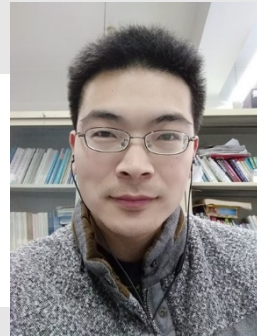


个人简历

个人概况

姓名: 吴志魏 性别: 男
毕业院校: 上海大学 专业: 凝聚态物理 最高学历: 硕士
手机: 15201933129 电子邮件: DreamingTech@163.com
地址: 宝山区南卫路 38 弄富泰苑 11 号



知识与技能

概述

分类	项目	说明
编程语言	Python	熟练掌握, 灵活应用
	R 语言	熟练掌握, 基本应用
数据库	Mysql	基本开发环境搭建, 基本增删改查, 联合查询
	MongoDB	基本操作
	Redis	基本开发环境搭建, 基本操作

爬虫应用

分类	项目	说明
解析库	urllib, requests, selenium	灵活应用
提取库	BS4, Xpath, Re, Jsonpath	灵活应用
数据存储	txt, csv, json, mysql	灵活应用
框架	scrapy	灵活应用
分布式	scrapy-redis 框架	理解及应用
	Docker	基本使用与分布式爬虫部署
反反爬	验证码	1. 使用第三方网站处理常见的验证码. 2. 使用开源代码处理倒立文字验证码
	代理	1. 自建动态 vps 代理服务器.
		2. 爬取开放代理并构建简单代理池. 3. 使用付费代理实现规模化数据爬取.

爬虫框架开发

项目	说明
爬虫框架设计	设计符合业务需求的爬虫开发框架, 主要包括请求管理, 异步任务设计, 数据解析与提取, 程序监控等组件

爬虫框架 主要组件	1. 单机内存及分布式持久化队列组件
	2. 基于 redis 的指纹过滤器, 基于 simhash 的过滤器, 基于 redis 的布隆过滤器
	3. 基于 requests, aiohttp, tornado.httpclient, selenium+Chrome 等模块实现的下载器
	4. 基于 asyncio, celery+eventlet/gevent 等模块实现的异步并发组件
	5. 基于 re, lxml, jsonpath, js2py/pyexecjs+node, splash 等模块实现的数据解析组件
	6. 根据具体业务设计数据清洗组件
	7. 基于 csv, json, txt 等的文件存储组件, 基于 mysql, postgresql, mongodb 等的数据库存储组件
	8. 基于 ELK 系统实现的监控组件

Docker 技术

分类	项目	说明
开发环境	Docker 开发环境的搭建与配置	熟练应用
Docker 基础	镜像管理, 容器及生命周期管理, 网络配置, 数据卷挂载, 私有仓库搭建	熟练应用
Docker 高级	使用 Dockerfile 创建镜像, 使用 Docker-Compose 创建容器	熟练应用
Docker 应用	基于 Docker 的单机及多主机 ELK 环境配置	基本应用
爬虫开发环境	基于 Docker 的 python 开发环境配置	熟练应用

数据分析

项目	说明
Excel	熟练掌握基本操作, 常用函数和数组函数, 掌握 Power Pivot, Power BI 和 Power Map 的基本使用和操作
统计	掌握基本概念, 基本的统计学原理
R 语言	掌握基本操作, 掌握绘图和基本数据分析软件包的使用和操作
数据分析	掌握数据清洗, 描述性统计分析及推断性统计分析的基本方法和 R 语言实现
数据挖掘	理解回归分析, 聚类分析, K 近邻, 贝斯网络, 决策树的原理和 R 语言实现

其它

项目	说明	
Linux	ubuntu, centos. 熟悉系统安装配置, 开发环境搭建, 常用操作.	
英语	四级成绩	518
	六级成绩	497
	口语	熟练进行日常交流
	阅读	无压力阅读英文开发文档
善于总结,	善于从看似无关的概念和内容中找到联系, 把看似无规则的内容总结整理为一	

善于整理	个完整的体系.
五笔	十年五笔经验

爬虫项目及代码展示

项目	代码
天善智能课程爬虫	https://github.com/dreamingtech/hellobi
豆瓣图书全站爬虫	https://github.com/dreamingtech/doubanbook_requests
selenium+chrome 下载 sciexp 整站视频	https://github.com/dreamingtech/sciexp
伯乐在线分布式全站爬虫	https://github.com/dreamingtech/jobbole_article
boss 直聘全站爬虫	https://github.com/dreamingtech/boss_zhipin
拉勾全站爬虫	https://github.com/dreamingtech/lagou_job
51job 全站爬虫	https://github.com/dreamingtech/51job
仿 scrapy 分布式爬虫框架	https://github.com/dreamingtech/spider_plus_fw

教育背景

本科学历: 2004 年 9 月 - 2008 年 7 月 河南大学 物理学专业
 硕士学历: 2009 年 9 月 - 2013 年 7 月 上海大学 凝聚态物理专业
 博士(经历): 2013 年 9 月 - 2016 年 9 月 上海大学 凝聚态物理专业

工作与实习

1. 2014 年 3 月 - 2014 年 12 月, 思博职业技术学院申洋学院任行政管理专业《领导科学》,《公共政策学》,《公共关系学》和《情报检索》代课教师.
2. 2016 年 9 月 - 2018 年 9 月, 上海杓衡信息科技有限公司, 负责机器翻译语料库的数据采集及数据清洗工作.
3. 2018 年 9 月 - 2018 年 11 月, 八维宏烨(上海)信息科技有限公司, 任 python 基础和 python 爬虫课程讲师.

个性特点

1. 做事比较稳重, 喜欢事事有条有理.
2. 细致认真, 很有耐心.
3. 乐于学习新知识, 新技术.
4. 一件事, 总是尽可能做到尽善尽美.
5. 善于总结, 善于整理, 善于从看似无关的概念和内容中找到联系, 把看似无规律的内容总结整理为一个完整的体系.

附: 部分工作笔记展示

https://github.com/dreamingtech/notes/blob/master/部分工作笔记展示_20190219.pdf
https://github.com/dreamingtech/notes/blob/master/伯乐在线全站爬虫_step_by_step.pdf
https://github.com/dreamingtech/notes/blob/master/爬虫笔记_基础_20190327.pdf