

《票票小程序的开发与实践》

项目开发文档

|  |  |
| --- | --- |
| 编 号： | 201813641033 |
| 立项级别： | 自治区级 |
| 项目类别： | 创新训练项目 |
| 项目负责人： | 黎波 |
| 所在系部： | 理工学院 |
| 联系电话： | 18577376354 |
| 指导老师： | 秦鹏、杨嫘 |
| 立项时间： | 2018年6月 |
| 完成时间： | 2019年6月 |

目录

[1. 可行性分析 4](#_Toc23565)

[1.1项目概述 4](#_Toc28915)

[1.2所建议系统的说明 5](#_Toc6644)

[1.2.1对所建议系统的说明 5](#_Toc7619)

[1.2.2数据流程和处理流程 5](#_Toc13330)

[1.3可行性研究说明 5](#_Toc29029)

[1.3.1技术可行性 5](#_Toc18522)

[1.3.2使用Java语言爬取与使用python语言爬取的方案对比 6](#_Toc10141)

[1.3.3项目涉及问题的解决方案 7](#_Toc27294)

[1.3.4.经济可行性 7](#_Toc6058)

[1.4法律可行性 8](#_Toc9483)

[1.5用户可行性 8](#_Toc6983)

[2. 需求分析 9](#_Toc2475)

[2.1需求概述 9](#_Toc1701)

[2.1.1目标 9](#_Toc17649)

[2.1.2运行环境 9](#_Toc338)

[2.1.3用户特点 9](#_Toc24976)

[2.1.4关键点 9](#_Toc20258)

[2.1.5约束条件 10](#_Toc12340)

[2.2需求描述 10](#_Toc20888)

[2.2.1功能需求 10](#_Toc6611)

[2.2.2性能需求 10](#_Toc3423)

[2.2.3输入输出要求 11](#_Toc6574)

[2.2.4数据管理能力要求 11](#_Toc7385)

[2.2.5环境需求 12](#_Toc17471)

[2.2.6可靠性要求 12](#_Toc32715)

[2.2.7安全保密要求 12](#_Toc10912)

[2.2.8用户界面要求 12](#_Toc31142)

[2.2.9资源使用需求 12](#_Toc9782)

[2.2.10软件成本消耗与开发进度需求 13](#_Toc31501)

[2.3尚未解决的问题 13](#_Toc13987)

[3. 概要设计 14](#_Toc14462)

[3.1 结构设计概述 14](#_Toc11336)

[3.1.1设计任务 14](#_Toc24392)

[3.1.2设计原则 14](#_Toc27617)

[3.1.3设计策略 15](#_Toc9016)

[3.2体系结构设计 15](#_Toc15364)

[3.2.1数据流设计 16](#_Toc29342)

[3.2.2主要接口设计 17](#_Toc21380)

[4.详细设计 18](#_Toc19360)

[4.1界面设计 18](#_Toc22541)

[4.2数据库详细设计 19](#_Toc17651)

[4.2.1数据库名 19](#_Toc13935)

[4.2.2关系模式 19](#_Toc11621)

[4.2.3数据库各表字段描述 21](#_Toc19057)

[4.2.4关键算法 28](#_Toc243)

[5. 测试报告 30](#_Toc14892)

[5.1 测试概述 30](#_Toc10035)

[5.2 详细的测试结果 30](#_Toc18743)

[5.2.1 测试方法说明 30](#_Toc20054)

[5.2.2 详细测试结果 30](#_Toc18982)

[5.2.2.1路径测试 30](#_Toc3868)

[5.2.2.2功能测试 31](#_Toc10594)

[6. 安装及使用 35](#_Toc8473)

[6.1 安装环境 35](#_Toc26172)

[6.2 安装过程 35](#_Toc29487)

[7. 项目总结 37](#_Toc21426)

[8. 致谢 38](#_Toc1734)

《票票小程序的开发与实践》

项目组成员：梁钰业、黎波、朱名康、罗鸿妮、杨飞燕

指导老师：杨嫘、秦鹏

【内容摘要】高速发展的社会节奏，电影的快速上映最能体现。于是电影网站也就随之出现，其中最具代表性的电影网站有：豆瓣电影，猫眼电影，淘票票电影，Mtime时光电影网站。每一部电影在电影网站中都会得到客观的评分，评分的高低会很大因数的影响观影人选择该电影与否，但由于不同的电影网站给出的评分大相径庭，所有观影人在选择电影时至少需要不少于两次的点击浏览，这样才能得到想观赏的电影评分，以此来做出对比后得到一个更客观的评分，为选择电影提供一个更客观的参考因数。本项目为便于观影人更高效便捷的得到电影评分以及电影相关信息而开发，主要使用爬虫来爬取数据后使用微信小程序来显示出来，数据库采用Mysql，本文详解了其开发过程

【关键词】 爬虫；电影；微信小程序；Mysql

# 1. 可行性分析

## 1.1项目概述

本可行性分析适用于《票票小程序的开发与实践》项目。《票票小程序的开发与实践》项目是基于网络爬虫实现抓取数据，使用微信小程序呈现数据的综合项目，本项目爬虫实现了动态与多线程爬取豆瓣电影网站、猫眼电影网站、淘票票电影网站以及Mtime时光电影网站上的电影信息。在信息爬取结束后对电影评分进行加权综合得出综合评分--票票分，使用户无需分别浏览多个网站来查看电影评分来选择电影，只需查看微信小程序端的票票分就可以了解一部电影评分高低以作为选择的参考因数，同时还能查看电影评论和剧情简介，大致了解电影的剧情。票票小程序还提供搜索功能，可搜索想看的电影的相关信息，也可点击热门搜索中的一项，来查看电影信息。

