



软件工程

Software Engineering

廖清

liaoqing@hit.edu.cn

计算机科学与技术学院



项目简介

- 项目目标：完成一个电影数据可视化分析系统（PC端or移动端均可），为业界人士提供当前电影现状分析的可视化系统。
- 项目占比：占课程总成绩40%
- 项目功能
 - 系统账户登入功能（必选）
 - 电影数据爬取功能（必选）
 - 电影数据可视化功能（必选）
 - 电影数据报表功能（必选）
 - 。。。 (可自行增加)



项目简介-功能描述

- 系统账户登入功能 (4分)
 1. 具备登记账户登入功能 (必选)
 2. 具备错误账户报错功能 (必选)
 3. 具备验证码功能 (可选)

票房预测系统

用户名:

密码:

验证码: 1 4 9 3

登录



项目简介-功能描述

●电影数据爬取功能（6分）

1. 能够爬取猫眼或豆瓣电影或其他影视网页的电影数据
2. 爬取数据的时间从2015年1月1日至2018年11月30日
3. 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，上映时间等。
4. 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（测试展示）。
5. 如数据爬取部分拥有持续爬取数据能力有加分。



项目简介-功能描述

●电影数据爬取功能细节思考

1. 票房单位怎样选择?
2. 一部电影多个题材, 多个演员, 数据如何存储?
3. 爬取数据是以页数编号爬取的, 如何将编号转化为时间段?
4. 网站的反爬虫机制

1	name	index	year	month	day	boxoffice	director	actor	genre
2	三生三世十里桃花	246896	2017	8	3	53457	赵小丁, 安东尼.拉默里纳拉	刘亦菲, 杨洋, 罗晋, 严屹宽, 李纯	爱情, 奇幻
3	杀破狼·贪狼	342176	2017	8	17	51713	叶伟信	古天乐, 吴樾, 林家栋, 克里斯·科林	剧情, 动作
4	心理罪	342381	2017	8	11	30034	谢东燊	李易峰, 廖凡, 万茜, 李纯, 张国柱	动作, 悬疑, 犯罪
5	侠盗联盟	1183619	2017	8	11	23692	冯德伦	刘德华, 舒淇, 张静初, 杨祐宁, 让·雷	动作, 冒险
6	二十二	615739	2017	8	14	17021	郭柯	毛银梅, 何玉珍, 韦绍兰, 李凤云, 张	纪录片
7	十万个冷笑话2	346308	2017	8	18	13285	卢恒宇, 李姝洁	王禎, 山新, 郝祥海, 李姝洁, 李璐	喜剧, 动画, 奇幻
8	鲛珠传	245938	2017	8	11	11363	杨磊	王大陆, 张天爱, 任达华, 盛冠森, 赵	喜剧, 动作, 爱情, 奇幻
9	赛尔号大电影6: 圣者无敌	1201161	2017	8	18	10211	王章俊	罗玉婷, 翟巍, 王晓彤, 孙晔	喜剧, 动画, 儿童, 奇幻



项目简介-功能描述

● 电影数据可视化功能（共16分）

➤ 能够动态展示每年/每季度/每月不同题材的票房份额

* 年份

2015

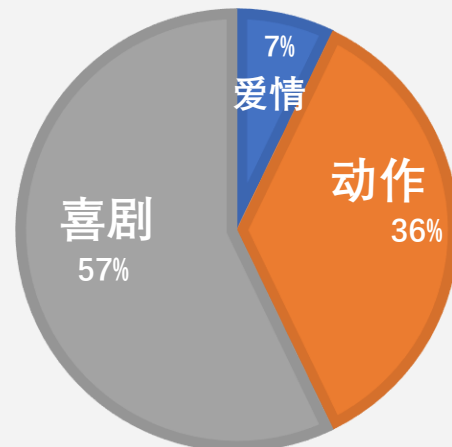
季度:

第一季

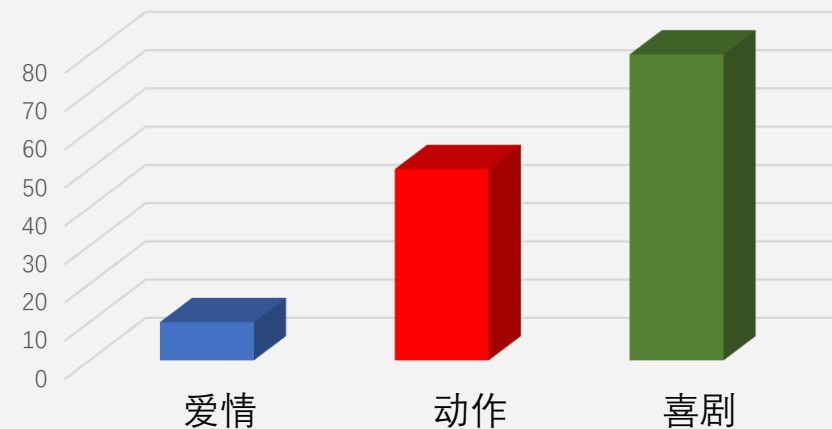
月份:

