# 个 人 简 历

## 个人概况

姓名: **吴志魏** 性别: **男** 1

毕业院校: **上海大学** 专业:  **凝聚态物理** 最高学历: **硕士**  1

手机: 15201933129 电子邮件:  DreamingTech@163.com 1

地址: **宝山区南卫路38弄富泰苑11号** 1

## 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. |

### 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 |

### 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等的数据器存储组件 |
| 8. 基于ELK系统实现的监控组件 |

### 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, 把看似无规则的内容总结整理为一个完整的体系. | |
| 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 |

## 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6月, 思博职业技术学院申洋学院任行政管理专业《领导科学》代课教师.

2. 2014年8月 - 2014年12月, 思博职业技术学院申洋学院任行政管理专业《公共政策学》,《公共关系学》和《情报检索》代课教师.

3. 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.

4. 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/%E4%BC%AF%E4%B9%90%E5%9C%A8%E7%BA%BF%E5%85%A8%E7%AB%99%E7%88%AC%E8%99%AB_step_by_step.pdf)

[https://github.com/dreamingtech/notes/blob/master/爬虫笔记\_基础\_20190327.pdf](https://github.com/dreamingtech/notes/blob/master/%E7%88%AC%E8%99%AB%E7%AC%94%E8%AE%B0_%E5%9F%BA%E7%A1%80_20190327.pdf)