项目支持方：自治区政府

开发方：《票票小程序的开发与实践》大创团队

项目用户：热爱看电影的全年龄人群

支持机构：广西师范大学漓江学院

## 1.2所建议系统的说明

### 1.2.1对所建议系统的说明

《票票小程序的开发与实践》主要是通过使用网络爬虫技术来实现其爬取功能的，网络爬虫技术是这个项目的关键点，项目使用Java后端+微信小程序前端+Mysql数据库来开发，在理论上只要实现了对网站的数据爬取并存入数据库中，这项目就已经完成了一半。开发本项目需要学习网络爬虫的相关知识，学会如何稳定高效爬取网站上的电影数据且不被封禁，其次需要了解Java相关开发知识，将爬取的数据做成接口给微信小程序前端使用，还需要掌握微信小程序开发相关知识，知道如何调用接口数据、页面设计、相关组件代码实现，了解SQL语言，了解数据库增删改查操作，知道如何建表、主键的使用，最好需要懂得服务器相关知识，懂得将java项目部署到服务器中，布置定时器来定时爬取网站上的数据，并在服务器部署Mysql数据来实现数据存储功能。

### 1.2.2数据流程和处理流程

数据流程以及处理流程：

（1）网络爬虫爬取电影数据电影网站数据

（2）Java后端获取数据库数据，处理数据使之成为Json格式，并把数据做成接口

（3）微信小程序前端读取接口中的Json格式数据并显示到页面中

## 1.3可行性研究说明

### 1.3.1技术可行性

（1）项目技术实现所涉及的技术问题：

①如何规避电影网站的防扒机制

②如何设计开发多线程的爬取方法

③如何清洗，组装成为json数据

④如何是实现小程序端的数据渲染

⑤用户使用查询功能时小程序如何反应更迅速

⑥当使用的jar包冲突时应该如何解决

⑦如何将项目部署至服务器上

1. 项目成员与具体开发环境要求
2. 项目成员

我们项目团队现有5人，我们成员为计算机科学与技术专业与信息管理与信息系统专业的学生，项目成员一起配合工作，一年时间足够时间去开发一个项目。其中黎波负责爬虫相关技术，罗鸿妮负责使用微信小程序开发工具来实现小程序前端数据呈现，梁钰业负责整体的项目部署，朱名康负责文档的编写，杨飞燕负责项目的经费管理，分工明确且大家在项目的相关技术领域获得过相应的奖项。具备开发项目能力。设备上，每个人都有一台笔记本电脑，且性能不低，每个人都可独立开发。一年项目时间，可边学边做，提高技术并完成任务，完全可以胜任项目需求。

②具体开发环境要求

设备要求：可流畅使用eclipse开发软件window系统电脑与微信开发者平台的Windows系统电脑。

软件要求：eclipse、微信开发者工具、navicat、FinalShell、阿里云。

开发要求：需要云服务器来部署项目。

环境要求：电脑安装JDK1.7版本以上、服务器。

经费要求：6000元经费。

### 1.3.2使用Java语言爬取与使用python语言爬取的方案对比

|  |  |  |
| --- | --- | --- |
|  | Java | Python |
| 页面下载能力 | 强大 | 一般 |
| Html分析工具 | 强大 | 一般 |
| 工程开发级别 | 大 | 中 |

二者相比之下，Java语言在页面下载能力、Html分析工具、工程开发级别三方面均优于python，所以采取Java语言开发爬虫方案。

### 1.3.3项目涉及问题的解决方案

|  |  |
| --- | --- |
| ①如何规避电影网站的防扒机制 | 构建ip池子与获取相关网页的实时cookie来模拟用户打开电影网站 |
| ②如何设计开发多线程的爬取方法 | 使用ExecatorEservice来实现爬虫的异步异步线程 |
| ③如何清洗，组装成为json数据 | 使用List<Map<String,String>> 数据类型来实现数据的清洗 |
| ④如何是实现小程序端的数据渲染 | 后台将数据以接口的形式传出，微信小程序端使用wx.requst()来获取后台接口数据 |
| ⑤用户使用查询功能时小程序如何反应更迅速 | 将原表数据穿入查询接口，而不是重新查询数据库中表的数据 |
| ⑥当使用的jar包冲突时应该如何解决 | 降低jar包本版 |
| ⑦如何将项目部署至服务器上 | 将项目以jar包的形式导出 |

### 1.3.4.经济可行性

（1）投资

成员调研费：2500元

项目设备：2000元

项目资料购买：500元

技术支持：1000元

（2）预期的经济效益

因《票票小程序开发与实践》项目为大创项目，未发布上市，故经济效益为0。

（3）市场预测

随着生活水平的不断提高，生活质量的提升，人们更注重精神享受，追求生活品质，观影人数会不断增加，而观影人数增加后，需要票票小程序这个集综合电影评分、电影信息为一体的小程序也会不断受到关注，潜在的用户数量很大。小程序前两年刚起步，目前还没有同类的小程序出现，可以先行抢占市场份额，竞争很小。且在日后的开发迭代中会使用分布式的开发平台将小程序开发成为一个电影评分的搜索引擎，使之更简便，更快捷，更有效率，更能抓住用户的心。

## 1.4法律可行性

因《票票小程序的开发与实践》为大创项目，且爬取评价并不受到企业禁止，也符合法律要求。但可能过于频繁爬取数据而导致被反爬虫封禁。

## 1.5用户可行性

此项目主要提供有观影需求的人群，人群并不受限制，有青少年、青年、壮年等等人群，他们在空闲时间会使用网络或前往电影院观影，故用户完全有可能使用本小程序来进行选择电影观看。

