

电影数据可视化系统

软件需求规格说明书

名字	学号	贡献描述
石嘉晗	SZ160110227	起草初稿、引言、软件一般性描述、系统划分
周宇飞	SZ160110128	数据报表子系统、登录子系统、其他需求描述、整理、完善
谢华鹏	SZ160110232	可视化子系统、数据报表子系统
王冠雄	SZ160110217	登录子系统、数据报表子系统、软件原型

文档变更历史记录

序号	变更日期	变更人员	变更内容详情描述	变更后的版本号
1	2018. 11. 1	石嘉晗	起草初稿	0. 1
2	2018. 11. 5	石嘉晗	引言、软件一般性描述部分	0. 2
3	2018. 11. 9	石嘉晗	系统划分	0. 3
4	2018. 11. 20	谢华鹏	可视化子系统	0. 4
5	2018. 11. 20	石嘉晗	爬虫子系统	0. 5
6	2018. 11. 21	周宇飞 谢华鹏 王冠雄	数据报表子系统	0. 6
7	2018. 11. 22	周宇飞 王冠雄	登录子系统	0. 7
8	2018. 11. 23	周宇飞	其他需求描述	0. 8
9	2018. 11. 29	王冠雄	软件原型	0. 9
10	2018. 12. 6	周宇飞	整理、完善、排版	1. 0
11	2018. 12. 20	石嘉晗 王冠雄	第一次修改	1. 1
12	2018. 12. 27	王冠雄 谢华鹏 石嘉晗	第二次修改	2. 0

目录

1. 引言.....	5
1.1 编写目的.....	5
1.2 读者对象.....	5
1.3 软件项目概述.....	5
1.4 文档概述.....	5
1.5 定义.....	6
1.6 参考资料.....	6
2. 软件的一般性描述.....	7
2.1 软件产品与其环境之间的关系.....	7
2.2 限制与约束.....	7
2.3 假设与前提条件.....	7
3.功能需求描述.....	8
3.1 系统的划分.....	8
3.2 可视化子系统的功能.....	9
3.2.1 可视化系统的用例图.....	9
3.2.4 可视化功能选择的用例描述.....	9
3.3 爬虫子系统的功能.....	11
3.3.1 爬虫子系统的用例图.....	11
3.3.2 用户请求电影数据的用例描述.....	12
3.3.3 管理员修改爬虫的用例描述.....	12
3.3.4 管理员维护数据库的用例描述.....	13
3.4 数据报表子系统的功能.....	13
3.4.1 生成数据报表的用例图.....	13
3.4.2 生成数据报表用例的描述.....	14
3.5 登录子系统的功能.....	15
3.5.1 登录子系统的用例图.....	15
3.5.2 用户登录注册的用例描述.....	15
3.5.3 管理员维护信息的用例描述.....	16

4. 其它需求描述.....	17
4.1 性能要求.....	17
4.2 设计约束.....	17
4.3 界面要求.....	17
4.4 进度要求.....	17
4.5 交付要求.....	18
4.6 验收要求.....	18
5. 软件原型.....	19
6. 附录（分工）	25

1. 引言

1.1 编写目的

本文档是哈尔滨工业大学（深圳）2018 秋季软件工程课程实践项目“电影数据可视化分析系统”的软件需求规格说明书。

1.2 读者对象

本文档可能的读者对象包括该系统的开发人员，课程实践项目的评分人员以及本系统的使用者等。

1.3 软件项目概述

- 项目名称：电影数据可视化分析系统
- 用户单位：电影市场相关从业人员以及电影爱好者
- 开发单位：哈尔滨工业大学（深圳）2018 秋软件工程课程第 4 组
- 大致功能和用途：
 - 功能包括账户登入功能、电影数据爬取功能、数据可视化功能以及电影数据报表功能。
 - 用途：帮助从业人员更好地掌握市场趋势，理解市场需求；帮助电影爱好者找到需要的数据。

1.4 文档概述

本文档首先给出系统的整体结构以及功能结构概貌，从整体架构上给出整个系统的轮廓。同时对功能需求，性能需求进行描述，以便用户与开发人员之间理解和交流。在第二章，给出对软件的一般性描述；第三章是对功能需求描述，包括系统的划分，子系统的功能等；第四章描述了其他需求，包括性能需求，设

计约束，界面要求等等；第五章展示了界面原型；第六章为附录。

1.5 定义

- 1) 系统：特指“电影数据可视化分析系统”。
- 2) 数据来源：系统中所有票房数据均来自艺恩网，其他数据如导演、题材等来自豆瓣网。
- 3) 数据库：本系统使用 Mysql 作为底层数据库。

1.6 参考资料

- [1] ROGER S P. 软件工程实践者的研究方法[J]. 1999
- [2] 史济民，顾春华，郑红. 软件工程：原理，方法与应用[M]. Gao deng jiao yu chu ban she, 2014.
- [3] 阿洛，诺伊施塔特，贵宾，等. UML 和统一过程：实用面向对象的分析和设计[M]. 机械工业出版社，2003.

2. 软件的一般性描述

2.1 软件产品与其环境之间的关系

本软件可独立地运行在 web 端，需要用户安装 chromium 内核浏览器作为支持，不依赖系统

2.2 限制与约束

对软件产品的设计提出以下的限制和约束

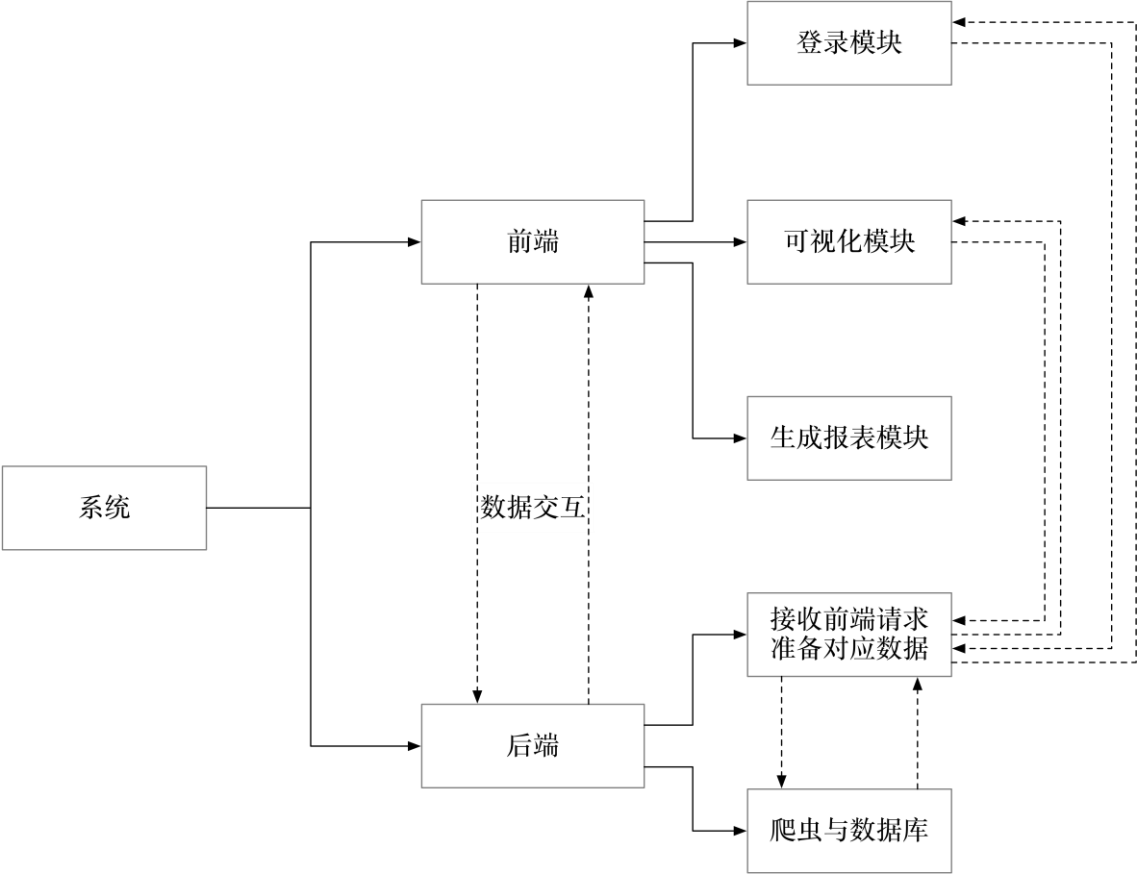
- 1) 界面友好，易于操作；
- 2) 结构灵活，易于维护；
- 3) 安全可靠。