08

2015年第一季度题材票房比例



2015年8月题材票房比例





项目简介-功能描述

● 电影数据可视化功能（共16分）

➤ 能够动态对比展示电影票房每年变化趋势

* 年份

2015

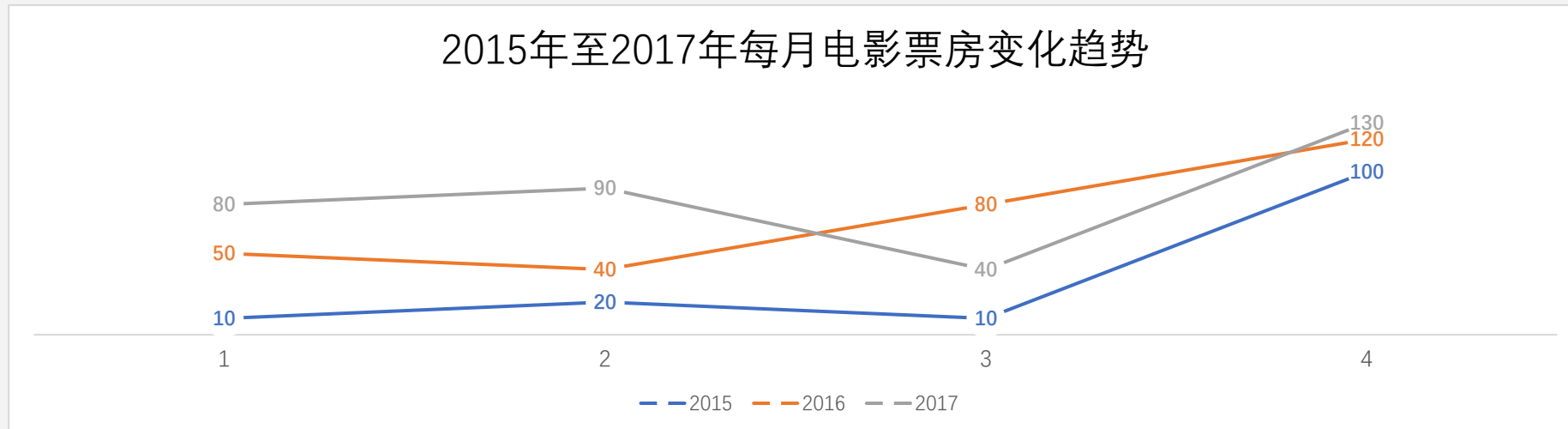
年份:

2016

年份:

2017

2015年至2017年每月电影票房变化趋势





项目简介-功能描述

● 电影数据可视化功能（共16分）

➤ 能够动态展示不同年份的前3、5、10名的top电影是什么

* 年份

2015



票房排名:

3





项目简介-功能描述

● 电影数据可视化功能（共16分）

➤ 能够找出某一年出演电影次数TOP的劳模男女演员

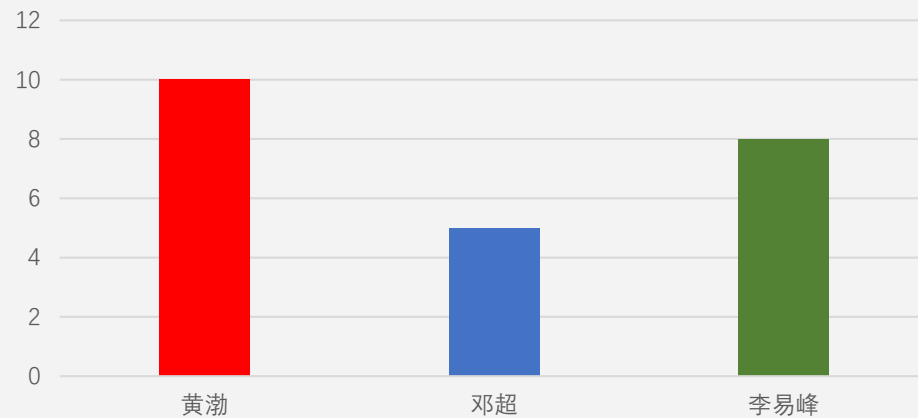
* 年份

2015

排名:

3

2015年TOP 3 劳模男演员



章子怡

杨幂

周冬雨



项目简介-功能描述

●电影数据可视化功能细节思考

- 用户需要输入什么信息？让用户选or让用户输入？
- 不符合输入情况时如何报错并提示用户？
- 选择合适的分析数据展示模式

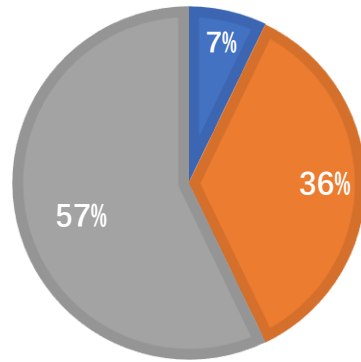


项目简介-功能描述

●电影数据报表功能（4分）

1. 通过勾选可视化功能部分将图片保存并展示在同一个文档中。

2015年第一季题材票房比例

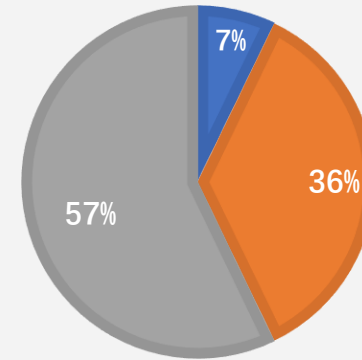


✓ 保存

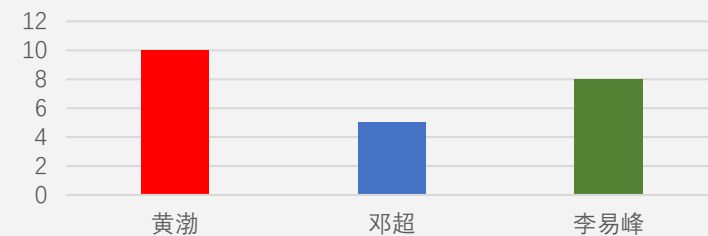
生成报告



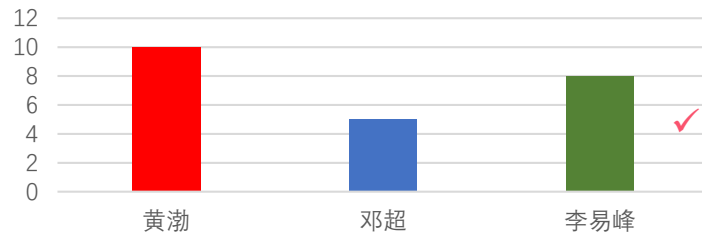
2015年第一季题材票房比例



2015年TOP 3 劳模男演员



2015年TOP 3 劳模男演员



✓ 保存



项目简介-分值分配

- 系统账户登入功能 (2-4分)
 - 电影数据爬取功能 (4-6分)
 - 电影数据可视化功能 (共8-16分)
 - 电影数据报表功能 (2-4分)
 - 项目过程中产生的报告 (5分)
 - 项目答辩Demo演示 (5分)
- 项目实践
共30分
- 项目展示
共10分

满足功能基本要求获得基础分。除要求功能外有扩展功能有加分 (至多5分)



自由探索

- 本次项目实践可以有最多三组同学自由选题，小组选题择优录取，自由探索项目若能完成，分值会比基础项目分值高**至多10分**。
- 录取依据包括且不仅限于：
 1. 项目定义是否明确
 2. 项目需求是否明确
 3. 项目功能设计是否明确
 4. 团队成员协作分工是否明确
 5. 项目创新程度
 6. 项目实施难度
 7. 提供文档越详细越有优先权



自由探索所需资料

- 不论是选择默认项目或是选择自由探索项目，均需于2018年10月8日前将选择结果告知助教**韩浩**。
- 选择自由探索项目的小组需于2018年10月9日前提交**额外项目资料**给助教**韩浩**，资料包括：
 1. 项目描述word/PDF 文档
 2. 项目介绍PPT
 3. 项目资料中需要介绍：
 - a) 项目实施的应用背景
 - b) 项目实施的迫切性
 - c) 项目需要完成的功能
 - d) 项目的人员安排
 - e) 项目的进度计划
 - f) 项目的验收指标

若项目失败，会丢失实践大部分分值，自由探索需谨慎！



课外加分

- <https://www.educoder.net/competitions/gcc-dev-2018>

OpenGCC

全国高校 绿色计算大赛

National colleges and universities Green computing contest

项目挑战组

报名截止时间：2018年11月10日24:00

模拟体验时间	预赛时间	决赛时间
2018年8月30日 至10月26日	2018年10月28日 9:00-24:00	2018年11月11日 9:00-24:00



课外加分

- <https://trustie.educoder.net/competitions/gcc-annotation-2018>

OpenGCC

全国高校 绿色计算大赛

National colleges and universities Green computing contest

开源标注组

报名截止时间：2018年10月30日24:00

模拟体验时间	正式赛时间
代码标注阶段： 2018年9月10日至9月25日	代码标注阶段： 2018年10月1日至10月31日
标注互评及报告撰写阶段： 2018年9月26日至9月30日	标注互评及报告撰写阶段： 2018年11月1日至11月10日



课外加分

- 三等奖 2分
- 二等奖 4分
- 一等奖 6分
- 特等奖 10分