# 2. 需求分析

## 2.1需求概述

### 2.1.1目标

高速发展的社会节奏，电影的快速上映最能体现。于是电影网站也就随之出现，其中最具代表性的电影网站有：豆瓣电影，猫眼电影，淘票票电影，Mtime时光电影网站。每一部电影在电影网站中都会得到客观的评分，评分的高低会很大因数的影响观影人选择该电影与否，但由于不同的电影网站给出的评分大相径庭，所有观影人在选择电影时至少需要不少于两次的点击浏览，这样才能得到想观赏的电影评分，以此来做出对比后得到一个更客观的评分，为选择电影提供一个更客观的参考因数。本项目为便于观影人更高效便捷的得到电影评分以及电影相关信息而开发。

### 2.1.2运行环境

安装微信APP并能运行微信小程序的智能设备。

### 2.1.3用户特点

（1）用户类型：青少年、青年、中年。

（2）用户特点：热衷观看电影且对电影优劣有所选择的观影用户。

### 2.1.4关键点

（1）关键技术：使用网络爬虫技术快速安全地爬取各大平台数据。

（2）关键算法：如何大规模的抓取数据当url挤爆内存的时候使用的数据库来装载url，每一个网站的cookie的时效性，所有当抓取的时候实现实时cookie的获取与加载去实现爬虫的遍历每一个webmagic爬虫去抓取页面时都得到200个ip的ip池的代理更好的伪装。

关键功能：获取爬取到的数据显示在微信小程序页面供用户浏览。

### 2.1.5约束条件

（1）经费限制：不高于6000元。

（2）开发期限：2018年6月至2019年5月。

（3）采用的技术：微信小程序前端开发、网络爬虫技术、JavaWeb开发技术。

## 2.2需求描述

### 2.2.1功能需求

（1）可以自动爬取淘票票电影网站、豆瓣电影网站、猫眼电影网站、Mtime时光电影网站最近上映的电影评分、电影名字、电影简介、电影评论等，将电影评分综合加权得出票票分，信息集中显示在页面上供用户浏览查看。

（2）推荐功能，将爬取的电影信息推荐到推荐页面上，为用户推荐信息。

（3）搜索和历史记录功能，可以搜索到用户需要查询的电影，并显示在页面上，用户搜索过后，可显示历史记录。

（4）选择排序功能，用户可以选择根据时间排序或者根据评分排序。

### 2.2.2性能需求

#### 2.2.2.1.精度需求

（1）输入数据精度：无输入精度要求。

（2）输出数据精度：电影评分要求保留小数点后一位。

（3）传输过程的精度：无要求。

#### 2.2.2.2.时间特性要求

（1）爬取数据时间要求：爬取数据速度要快，爬取一部电影数据不得高于15s。

（2）处理时间要求：搜索处理时间不得高于3s。

（3）计算时间要求：票票分计算时间不得高于1s。

#### 2.2.2.3灵活性要求

当用户使用不同分辨率智能设备时，能正常显示小程序页面。

### 2.2.3输入输出要求

（1）搜索数据输入：

数据格式：String字符型

数据数值范围：无

数据精度：无

（2）搜索数据输出：

数据格式：Json格式

数据数值范围：无

数据精度：无

（3）正常搜索结果输出：输出所搜索电影的电影数据。

（4）异常输出：控制台输出报错。

### 2.2.4数据管理能力要求

无需要管理的问卷和记录个数要求。

### 2.2.5环境需求

|  |  |
| --- | --- |
| 硬件要求 | （1）机型要求：可运行微信小程序的智能手机。  （2）外部设备：无要求。  （3）数据通信接口：无要求。 |
| 软件要求 | （1）操作系统：可运行微信小程序的手机系统，如IOS、Android等。  （2）数据库管理系统：无要求。  （3）网络软件：无要求。 |
| 技术要求 | 掌握使用微信打开微信小程序方法。 |
| 网络要求 | 至少2G网络。 |

### 

### 2.2.6可靠性要求

较低的可靠性要求。

### 2.2.7安全保密要求

无特殊安全保密要求。

### 2.2.8用户界面要求

用户界面要求简洁大方，使用方便，模块排版符合人类正常审美，字体不要过于浮夸，图片清晰不失真。

### 2.2.9资源使用需求

（1）开发所需资金不超过6000元。

（2）软件运行时不耗费过多手机资源，不给手机带来严重负担。

### 2.2.10软件成本消耗与开发进度需求

|  |  |
| --- | --- |
| 软件开发进度要求 | 开发费用要求 |
| 前期开发（2018年6月-2018年8月） | 不高于2000元 |
| 中期开发（2018年9月-2018年10月） | 不高于2000元 |
| 后期开发（2018年11月-2019年5月） | 不高于2000元 |

## 2.3尚未解决的问题

（1）未实现登录注册功能。

（2）用户没办法收藏想看的电影。

# 3. 概要设计

## 3.1 结构设计概述

### 3.1.1设计任务

（1）将票票小程序需求分析转变为一种软件结构.

（2）建立票票小程序的逻辑模型。

（3）确保票票小程序的所有需求都被分配给软件部分，进行下一步细化以便于后面的详细设计。

### 3.1.2设计原则

（1）总体原则：由粗到细的原则，互相结合的原则，定性分析和定量分析相结合的方法，分解和协调的方法和模型化方法；

（2）要系统考虑系统的一般性、关联性、整体性和层次性；

（3）分解协调：目的是为了创造更好的系统。系统分解是指将一个复杂的系统分解为若干个子系统，系统协调一是系统内协调，即根据系统的总结构、总功能、总任务和总目标的要求，使各个子系统之间互相协调配合，在各个子系统局部优化基础上，通过内部平衡的协调控制，实现系统的整体优化；

