

庞彬

留一片空白，随时浓墨重彩。

男 | 24岁 | 13939105640 | 13939105640@sina.cn | 两年工作经验

求职意向



Python爬虫 全职 江苏南京 薪资面议 随时到岗

教育背景



2012.9-2016.6

南阳理工学院

软件工程移动

主修课程：
计算机网络原理、C++、Mysql数据库原理、计算机操作系统、Java、Python、Android、IOS操作系统、嵌入式、系统分析与设计、Html5

工作经验



2015.10-2017.11

深圳市百诺信科技有限公司

Python 爬虫工程师

1、负责设计和开发网络爬虫系统，进行多平台信息的抓取和分析工作；
2、负责网页信息数据抽取、清洗、消重等工作，提升平台的抓取效率；
3、负责爬虫核心算法和策略优化，采集系统的调度策略；
4、实时监控爬虫的进度和警报反馈。

专业技能



1. 计算机专业，具有良好的计算机基础与逻辑思维能力；
2. 熟练使用Scrapy、Django等python主流框架，实现数据发掘和后台开发；
3. 熟悉Mysql、Mongodb、Redis等数据库，熟悉Linux平台开发；
4. 能爬取各种主流网站和App，例如：微信公众号、内涵段子（App）等；
5. 熟练使用Selenium、Xpath、Bs4、正则表达式等网页提取技术，对html标签精准解析；
6. 了解Nginx、vsftpd、celery、负载均衡、分布式等运维技术；
7. 对搜索引擎、数据分析、数据展示、有一定的了解和使用经验；
8. 了解Bootstrap、HTML、Jquery等前端开发技术结合Django将数据前端展示；
9. 熟练使用Git、SVN进行源代码管理，协助他人

项目经验



去哪网，携程网团购信息爬虫

责任描述：

1.负责分析去哪网，携程网团购验证机制，网页结构，制定爬取策略；
2.负责使用scrapy框架编写爬虫代码，对各大旅游城市团购数据进行采集，包括酒店，美食，旅游，门票数据；
3.使用redis部署分布式爬虫任务，充分利用服务器资源，提高爬取效率；
4.使用BloomFilter做去重工作，优化去重任务，提高空间利用率；
5.负责编写shell脚本，定时开启爬虫任务，并编写爬虫警报脚本代码，及时发现爬虫异常；
6.负责爬虫代码的维护和优化。

新浪微博爬虫

责任描述：

- 1.根据需求，负责使用多个微博账号进行微博数据爬取；
- 2.负责微博账号登陆验证等模块的开发，使用云打码工具进行验证码识别，根据不同验证码类型执行验证操作；
- 3.研究新浪微博反爬虫策略，使用redis存储用户cookie值，维护cookie可用性；
- 4.负责将爬取数据去重，并将数据异步写入数据库；
- 5.负责代码的维护和优化。

美团外卖数据爬虫

根据需求，对美团外卖的商家数据进行抓取，主要针对优选商家，精产品，优惠计划，评论数据进行提取。

责任描述：

- 1.负责分析美团外卖网站结构，制定爬取策略；
- 2.负责用scrapy编写爬虫代码，通过api接口获取街道地址的经纬度；
- 3.通过xpath提取美团外卖中的店铺信息，如：商家信息，订单数量，商品价格，商品种类等数据，并存入redis数据库；
- 4.负责部署爬虫项目，并对项目进行监控，异常错误发送邮件提示；
- 5.负责爬虫代码的维护和优化。

国家企业信用信息公示系统爬虫

根据需求，对国家企业信用信息公示系统数据进行抓取，主要针对各公司注册状态，统一社会信用代码，成立时间，法定代表人等数据进行提取。

- 责任描述：
- 1.负责分析该网站的网站结构，指定爬取策略；
 - 2.负责用selenium模拟用户行为，解决滑动验证码的验证机制，获取详情页数据；
 - 3.通过scrapy利用xpath对详情页进行解析，批量提取详细数据；
 - 4.利用布隆去重机制，对爬取的数据去重，并通过异步写入的方式存入数据库；
 - 5.负责代码的维护和数据库的优化。

自我评价



- 1.热爱生活、热爱编程，对新事物接收能力比较强；
- 2.适应能力强、工作细致谨慎，有进取精神和吃苦耐劳精神；
- 3.喜欢浏览博客和技术文档，经常浏览Github、推酷、微信公众号学习Python的相关知识，了解python发展新动向；
- 4.具有良好的团队意识，喜欢钻研与探究，能够适应短期出差。
- 5.能够根据需求学习新的知识技能。