

联系方式

- 手机: 18205089562
 - Email: derekdongsir@163.com
-

个人信息

- 董洪伟/男/1993
 - 本科/南京理工大学智能科学与技术系
 - 工作年限: 3年
 - 期望职位: Python开发工程师
 - 期望薪资: 面议
 - 期望城市: 北京
 - 目前状况: 离职(随时可到岗)
-

工作经历

南京知行科技有限公司 Python 爬虫工程师 2017年8月 ~ 2019.6

工作描述:

1. 根据采集需求书利用python爬虫技术爬取目标网站的数据
 2. 参与异步爬虫,分布式爬虫等爬虫框架的设计,针对遇到的反爬制定反反爬手段
 3. 通过大数据技术对采集的数据进行清洗
-

南京杰佳科技有限公司 Python开发工程师 2016年7月 ~ 2017年8月

1. 参与公司 Web 项目部分业务的需求分析,逻辑流程设计;
 2. 参与项目部署,大数据环境搭建,大数据清洗等技术的使用.
-

项目名称: 各类公示网站信息爬取 项目周期: 2018年6月 ~ 至今

项目描述: 公司主营业务,对政府,机构网站的公示数据进行爬取

主要工作:破解各种高难度反爬,例如滑块验证码,js加密,app采集等,设计编写aiohttp异步爬虫框架,通过redis-scrapy部署分布式爬虫,采集数据量达到每月上亿规模,通过大数据技术对数据进行清洗.

项目名称: 各类社交网站信息爬取 项目周期: 2017年11月 ~ 2018年6月

项目描述:对人人网,微博,世纪佳缘,百合网等社交网站用户信息进行爬取

主要工作:熟悉掌握使用scrapy进行大规模数据采集,解决采集过程中遇到的图片验证码,动态cookies,自定义字体等反爬手段,数据采集量达到每月6000万;后期参与公司内部分布式爬虫框架的设计和编写工作.

项目名称： 人才网站简历数据爬取 项目周期： 2017年8月 ~ 2017年11月

项目描述:根据公司的采集需求表对企业,政府的人才简历数据进行爬取

主要工作:使用requests爬取某些反爬难度较低的网站,通过编写ip代理池,加载随机headers等手段破解简单反爬机制,信息采集量达到每月2000万;

项目名称： 书虫图书推荐网站 项目周期： 2016年10月 ~ 2017年8月

项目描述: 在大数据环境下对网络图书资源进行整理和分类,通过对用户访问信息的统计分析,为用户精准推送其感兴趣的图书资讯.

主要工作:

- 1.参与项目部署,使用nginx,uwsgi做动静分离和负载均衡,使用redis做缓存;
2. 利用hadoop搭建HDFS和hbase集群,收集用户的访问信息;
3. 使用hive,sqoop技术对用户访问日志进行分析,并将处理后的结果转储到mysql数据库

项目名称： 某用人单位员工管理系统 项目周期： 2016年7月 ~ 2016年10月

项目描述: 该管理系统提供管理员登录和员工登录接口,管理员可对员工信息,员工账户及部门信息进行增删改查,员工可通过注册登录查询个人账户信息;

主要工作:

1. 参与了项目的需求分析讨论;
2. 参与登录注册模块代码的编写,包括图片验证码,手机号码登录,邮箱激活账户,异步请求返回账户是否已注册等;
3. 参与员工信息分类展示,批量删除,账户信息修改等功能的设计和编写,利用Ajax对页面信息进行动态更新

技能清单

- 熟练掌握 Python基本语法,OOP 编程思想,多进程,多线程,协程等;
- 熟练掌握 html,css,js,jq,Ajax,bootstrap框架等前端技术,熟练使用jq进行dom操作 ;
- 熟练掌握关系型数据库mysql的使用, 对sql优化, 事务控制, 索引有深入理解
- 熟练掌握django的使用和MTV设计模式, 能够快速进行web后端程序的开发
- 熟练掌握Linux的使用, 能够熟练使用uwsgi,nginx做负载均衡和动静分离
- 熟练掌握redis的使用及其持久化机制, 能够快速搭建分布式redis集群、
- 熟练掌握git工具进行版本控制和团队协作开发
- 熟练掌握hdfs做分布式存储, hive做数据清洗, sqoop和mysql做数据交互。
- 熟练掌握 Scrapy 爬虫框架,以及运行开发流程以及使用scrapy-redis做分布式爬虫
- 熟练掌握aiohttp异步请求框架的使用, 并对其封装
- 参与过分布式爬虫的设计和编写
- 熟练掌握常见的反爬机制和破解方式, 如滑块验证码, js 加密, 自定义字体等
- 熟悉word,excel等离线数据的清洗,熟悉python-docx pywin32 pandas numpy等模块
- 熟悉spark做数据去重和合并
- 熟练掌握使用pyecharts做数据可视化和数据报表
- 了解机器学习, 对线性回归算法, 逻辑回归算法, 聚类算法等有一定的了解

- 熟悉常见数据结构，如栈，队列，树等，以及常见排序，查找算法
- 熟悉http/tcp协议等相关知识

致谢

感谢您花时间阅读我的简历，期待能有机会和您共事。