

### 个人信息

手机: 18810384469

邮箱: [forward\\_lei@163.com](mailto:forward_lei@163.com)

微信: 52209153

现居: 北京

学历: 本科

工作: 2.5 年

### 职业技能

- Scrapy 框架和分布式爬虫框架 redis-scrapy 组件;
- 爬虫相关 Requests re Xpath BeautifulSoup 模块的使用;
- 掌握常用的反爬机制,熟悉 selenium 动态数据抓取;
- 了解 JS 解密,掌握线程,多进程,协程;
- 掌握抓包工具 Fiddler;
- HTTP 协议、Websocket 协议、TCP、UDP 协议;
- Numpy、Pandas、Mulplotlib 做简单的数据分析处理;
- MySQL、Redis、MongoDB 的使用;
- MySQL 中的事务、索引,具有一定的数据库优化能力;
- Redis、主从同步、持久化配置;
- Django、Flask 框架,理解中间件, cookie 和 session 原理;
- 使用 Django-rest-framework 框架,遵循 RESTFUL API 接口开发规范;
- HTML、CSS 的前端技术,及 Bootstrap、jQuery、Vue 的应用;
- Linux 系统常用操作指令,熟悉 git 的使用;
- Nginx 负载均衡、UWSGI 的配置及使用;

### 工作经历

2018/02—2019 至今

北京传臻科技有限公司 | python 爬虫工程师

2016/05 -- 2018/01

北京茵普兰科技发展有限公司 | python 工程师

### 项目经验

#### ● 艾比格特/iBIG OS

项目描述: iBIG OS 是一款无限移动硬盘,能够与各种移动设备和常用 APP 互通,整合数据; 其中该 app 中有电影功能,通过手机可以给无线硬盘进行电影离线下载,为了提高用户体验度,分别爬取各个网站的优质电影,从而提高用户认可度.

负责内容:

使用 Scrapy-redis 分布式爬取数据,对数据进行本地保存;  
通过 mysql 对电影图片,url 链接,名字等描述进行存储;  
对下载器中间件中做 UA 池和 IP 代理池进行配置;  
对编码问题进行处理;

项目心得: 处理中间件 UA 池 及 IP 池,加强爬取速度,在 setting 进行配置如处理增加并发、下载超时、禁止重试等提高爬取效率.

#### ● 爬取房产信息

项目描述: 为获取北京二手房各个地段交易信息,主要爬取我爱我家和链家 2 家交易信息,为客户提供更直观的信息展示.

负责内容: 对近半个月房产数据进行爬取分析,利用 xpath 解决网站数据采集,并用 crawls spider 做全站数据爬取,利用 redis 把爬取的 url 进行指纹处理,做到增量式爬虫,防止爬取重复信息,保证数据的一致性.

项目心得: 对于网站页面分析及反爬处理更得心应手,在运用 scrapy 框架,理解框架流程更加深刻.

- **iBIG Stor**

项目描述: iBIG Stor 是为打造智能互联家族而生的操作平台, 为移动存储产品和整个智能设备产业链的互动和链接创造可能性.

负责内容:

基于 rest-framework 编写后端接口;

登陆注册模块:注册利用阿里云发送验证码;

购物车模块:主要将所选的购物数据存放 redis 数据库进行操作;

支付模块: 采用的常用的支付宝和微信,对优惠券做了处理;

推送模块: 利用微信接口给用户推送信息;

其他模块: 编写首页,开放平台,商场等页面接口;

项目心得: 项目基于前端 vue,后端 django 开发的,做到前后端分离,对于短信,支付,微信推送调用第三方接口,提高了自己看文档习惯,并且把这些接口封装成组件,适用于各大框架.

- **小飞侠**

项目描述: 主要关注对象是 5 岁以下的儿童,让玩具成为父母间沟通的桥梁,用智能玩具成为儿童的乐趣,依靠孩子的语音指令让玩具做出相应的响应.

负责内容:

利用 scrapy 框架网上采集数据,利用 mongodb 数据库进行数据存储;

手机 app 用 MUI 前端布局,使用 HTML5 做系统硬件驱动,如摄像头,麦克风;

利用 flask 框架做后端处理;

调用百度 ai 中的做语音识别和语音合成;

调用图灵机器人接口做聊天对话;

利用 websocket 处理 app 与玩具之间通讯,及利用 redis 做消息缓存;

项目心得: 对 websocket 协议,nlp(自然语音处理),pypinyin,genism,jieba 分词有了更深得理解

- **蓝天系统**

项目描述: 该项目是为了完善销售流程的规范化及标准话,并对员工进行监管.主要用于公司销售、仓管、售后等人员提供公平的处理平台.为销售人员提供客户关系管理,对于新客户信息的录入,定期对客户的跟进和回访.为仓管和供应商提供关系,以及产品录入.

负责内容: 为避免销售之间抢单产生不和谐因素,实行公户和私户管理,公户放未指定的批发商信息可以进行 2 种方式分配到私户,一种是主管进行主动分配,第二种是销售自己申请.私户是存放自己的客户,在申请客户时基于数据库的事务/锁保证数据安全,从而保证安全见客户申请到自己的账户,此外做模糊查询能快速找到自己目标对象.

项目中继承了一个权限组件,主要用于对各种角色只能访问自己所有有的权限,做到无跨级操作.技能点就是用 url 做权限, 利用 Django 中间件和 Session 实现权限的管理, 自定义分页组件, 实现不同职位的不同权限, 保证了公司数据的安全性; 用户权限存放在 session 中, 用户访问时需要在中间件进行校验(白名单).

项目心得: 使用 querydict 管理 url 里的 query 参数,并对 mysql 事务有深刻定义.

## 自我评价

生活态度: 对于技术无比热爱,对于技术喜欢探索,并且定期归纳总结,不断完善自己的知识结构

工作态度: 工作认真负责,喜爱热爱爬虫,对于代码精益求精,可适应高强度工作.

性格方面: 性格随和,能快速融入团体,喜欢长跑,喜欢篮球,热爱运动.