《Python程序设计》大作业

实验报告

实验题目： 高校就业信息网爬取

姓名： 林亦旸

学号： 2020212142

日期： 2022/1/07

1. **需求分析**

本次作业分为两个部分进行。第一部分，是爬取北邮就业信息网，电子科技大学研究生就业信息网和西电就业信息网的招聘信息。第二部分是对爬取下来的数据做数据处理，得到最受关注几条的招聘信息。

首先，对需要爬取的网站进行页面分析。北邮就业信息网是动态页面，需要用selenium模拟用户打开页面或用json文件爬取。发现北邮就业信息网的翻页是通过脚本控制的，不能靠修改url实现翻页。而我们需要进入招聘信息二级页面进行爬取，这样会导致翻页无法正常进行。除非用一个浏览器爬取目录，另外一个爬取二级页面。但这样做多占用系统资源。故在源代码中查找读取json文件的脚本，找到获取json的方法。用json文件爬取，不仅速度快，内容也好解析，对网站服务器造成的负载也小。因此对于成电就业网我也采取了这种爬取方式。

1. **模型设计**

本次作业分为两个部分进行

1. **系统设计**
2. **数据分析**
3. **作业小结**