# 项目总结报告

豆瓣电影数据可视化

小组名称： 第五组

学员姓名： 唐浩彬

目录

[项目总结报告 1](#_Toc17289904)

[1. 项目基本信息 2](#_Toc17289905)

[2. 项目成果 2](#_Toc17289906)

[3. 开发工具与环境 2](#_Toc17289907)

[4. 项目需求变更 3](#_Toc17289908)

[5. 项目计划和进度实施 3](#_Toc17289909)

[6. 经验总结 3](#_Toc17289910)

[7. 教训总结 4](#_Toc17289911)

[8. 其他改进意见 4](#_Toc17289912)

# 项目基本信息

(主要介绍一下豆瓣影评数据分析的基本需求等介绍)

本项目主要实现将豆瓣电影网中的每年的电影数据—名称、评分、国家、年份等信息爬取到然后以柱状图、饼状图、折线图的形式展现出来。

# 项目成果

(主要说明一下自己所负责的页面模块开发完成之后的页面效果以及文件信息等)

我主要负责爬取2011-2012年的所有电影数据，爬取成功后用nodeJS设置爬虫然后将数据传入到MongoDB的数据库中保存，然后将数据以JSON的文件形式导出。

# 开发工具与环境

(主要说明自己所负责的当前页面开发说使用的技术，以及每种技术的作用是干什么)

使用cheerio解析页面数据信息。使用async进行同步进行避免异步导致的IP被封禁。使用MongoDB实现数据的存储。

# 项目需求变更

(主要说明在项目开发过程中项目出现的变更需求，出现变更的原因，以及变更导致的结果和影响等)

需要爬取的数据需求变更：

爬取名字、评分、国家、年份、评论人数

=>爬取名字、评分、国家、年份、导演

=>爬取名字、评分、国家、年份

变更原因：电影子页面没有评论人数、图表中没有需要导演数据的内容

导致的结果和影响：修改爬虫代码cheerio部分

需要存储的数据需求变更：

存储所有电影 =》 只存储评分、年份不为空的电影数据

变更原因：存储空评分或年份不利于数据统计

导致的结果和影响：修改爬虫代码cheerio部分及存储数据库部分

# 项目计划和进度实施

(主要说明项目开发的时间计划和完成进度)

小组每个成员负责爬取2年的数据，最后组长汇总进行制表。

# 经验总结

（项目实施过程当中积累的经验）

由于对基础知识不熟悉导致边学边做使用了很多时间，一切事物基础最重要。同时，不会的知识点问组员，团队的互相合作也可以加快进程。

# 教训总结

（项目实施过程当中的教训总结信息，由于技术或者管理等因素导致的项目不能按时交付以及后期集成的问题）

因为自身基础不牢靠导致团队工作时间延后，打牢基础最重要。

# 其他改进意见

(对项目的开发过程有其他可以优化和改进的地方)