（4）屏蔽抽象：从简单的框架开始，隐含细节；

（5）一致性：统一的规范、统一的标准、统一的文件模式；

（6）每个模块应当有一个统一命名的容易理解的名字；

（7）编码：由外向内（界面－>核心）；

（8）面向用户：概要设计是对于按钮按下后系统“怎么做”的简要说明；

（9）模块、组件的充分独立性、封闭性；

（10）同时考虑静态结构与动态运行；

（11）每个逻辑对象都应当说明其所处物理对象（非一一对应）；

（12）每个物理对象都有合适的开发人员，并且利于分工与组装。

### 3.1.3设计策略

（1）降低模块的接口复杂度和冗余度，提高协调性。

（2）票票小程序皆为单入口，单出口的模块。

（3）使票票小程序任意模块的作用域在其控制域内。

（4）减少扇出，追求高扇入。

## 3.2体系结构设计

根据项目的需求，本作品设计的数据流图如图1所示，通过复审和精化数据流图，得到如图2所示的精化软件结构图。

### 3.2.1数据流设计

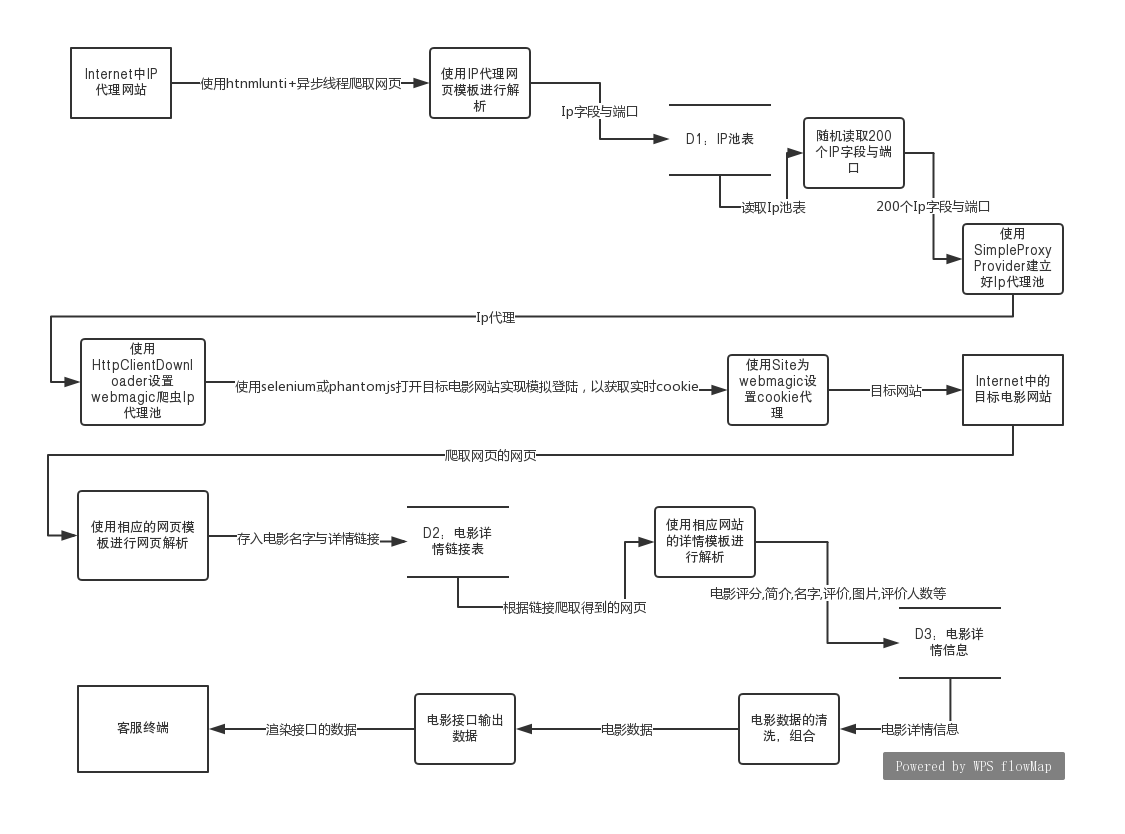


图1 作品数据流图

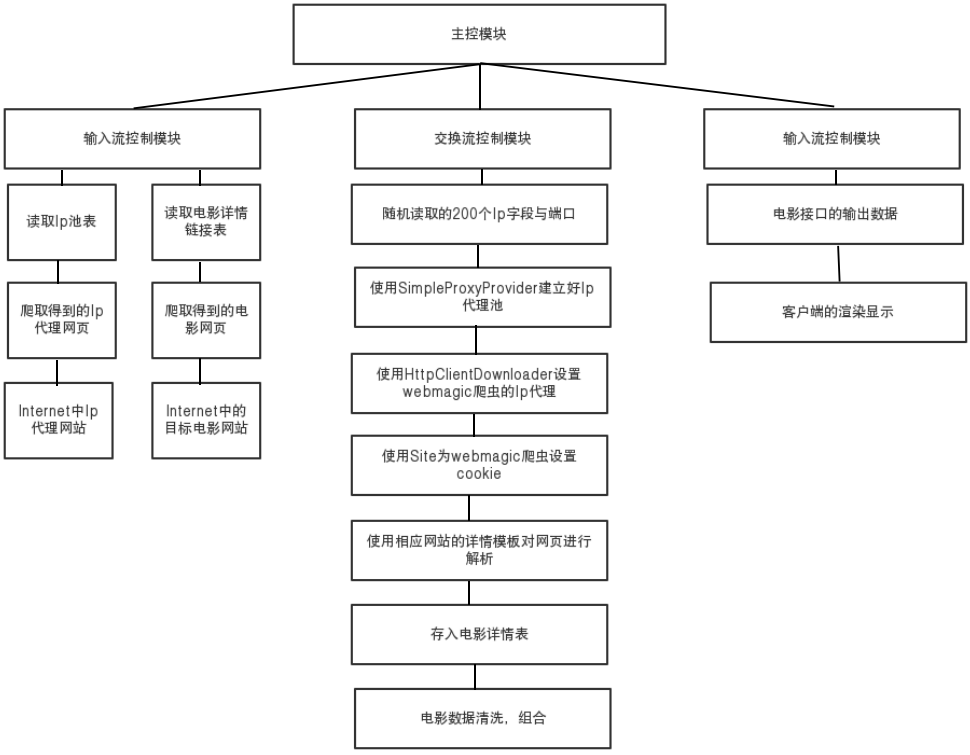


图2 精化软件结构图

### 3.2.2主要接口设计

（1）用户接口

①电影渲染接口：类型为List<Map<String,String>>类型，字段为电影名称，电影类型，电影年份，电影上映时间，电影简介，电影评论，各个网站的电影评分与评论人数。

②电影查询接口：类型为List<Map<String,String>>类型，需得到用户输入的电影名称，返回字段为电影名称，电影类型，电影年份，电影上映时间，电影简介，电影评论，各个网站的电影评分与评论人数。

（2）外部接口

