**中国科学技术大学软件学院**

**工程实践结题报告**

**项目题目： Java留学生信息管理与分析系统**

**指导老师： 张曙**

**研究领域： J2EE软件开发**

**小组成员： 罗迪 随哲朝 章文**

**成员学号：SA18225272 SA18225324 SA18225505**

**中国科学技术大学软件学院**

目录

[摘要 1](#_Toc12920299)

[Abstract 2](#_Toc12920300)

[一、项目工程结果 3](#_Toc12920301)

[1,具体产品 3](#_Toc12920302)

[2,系统功能 3](#_Toc12920303)

[二、开发工作评价 5](#_Toc12920304)

[1,开发效率 5](#_Toc12920305)

[2,产品质量 5](#_Toc12920306)

[3,技术方法 5](#_Toc12920307)

[4,主要问题及解决办法 5](#_Toc12920308)

[三、不足和展望 6](#_Toc12920309)

[1,不足 6](#_Toc12920310)

[2,展望 6](#_Toc12920311)

摘要

经过两个学期的长期工作，工程实践从初期开题的设计到最后的结题阶段，主要完成的任务量如下：

1. 实现了一套低耦合、可扩展性良好的Web系统，其中包含后端管理平台与

前端Web网站.

(2) 基本完成了开题初期设计的所有功能，包括个人评估、理想院校、院校搜索三个功能，而且还额外添加了排名的展示，以及一些非功能需求。

(3) 技术上前端运用了Bootstrap框架，后端使用的是SSM的整体框架实现，Spring+SpringMVC+Mybetis的Java EE企业级框架搭建整个系统，使用AJAX实现在后台与服务器进行信息交互，使网页能够异步更新，通过Python的Requests和Beautifulsoup实现数据爬取和过滤处理。

(4) 在前端设计上，我们做到了设计规范的一致性，友好清晰的交互页面。

(5) 在安全方面，摒弃了传统的Session验证方式，用拦截器作为权限管理。

**关键词：Java信息管理与分析系统; Bootstrap;SSM;Session**

Abstract

After two years of long-term work, the main tasks of our engineering practice from the initial design to the final stage of the project are as follows:

(1)We have implemented a low-coupling, scalable Web system that includes a back-end management platform and a front-end Web site.

(2)The system basically realizes all the functions of the initial design, including the four functions of personal evaluation, ideal colleges, and college search rankings, as well as some non-functional requirements.

(3)The front end of our system uses the Bootstrap framework, the back end of the system uses the overall framework implementation of SSM, and the Java EE enterprise level framework of Spring+SpringMVC+Mybetis builds the entire system. The AJAX implementation interacts with the server in the background to enable the web page to be updated asynchronously. Finally, data crawling and filtering are handled by Python's Requests and Beautifulsoup.

(4) In front-end design, we have achieved the consistency of design specifications, friendly and clear interactive pages.

(5) In terms of security, the system abandons the traditional session authentication method and uses the interceptor as the rights management.

