# 豆瓣内容偏好分析项目

2019/10/23

## 项目概述

基于豆瓣网站内容分析当前热门电影、音乐、书籍的类型和用户偏好。通过该项目学习数据抓取和用户分析。

## 技术标准和规范要求

### 数据工程师

爬取豆瓣网站的分类：电影，音乐，读书

**电影信息：** 导演，编剧，主演，类型，制片国家/地区，语言: 英语，上映日期，片长，又名: 双子任务，IMDb链接，top 10 的影评

**音乐信息：**主要的基本信息和 top10短评 乐评信息

**书籍信息：**主要的基本信息和 top10短评 书评信息

**存储：**mysql 设计存储相关库和表结构

## JAVA工程师、算法工程师

1. 查询出当前一周评论最多的50个亚洲区域视频降序排列
2. 最火的读书类目是哪个，将这个类别的评论最多的10本书籍降序排行（要求：封面图、书籍名、作者、简介）
3. 做一个音乐评论热度榜（前50），且有导出excel功能,（导出字段：音乐名、演奏者、流派，发行时间）
4. 近30日最受女性欢迎的书籍分类有哪些，请给出相应的算法模型
5. 要求使用springBoot架构，通过API接口和前端工程师交互

### UI工程师、前端工程师

1. 设计一个榜单页面，要求展示以上三个排行
2. 页面要支持移动端适配
3. 音乐评论榜有可下载按钮，下载excel文件
4. 每个榜单前三名要求有金银铜三个徽章提示
5. 前端工程师要求使用VUE前端架构和ES6,可以使用jquery

### 测试工程师

根据项目要求，编写测试用例，保证项目在规定期限完美交付

## 项目节点

11.1-11.7 11.8-11.22 11.23 11.25

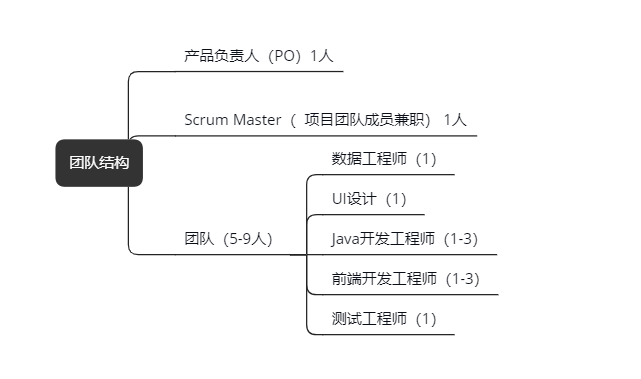
|  |  |  |  |
| --- | --- | --- | --- |
| 迭代一（11.7） | 迭代二（11.22） | 项目验收（11.23） | 最终验收（11.25） |
| 1. 表设计 2. 基础数据 3. 页面原型和设计稿 4. 前后端框架演示 5. 项目组成员和负责人 6. 迭代总结 | 1. 榜单数据展示 2. 算法公式 3. 文件可下载 4. 迭代总结 | 1. 项目演示 2. 答辩PPT | 项目总结文档 |
|  |  |  |  |

## **四、验收标准**：

1. 前端页面符合W3C标准，支持移动端适配
2. 正常查询电影、书籍、音乐三个排行表单
3. 音乐榜单可以下载excel文件
4. 前端页面榜单要求展示缩略图
5. 给出相应算法公式
6. 项目答辩PPT和功能演

## 五、其他

###### scrum项目团队



###### 未来可期

项目二、基于方舟argo实现网站的千人千面

如何像淘宝一样根据用户偏好，智能展示用户所偏爱的内容，让以前的人找物，改为物找人？方舟argo是一款免费的智能用户行为分析工具，我们基于它强大的分析能力，通过人群偏好，去智能改变网站内容和展现方式。

<https://argo.analysys.cn/>

项目三、视频智能内容提取

大家都会有这个痛点，录好了视频，但是相应的字幕要通过录音或者笔记来实现，这是很浪费时间和人力的工作，如何通过视频自动提取相应的字幕，实现字幕一键导出。

##### 附件一：项目成员表

## 项目成员表

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 一、项目基本情况 | | | | | | | |
| 项目组名称 | | 在云影湖偷钓锦鲤 | | 项目组编号 | | 1 | |
| ScrumMaster | |  | | 制作时间 | | 2019.11.1 | |
| 二、项目组成员 | | | | | | | |
| 成员姓名 | 项目角色 | | 专业 | 开始日期 | 结束日期 | | 投入频度 |
| 陈双 | 数据工程师 | | 计算机科学与技术 | 2019.11.1 | 2019.11.25 | |  |
| 蔡坤学 | 后台java工程师 | | 计算机科学与技术 | 2019.11.1 | 2019.11.25 | |  |
| 李洲 | 后台Java工程师 | | 计算机科学与技术 | 2019.11.1 | 2019.11.25 | |  |
| 罗振炼 | 数据工程师 | | 网络工程 | 2019.11.1 | 2019.11.25 | |  |
| 张坤 | 前端工程师 | | 软件工程 | 2019.11.1 | 2019.11.25 | |  |
| 覃樟 | 数据工程师 | | 计算机科学与技术 | 2019.11.1 | 2019.11.25 | |  |
| 刘洪冰 | 数据工程师 | | 软件工程 | 2019.11.1 | 2019.11.25 | |  |
|  |  | |  |  |  | |  |
|  |  | |  |  |  | |  |

###### 附件二：项目总结报告

豆瓣内容偏好分析项目总结报告

目 录

1 项目基本信息

2 项目完成情况

2.1 目标完成情况

项目目标：

实际完成情况：

2.2 任务完成情况

//如给甲方提供的材料中体现了这部分内容，可直接引用，不必复制

3 项目实施总结

3.1 项目工作量说明

//相对于立项时的标准工时规划，描述工作量变化情况。可描述，可用表格形式表示。

3.2 项目进度说明

//项目进度完成情况（延期、提前）及原因说明，可总体说明，可分里程碑说明。

3.3 项目风险及解决

//描述项目识别的风险，采取措施及成效

3.4 客户满意情况说明

//客户满意度情况说明，包括投诉与表扬

3.5 实施情况总结本项目特有的难题、解决方式。

//对项目管理水平、需求控制、风险控制、干系人关系管理等工作进行总结。针对项目组。

4 项目成果总结

//针对项目成果有价值部分的说明

5 经验与教训

5.1 项目成功的经验

//项目成功做法、优秀实践等，用于分享

5.2 项目失败的教训

//项目惨痛的教训、有改进的建议

6 问题与建议

6.1 项目维护推广建议

//由项目经理做出后续维护工作安排，包括维护工作类别（上门服务、bug修改等）、周期、维护阶段负责人、预计工时。

6.2 对本项目过程的改进建议

6.3 项目遗留问题及处理方案