2.3 假设与前提条件

- 1) 假设用户只查询 2015-2018 年之间的数据信息；
- 2) 假设用户未登录，也可以查询一部分信息。

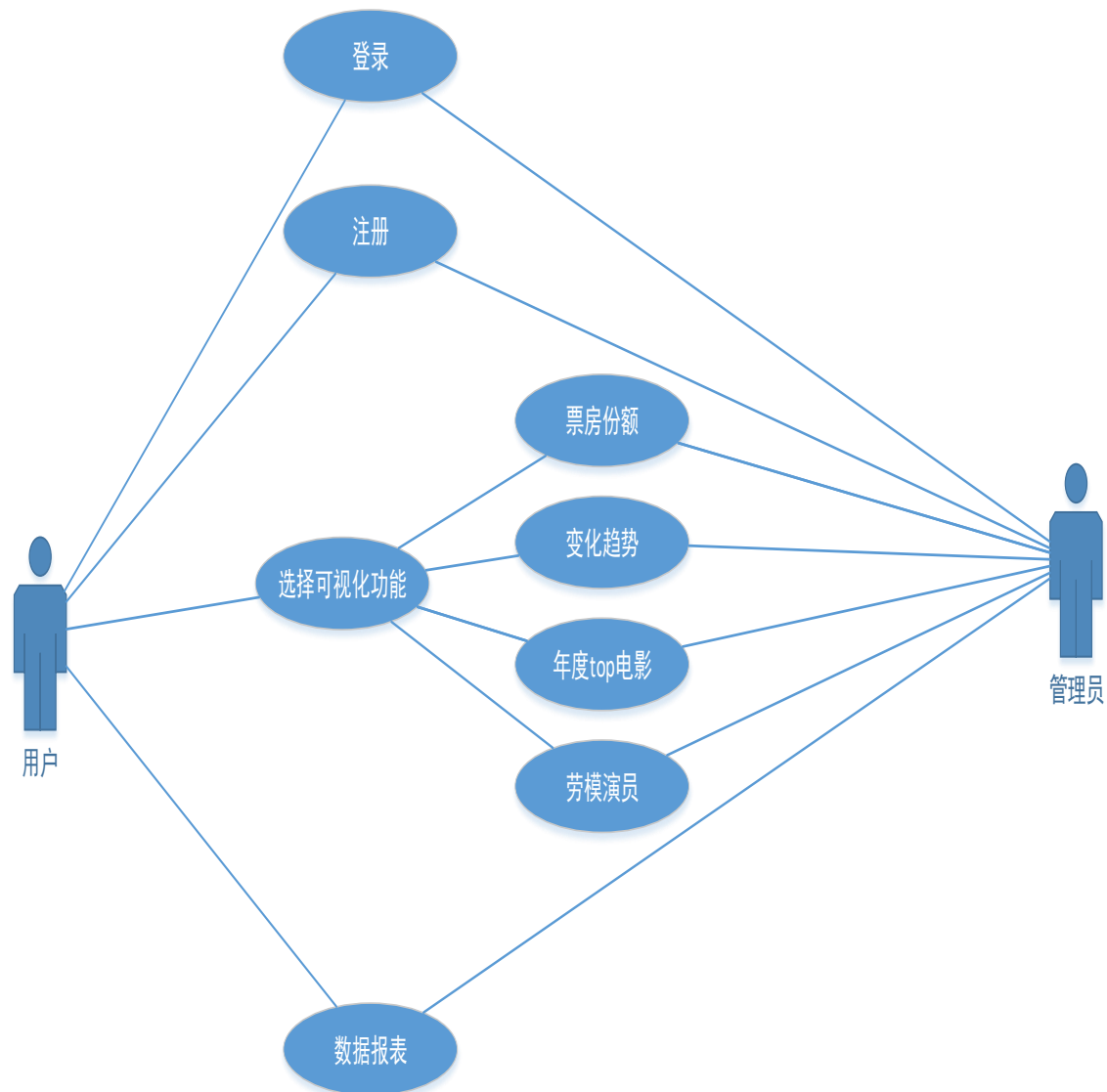
3.功能需求描述

3.1 系统的划分



3.2 可视化子系统的功能

3.2.1 可视化系统的用例图



3.2.4 可视化功能选择的用例描述

1. 目标：本用例允许合法用户选择可视化功能
2. 事件流
 - 1) 常规流程

当用户选择可视化功能时，本用例开始执行

 - i. 用户选择想要的可视化功能
 - ii. 系统查询数据库，利用数据库的数据建图并显示
 - iii. 如果无可用数据则返回错误信息

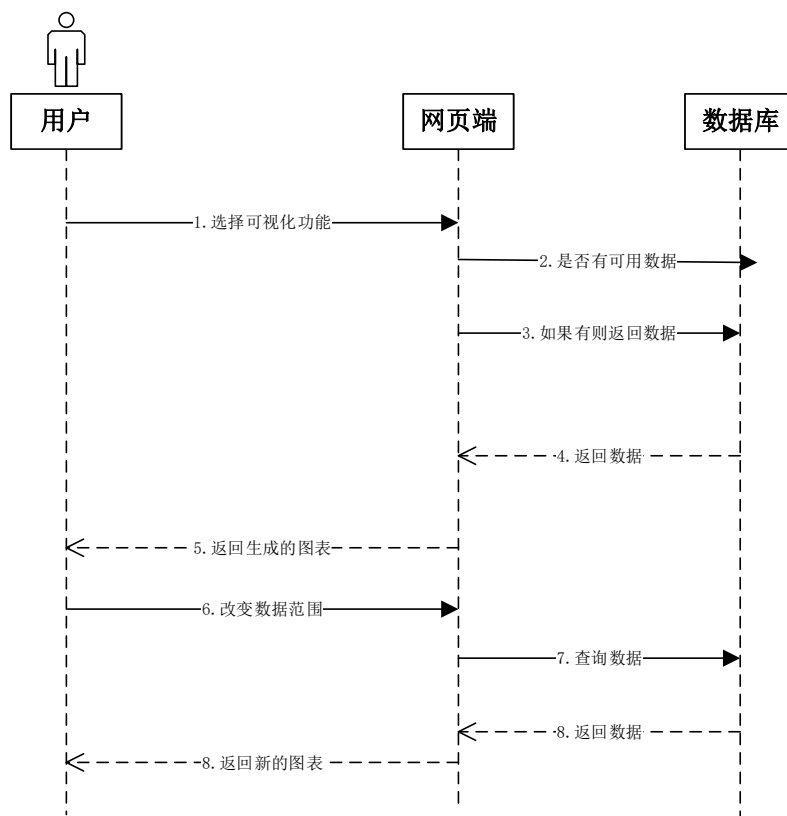
2) 扩展流程

(1) 用户更改图表数据范围

系统把更改后的数据传到后台，返回新的数据并生成图表

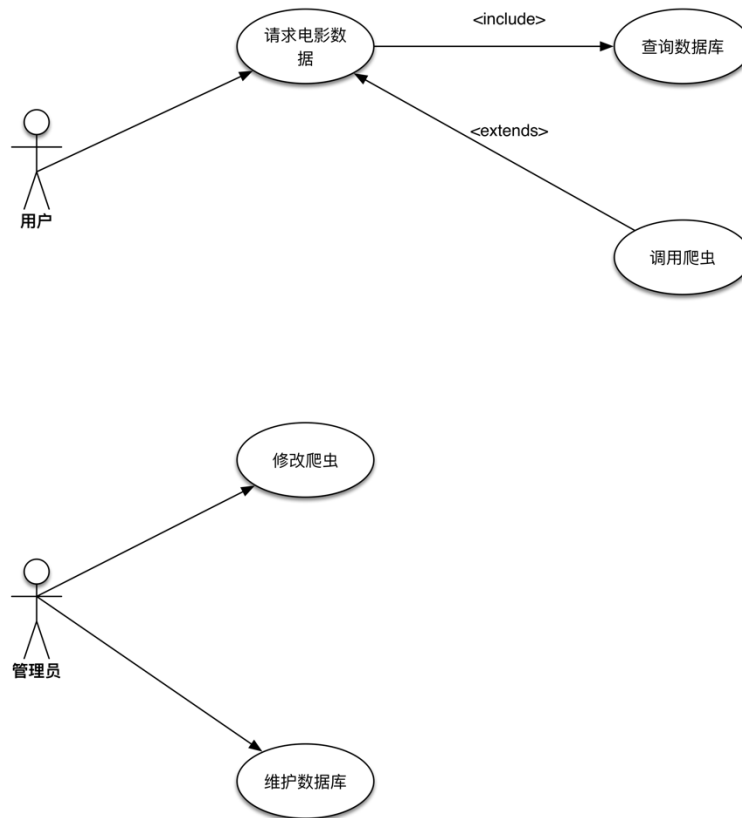
3. 前置条件：用例开始前，数据库不能为空

4. 后置条件：如果用例成功，用户得到想要的可视化结果