**Keyword: Java Information Management and Analysis System，Bootstrap，SSM，Ses****sion**

一、项目工程结果

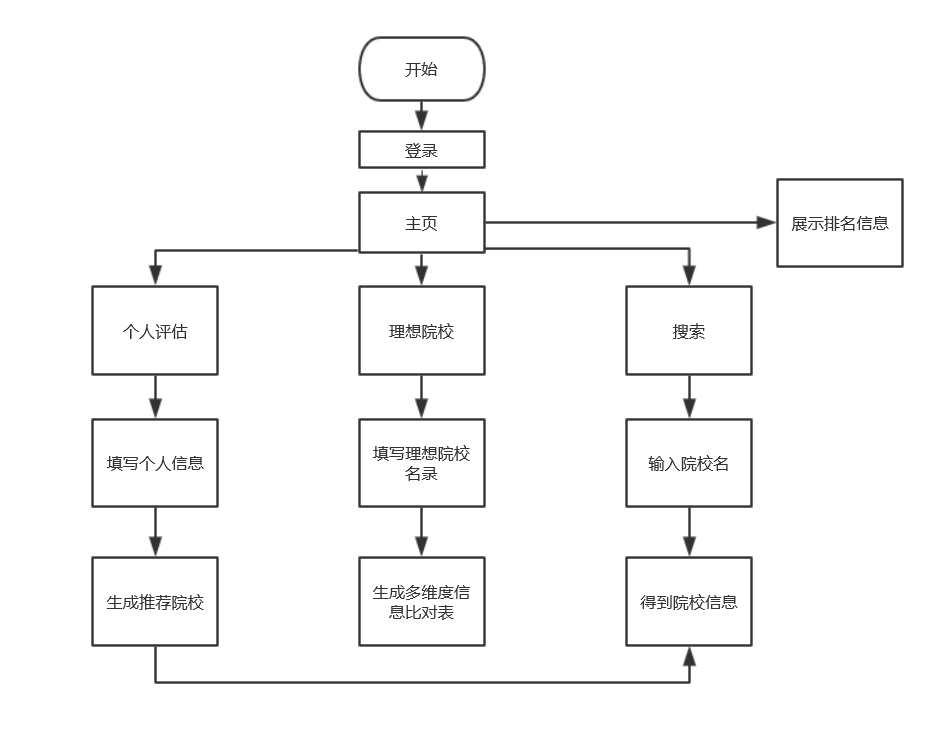
1,具体产品

数据库：MySQL，包含英美澳香港四个国家地区各自的学校表，各个国家热门专业排名表，用户信息表，爬虫

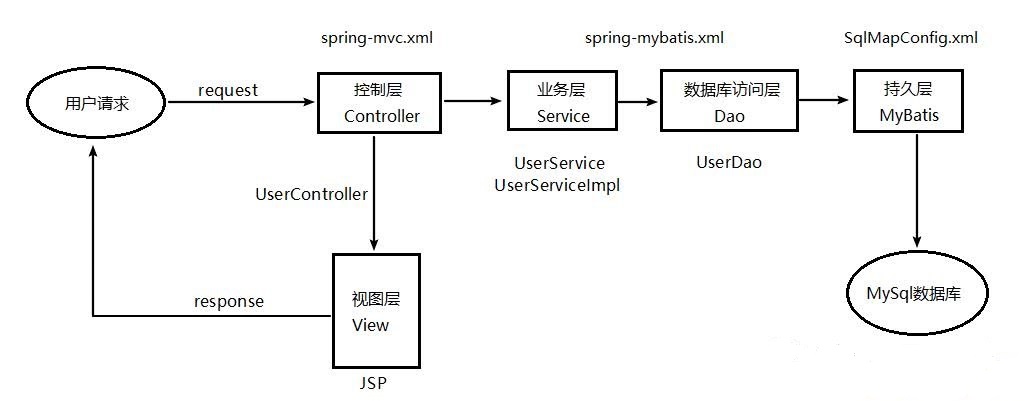
后端：SSM框架为核心，SpringMVC用于处理用户请求、控制层和视图层之间的交互；Spring利用IOC容器装载bean；MyBatis这一基于Java的持久层框架，使用简单的XML或注解用于配置和原始映射，将接口和Java的POJOs映射成数据库中的记录，完成数据的管理。

前端：使用Bootstrap前端开发工具包框架，利用Bootstrap的JS、css等显示出相应的效果出来，解决了众多兼容性问题，快速展示学校排名和专业排名等数报表页面。

