

个人简历

赵海博

热爱技术，敢于挑战

男 | 23岁 | 本科 | 河南洛阳 | 17839905883 | zds1233@163.com |

● 求职意向

Python开发、爬虫工程师 全职 薪资面议 随时到岗

● 教育背景

2012.9-2016.7

黄河科技学院

物联网工程

所修专业：计算机组成原理、数据库原理、计算机网络、软件工程、C程序设计、Java程序设计、数据结构、算法分析、操作系统、传感器网络、单片机原理等

● 工作经验

2015.9-2017.10

深圳市百诺信科技有限公司

Python爬虫工程师

行业类别：计算机产品研发 | 企业性质：民营

工作描述：

1. 参与公司平台产品的核心系统研发；
2. 设计分布式爬虫对需求网站的数据进行分布式爬取，协助建立数据模型，对数据进行挖掘，优化及统计；
3. 优化抓取策略和算法，提升爬虫的抓取效果；
4. 结合数据，使用Django对项目进行二次开发。
5. 同时负责项目及Mysql、Redis等数据库的性能优化提升；
6. 负责基于linux环境下的项目部署；
7. 项目定期维护及优化。

● 项目经验

项目名称：新浪微博爬虫

项目简介：根据公司需求，爬取新浪微博用户数据，研究新浪微博反爬虫策略，制定相关应对方案，研究验证

策略，使用云打码平台识别各种验证。

责任描述：

- 1.根据公司需求，负责使用多个微博账号进行微博数据爬取；
- 2.负责微博账号登陆验证等模块的开发，使用云打码工具进行验证码识别，根据不同验证码类型执行验证操作；
- 3.研究新浪微博反爬虫策略，使用Redis存储用户cookie值，维护cookie可用性；
- 4.负责将爬取数据的去重，并将数据异步写入数据库；
- 5.负责代码的维护和优化

项目名称：去哪儿网，携程网团购信息爬虫

项目描述：根据公司需求，对去哪儿网，携程网团购数据进行采集，每天定时开启采集任务，并对数据进行筛选和去重。

责任描述：

- 1.负责分析去哪儿网，携程网团购验证机制，网页结构，制定爬取策略；
- 2.负责使用scrapy框架编写爬虫代码，对各大旅游城市团购数据进行采集，包括酒店，美食，旅游，门票数据；
- 3.使用Redis部署分布式爬虫任务，充分利用服务器资源，提高爬取效率；
- 4.使用BloomFilter做去重工作，优化去重任务，提高空间利用率；
- 5.负责编写shell脚本，定时开启爬虫任务；
- 6.负责编写爬虫警报脚本代码，方便爬虫中断或失败，及时发现问题；
- 7.负责爬虫代码的维护和优化。

项目名称：电影天堂全站爬取

项目描述：根据公司需求，对电影天堂内所有类别信息进行爬取，提取特定信息数据，并使Elasticsearch+Django编写搜索服务器。

责任描述：

- 1.根据公司需求，负责爬取电影天堂所需信息；
- 2.负责对爬取的数据进行清洗去重工作并存入数据库并保存为Excel表格，爬取完成或程序出错，通过邮件通知开发人员；
- 3.使用Redis部署分布式爬虫，充分利用服务器资源，提高爬取效率；
- 4.使用BloomFilter去重，优化数据存储；
- 5.使用Elasticsearch+Django编写搜索服务器，将所有爬取的数据合并，方便进行搜索查找、详情查看、点击下载；
- 6.负责代码的维护和优化。

项目名称：制药物联网管理平台

项目描述 制药物联网管理平台为满足从原材料供应到销售终端接收的实体流动过程，方便了在原辅材料采购、药品制造和分销过程中的任何一个时刻对物理对象进行识别，将药材运输、药品加工、药品包装、药品销售等各项生产环节实时有效且有序的结合。

- 1.产品目录：对每种商品进行详细的介绍，包含药品的含量，功能，保质期等信息；
- 2.生产监管：从药材到饮片，饮片到净药材，净药材到药品每个环节的数据信息记录和视频监控；
- 3.药品生产进度控制：对每种药品每天的实际生产量、划生产量、剩余量等信息的统计；
- 4.药品库存管理：实现对原药材、饮片、净料、药品的产生码、扫码入库、扫码出库以及库存量的查询；
- 5.药品销售管理：对每种药品的销售信息进行统计；
- 6.基础设置：实现用户的增加、权限的分配等功能；

职责描述：

- 1.参与了系统需求的讨论、功能的划分和数据库的设计的讨论工作等；
- 2.实现产品目录、生产监管、药品生产进度控制、药品库存管理功能的代码编写和测试；
- 3.编写系统的操作手册。

● 技能特长

- 1.计算机专业，具有良好的计算机基础与逻辑思维能力以及代码编写规范；
- 2.熟练使用Scrapy、Django等Python主流框架，实现数据发掘和后台开发；
- 3.熟悉Mysql、Mongodb、Redis等数据库，熟悉Linux平台开发；
- 4.拥有各种主流网站的爬虫经验，包括但不限于微信，微博类爬虫；
- 5.熟练使用Selenium、Xpath、Bs4、正则表达式等网页解析技术；
- 6.了解Nginx、vsftpd、celery、负载均衡、分布式等运维技术；
- 7.对搜索引擎、数据分析、数据展示、有一定的了解和使用经验；
- 8.了解Bootstrap、HTML、Jquery等前端开发技术结合Django将数据前端展示；
- 9.熟练使用Git、SVN进行源代码管理，协助他人开发。

● 自我评价

- 1.我热爱编程，有良好的分析问题、解决问题的能力；
- 2.能吃苦耐劳、待人热情、真诚，有较强的适应能力和团队协作精神；
- 3.有较强的自我学习能力，空闲时间喜欢浏览技术博客，github等网站；
- 4.性格外向随和，能接受高强度出差和加班，能快速适应新环境。