

基本信息

- 李勃旺-男-1996/10
- 天津工业大学(双一流一本) 复合材料科学与工程&工学学士 **自学编程**
- 电话: **18990284571**
- 邮箱: 821346679@qq.com & qq8212346679@gmail.com
- 工作年限: 一年半

求职意向

- 工作地点: **北京, 上海, 成都**
- 期待职位: **Python爬虫开发工程师, Python开发工程师**

个人技能

- 基础方面
 - 掌握 Python 中的闭包, 装饰器, 生成器, 多线程, 多进程, 协程等并能灵活运用以解决编程问题
 - 对 Java, JavaScript, Go, 汇编语言 等其他编程语言有一定的了解并做到基本的阅读
 - 了解 Web 前端, 可读写基本的 Js/Css/Html 代码
 - 对 Pandas, Numpy, Matplotlib 等数据分析库有初步的掌握
 - 熟练使用 Mysql, Mongodb, Redis, 能设计数据库, 对索引有深入理解, 能手写高级的 SQL 语句
 - 能使用 Flask, Django 做基本的网页搭建
 - 熟悉 Linux, 使用过 Shell 编程
 - 有一定 算法基础, 熟悉常用的 数据结构 与 算法, 如队列、栈、链表、二叉树、各类排序等
 - 熟练使用 GitHub 管理代码, 参与过 GitHub 开源项目的代码贡献, 对自身代码结构有非常高的要求
 - 对大数据 Hadoop, Kafka, Spark, Hive 等相关技术有基本了解, 并在持续学习相关技术
 - 有良好的自学能力和英语能力, 能阅读英文原文官方文档并使用 Google, Stack Overflow 进行问题搜索
 - 逻辑能力较强, 独立负责过很多项目, 大学高等数学都是 90 分以上; 自学能力较强, 所有计算机知识都是自学, 并将一直自学下去
- 爬虫方面:
 - 爬取(Js逆向, 安卓逆向)过的站点或APP(部分):
 - B站, 知乎, 抖音, 烯牛数据, YouTube, 微博, CSDN 等
 - 熟悉常用的网络协议如 TCP/IP 协议、HTTP(S)协议, 知道各协议的区别与作用以及网络通讯的过程
 - 熟练掌握 Xpath, Re, Requests, Selenium, Scrapy, Scrapy-Redis 等爬虫基本库和框架以及分布式爬虫
 - 对网页加密数据有丰富的 Js逆向 能力
 - 有一定的安卓逆向经验

工作经历

• 北京创新乐知网络技术有限公司(CSDN)

- 职位: 爬虫工程师
- 工作时间: 2020/06/01 --- 至今
- 工作内容:
 - 负责把博客从外部站点搬运(经博主同意, 文章格式排版正确)到 csdn.net 的业务
 - 负责对需求的外部站点视频和图文抓取
- 取得成果:

- 重构和优化 Github , Sina , Ochina , Stackoverflow , Zhihu , 微信公众号 项目的逻辑结构, 使之前繁杂的逻辑变得清晰, 可维护性变高, 如之前需要 300 行的逻辑再重构后仅需要 80 行
- 每日可从其他站点更新上万条视频信息到公司数据库, 以达到公司内容泛化的要求

• • 天津尚仪企业咨询管理有限公司

- 职位: 爬虫工程师
- 工作时间: 2019/06/10 --- 2020/05/17
- 工作内容:
 - 负责公司的数据库维护
 - 负责公司所有的数据更新, 维护和提取等工作
 - 负责公司所有的爬虫项目架构设计优化, 维护使其稳定运行
- 取得成果:
 - 把上一任留的 Java 项目改成 Python 实现
 - 通过优化代码逻辑, 使每日更新数据时间由原来的近 2小时 缩短至 5分钟 左右
 - 提供公司后端业务需要的多种爬虫接口

项目经验

• • 博客搬家项目(CSDN)

- 技术栈: Js逆向 + Scrapy + Xpath高级 + 递归算法 + JavaScript + Appium + Mysql + Redis + AnyProxy
- 职务: 参与者
- 项目描述: 对多个其他站点的博客进行文章格式清洗后迁移到公司博客主页(csdn.net)

• • Blink 项目(CSDN)

- 技术栈: Js逆向 + 安卓逆向 + Redis + Mysql + Scrapy
- 职务: 负责人
- 项目描述: 对其他站点如: 知乎, 抖音, 微博, B站 的一些符合要求的内容进行搬运, 在公司平台上进行内容泛化处理

• • 天津市各个政务平台信息汇总的爬虫项目(天津尚仪企业咨询管理有限公司)

- 技术栈: Mysql + Xpath + Requests + 多线程 + Django + 前端
- 职务: 负责人
- 项目描述: 爬取需要监视的多个网站的最新公告, 每日更新到数据库, 并通过 Django 部署到线上向公司内部展示

• • 给后端提供的API(天津尚仪企业咨询管理有限公司)

- 技术栈: Mysql + Redis + Selenium + Js逆向 + Flask + 分布式 + 验证码模拟登录
- 职务: 负责人
- 项目描述:
 - 专利信息查询API: 使用 Js逆向 + 其他技术爬取目标网站信息, 并用 Flask 部署到公司服务器上, 作为接口返回给后端, 通过专利号就可以查询到专利的所有信息
 - 企业名称查询专利API: 通过模拟登录 + 其他技术爬取目标网站信息, 通过企业名称查询企业拥有的所有专利信息
 - 商标信息查询API: 通过 Selenium + 其他技术爬取目标网站信息, 通过商标号能查询商标的所有信息