2,系统功能

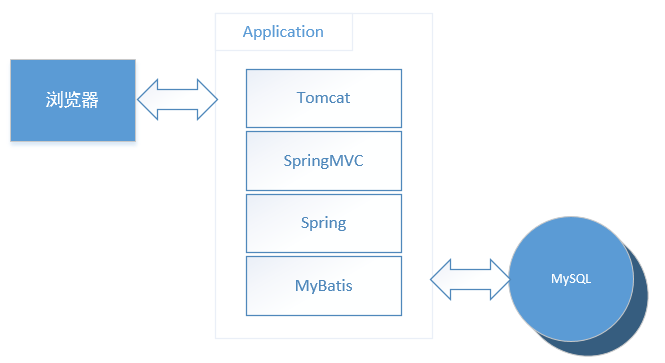


（1）技术架构

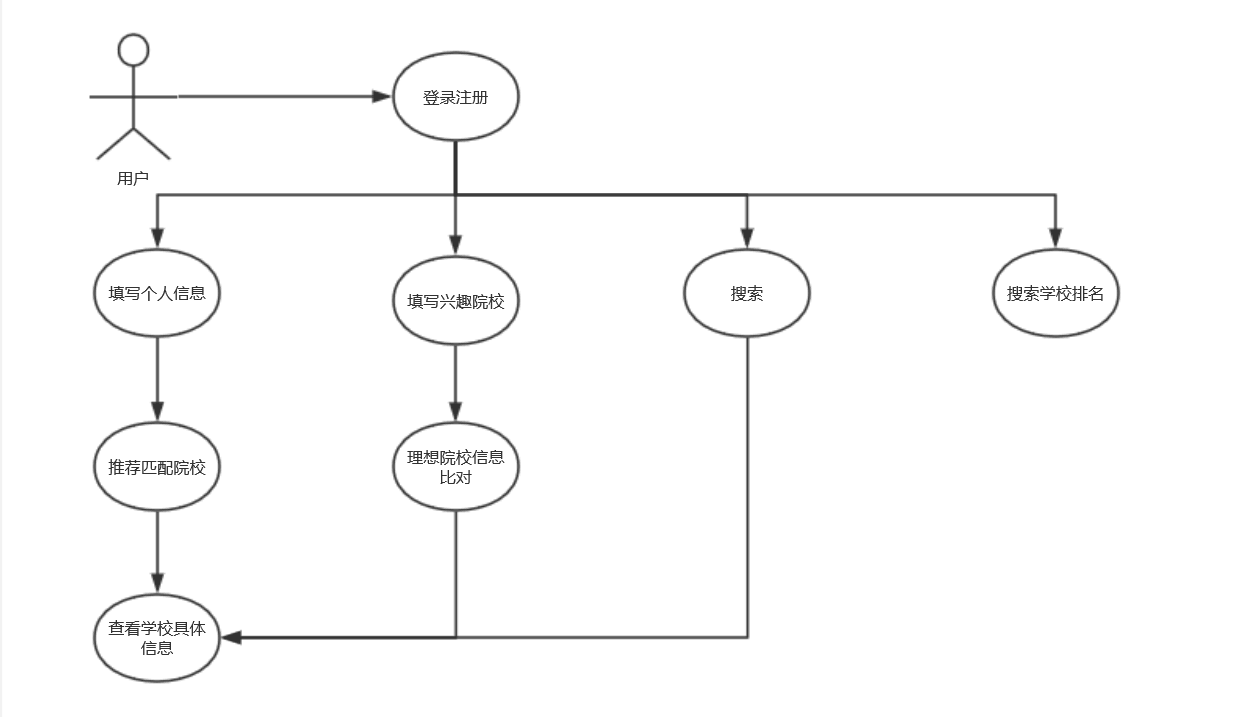


（2）基本流程

数据流图：



用例图：



工程实践进度:

|  |  |
| --- | --- |
| **时间** | **工作** |
| 2018.10-2018.11 | 文献调研、需求分析及开题 |
| 2018.11-2018.12 | 系统设计以及技术路线选定 |
| 2018.11-2019.1 | 爬取、处理和分析数据 |
| 2019.03-2019.05 | 数据库设计、界面设计 |
| 2019.06-2019.07 | 核心算法优化、系统测试和材料整理 |