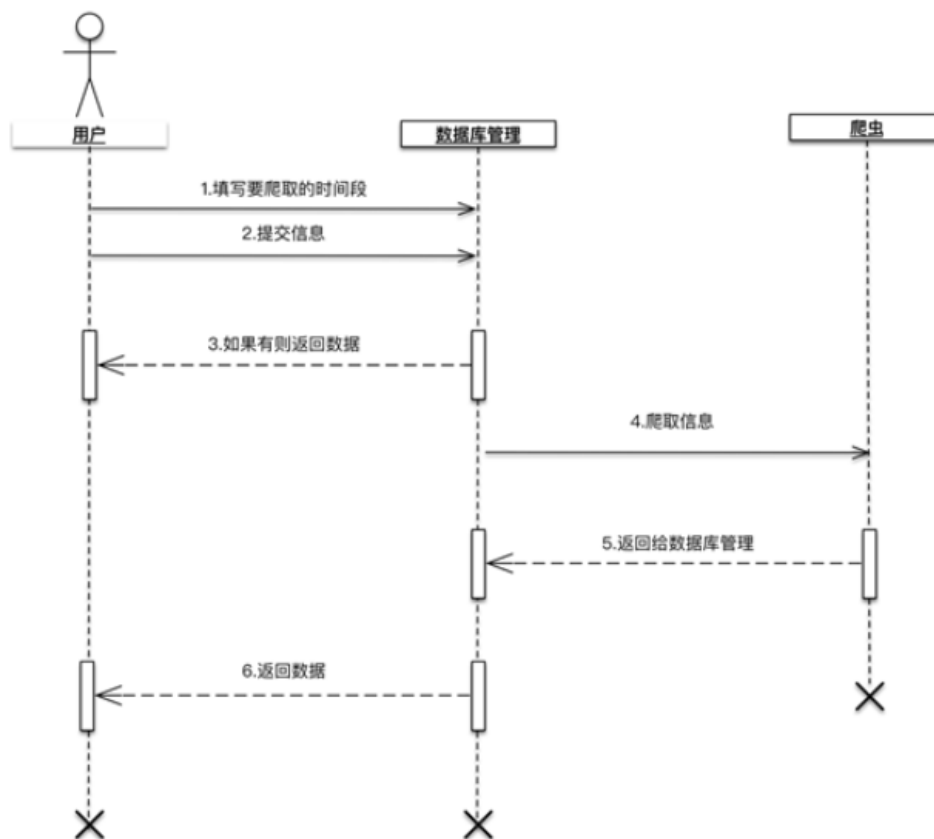


3.3 爬虫子系统的功能

3.3.1 爬虫子系统的用例图



3.3.2 用户请求电影数据的用例描述

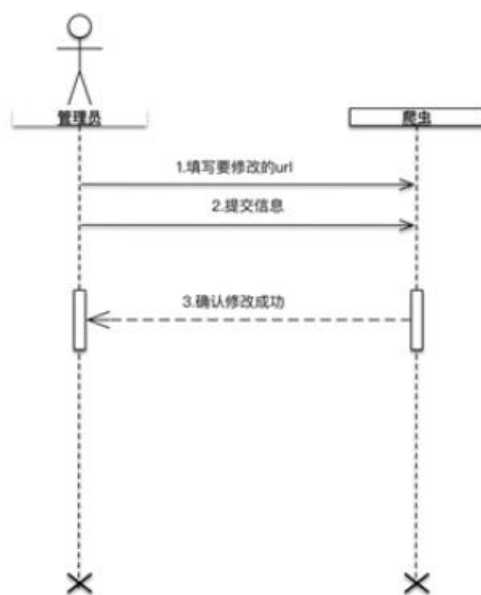


1. 目标：本用例允许合法用户获得电影数据
2. 事件流
 - 1) 常规流程
当用户请求数据时，本用例开始执行
 - i. 用户输入需要的日期范围
 - ii. 系统查询数据库，如果有符合要求的数据则返回
 - iii. 如果无则调用爬虫，爬取需要的数据并写入数据库
3. 前置条件：用例开始前，用户必须登录成功
4. 后置条件：如果用例成功，该用户将获得需要的数据信息

3.3.3 管理员修改爬虫的用例描述

1. 目标：本用例允许管理员修改爬虫
2. 事件流
 - 1) 常规流程
当管理员需要修改爬虫时，本用例开始执行
 - i. 管理员修改爬取的 url
 - ii. 确认修改
3. 前置条件：用例开始前，管理员必须登录成功
4. 后置条件：如果用例成功，爬虫爬取的目标将被修改