① 项目爬虫与数据库接口：存在三种类型的接口，如下，

第一：电影详情链接接口为Movie\_Link 类型其字段为电影名称与电影详情链接。

第二：电影详情接口为Movie类型其字段 为电影名称，电影类型，电影年份，电影上映时间，电影简介，电影评论， 各个网站的电影评分与评论人数。

第三：Ip池接口为IP类型其字段为IP 字段与IP端口。

② 项目爬虫与Internet接口:电影页面接口类型为String类型的htmlPage其中字段为电影页面的DOM。

③ 项目爬虫与客户端接口:类型为List<Map<String,String>>类型，字段为电影名称，电影类型，电影年份，电影上映时间，电影简介，电影评论，各个网站的电影评分与评论人数。

（3）内部接口

① 随机的200个Ip字段与端口接口：类型为Proxy[]类型字段为Proxy。

② 实时Cookie代理接口：类型为Set<Cookie>类型，字段为Cookie。

③ 解析网页详情接口：类型为void类型，字段为htmlPage。

④ 解析网页详情链接接口：类型为void类型，字段为htmlPage。

⑤ 电影数据清洗组合接口:类型为List<Map<String,String>>类型，字段为电影名称，电影类型，电影年份，电影上映时间，电影简介，电影评论，各个网站的电影评分与评论人数。

# 4.详细设计

## 4.1界面设计

客户终端作品采用微信小程序来呈现数据，各个页面的设计如图3至图8所示。

图3 首页 图4 推荐页

图5 详情页 图6 搜索页

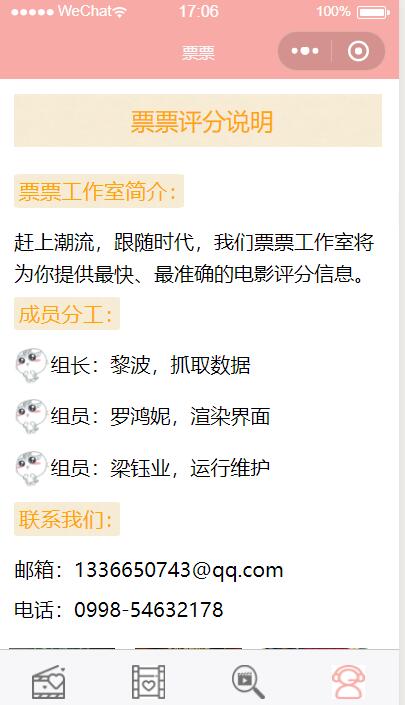
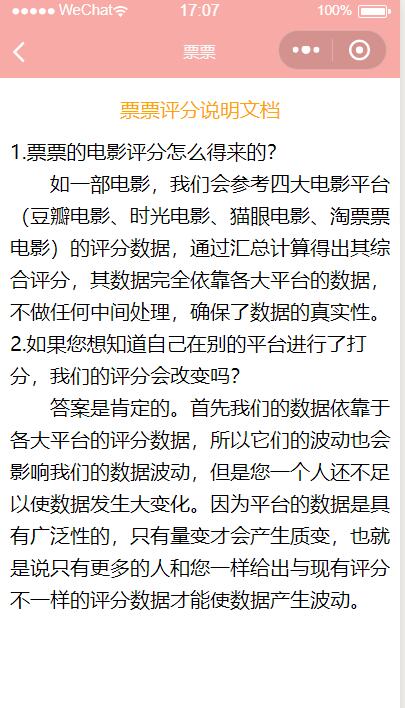
 