此次工程实践的进度与开题初期预想的时间相当，实践流程进度推进也十分顺利，基本实现了预设的各个阶段的目标。

二、开发工作评价

1,开发效率

在这个工程实践项目中，我们预先拟定的计划是一个同学负责前端、一个同学负责后端、一个同学负责数据库。但在执行层面遇到了困难，因为发现最初的系统设计耦合度很高，导致各个环节不能分的很清楚，变成了有些页面是这个做的，有些页面是那个做的，有的算法是这个写的，有的数据是那个爬取的。这种混乱，应该变成深切的教训，这一切的开端应该追溯到不甚明确的最初系统设计。总体而言，虽然最后比较圆满地完成了任务，但是最初的系统设计如果能做得更好，开发效率应该可以大幅上升，这也从另一方面让我们知道了开题的重要性。

2,产品质量

虽然之前有一些混乱，但在测试阶段，代码表现还是不错的。并未发现明显的问题和无法更改的设计错误，各项环节运行良好。

3,技术方法

通过Spring的IOC特性，将对象之间的依赖关系交给了Spring控制，方便解耦，简化了开发；通过Spring的AOP特性，对重复模块进行集中，实现事务，日志，权限的控制，提供了对其他优秀开源框架的集成支持，SpringMVC是使用了MVC设计思想的轻量级web框架，对web层进行解耦，使我们开发更简洁。

bootstrap优势：简单灵活可用于架构流行的用户界面和交互接口的html、css、javascript工具集。基于html5、css3的bootstrap，具有大量的诱人特性：友好的学习曲线，卓越的兼容性，响应式设计，12列格网，样式向导文档。

4,主要问题及解决办法

(1)分页选择问题

后端开发分页时考虑了几种方案，一是直接从数据库中分页，二是前端分页，为保持前端页面的整洁统一，采用了自定义函数进行分页，也便于开发者使用。

(2)模糊查询展示问题

在条件不完全的情况下进行查询，1.用LIKE代替= 2.使用通配符代表在该位置存在任意一个字符,在该位置一定有一个字符,%代表在该位置存在任意多个字符，3利用contact函数连接字符串，4采用自定义分页函数展现给用户。

(3)个人评估功能实现—聚类算法遇到的问题

1.聚类特征的选取

特征选取是很难做决定的，因为不确定哪些特征影响比较小可以拿掉，哪些特征比较关键不可以拿掉。在这个情况下，一开始试着选取了世界排名、区域排名等信息加入聚类分析。后来发现世界排名比较特殊，很多区域排名相近的学校，世界排名差异很大，导致聚类结果出现偏差。于是就将世界排名这项去除，聚类才得到了比较好的结果。

2.缺失数据点的处理

k-means算法中，对缺失数据是非常敏感的，这里采用了平均值填充的方式解决。将相近学校的数据取平均填入缺失数据点中。

3.用户数据缺失问题的处理

这个问题仔细思考之后，发现6维特征中，用户缺失哪一条就忽略哪一条不计算其距离就可以圆满解决，但是代价是推荐精度的降低，用户至少在成绩排名上的数据真实才能够做出较好的聚类，这个问题可能没法避免。因为算法不能无中生有，确实需要一些关键数据才能开始匹配。

三、不足和展望

1,不足

页面设计不统一：因为多个页面的编写任务分配到各个组员，因此页面的样式风格不一致，不能达到比较美观的统一页面。

部分功能比较独立，互相之间缺少联系；各个功能模块的细节可以更加优化。

2,展望

该留学生信息管理与分析系统还缺少信息社区功能模块，在未来该系统可以在原有的基础之上添加留言板、评论回复，还有对一些相关网站信息的爬取，将这些的链接添加在页面上。

该系统将会在未来完善这些功能，届时我们的信息与分析系统会成为意愿出国学生的得力助手，管理分析更加精准，更加便捷地服务出国留学人员。