3.3.4 管理员维护数据库的用例描述



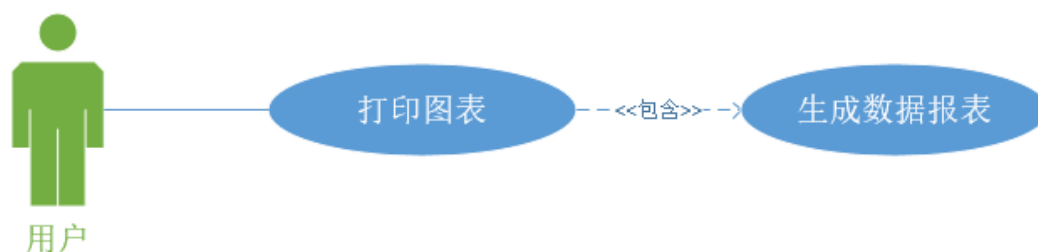
1. 目标：本用例允许管理员维护数据库中数据
2. 事件流
 - 1) 常规流程

当管理员需要维护数据库时，本用例开始执行

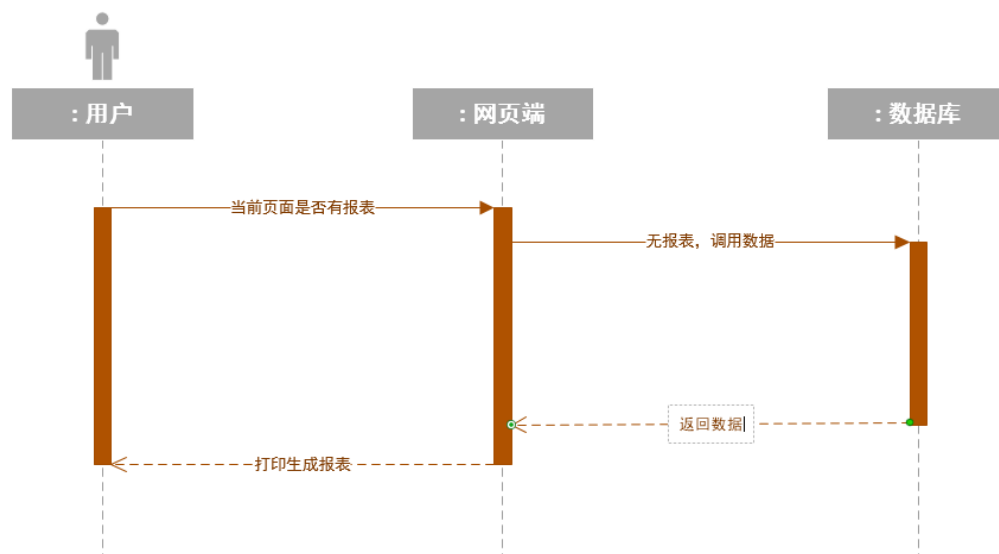
 - i. 管理员对需要修改或删除的数据进行操作
 - ii. 确认修改
3. 前置条件：用例开始前，管理员必须登录成功
4. 后置条件：如果用例成功，数据库中内容将更新

3.4 数据报表子系统的功能

3.4.1 生成数据报表的用例图



3.4.2 生成数据报表用例的描述



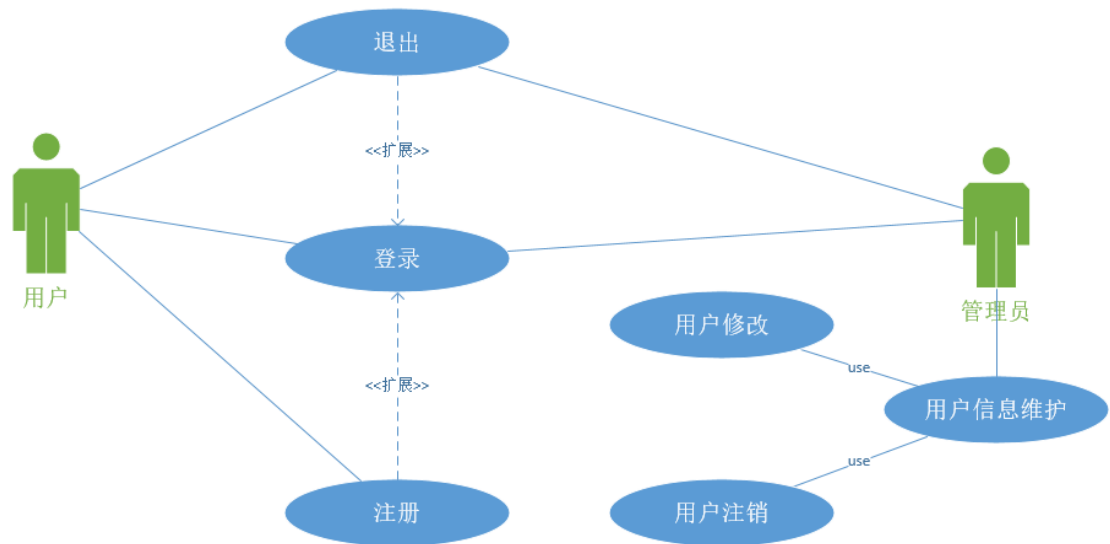
1. 目标：本用例允许合法用户选择生成数据报表
2. 事件流
 - 1) 常规流程

当用户选择生成数据报表时，本用例开始执行

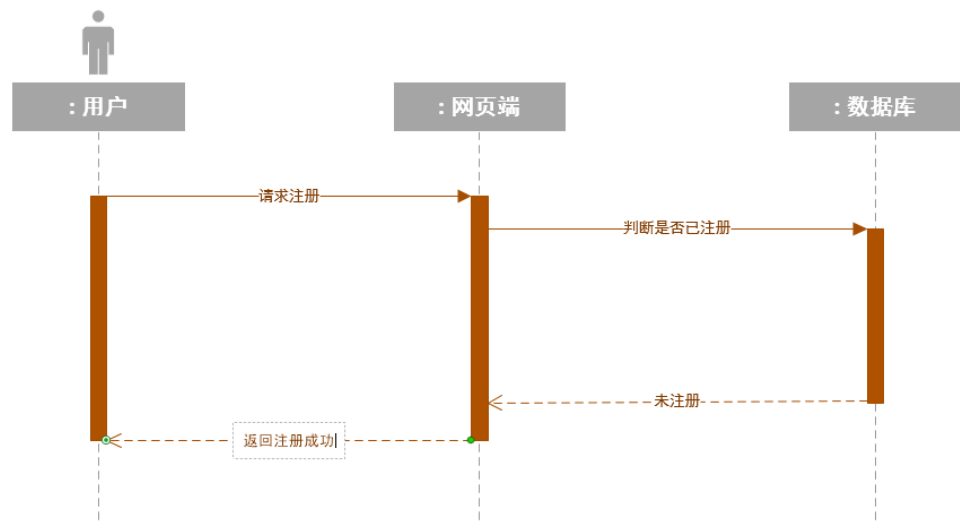
 - i. 用户选择想要生成的图表
 - ii. 系统将用户选择的图表汇总在新页面中。
 - iii. 用户确认后选择导出 pdf
 - iv. 系统将新页面的所有图表导入 pdf 中
 - v. 如果失败则返回错误信息
3. 前置条件：用例开始前，已有图表生成
4. 后置条件：如果用例成功，用户得到想要的的数据报表