图7 关于页 图8 评分细则页

搜索页，如图6所示。

图6 搜索页

## 4.2数据库详细设计

### 4.2.1数据库名

数据库名称：Li。

### 4.2.2关系模式

（1）豆瓣新电影链接表（电影名字，电影链接）；

（2）豆瓣新电影详细信息表（电影名字，电影评分，电影评论人数，电影图片路径，电影网站路径，电影简介，电影评论，电影类型，电影上映年份，电影上映日期）；

（3）豆瓣旧电影链接表（电影名字，电影链接）；

（4）豆瓣旧电影详细信息表（电影名字，电影评分，电影评论人数，电影图片路径，电影网站路径，电影简介，电影评论，电影类型，电影上映年份，电影上映日期）；

（5）猫眼新电影链接表（电影名字，电影链接）；

（6）猫眼新电影详细信息表（电影名字，电影评分，电影评论人数，电影图片路径，电影网站路径，电影简介，电影评论，电影类型，电影上映年份，电影上映日期）；

（7）猫眼旧电影链接表（电影名字，电影链接）；

（8）猫眼旧电影详细信息表（电影名字，电影评分，电影评论人数，电影图片路径，电影网站路径，电影简介，电影评论，电影类型，电影上映年份，电影上映日期）；

（9）时光新电影链接表（电影名字，电影链接）；

（10）时光新电影详细信息表（电影名字，电影评分，电影评论人数，电影图片路径，电影网站路径，电影简介，电影评论，电影类型，电影上映年份，电影上映日期）；

（11）时光旧电影链接表（电影名字，电影链接）；

（12）时光旧电影详细信息表（电影名字，电影评分，电影评论人数，电影图片路径，电影网站路径，电影简介，电影评论，电影类型，电影上映年份，电影上映日期）；

（13）淘票票新电影链接表（电影名字，电影链接）；

（14）淘票票新电影详细信息表（电影名字，电影评分，电影评论人数，电影图片路径，电影网站路径，电影简介，电影评论，电影类型，电影上映年份，电影上映日期）；

（15）淘票票旧电影链接表（电影名字，电影链接）；

（16）淘票票旧电影详细信息表（电影名字，电影评分，电影评论人数，电影图片路径，电影网站路径，电影简介，电影评论，电影类型，电影上映年份，电影上映日期）；

（17）Ip池表（ip地址，端口）。

### 4.2.3数据库各表字段描述

作品采用关系型数据库存储数据，表结构如下表所示。

表1 douban\_movie\_link（豆瓣新电影链接表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| link | varchar | 255 | 是 | 否 |

表2 douban\_old\_movie\_link（豆瓣旧电影链接表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| link | varchar | 255 | 是 | 否 |

表3 maoyan\_movie\_link（猫眼新电影链接表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| link | varchar | 255 | 是 | 否 |

表4 maoyan\_old\_movie\_link（猫眼旧电影链接表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| link | varchar | 255 | 是 | 否 |

表5 mtime\_movie\_link（时光新电影链接表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| link | varchar | 255 | 是 | 否 |

表6 mtime\_old\_movie\_link（时光旧电影链接表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| link | varchar | 255 | 是 | 否 |

表7 taopiaopiao\_movie\_link（淘票票新电影链接表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| link | varchar | 255 | 是 | 否 |

表8 douban\_movie（豆瓣新电影详细信息表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| socre | varchar | 255 | 是 | 否 |
| count\_of\_man | varchar | 255 | 是 | 否 |
| img | varchar | 255 | 是 | 否 |
| url | varchar | 255 | 是 | 否 |
| simplictiy | text | 0 | 是 | 否 |
| content | text | 0 | 是 | 否 |
| class | varchar | 255 | 是 | 否 |
| year | varchar | 255 | 是 | 否 |
| p\_date | varchar | 255 | 是 | 否 |

表9 douban\_old\_movie（豆瓣旧电影详细信息表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| socre | varchar | 255 | 是 | 否 |
| count\_of\_man | varchar | 255 | 是 | 否 |
| img | varchar | 255 | 是 | 否 |
| url | varchar | 255 | 是 | 否 |
| simplictiy | text | 0 | 是 | 否 |
| content | text | 0 | 是 | 否 |
| class | varchar | 255 | 是 | 否 |
| year | varchar | 255 | 是 | 否 |
| p\_date | varchar | 255 | 是 | 否 |

表10 maoyan\_movie（猫眼新电影详细信息表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| socre | varchar | 255 | 是 | 否 |
| count\_of\_man | varchar | 255 | 是 | 否 |
| img | varchar | 255 | 是 | 否 |
| url | varchar | 255 | 是 | 否 |
| simplictiy | text | 0 | 是 | 否 |
| content | text | 0 | 是 | 否 |
| class | varchar | 255 | 是 | 否 |
| year | varchar | 255 | 是 | 否 |
| p\_date | varchar | 255 | 是 | 否 |

表11 maoyan\_old\_movie（猫眼旧电影详细信息表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| socre | varchar | 255 | 是 | 否 |
| count\_of\_man | varchar | 255 | 是 | 否 |
| img | varchar | 255 | 是 | 否 |
| url | varchar | 255 | 是 | 否 |
| simplictiy | text | 0 | 是 | 否 |
| content | text | 0 | 是 | 否 |
| class | varchar | 255 | 是 | 否 |
| year | varchar | 255 | 是 | 否 |
| p\_date | varchar | 255 | 是 | 否 |

表12 mtime\_movie（时光新电影详细信息表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| socre | varchar | 255 | 是 | 否 |
| count\_of\_man | varchar | 255 | 是 | 否 |
| img | varchar | 255 | 是 | 否 |
| url | varchar | 255 | 是 | 否 |
| simplictiy | text | 0 | 是 | 否 |
| content | text | 0 | 是 | 否 |
| class | varchar | 255 | 是 | 否 |
| year | varchar | 255 | 是 | 否 |
| p\_date | varchar | 255 | 是 | 否 |

表13 mtime\_old\_movie（时光旧电影详细信息表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| socre | varchar | 255 | 是 | 否 |
| count\_of\_man | varchar | 255 | 是 | 否 |
| img | varchar | 255 | 是 | 否 |
| url | varchar | 255 | 是 | 否 |
| simplictiy | text | 0 | 是 | 否 |
| content | text | 0 | 是 | 否 |
| class | varchar | 255 | 是 | 否 |
| year | varchar | 255 | 是 | 否 |
| p\_date | varchar | 255 | 是 | 否 |

