天津工业大学计算机科学与技术学院

实训总结报告



　项目名称: 基于豆瓣资源的个性化信息整合库平台

　合作培养企业：天津市融创软通科技股份有限公司

　专业方向：大数据方向

　项目负责人：何晶

　专业班级: 软件工程Z1801

姓名：史芯源

　实训起止日期：2020.6.15-2020.7.15

**1 引言** （一级标题，小三黑体，居中）

1.1项目概述（二级标题，四号黑体，左对齐）

说明项目目的、背景、意义及参考资料（包括企业提供的素材、DEMO）等；

1.2 需求分析

为什么做这个题目，有什么实际价值

1.3 运行环境

软、硬件环境

（正文，小四号宋体，1.25倍行距）

**2 项目设计**（小三黑体，居中）（以个人主要负责内容为主）

2.1 设计思路

首先作为一个数据分析后台管理，第一步我们要获取数据，这边我们使用了scrpay框架进行数据爬取，我个人没有使用scrpay框架直接写了爬虫代码，设置了动态IP和用户协议来防止IP被封，提高爬取速度。然后就是前端界面，这块我们使用了layui框架，我构建了welcome界面，以及电视剧这部分。后端用java实现，创建了一套从数据库到前端界面的代码包括：controller（js文件调用，可以将所需数据push到前端）,mapper（从数据库提取数据的主要模块，撰写sql语言的地方）,entity（所需数据）,impl,service（接口）,

这一套流程可以将所需要的数据封装push到前端，供给之后之用的echarts图表，echarts作为数据展示的主要工具，进行数据可视化。

前端只写了两块包括数据查询/修改（因为爬取的数据没有修改点，这边电视剧模块没有写修改的功能，但是在用户那块进行了编写），以及数据可视化。

2.2 模块功能介绍

主要分为两部分：

第一部分：

数据查询功能，可以将爬虫爬取到的豆瓣电视剧信息在前端界面展示，可以针对导演名称，以及电视剧名称进行查询（这里可以添加更多的查询条件）。

第二部分：

数据分析/可视化：

第一个图表：

这里对爬取的电视剧评分进行的统计，由高到低的进行了数据展示，但是9.0之后的数据进行隐藏，全部展示略显拥挤。

第二个图表：

这个图表是针对电视剧发布时间进行的统计，统计了近年来电视剧的发布情况，得出近年来电视剧生产数量数年增加，2020年仅半年就有追平2019年的趋势，说明整个中国电视剧产业走向良好

第三个图表：

对爬取的所有电视剧星级信息进行了整理，对不同星级的占比进行了曲线展示，数据可得高星级还是少数，并且呈递减趋势。

第四个图表：

对我所爬取的数据进行了简单的分类统计，统计了不同类型的电影各有多少。

第五个图表：

对爬取的导演进行了数据汇总，汇总爬取到的数据有多少是他们创作的。分析导演的产出率。

2.3 模块结构图

2.4 程序流程图

2.5 功能设计分工

需详细标注每部分的内容，具体任务及任务安排等。

**3 详细设计**（小三黑体，居中）

详细写明项目的设计（系统设计、数据库设计）等，并写明对所涉技术、核心功能等描述；

**4 实现方案**（小三黑体，居中）

详细写明项目功能实现描述、核心代码及实现结果等；

并对调试好的程序，从执行程序弹出界面开始，每一步操作截一个图，并附加说明（图片格式为JPG格式）。

**5 系统测试**（小三黑体，居中）

详细写明测试方法、测试描述（bug与修正方法等）、预期结果等；