3.5 登录子系统的功能

3.5.1 登录子系统的用例图



3.5.2 用户登录注册的用例描述



1. 目标：本用例允许用户进行登录、注册。

2. 事件流：

常规流程

当前状态为未登录时，本用例开始执行。

- (1) 系统要求用户登录或注册
- (2) 用户选择注册账号，若账号未被注册，则注册成功
- (3) 用户选择登录账号，若信息一致，用例结束

扩展流程

(1) 注册账号已存在

在主流程中，注册账号已存在，系统显示错误信息，用例结束

(2) 登录账号信息不匹配或不存在

在主流程中，如果登录账号信息不匹配或不存在，系统显示错误信息，用例结束

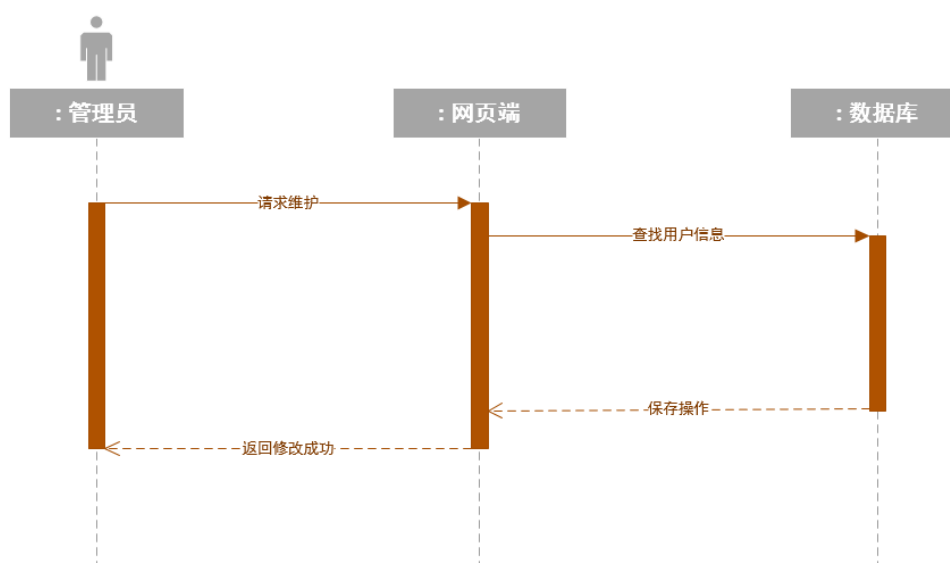
(3) 游客登录

用户选择游客登录，系统默认用户名为游客，进入系统

3. 前置条件：用例开始前，用户处于未登录状态

4. 后置条件：如果用例执行成功，用户处于登录状态，否则系统状态不变

3.5.3 管理员维护信息的用例描述



1. 目标：本用例允许管理员维护用户信息。

2. 事件流：

常规流程

当前状态为管理员已登录时，本用例开始执行。

(1) 管理员选择某一用户

(2) 选择修改或者注销

(3) 保存操作

扩展流程

(1) 管理员登录已过期

在主流程中，管理员登录已过期，系统显示错误信息，用例结束

(2) 登录账号信息不匹配或不存在

在主流程中，如果登录账号信息不匹配或不存在，系统显示错误信息，用例结束

3. 前置条件：用例开始前，管理员处于登录状态

4. 后置条件：如果用例执行成功，操作已保存，否则系统状态不变

4. 其它需求描述

4.1 性能要求

运行效率：较高

响应速度：可在用户提交数据请求的 30s 内返回输出结果，爬虫请求视数据规模而定

4.2 设计约束

1. 开发工具：PyCharm Professional, Navicat Premium, MySQLWorkbench, Visual Studio, Adobe Dreamweaver
2. 运行环境：任何基于 Chromium 内核的浏览器均可
3. 安全性：允许用户有常规的输入错误，避免造成内存溢出等系统安全问题
4. 可靠性：可根据正常的用户输入得到正常的结果输出，数据来源均为网络上爬取的或者已经爬取好存放在数据库中的数据

4.3 界面要求

用户界面能够对用户的操作作出恰当的反应，并帮助用户完成需要的工作。通过界面视觉减少用户的记忆负担，用户能以一致的方式获取信息。

4.4 进度要求

- | | |
|-----------------------------|--------------------|
| ■ 第一周
需求分析 | ■ 第四周
代码编写、软件实现 |
| ■ 第二周
体系结构设计、数据结构设计、组件设计 | ■ 第五周
软件验证 |
| ■ 第三周
算法设计、接口设计 | ■ 第六周
软件维护 |

4.5 交付要求

1. 软件的源代码
2. 运行环境搭建
3. 软件的说明文档
4. 总结 ppt 以及 demo

4.6 验收要求

1. 系统账户登入功能：

能正常完成已注册用户的登录和退出，具备错误账户报错功能，能实现未注册账号的注册，可增加验证码等其他功能，增加游客登录功能，同时具备一定的输入容错能力。

2. 电影数据爬取功能：

能够爬取猫眼或豆瓣电影或其他影视网页的电影数据，爬取数据的时间从 2015 年 1 月 1 日至 2018 年 11 月 30 日，爬取数据需包含但不限于电影名、电影题材、电影演员、电影导演、当前累计票房等。数据爬取功能需满足登录用户在系统中输入一段时间，系统能够将时间段内爬取返还给用户。另外实现多种格式数据的自动拼接，有一定的抗反爬能力。

3. 电影数据可视化功能&电影数据报表功能：

能够动态展示每年/每季度/每月不同题材的票房份额，能够动态对比展示电影票房每年变化趋势，能够动态展示不同年份的前 3、5、10 名的 top 电影是什么，能够找出某一年出演电影次数 TOP 的劳模男女演员。能够返回以上功能的正常输出并将其可视化，自动生成合适的图表，并允许用户能够保存并打印生成的图表。

4. 软件拥有一定的安全性和可靠性。

5. 软件原型

数据库设计原型:

1. 票房表

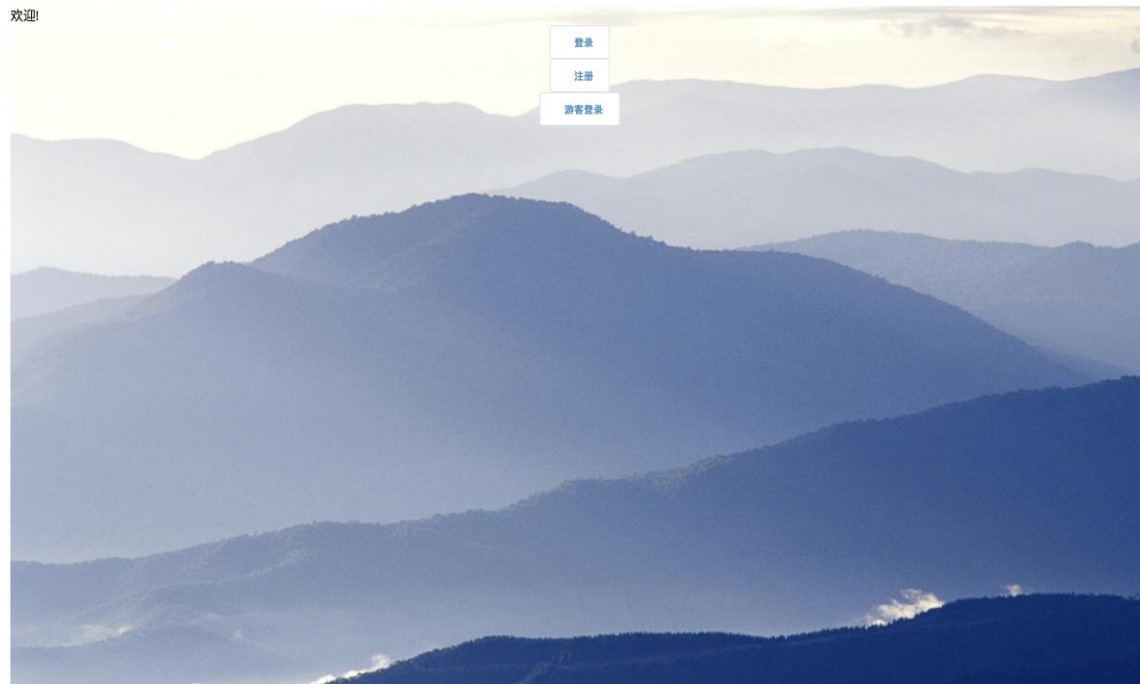
Rank	MovieName	BoxOffice	ShowCount	AudienceCount	SumBoxOffice	ServicePrice	AvgShowPeople
1	捉妖记	2147483647	1574060	65626600	2440020000	-	41.69
2	速度与激情7	2147483647	1478780	62445600	2426590000	-	42.23
3	港囧	1613536441	1220390	49106700	1614100000	-	40.24
4	复仇者联盟2：奥创纪元	1464391392	1282620	36598000	1464390000	-	28.53
5	夏洛特烦恼	1441595285	1322820	44762700	1441610000	-	33.84
6	侏罗纪世界	1420697319	1133070	37102400	1420730000	-	32.75
7	寻龙诀	1373912754	836389	37798600	1682740000	-	45.19
8	煎饼侠	1160309067	902217	35511400	1162760000	-	39.36
9	澳门风云2	974846563	680599	24858200	974856000	-	36.52
10	西游记之大圣归来	956428853	802488	27586300	956521000	-	34.38
11	碟中谍5：神秘国度	869768902	1016320	26529700	869769000	-	26.1
12	霍比特人3：五军之战	766497841	644272	19182200	766498000	-	29.77
13	天将雄师	744176999	482516	18308200	744177000	-	37.94
14	终结者：创世纪	725358336	824331	19725900	725359000	-	23.93
15	狼图腾	699602390	519763	17817200	704245000	-	34.28
16	九层妖塔	682929803	654360	19880400	682930000	-	30.38
17	蚁人	671755014	831685	18603700	671764000	-	22.37

2. 信息表

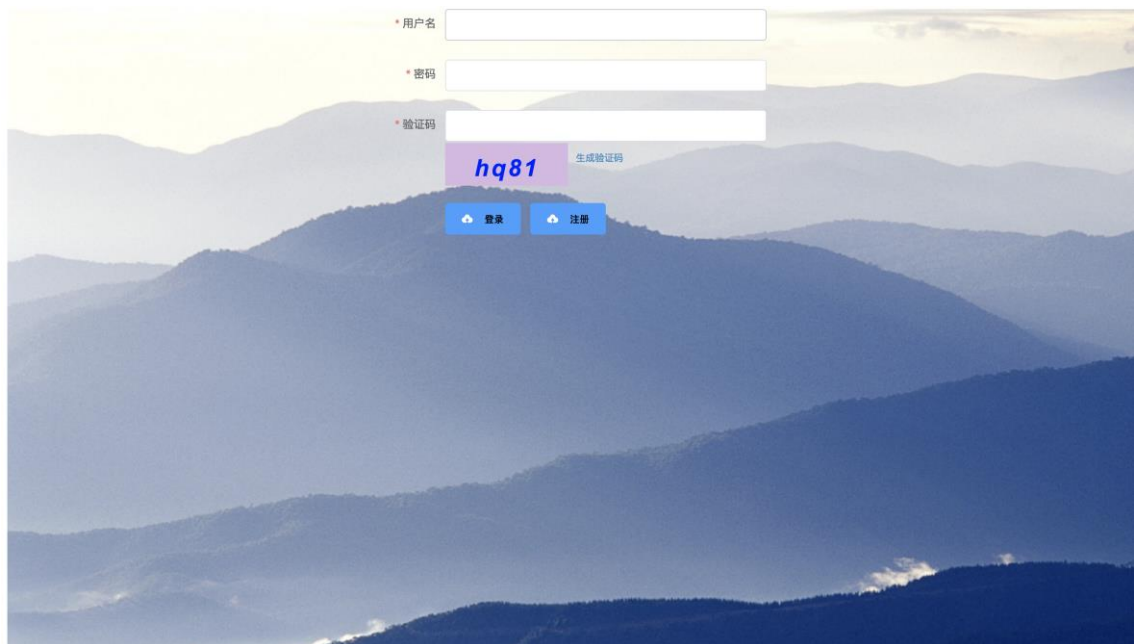
MovieName	Genre1	Genre2	Genre3	Director	CAST1	CAST2	CAST3	CAST4	CAST5	Date	YEAR	DOUBAN_SCORE
神奇女侠	动作	奇幻	冒险	派蒂·杰金斯	盖尔·加朵	克里斯·派恩	康妮·尼尔森	罗宾·怀特	丹尼·赫斯特	2017-06-	2017	7.1
第三度嫌疑人	剧情	悬疑	犯罪	是枝裕和	福山雅治	役所广司	广濑铃	满岛真之介	市川实日子	2018-03-	2018	7.1
老阿姨	剧情	-	-	雷献禾	陶慧敏	李雪健	陶海	-	-	2016-07-	2016	7.1
聊斋变异	剧情	悬疑	惊悚	陆宣	萧赫	谈苏阳	郁晓东	徐蒙源	龚芳妮	2016-12-	2016	7.1
英伦对决	动作	犯罪	-	马丁·坎贝尔	成龙	皮尔斯·布鲁斯南	刘涛	梁佩诗	夏莉·墨菲	2017-09-	2017	7.1
茶界中国	纪录片	-	-	刘嘉	-	-	-	-	-	2017-08-	2017	7.1
记忆大师	剧情	悬疑	惊悚	陈正道	黄渤	段奕宏	徐静蕾	杨子姗	梁杰理	2017-04-	2017	7.1
阿唐奇遇	动画	奇幻	冒险	王微	石磊	原泽宇	李冠霖	洪晃	米粒	2017-07-	2017	7.1
鲁豫有约大咖一日行	纪录片	真人秀	-	-	蒋欣	朴树	杨洋	李宇春	陈奕迅	2017-08-	2017	7.1
鲨滩	惊悚	冒险	-	佐米·希尔拉	布蕾克·莱芙莉	奥斯卡·贾恩那达	布莱特·卡尔	塞多娜·莱夫	-	2016-09-	2016	7.1
八月	剧情	-	-	张大磊	孔维一	张晨	郭燕芸	-	-	2017-03-	2017	7.1
关于他们的一些事	剧情	短片	-	邹桥	陈佳玮	张婧茹	叶浩伟	王泽霖	-	2017-05-	2017	7.1
厕所英雄	剧情	喜剧	-	什里·那拉扬	阿克谢·库马尔	阿努潘·凯尔	布米·佩德	拉什·沙玛	萨钦·克德	2018-06-	2018	7.1
哆啦A梦：新·大雄	动画	-	-	八锹新之介	水田山葵	大原惠美	嘉数由美	木村昴	关智一	2016-07-	2016	7.1
哪一天我们会飞	剧情	爱情	-	黄修平	杨千嬅	林海峰	苏丽珊	游学修	吴肇轩	2016-03-	2016	7.1
回家吃饭	真人秀	-	-	-	王小丫	-	-	-	-	2015-04-	2015	7.1
声来如此	短片	-	-	Austin Pete	袁靖童	-	-	-	-	2017-01-	2017	7.1
少奶奶的御花园	脱口秀	-	-	-	肖骁	-	-	-	-	2015-02-	2015	7.1
川流不息	纪录片	-	-	焦波	-	-	-	-	-	2018-05-	2018	7.1
忌日快乐	悬疑	惊悚	恐怖	克里斯托弗·史密斯	杰西卡·罗伯森	伊瑟尔·布罗萨德	露比·莫迪尼	查尔斯·艾特	劳拉·克利夫	2018-02-	2018	7.1
恋爱回旋	喜剧	运动	-	石川淳一	新垣结衣	瑛太	广末凉子	苍井优	濑户康史	2018-03-	2018	7.1
战狼2	动作	-	-	吴京	吴京	弗兰克·格里罗	吴刚	张翰	卢靖姗	2017-07-	2017	7.1
扑通扑通我的人生	剧情	家庭	-	李在容	宋慧乔	姜栋元	曹成穆	白日燮	许俊硕	2015-03-	2015	7.1
星球大战7：原力觉醒	动作	科幻	冒险	J·J·艾布拉姆斯	黛西·雷德利	约翰·博耶加	哈里森·福特	多姆纳尔·格利森	亚当·德赖弗	2016-01-	2016	7.1
星球大战8：最后绝地武士	动作	科幻	冒险	莱恩·约翰逊	马克·哈米尔	凯丽·费雪	黛西·雷德利	亚当·德赖弗	约翰·博耶加	2018-01-	2018	7.1
春娇救志明	剧情	喜剧	爱情	彭浩翔	余文乐	杨千嬅	蒋梦婕	曾国祥	秦沛	2017-04-	2017	7.1

