

个人资料

姓名:	乔艳冰	籍贯:	河南省商丘
年龄:	24	民族:	汉
性别:	男	毕业院校:	郑州轻工业大学
学历:	本科	专业:	网络工程(物联网技术)
移动电话:	17634329307	电子邮件:	qiaoyanbing1@163.com
QQ:	599426441		

求职意向

应聘职位:Python 工程师

专业技能

- 1:具有良好的编程功底,精通 Python 编程,良好的代码风格。
- 2:熟练掌握运用 Python 采集各类网站数据,多线程,多进程和队列的使用。
- 3:熟练使用正则表达式、CSS、Xpath 等网页信息抽取技术,熟悉 http 协议。
- 4:精通使用 Selenium、Phantomjs, Splinter 等工具的使用以及自动化设置。
- 5:熟练掌握 Scrapy 框架完成数据的采集,分布式爬虫的设计与实现。
- 6:熟悉利用 Redis 对 Scrapy 控制爬取数据的分布式的操作。
- 7:了解 Cookie 的登录原理,分析多个网页的抓取规则,以及一些数据清洗。
- 8:熟悉 Django Web 开发框架, Nginx 项目部署以及了解 docker 的使用。
- 9: 熟练掌握 HTML, JSP, JavaScript, CSS 语言, 熟练使用 AJAX, 以及熟悉前端 bootstrap 的一些使用。
- 10:熟练掌握 Linux 常用命令的使用以及开发环境, 熟悉 Docker 容器。
- 11:熟练使用 MySQL 数据库编程, 熟悉 Redis 的使用。
- 12:熟练掌握 Java 编程语言, 熟练使用 Spring、Spring MVC、Mybatis、Spring Boot。
- 13:了解关于 Python 机器学习的一些算法并且对机器学习和人工智能领域很感兴趣。

实习经历

2017.4 - 2018.2 上海润欣科技股份有限公司

Python 工程师

- 1:负责采集新闻网站数据的采集的,以及新闻的一些清洗入库工作。
- 2:负责一些器件型号等各种信息的采集工作,IP 代理的爬取及使用。
- 3:使用 Selenium、Phantomjs 完成一些数据的采集以及自动化设置。
- 4:通过 API 完成对流量宝的自动化设置,以及维护、改善现有的爬虫的运行。
- 5:使用 Scrapy-Redis 分布式对目标网站数据的抓取。
- 6:通过使用 Fiddler 抓包对手机 APP 数据的抓取。

实践经验

电子元器件和新闻爬虫系统

基于 Scrapy 爬虫对电子元器件的详细信息的抓取以及各大新闻网站的新闻信息的抓取，通过对目标网站的分析，使用内置 Xpath、Css 对网页内容结构化的提取，其中也包括对从动态网页的分析，对网页中特殊字段使用了正则表达式进行采集，在数据清洗的过程中也使用了正则表达式和一些字符串内置函数对数据进行处理，存入数据库中，其中使用了自己自建的 IP 代理池，大部分都是从网上抓取过来的，但是不是很稳定。

垂直搜索引擎系统（Django+Elasticsearch+Scrapy-Redis）

基于 Django 的垂直搜索引擎，利用 Scrapy 结合 Redis 实现分布式的数据抓取，抓取了部分网站的数据，上传至 Elasticsearch，利用 Elasticsearch 搜索服务器搭建了一个搜索引擎，主要通过 Django 与 Elasticsearch 交互，完成搜索功能，使用 Nginx 进行项目的部署。

豆瓣推广系统

主要调用了 Selenium 接口控制浏览器模拟人为操作。在豆瓣的各个小组内推广自己的产品，其中包括豆瓣账号注册模块，手机号由第三方提供，使用云打码来破解验证码，一个 IP 对应一个账号来防止被封；小组添加模块，每个账号可以指定添加任意一个或多个小组；小组发帖模块，在自己所添加的小组，进行指定内容的发帖；小组顶贴模块，在小组内，对指定标题进行顶贴，从而达到推广自己的产品。

自我评价

自己平时抓取的网站也有很多比如知乎、豆瓣、京东、手机 APP、今日头条等等网站，其中自己遇到过反爬措施有 User-Agent、IP、验证码、登录、加密等等，自己也找出了相应的解决办法。同时自己具有较强的沟通能力、学习能力和团队合作能力，能够勇于接受挑战，渴望学习新的知识，优秀的分析问题和解决问题的能力，对解决具有挑战性的问题充满激情。

教育背景

2014.9 - 2018.6

郑州轻工业大学