

联系方式

- 手机：15868759135
 - Email：zhujingdi1998@gmail.com
 - QQ/微信号：986494553/b986494553
-

个人信息

- 朱静迪/男/1998
 - 本科：温州大学城市学院计算机系
 - 技术博客：<http://xiantang.info>
 - Github：<https://github.com/xiantang>
 - 期望职位：Java后端实习生/数据采集实习生
 - 期望城市：杭州/上海
-

开源项目和作品

Java项目

京东全站商品爬虫

- 简介：
 - 基于 `scrapy-redis` 和 `Spring` 的分布式价格监控平台。
- 技术细节：
 - 使用RabbitMQ作为消息队列，通过Quartz定时扫描数据库，发送邮件提醒用户价格更新。提高性能用redis进行缓存减少数据库访问压力。
 - 实现 `Spring-anti-spider` 插件，为恶意采集设置三个等级，分别是限制接口每分钟调用次数，检查UserAgent，用户IP封禁，并运用到项目中。

- 重构爬虫的 pipeline，使用**有限状态机**省去了繁杂的 if-else 代码。通过阅读和修改 scrapy-redis 源码，将原来基于 Redis set 的去重对列，改为基于布隆过滤器的实现，提高了去重的效率。
- 使用 scrapyd 的 api 编写脚本控制爬虫的状态，达到增量爬取的效果。
- 使用 docker 进行爬虫的批量部署。
- 总结：
 - 爬虫部署在三台 ubuntu 云服务器上，并稳定爬取。
 - 只需三天就可以完成对京东的 1100w 商品数据的价格监控，每日能够记录 3w+ 条价格更新。

共享单车微信小程序

- 简介：
 - 基于 Spring Boot 和腾讯地图的共享单车小程序。
- 技术细节：
 - 调用聚合数据的短信 api 发送短信，使用 Redis 对用户手机验证码进行为期 60s 的缓存。
 - 利用 MongoDB 的地理位置索引 2dsphere，对球面上附近地点的单车进行查询，实现在加载地图和移动地图的时候，加载附近的单车。
- Github：[共享单车微信小程序](#)

Android仿知乎日报

- 简介：
 - 一个轻量版的知乎日报，加载更快，体积更小。
- 负责模块：
 - 产品数据库设计
 - 首页模块设计
 - 重构项目
- 技术细节：
 - 使用 fiddler 确定知乎日报接口。
 - 使用 jsoup 对知乎接口返回的 HTML 进行裁剪。
 - 使用开源的 RecyclerView 代替原生的 ListView，利用 volley 异步请求库，进行数据获取。
- Github：[Android仿知乎日报](#)

Python项目

个人技术博客

- 简介：
 - 使用 `django` + `bootstrap` 搭建的个人博客。
- 技术细节：
 - 前端使用 `edit.js` 库，提升用户的写博文时的体验，后端使用 `markdown` 库，将 `Markdown` 文档渲染为 `HTML` 进行展示。
- Github:[个人技术博客](#)

工作以及开发经历

2018.07-2018.09 我和我们文化传播

- 简介：
 - 负责豆瓣全站电影数据增量采集。
- 技术细节：
 - 使用芝麻代理并编写代理池，解决豆瓣IP封禁。
 - 优化爬虫架构，将评论爬取和详细页面爬取模块解藕。
- 总结：
 - 爬取 `7w` 电影数据，以及 `500w` 电影评价。
 - 使用 `aiohttp` 以多进程的方式编写爬虫框架，增量爬取周期从 `7` 天降低到 `3` 天。

技术文章

- [爬虫遭遇状态码521陷阱 破解js加密cookie](#) 收获 73 个赞同，330 个收藏。

获得奖项

- 2018第一届全国大数据竞赛二等奖
- 2018年学业三等奖学金

技能清单

以下均为我熟练使用的技能

- Web开发：Java/Python
- 后端框架：Spring Boot/Django/SSM
- 项目构建：maven
- ORM框架：mybatis
- 前端框架：Bootstrap
- 数据库相关：Mysql/Redis/sqlserver
- 版本管理、文档和自动化部署工具：git
- 数据采集：scrapy/selenium/requests
- 运维：linux(terminal) , docker

我的优势

代码之内

- 热爱开源，Github 2018年 贡献 990+ Commits。
- 喜欢总结，[个人博客](#)拥有博文 40 余篇，方向包括但不限于 machine-learning , 算法 , Linux 。
- 有刷题的习惯，在 leetcode 上 Accepted 的题目 150+ 。
- 具有自我驱动能力，由于有较扎实的CS基础，学习语言和新技术较快，遇到问题会先思考再通过 Stack Overflow / Google 来解决问题。

代码之外

- 喜欢健身，在钻研技术的同时保持良好的体魄。
 - 热爱阅读，喜欢阅读经典的技术大部头来提升水准。
-

致谢

感谢您花时间阅读我的简历，期待能有机会和您共事。