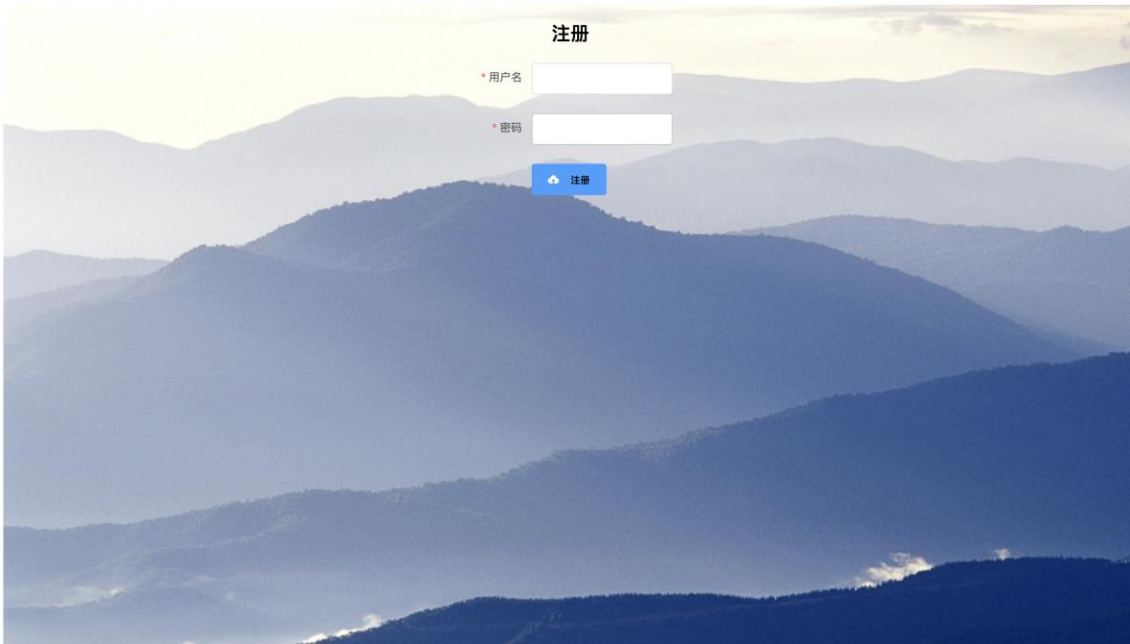
用户界面设计最终版:

1. 主界面



2. 登录注册界面





3. 功能选择界面



4. 数据可视化界面

4.1 年/月和年/季度不同题材票房份额显示

年/季度:

票房份额

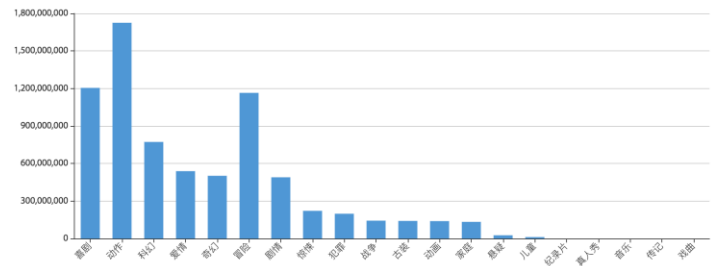
年份 2015 季度 1

选择的是: 2015 选择的是: 1

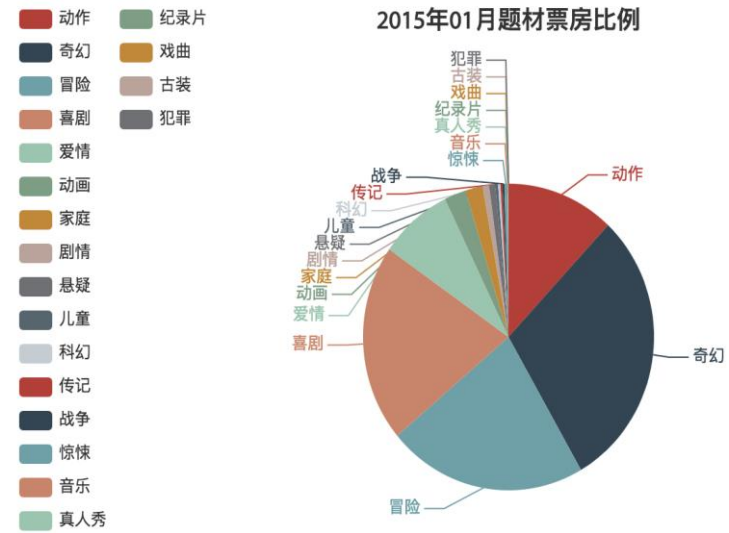
月份 01

选择的是: 01

2015年第1季度题材票房比例



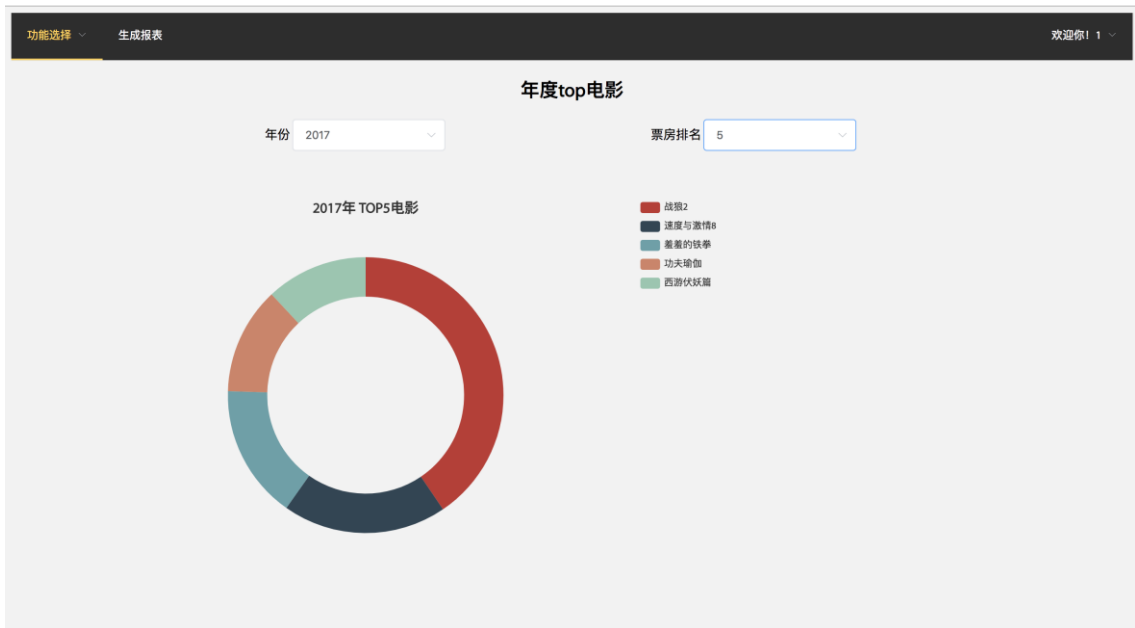
年/月:



4.2 年/月票房趋势显示



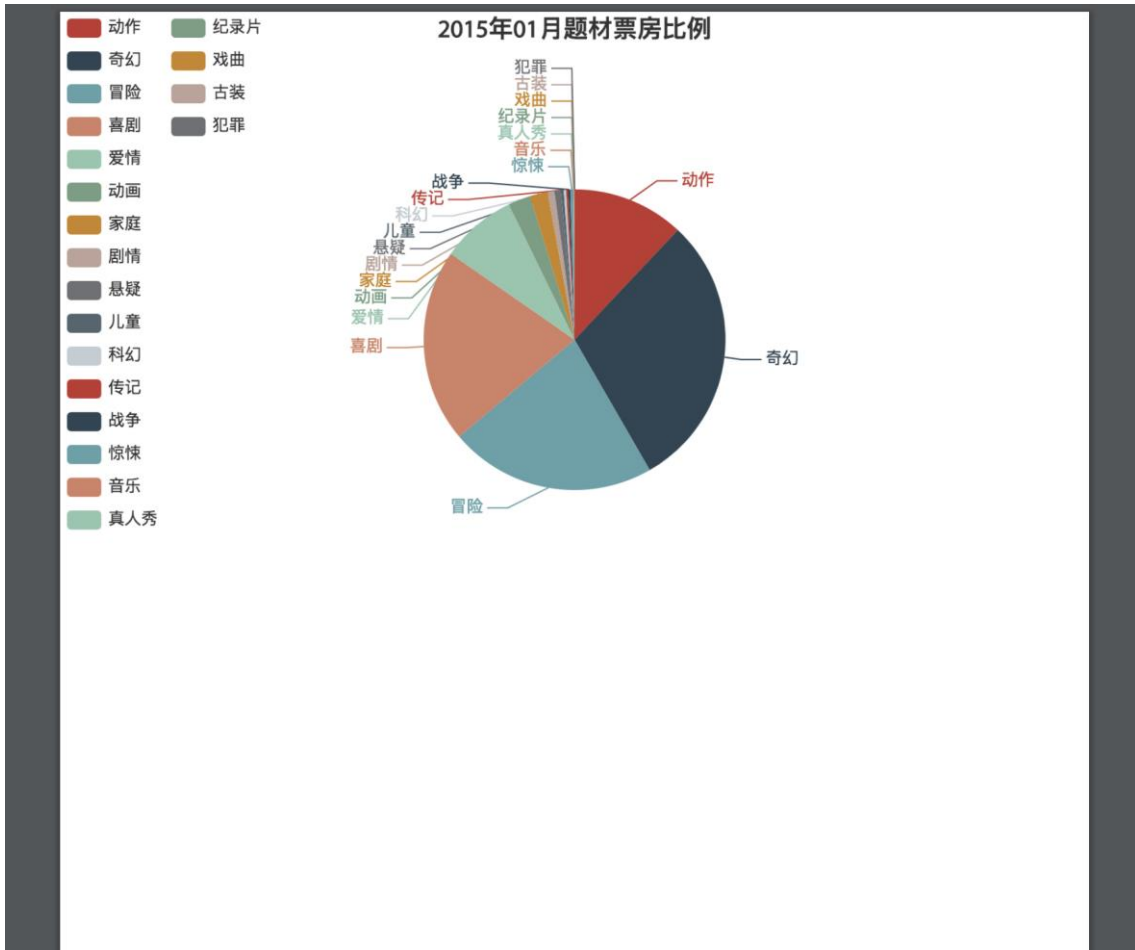
4.3 年度 top 电影显示



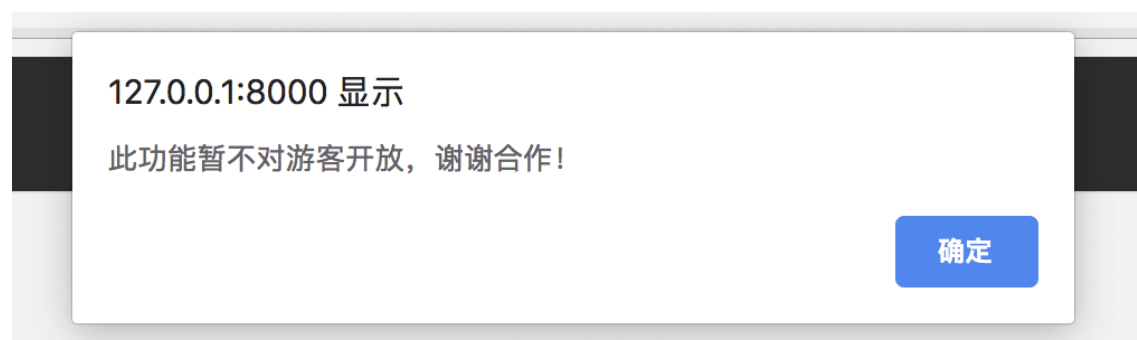
4.4 年度 top 男女演员显示



4.5 报表生成功能



4.6 游客功能限制（只可以观看默认图表，不可以生成报表）



6. 附录（分工）

分工情况：

名字	学号	分工情况
石嘉晗	SZ160110227	管理数据库、爬虫功能
周宇飞	SZ160110128	后端功能实现、前后端整合
谢华鹏	SZ160110232	前端架构、前后端整合
王冠雄	SZ160110217	后端功能实现、前后端整合