表14 taopiaopiao\_movie（淘票票新电影详细信息表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| name | varchar | 255 | 否 | 是 |
| socre | varchar | 255 | 是 | 否 |
| count\_of\_man | varchar | 255 | 是 | 否 |
| img | varchar | 255 | 是 | 否 |
| url | varchar | 255 | 是 | 否 |
| simplictiy | text | 0 | 是 | 否 |
| content | text | 0 | 是 | 否 |
| class | varchar | 255 | 是 | 否 |
| year | varchar | 255 | 是 | 否 |
| p\_date | varchar | 255 | 是 | 否 |

表15 ip\_chi（IP池表）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 类型 | 长度 | 允许空值 | 是否主键 |
| ip | varchar | 255 | 否 | 是 |
| port | varchar | 255 | 是 | 否 |

### 4.2.4关键算法

此算法考虑的最主要的方向是如何大规模的抓取数据当url挤爆内存的时候使用的数据库来装载url，每一个网站的cookie的时效性，所有当抓取的时候实现实时cookie的获取与加载去实现爬虫的遍历每一个webmagic爬虫去抓取页面时都得到200个ip的ip池的代理更好的伪装。核心流程如图9所示。

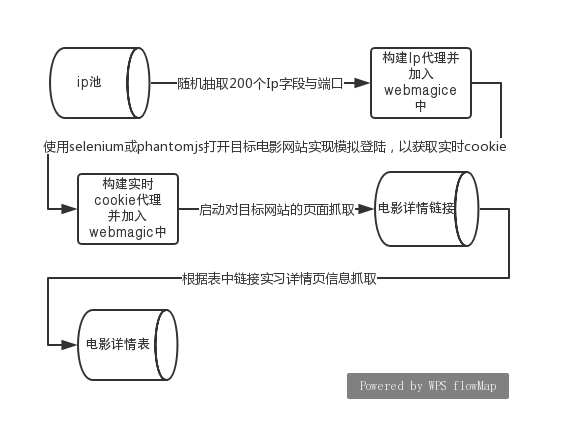


图9 算法核心流程图

爬取的规则如下，启动selenium或是phantomjs来实现模拟登陆（与实时点击）后来获取相关网站的实时cookie值并加载如爬虫的伪装头中，通过htmlunit框架加上异步线程抓取而来的Ip代理字段与端口并存入数据库中Ip\_chi表，从中随机抽取200个Ip构建Ip代理这样就可以使爬虫在每一次对网页进行爬取时拥有200个的Ip代理池，通过以上步骤就伪装好了爬虫。第一次爬虫启动时是为了遍历网站的目标页面中详情页的详情链接与电影名字并存入数据库，第二次启动爬虫是通过数据库中的详情链接和开启20~30的异步线程打开详情页链接抓取页面获取相关信息。

# 5. 测试报告

## 5.1 测试概述

本测试报告用于《票票小程序的开发与实践》项目记录测试过程，总结测试情况，分析测试数据，归纳测试工作进行过程中暴露的问题与遗留的风险，确保项目质量。

## 5.2 详细的测试结果

## 5.2.1 测试方法说明

作品主要采用结构化测试的路径测试方法和功能测试方法。路径测试方法的优点在于降低了测试用例设计的难度，只要搞清楚各个种流程，就可以设计出高质量的测试用例来，而不用太多的测试方面的经验；在测试时间较短的情况下，可有有点放矢选择测试用例，而不用完全根据经验取舍。功能测试主要是为了测试每个功能是否能够正常使用，完全不考虑程序内部结构和内部特性，只检查程序功能是否按照需求分析的规定正常使用。

## 5.2.2 详细测试结果

### 5.2.2.1路径测试

（1）在构建ip池的时候采用里200个异步线程对IP代理网站进行网页抓取，导致对方服务器返回503导致使没有完全抓取IP代理网站的ip，如图10所示。



图10 web返回503错误图

解决方法：设置合理的线程个数与采用合理的抓取次数。

（2）在抓取好的网页解析进行解析后没有将评论存入数据库产生的原因为数据库的content字段的编码格式不正确，无法插入含有表情包的评论。如图11所示。

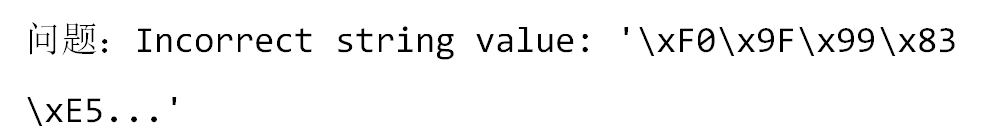


图11 字段编码格式错误图

解决方法：修改mysql的编码格式。

（3）电影查询接口的查询速度啊过慢，使用户体验感不好。

解决方法:将原List中的内容再度传入查询。

（4）当爬取的次数过多模拟登陆帐被暂时查封，导致抓取不完全，

解决方法：多申请，相应网站的账号用户当被查封时使用其他的账号登陆。

（5）多次的jar冲突，使项目在代码无法启动。

解决方法：尽量使用交旧的版本jar包。

（6）8080端口在启动时被占用。

解决方法：先查询出那个线程占用，如图12所示。

IMG_256

图12 线程占用图

在杀死其使其释放8080端口，如图13。

IMG_256

图13 杀死线程图

### 5.2.2.2功能测试

（1）首页模块

①测试用例

点击底部导航栏首页图标。

②测试结果

首页正常显示电影详细信息。

③测试分析。

首页显示电影功能正常。

（2）推荐模块

①测试用例

点击底部导航栏推荐图标。

②测试结果

推荐页正常显示推荐电影。

③测试分析。

正常显示推荐电影，推荐功能正常。

（3）搜索模块

①测试用例

|  |  |
| --- | --- |
| 序号 | 测试用例 |
| 1 | 大侦探皮卡丘 |
| 2 | 11111 |
| 3 | 反贪风暴4 |
| 4 | Asdaf |
| 5 | 复仇者联盟 |

