# 项目技术说明书

豆瓣电影数据可视化

小组名称： 第五组

学员姓名： 郑钰淇

目录

[项目技术说明书 1](#_Toc17289747)

[1. 项目介绍 2](#_Toc17289748)

[2. 需求分析 2](#_Toc17289749)

[3. 技术选型 3](#_Toc17289750)

[4. 爬虫开发和优化 3](#_Toc17289751)

[5. 数据清洗和存储 4](#_Toc17289752)

[6. 数据查询和可视化展示 5](#_Toc17289753)

[7. 开发环境 5](#_Toc17289754)

[8. 产品原型 6](#_Toc17289755)

# 项目介绍

基于Node.js从豆瓣网上爬取2009-2019的电影数据，筛选清洗后将电影名，评分，上映国家和上映日期存入mongo数据库，通过echarts渲染展示出影评分和电影数量分析、近十年来每年发行的电影分析、2019各个国家发行的电影数量、评分在5分一下的电影名称等数据可视化的展示，最终实现豆瓣电影数据可视化分析平台的搭建。

# 需求分析

准备好一个mongo数据库服务端、爬虫代码。

用爬虫爬取豆瓣网所需要的电影名称、电影评分、电影国家、电影上映日期。

# 技术选型

Node.js：前端应用开发方法

