**一、目的**

本文从拉勾网爬取全国互联网公司部分职位信息,并持久化存储到Mysql数据库中,之后进行前端页面展示。

**二、项目干系人**

刘瑞杰

付广鑫

张策

**三、用到的软件及技术**

Python版本：Python3.6

Django ：数据的提供与展示

Mysql ：持久化存储

Requests ：下载网页

Time ：暂停进程

**四、数据需求**

岗位名称：python开发工程师、运维工程师等岗位

岗位类型：前端、后端、区块链等工作类型

薪资水平：显示岗位的薪资区间

工作经验：显示面试此岗位所需要的工作年龄

教育经历：显示面试此岗位所需要的学历水平

个人技能：显示面试此岗位所需要的技能水平

公司详情：面试的公司名称、类型、规模人数、所在城市等信息

发布时间：显示此条招聘信息的发布时间

其他信息：职位诱惑、职位描述、公司地址

**五、项目功能**

1. 将拉勾网爬取到的信息，展示到前端页面

2. 按照薪资进行排序

3. 数据的横向竖向显示

4. 控制页面字段的显隐状况

5. 实现数据的分页功能，并且每页显示10条数据

**六、爬取与展示步骤**

1. 使用requests模块进行模拟请求获取cookie

2. 携带cookie进行第二次访问，发起POST请求，爬取到当前页面的pisitionid,contentid，及其他岗位信息

3. 拼接职位详情页的url，url = 'http://www.lagou.com/jobs/%s.html' % pisitionid，发送GET请求获取职位详情页信息

4. 拼接公司详情页的url，url = 'http://www.lagou.com/gongsi/%s.html' % companyid，发送GET请求获取公司详情页信息

5. 将爬取到的需要数据持久化存储到Mysql数据库

6. 使用Django框架进行数据展示

**七、项目不足**

1.薪资是升序排列的，升降顺序功能未实现，显示升序是在数据库中怕排好的顺序

2.数据爬取采用单线程，未应用到多线程，爬取效率低

**八、成果与展示**

****



