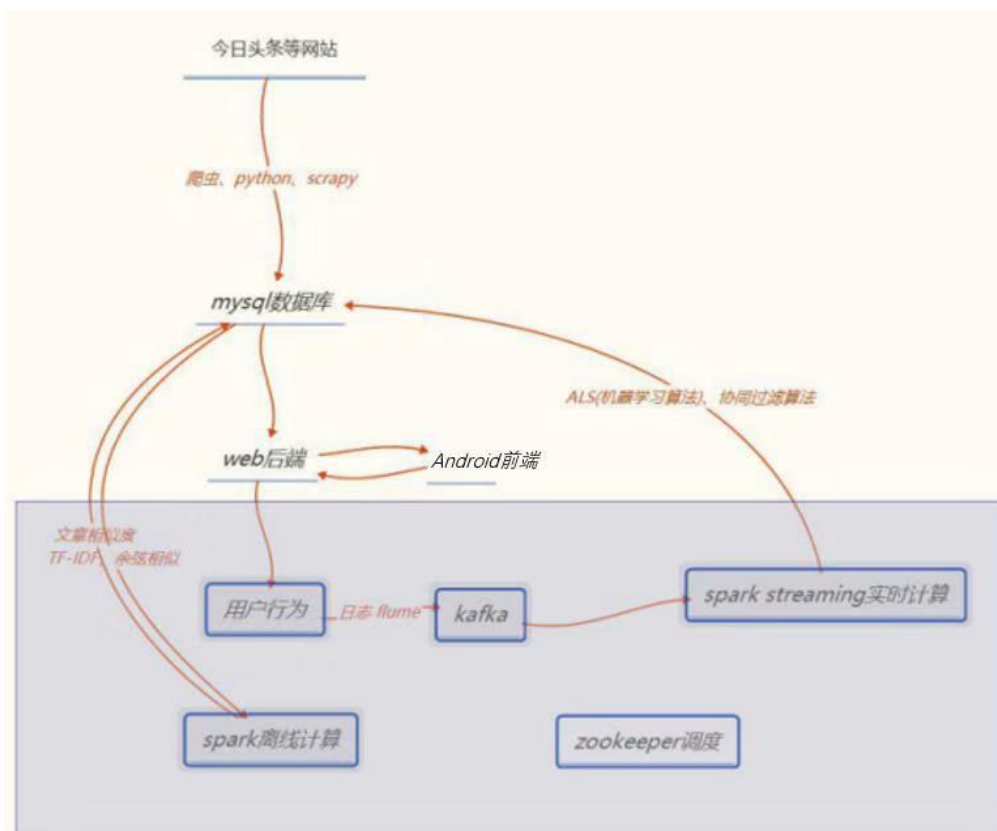
##### 2.后端:

- 1)主要开发语言: java, python
- 2) 软件架构: springboot+mysql+maven, scrapy
- 3)开发工具: IntelliJ idea, atom

推荐系统

- 1) 主要开发语言: scala, shell
- 2)软件架构: hadoop, spark, flume, kafka, hive, zookeeper
- 3)开发工具: intellij idea

实现流程:



---

## 第五章 其他

(本章节主要描述前文未涉及的内容, 如作品的工业设计、创意设计、作品的成本计算等)

(1) 作品推广及运营模式:

推广:

- ①通过向同学, 微信好友发送文章链接, 在页面提示安装 app
- ②通过微信群, 微信公众号进行宣传. 在校内派发宣传单
- ③通过微博, b 站等热门社交媒体进行宣传

运营:

- ①系统中的资讯均来自网络, 暂不需要人工运营。后续开发, 维护工作均开发组成员承担。
- ②运维方面: 编写 shell 脚本, 设置定时任务运行 shell 脚本, 实现集群进程的启动和停止等。降低人力成本。

(2) 开发日程:

- ①10 月 16 号前完成项目开发方案及爬虫, 文章推荐, web 网站的开发
- ②10 月 30 号前完成安卓的 UI 设计
- ③12 月的 16 号前完成安卓 app 客户端的代码编写, 完成所有模块的功能测试并上线
- ④12 月的 16 号后对产品进行推广, 更新

(3) 成本:

- ①费用主要来自于服务器的租赁费用
- ②集群的规模视用户量而定

## 参考文献

(1)springboot 官网;

<https://spring.io/projects/spring-boot>

(2)log4j 官网

<https://logging.apache.org/log4j/2.x/>

(3)hadoop 官网

<https://hadoop.apache.org/>

(4)spark 官网

<https://spark.apache.org/>

(5)kafka 官网

<https://kafka.apache.org/>

(6)flume 官网

<https://flume.apache.org/>

(7)zookeeper 官网

<https://zookeeper.apache.org/>

---

(8)hive 官网

<https://hive.apache.org/>

(9) <<spark 编程基础>>

(10) 安卓开发者官网

<https://developer.android.google.cn/>