

地址：广东省深圳市

邮箱：[805083697@qq.com](mailto:805083697@qq.com)

年龄：22岁

电话：13924668380

姓名：陈子杰

求职方向：python爬虫,数据挖掘，数据分析，开发岗位

目前工作状态：随时上岗

期望薪资：7000-9000

教育背景

2014.9-2018.6 北京师范大学珠海分校 数字媒体技术

主修课程：

C#程序设计实践，数据结构，数据库应用技术，计算机网络，Web开发项目实践（Asp.net），大数据处理与数据挖掘，数字图像处理，三维建模与动画，视频剪辑

奖项证书：

第十四届“挑战杯”广东大学生学术科技作品竞赛三等奖

信息技术学院三等奖学金

大学英语四级证书

技能特长

掌握requests向网站发送请求，获取响应后处理网页数据

掌握使用beautifulsoup、xpath、正则表达式提取网页元素并处理

熟悉使用Mysql、MongoDB、Redis的在python中增删改查的应用。

熟悉scrapy框架流程，懂得如何使用框架采集数据，并保存数据。

了解Selenium+PhantomJS模拟登陆并提取网页元素。

了解处理基本的反爬虫手段，如随机cookies，ip代理，User-agent和验证码问题。

熟练HTML，CSS基础知识，使用过bootstrap框架开发网站前端，使用过ASP.NET开发过网页后端，对Javascript/Jquery有一定的了解。

掌握Python语言基本语法，函数式，面向对象编程，会用pandas，matplotlib进行数据分析和实现数据可视化。

掌握利用python等库对数据进行去重、清洗、整合并保存。

掌握excel操作，熟悉使用数据透视表和图表工具，懂得使用vlookup等函数对数据进行操作。

了解部分机器学习算法，如KNN，决策树，朴素贝叶斯，随机森林等。

熟悉使用Photoshop，Axure软件进行UI交互设计。

自我评价

● 自我学习能力强，懂得利用搜索引擎解决难题，喜欢从实战中学习经验。

● 对自己要求严格，做事情讲究高质量，乐于接受新技术。

● 为人友好，能快速融入团队，懂得协作。

工作经验

2018.8-2018.11 新潮文化传媒公司 职位：数据分析员

● 工作职责：负责每日销售经营部数据统计，汇总，并在系统中做出增删改查。进行日常业务相关文件数据的录入、监管、后期处理。

● 主要项目描述：该项目是利用python的pandas，numpy库开发一个脚本，能对大量的数据进行简化和数据处理并且存储至相应的表。该脚本将本需2小时制作时间减少至10分钟。

● 责任描述：该项目我使用python结合numpy、pandas包，对数据实现引入、将excel数据转换为矩阵、根据每个表的要求进行预处理和数据清洗，数据处理完后保存并填充到相应表中。

项目经验

其他部分爬虫项目：https://github.com/armilkchen

2019.1-2019.2 利用scrapy对新浪微博进行数据采集

● 项目描述：该项目是通过爬虫技术对新浪微博进行采集，以几个新浪微博账号为起始点，爬取他们各自的粉丝和关注列表，然后获取粉丝和关注列表的粉丝和关注列表，以此类推，实现递归爬取。

● 责任描述：该项目我使用了Scrapy框架来进行爬取，首先通过分析网页找到需爬取的元素源头，通过requests请求获取response并分析，然后提取相应的信息，并对某些数据进行清洗，最后储存至MongoDB。因为新浪微博有反爬取机制，我使用了之前完成的cookies池和IP代理池，每次都随机抽取一个cookie和ip来发出请求来应对反爬。

2019.1-2019.2 豆瓣图书的数据采集

● 项目描述：该项目是通过数据采集把豆瓣图书每个分类的每本书籍的书名、作者、价格等数据爬取并储存到Mysql中。

● 责任描述：该项目首先我通过分析网页发现每个分类的网址区别，根据不同分类生成不同的网址，再将网址进行二次爬取将每本书的元素提取出来。该项目我使用的是requests+beautiful soup来进行网页分析提取，因为豆瓣禁止同一ip频繁的爬取的反爬机制，我先是利用time.sleep函数进行微调，设置每爬取一个分类随机休息5-8秒，爬取的数据有变多但是还是会被封IP，最后只能使用代理IP来进行爬取。

2018.7-2018.8 对数据分析师职位信息进行数据分析

● 项目描述：该项目是通过网上招聘APP中关于“数据分析师”这个职位的信息来当做原始数据，进行了一次数据分析，可以分析得到该职位受到哪些因素影响。

● 责任描述：该项目我使用了两种方法实现，一种是使用Excel的数据透视表结合图表工具进行分析。另一种方法是使用python结合numpy、pandas，matplotlib包完成对数据实现分析、清洗、可视化。

2018.6-2018.7 kaggle泰坦尼克号生存预测

● 项目描述：该项目是通过机器学习算法去根据官方给出的训练数据生成合适的模型并预测测试数据中乘客的生存状况。

● 责任描述：使用python结合numpy、scikit-learn、pandas，matplotlib包实现数据分析，数据清洗，数据可视化，再使用相应算法的如随机森林、线性回归等机器学习算法去预测乘客的生存状况获取结果。

2017.9-2018.5 The Robbers-html5双人闯关游戏 项目组长

● 项目描述：该项目是一个html5双人闯关小游戏，需要两名玩家控制两个角色去躲避陷阱

障碍物，并在规定时间内获取足够钻石。

● 责任描述：使用Construct2可视化编程引擎去完成整个游戏的代码和逻辑实现，包括每个物件的交互设计，人物状态转换等。

2017.9-2018.5 Now You See Me-VR技术治疗恐高症 项目核心成员

● 项目描述：是利用VR设备和心跳脉搏检测设备去治疗患有恐高症的人群。用户可通过VR设

备进入3D场景，场景会让用户感到高度落差，从而检测用户心跳脉搏等数值。

● 责任描述：参与项目整体流程设计以及重大方案的讨论和审定。进行VR场景的设计以及搭

建，解决一系列Maya与Unity软件之间接口出现的难题。

● 获得成就：该项目获得第十四届“挑战杯”广东大学生学术科技作品竞赛三等奖

2017.3-2017.6 电影网站前后端制作 项目组长

● 项目描述：该网站可实现用户搜索电影，查看电影的详情，获取喜爱电影的推荐，并可注

册登录和修改个人信息。管理员可实现对电影和用户的资料管理包括增加、删除、修改和查

询。

● 责任描述：负责前后端的页面展示和编写。前端使用bootstrap框架加html+css所完成

的，后端使用asp.net框架完成。

2017.3-2017.6 个人记账APP制作 项目组长

● 项目描述： 用户可在此程序实现月结算、收入开支的计算、数据的写入、修改、删除、分

类、备注等功能。

● 责任描述：负责整个项目的代码，逻辑设计。使用C#语言在Visual Studio 2010开发完

成

2015.9-2016.2 美食网站前端页面制作 项目组长

● 项目描述：该网站可获得各种类型美食的详细制作方法和附近推荐的美食餐厅的详细情

况。

● 责任描述：负责前端的页面展示与编写。该网站可实现网站的基本交互操作，用户注册登

录，网页跳转，图片轮播，导航下拉菜单，个人信息的修改等功能。网站是使用html+css

制作的静态网页，图片轮播使用jquery技术完成。