



黄章健

26岁 男

求职意向：Python工程师

手机：18850383607

邮箱：hzjkaka@163.com

基本信息

居住地：安徽 六安

民族：汉

求职意向

期望职位：Python工程师

期望薪资：8000-12000元/月

求职状态：随时入职

资格证书

计算二级(java)

2016-10 永久有效

普通话二乙

2014-10 永久有效

英语(CET-6)

职业技能

Xpath、bs4、正则和PyQuery的数据解析 熟练

Scrapy和Scrapy-Redis 熟练

基于Cookie和JWT的网站登录原理 熟练

动态信息的爬取以及Selenium的使用 掌握

MySQL和MongoDB的数据存储 掌握

使用Charles进行APP数据采集 掌握

处理验证码以及使用打码平台 掌握

动态User-Agent、IP的设置 熟练

Python多线程加速爬取 熟练

Scrapy+Gerapy 分布式管理 掌握

兴趣爱好

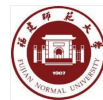
足球(挚爱) 读书(记笔记) 骑行

福建省“双一流”建设高校软件工程(信息安全方向)统招硕士毕业,2年Python脚本编写经验,尤其擅长Python网络爬虫。熟悉基本的Python知识,熟练使用爬虫框架Scrapy以及使用Scrapy-Redis搭建分布式爬取环境和掌握着APP端数据采集方法,且有着各类反爬问题和数据解析问题的处理经验,对捕捉最新Python数据应用技术充满热情。

教育背景

2017-09 至
2020-06

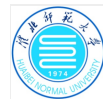
福建师范大学
软件工程·硕士



成绩专业排名前4,国家基金面上项目“5G跨云模式的数据隐私保护与抗mitm攻击方法研究”项目组成员

2013-09 至
2017-06

淮北师范大学
数字媒体技术·本科



成绩专业排名前10,虚拟校园设计项目组成员

项目经历

2018-12 至
2020-04

国家基金项目：电商数据的爬取与隐私保护

--项目概述：对电影评分数据进行隐私建模以及设计隐私保护算法。

--主要职责：1.参与项目数据集(豆瓣等)采集的需求分析,并编写相关爬虫;
2.对反爬策略进行分析应对和对数据设计隐私保护算法。

--主要技术：1.采集框架：Scrapy+Charles+Chrome;
2.数据解析：Xpath、jsonpath和正则表达式;
3.反爬：封IP(IP代理池和设置下载延迟);验证码(PIL+OCR
,OpenCV+Selenium);字体(手机端绕过且用Fonttools建立映射表);
4.数据存储：MongoDB和MySQL;
5.算法设计：基于本地差分完成隐私建模和数据的混淆;

--项目收获与成果：1.深度参与到7人的研究团队中,通过周例会提升了协作效率;2.完成两种基于评分数据的隐私保护算法的设计和一篇CCF C类论文的投稿发表。

2019-09 至
2019-10

Python数据分析：微博和豆瓣评论数据分析

--项目概述：爬取关于隐私保护的微博的评论和豆瓣中电影短评数据进行分析。

--主要技术：1.采用Scrapy中CrawlSpider爬取简书中文本挖掘的相关知识;
2.爬取：requests+Xpath+正则;
3.反爬：下载延迟,携带cookie登陆;
4.存储：保存为CSV文件;
5.分析：SnowNLP(情感分析),jieba+wordcloud(词云制作)。

--结果：得到微博和豆瓣短评的词云图以及关键字条形图。

2019-08 至
2019-09

Python数据采集：房源网站数据采集

--项目概述：爬取链家 and 房天下 的房源数据。

--主要技术：1.采用Scrapy框架爬取链家上海的租房数据;
2.采用Scrapy-Redis搭建分布式环境爬取房天下中各城市新房数据;
3.反爬：设置下载延迟;使用下载中间件设置动态UserAgent和IP;
4.数据存储：提取的数据存储到MongoDB中。

在校经历

奖学金

2017-12

研究生：学业奖学金三等(3次),华为杯研究生数学建模大赛成功参与奖

2014-11

本科：暑期实践先进个人;校级三好学生;校级三等人民奖学金