



马晓腾

爱好技术，喜欢运动，实践做一个终身学习者

25岁 | 本科 | 北京 | 18516527277 | jizhuwo00@hotmail.com |

求职意向

Python 开发工程师 北京 12K-14K 1个月内到岗

教育背景

2010.9-2014.7 太原科技大学 安全工程

工作经验

2017.10-至今 云孚科技(北京)有限公司 Python工程师

●爬虫方面

开发技术: Requests / Scrapy / Selenium / Splash / Beautifulsoup / Xpath / MongoDB / MySql / Redis / Charles

- 负责公司爬虫开发任务的汇总和分配管理
- 对新的爬虫任务制定开发计划，实施反爬破解，确定爬虫方案，以及脚本编写和测试工作
- 解决爬虫运行过程中遇到的反爬障碍
- 开发和维护爬虫接口，为NLP部门提供稳定的应用接口
- 根据NLP部门的需求提供数据清洗服务
- AWS运维和爬虫部署，负责与海外客户业务沟通

●WEB项目开发方面

开发技术: Python / Django / MySQL / RESTful / Nginx / uWSGI

- 参与中文语义标注平台的开发, 负责项目后端功能实现, 包括数据库表设计, API编写, 以及与前端的交互对接等
- 币小区和DTP对话系统的二次开发, 新增功能的实现和代码重构

2014.8-2017.9 克模塑胶(上海)有限公司 QEHS工程师

- 负责企业CMS、安全、环境和质量管理系统的建立和维护。
- 负责应对安监、环保等部门的检查接待工作，负责应对客户供应链的各种审查稽核工作。
- 参与供应商的体系审核和评选工作。

项目经验

2017.11-至今

爬虫项目

爬虫工程师

实施的爬虫项目较多，主要有以下几种：

- 1. 数据量大, 持续时间久的，包括政务新闻采集爬虫系统, Tabelog(日本版大众点评), Wikipedia和Google搜索结果批量抓取, 百度百科和环球军事武器库等
- 2. 采用分布式和爬虫框架开发部署：使用Scrapy, Scrapy-Redis开发爬虫, 使用Redis和Kafka编写通用任务队列模块
- 3. 跨语言的精准抓取：目标数据包含全球多个国家, 包括Baemin(韩国), Tabelog(日本), Deliveroo(澳大利亚), Hungerstation(中东), Ifood(巴西), Sindelantal(墨西哥)等
- 4. 国内反爬比较积极的网站抓取：大众点评，美团，饿了么等美食网站，以及微信公众号文章的抓取
- 5. 其他爬虫技术: 使用Charles/Wireshark 抓取App的数据接口(Zomato, Carriage等), 使用反编译获取APP接口(Jahez)，破解极验滑动验证码，使用Splash代替Selenium+Chrome+Chromedriver的方案来解决JS动态加载的页面，scrapy部署等

2018.6-2018.12

中文语义标注平台

Python 后端工程师

描述：中文分词标注平台，可以分词标注和实体标注。

项目链接: <https://labelme.yunfutech.com/>

工作内容:

- 1. 数据库表设计，对整个项目的数据流进行控制
- 2. 业务流程api编写
- 3. 与前端的交互和对接

自我评价

- 爱好跑步和游泳
- 拥有较强的自学能力,能通过Google, Stack Overflow, Github快速解决问题 和掌握新技术
- 很强的求知欲和动手能力,喜欢尝试新技术新工具
- 遵循Learning by doing, 将工作总结和学习内容汇总，部署在个人的云服务器上(book.maxiaoteng.xyz)