②测试结果

|  |  |
| --- | --- |
| 测试用例 | 测试结果 |
| 大侦探皮卡丘 | 无搜索结果 |
| 11111 |
| 反贪风暴4 |
| Asdaf |
| 复仇者联盟 |

③测试分析

在搜索栏搜索电影不显示电影信息，搜索功能不能正常工作。

（4）历史记录模块

①测试用例

历史记录用例：

|  |  |
| --- | --- |
| 序号 | 测试用例 |
| 1 | 爱宠大机密 |
| 2 | Xxxx |
| 3 | 战狼 |
| 4 | Asdaf |
| 5 | 复仇者联盟 |

清空历史记录：

|  |  |
| --- | --- |
| 序号 | 测试用例 |
| 1 | 点击清空历史记录按钮 |

②测试结果

历史记录：

|  |  |
| --- | --- |
| 测试用例 | 测试结果 |
| 爱宠大机密 | 均显示在历史记录中 |
| Xxxx |
| 战狼 |
| Asdaf |
| 复仇者联盟 |

清空历史记录：

|  |  |
| --- | --- |
| 测试用例 | 测试结果 |
| 点击点击清空历史记录按钮 | 历史记录被清空 |

③测试分析

在搜索栏搜索过的电影名字都会显示在搜索历史记录中，历史记录模块无误。点击清空历史按钮后，历史记录消失，清空历史记录功能无误。

# 6. 安装及使用

## 6.1 安装环境

|  |  |
| --- | --- |
| 硬件要求 | （1）Cpu i3版本以上  （2）内存8g以上（包括8g） |
| 软件要求 | 1. MySQL5版本以上 2. jdk7版本以上 3. 最新版本的微信小程序开发工具 4. 使用与jdk版本的相对应的eclipse 5. 使用微信6.6.1以上的微信客户端 |

## 6.2 安装过程

（1）使用mysql8.0的数据库在配置文件中使用常规即可但编码格式需使用Utf8mb4。

1. 设置root用户其密码为124256，并新建数据库li再将database文件夹中数据库的表导入li中。
2. 使用eclipse导入d这个maven工程,为爬虫源码。
3. 使用eclipse导入mao这maven工程，为用户接口。
4. 使用微信小程序开发工具导入票票小程序项目。

使用过程：

爬取使用过程

（1）使用Douban\_Spider可以爬取豆瓣新电影链接。

（2）使用Douban\_Spider\_old可以爬取豆瓣旧电影链接。

（3）使用Douban\_son\_spider\_new\_movie可爬取豆瓣新电影详情。

（4）使用Douban\_son\_spider\_old\_movie可爬取豆瓣旧电影详情。

（5）使用Maoyan\_Spider可以爬取猫眼电影的新旧电影链接。

（6）使用Maoyan\_Son\_Spider\_new 可以爬取猫眼新电影的电影详情。

（7）使用Maoyan\_Son\_Spider可以爬取猫眼旧电影的电影详情。

（8）使用Mtime\_Spider可以爬取Mtime电影的新旧电影链接。

（9）使用Mtime\_Son\_Spider 可以爬取Mtime新电影的电影详情。

（10）使用Mtime\_Spider\_old可以爬取Mtime旧电影的电影详情。

（11）使用Taopiaopiao\_Spider可以爬取Mtime电影的新电影内容。

用户接口使用过程

使用App启动用户接口。

渲染界面使用过程

打开微信小程序开发工具就可看到数据呈现再客户终端的排列形式。

# 7. 项目总结

在项目的协调方面我们分工明确，黎波负责接口的编写与爬虫的实现，罗鸿妮负责前端的界面的编写与修改文档的工作，梁钰业负责服务器数据库的搭建与接口的部署，各司其职。

在下一个版本中或者今后的版本中，我们更倾向于使用分布式矩阵计算和非关系型数据库（key-value数据库），提升项目开发速度。保证更多的时间用于算法的优化。目标是做成一个关于电影的搜索引擎，为大家的生活代来便利。（推广关于评分）

在日后的发展中可将本作品向其他关于评分对比的事物中，例如美食评分对比，书籍评分对比，旅游景点对比中。

# 8. 致谢

一路开发下来遇到了很多困难而无力解决，特别感谢帮助我们团队耐心解决问题、辛苦奉献的老师们，感谢为我们解惑寻找Bug的同学，感谢相关参考文献作者的知识奉献。谢谢他们的付出，没有他们就没有今日的票票小程序。

**【参考文献】**

[1]罗刚．自己动手写网络爬虫(修订版)[M]．第一版．北京：清华大学出版社．2016

[2]罗刚．网络爬虫全解析技术、原理与实践[M]．第一版．北京：电子工业出版社．2017

1. 任永昌．软件工程[M]．第一版．北京：清华大学出版社．2012

**Ticket ticket small program development and implementation**

Project team members: liang yuye, li bo, zhu mingkang, luo hongni, Yang feiyan

Instructors: Yang lei, qin peng

[abstract]The rapid development of the social rhythm, the rapid release of the film can best reflect.Then the film website also appears subsequently, among them the most representative film website has: douban film, cat eye film, tao piaodian film, Mtime film website.Every movie in the movie website will get objective ratings, scoring the impact will be a big factor of moviegoers to choose the movie or not, but as a result of different film website ratings, given all the extent when choosing movies take at least two click browse, in this way can we get to watch movie score, in order to make a comparison after get a more objective grading, provides a more objective reference for choosing film factor.This project is developed for the convenience of moviegoers to get movie scores and movie-related information more efficiently and conveniently. It mainly USES crawler to crawl the data and then USES WeChat small program to display it. The database USES Mysql

[Key words]: crawler, movie, WeChat applet, Mysql

l