

求职简历

个人信息

姓 名:	仲俊	性 别:	男
出生日期:	1993 年 1 月 22 日 (25)	居 住 地:	深圳 宝安
工作年限:	3 年	电子邮件:	564521939@qq.com
联系电话:	150 1355 0975	户 口:	湖北 黄冈



求职意向

工作岗位: Python 爬虫工程师 薪资: 面议

个人技能

- 熟悉 Python 的基础语法和常用的数据结构, 了解并努力遵守 PEP 8 编码规范。
- 会配置爬虫项目需要的 Windows 环境。了解一些 Linux 知识, 能配合公司学习和使用 Linux 环境。
- 熟悉 scrapy 爬虫框架, 能够完成常见的爬虫项目, 并且能够根据需要编写合适的 downloader 和 spider 中间件。
- 熟悉 HTML, XML 等网页标记语言, 了解 CSS, JavaScript, AJAX, JSON 等 web 前端知识。
- 了解 HTTP/HTTPS 协议, 熟悉 HTTP 的首部字段和状态码信息。了解一些抓包工具, 如 Charles, Fiddler。
- 能够熟练使用正则表达式、XPath 和 CSS 选择器从结构化和非结构化的数据中获取信息。
- 会使用 selenium-scrapy 爬取动态页面, 会基于 cookie 的页面登录爬取。
- 会 MongoDB 和 MySQL 存储爬取的数据。熟悉这两种数据库的增删改查。
- 以上技能均个人自学, 有一些会用但不熟的地方, 期望能在贵公司努力的工作, 更好的进步。

项目经验

1. 使用 scrapy 框架爬取了你我交流的社区的所有帖子主题及其链接。
你我交流社区的帖子主题页总共有 448 页, 需要进行翻页处理, 防爬措施采用了检测 User-Agent。整体上获取信息比较容易, 需要注意的时网页结构的前后不一致, 需要采取不同的提取信息 xpath 路径。项目中使用了 Faker 中间件来切换 User-Agent, 设置请求延迟时间为 5s, 来规避防爬措施和减轻你我社区的服务器压力。github 地址: https://github.com/lanluyu/niwo_bbs。
2. 使用 selenium-scrapy 爬取东方财富网的个股信息。
爬取了东方财富网的千股千评页面的所有个股信息。其个股信息采用了 js 加载的, 所以这里采用了模拟浏览器的工具 Selenium 结合 Scrapy 框架的方式来爬取股票信息。Selenium 采用了 headless mode 无头模式和无图片模式提升了爬取速度。最终爬取了 3659 条全部股票信息, 爬取的数据使用 MongoDB 存储。github 地址: <https://github.com/lanluyu/dongfang>

教育经历

2011 /9 -- 2015 /6 武汉生物工程学院 本科 机械设计制造及其自动化
英语 CET-4 (422) 读写良好, 听说一般。

个人介绍

- 1) 学习能力强, 自学计算机, 善于逻辑思考, 能够从整体上分析、把握复杂的问题, 对细节难题具备 快速钻研和解决的能力。
- 2) 有 3 年的工作经历与社会阅历, 能快速地融入团队, 与团队合作解决工作中遇到的困难和问题。
- 3) 良好的表达与沟通能力, 能够积极主动的完成上级安排的工作任务, 对工作尽心尽责。
- 4) 个人也对金融业十分感兴趣, 爬虫爬取的金融数据能够帮忙发现一些金融规律。两者结合能创造更多价值。