

段浩鑫_python爬虫_3年

电话：185-2630-3496 | 邮箱：18526303496@163.com | 地址：北京

微信：1852603496

年龄：23岁 | 性别：男 | 民族：汉 | 婚姻状况：未婚

当前状态：离职 | 求职意向：python爬虫工程师 | 期望薪资：10-20

个人总结

回首过去，工作已近乎三年有余。有成长也有不足。工作中更善于思考乐于与同事共同完成相关任务。敢于承担，勇于面对困难，对工作热诚有责任心。对生活饱含热情，也曾年幼无知茫然过，现在更懂得珍惜生活，认真对待事物，平心静气享受生活。是态度和思想的转变成就了现在的我。虽无惊人之才华，但现在的我足以面对工作生活中的种种问题。对未来的生活充满信心，也将会更努力完善更美的人生。

专业技能

- 1、熟悉Python语言，熟练使用Python标准库和流行的第三方库
- 2、熟悉面向对象的设计思想,熟悉 Django Restful API 接口规范，以及 Django-rest-framework 框架；
- 3、熟练掌握mysql、mongodb、redis数据库操作
- 4、熟练掌握jQuery，html，css，bootstrap，vue，mui，Ajax等前端技术
- 5、熟练掌握Django/Flask/Tornado等常用开发框架
- 6、熟练 Scrapy 框架，掌握 Scrapy-Redis 搭建分布式
- 7、熟悉正则表达式和XPath，能够从结构化或非结构化数据中获取信息；
- 8、熟练掌握抓包工具 Fiddler，熟练掌握Pndas对数据进行清洗
- 9、熟悉应用Selenium+PhantomJS实施动态HTML抓取
- 10、熟练掌握Linux运维经验，能独立安装、配置、搭建开发环境，编写shell脚本；
- 11、熟悉常见反爬机制，验证码识别，IP代理池、应用Ip池、headers认证和cookie等
- 12、熟悉版本控制工具 Git、以及 Gitlab 的管理维护；
- 13、具备强烈的责任心，对工作有激情，良好的沟通能力，良好的团队合作精神，能够承受一定的工作压力

工作经历

北京量子伟业科技股份有限公司 - python开发工程师 华北-开发项目部 2017年3月 - 2017年11月

- 1、参与Web后台系统架构设计、开发、优化和维护；
- 2、参与消息系统的架构设计、开发、优化和维护；
- 3、编写相关技术文档，编写核心测试案例；
- 4、跟踪和调研最新技术热点，推进新技术在研发中使用；
- 5、对产品进行持续优化
- 6、配合项目的整体计划和安排，参与核心技术问题的攻关，系统优化，协助解决

北京创智信科科技股份有限公司 - python开发工程师 研发部 2017年11月 - 至今

- 1、负责公司项目后台的维护、性能优化、后期迭代。
- 2、负责所属模块技术文档的编写及维护、项目需求报告。
- 3、负责数据爬虫后台程序的开发，负责爬取数据的简单清洗工作。对在线产品的维护，及部分模块的开发。
- 4、设计和实现公司相关产品的后台架构设计以及相关模块的开发。
- 5、快速定位及解决开发、运行过程中的技术问题。
- 6、参与项目需求分析，业务逻辑流程与设计。
- 7、规范文档的编写、维护，以及其他与项目相关的研发工作。

开源项目及作品

吉屋网站房产数据爬取 爬虫工程师 2017年6月 - 2017年7月

- 1、使用 Scrapy 以及 Scrapy-Redis 组件来构建分布式爬虫系统

- 2、使用 lxml, bs4 等解析库对页面进行解析
- 3、编写高匿代理中间件和客户登录中间件
- 4、将爬取到的数据持久化到 redis数据库中
- 5、使用 User-Agent 设置用户头信息
- 6、根据页面不同的url和数据, 编写规则和非规则爬

招聘网站信息爬取 (含app) 爬虫工程师

2017年10月 - 2017年11月

- 1、使用 Scrapy_Redis 爬虫框架实现分布式爬虫, 针对特定网站的信息进行采集。
- 2、使用 Selenium + PhantomJs 对部分动态网站的信息进行采集。
- 3、使用 Xpath 来定位数据, 提取有效内容
- 4、使用 sha1 加密的方法将数据存入 redis 中实现数据和 url 去重
- 5、使用 Mysql 进行信息存储, 针对异常进行相应处理
- 6、无忧智联登陆界面验证码的突破以及自动化实时抓取
- 7、在远端服务器上分布式抓取部署
- 8、对数据进行持久化抓取以及实时抓取

养养宠物 程序开发

2017年12月 - 2018年6月

- 1、基于Flask轻量级Web框架开发, 从而减少了开发周期。
- 2、使用Flask-Migrate插件来实现数据库迁移。
- 3、在页面样式上大多用Flask-bootstrap引入cdn, 减轻前端页面开发带来的成本。
- 4、渲染模板使用的是jinja2, 因为它形式简单, 并且包含响应文本。
- 5、采用Nginx进行反向代理非负载均衡来提高并发量。
- 6、采用cookie实现页面的登录跳转以及使用session保存用户登录信息, 判断用户状态。

开发环境: CentOS6.9 Python3.6 + PyCharm

应用技术栈: Flask + MySQL + Nginx + uWSGI + Ajax + jQuery

狗命网 程序开发

2018年6月 - 2018年11月

技术概况:

- 1、使用Celery消息队列来异步发送短信激活账户。
- 2、使用redis来缓存购物车信息和最近浏览商品信息, 以及频繁访问的信息。
- 3、通过Whoosh + jieba分词来实现搜索功能的实现。
- 4、使用djano-restframework-jwt插件实现鉴权功能。
- 5、使用Vue实现前端功能, 通过axios发送请求。
- 5、使用xadmin搭建网站的后台管理系统。
- 6、GenericViewSet和Router完成API接口和URL配置。
- 7、采用Nginx进行反向代理和负载均衡配置。
- 8、集成第三方SDK登录以及通过WebServer调用支付宝第三方接口

责任描述: 登录注册模块, 购物车模块, 用户中心模块, 根据业务逻辑设计部分表结构

开发环境: Linux7.3 + python3.6 + PyCharm

应用技术栈: Django + Django-restframework + MySQL + Redis + Nginx + uWSG

萌宠比价网 爬虫开发

2018年11月 - 2018年12月

技术描述:

- 1、创建IP代理池, 解决IP被限制的问题, 并定期维护无效IP。
- 2、使用fake_useragent库, 随机更换每次请求的User-Agent。
- 3、设置禁用cookie, 使服务器无法根据Cookie判断爬虫是否访问过。
- 4、使用xpath, re等模块来提取数据。
- 5、使用pandas、numpy、Series清洗数据

教育经历

河北科技大学 - 计算机科学与技术 本科 计算机学院

2012年9月 - 2016年10月