Cheerio：实现HTML数据解析

MongoDB：对需要的数据进行存储

eCharts：数据可视化展示

superagent：客户端请求代理模块（支持HTTPS）

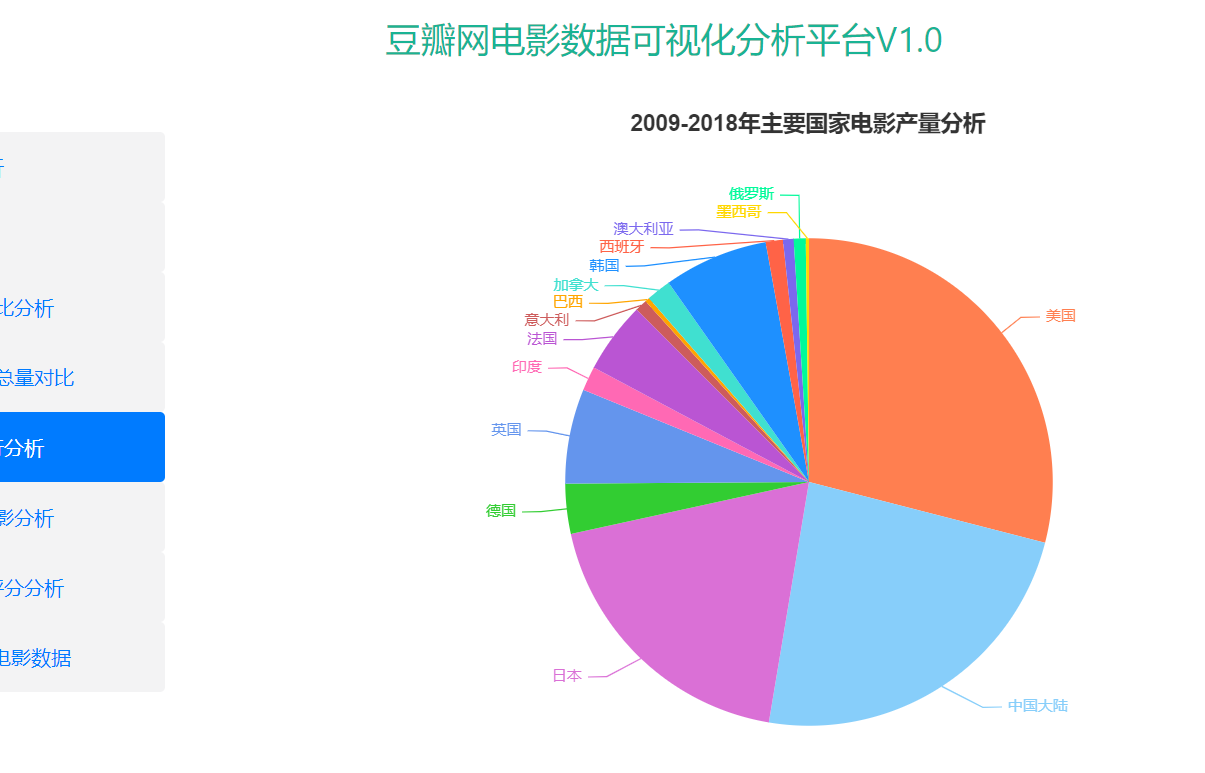
# 爬虫开发和优化

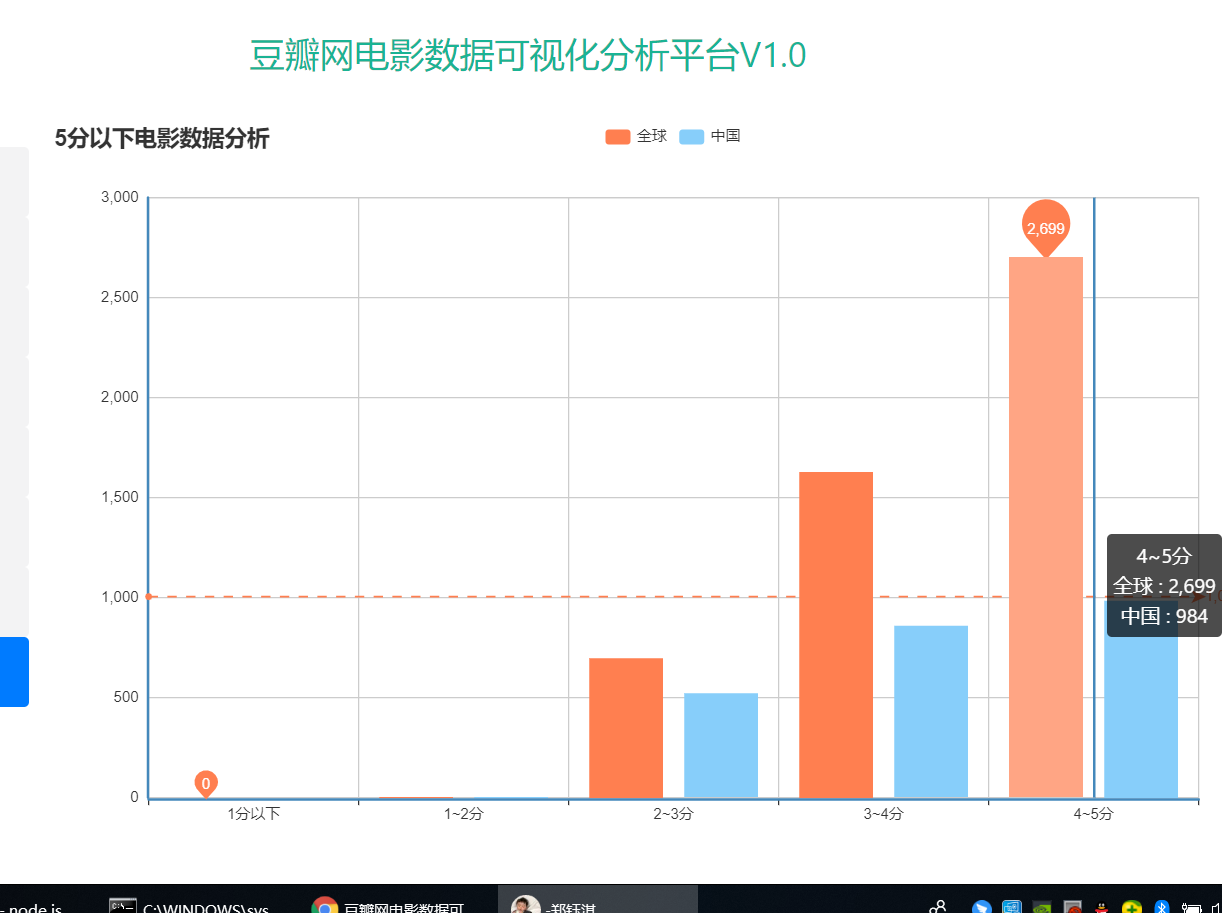
1. 连接数据库并建立model
2. 先爬取外层大的url，筛选出电影名称和评分以及电影的详细页面的url
3. 然后进入子页面使用cheerio解析HTML并配合正则爬取所需的电影上映日期和国家
4. 将爬取下来的四个值存入事先准备好的mongo数据库中
5. 重复前面三步直至所有数据爬取完毕

# 数据清洗和存储

# 

# 数据查询和可视化展示





# 开发环境

IDE：VS Code ：功能较为强大，插件丰富，调试方便

# 产品原型

(主要说明自己所负责的的模块最终的呈现效果)

|  |  |
| --- | --- |
| 豆瓣电影数据可视化分析平台V1.0 | |
| 2013-2014的电影数据爬取 | echarts展示 |
|  |
|  |
|  |
|  |
|  |
|  |
| 自己自由发挥 |