|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | **个人信息**  姓名：王盛民  电话：18911831282  邮箱：[2541884711@qq.com](mailto:2541884711@qq.com)  男|(1987)现居住北京|8年工作经验 | | |  |  |  |  | | --- | --- | --- | --- | | |  |  | | --- | --- | | QQ号： | 2541884711 | |  | | |  |  | | --- | --- | | 身高： | 175cm | |  | | |  |  | | --- | --- | | 政治面貌： | 共青团员 | |  | |  |  | | --- | | **求职意向** | | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | |  |  | | --- | --- | | 期望薪资： | 面议 | | |  |  | | --- | --- | | 地点： | 北京 | | | |  |  | | --- | --- | | 职能/职位： | 爬虫工程师 | |  | | |  |  | | --- | --- | | 到岗时间： | 一周内 | | |  |  | | --- | --- | | 工作类型： | 全职 | | | |  |  | | --- | --- | | 自我评价： | 擅长做抓取，抽取，数据处理工作, 喜欢独立性和创新性的研究  熟悉常见的反爬策略，加密，混淆，多层跳转，加壳，验证码，滑动组件，ip帐号鉴权（单位时间内单位变量访问次数）并有一定的研究，  熟悉 springmvc,springbot 开发框架，  熟悉mysql,redis,数据存储软件，  熟悉httpclient,httpunit,selenium,httphttpclean,jsoup,JDom，Tidy java网络和网页解析组件，  熟悉python，scrapy 网络技术和框架，  熟悉dubbo,thrift rpc调用，  熟悉hadoop hbase zk solr 大数据和搜索技术。 | | | |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **工作经验**   |  |  | | --- | --- | | 2018/11-2018/1 | **爬虫开发** | | 惠盛科技 [0年] | | | |  |  | | --- | --- | | 工作描述： | 体育咨询、视频信息抓取，分布式爬虫设计，网站结构分析，代理服务部署。 | | | | | | |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  | | --- | --- | | 2015/5-2018/8 | **爬虫工程师** | | 华道征信科技公司 [4年] | | | |  |  | | --- | --- | | 工作描述： | 主要是支撑华道个人征信业务的数据采集工作，图片验证码识别，分布式爬虫，和文本分块主题抽取，反爬虫技术，任务定义划分分配验证。 | | | | | |  |  | | --- | --- | | 2013/5-2015/5 | **数据采集与分析** | | 光芒技术 [2年] | | | |  |  | | --- | --- | | 工作描述： | 垃圾短信文本收集和分类分析，短信通道管理项目，短信统计分析 | | | | |  | | |  |  | | --- | --- | | 2010/9-2013/5 | **软件工程师** | | 灵动快拍 [3年] | | | |  |  | | --- | --- | | 工作描述： | 公司的图书数据、图片信息采集 | | | | |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **项目经验**   |  |  | | --- | --- | |  | **八斗体育** | | |  |  | | --- | --- | | 项目描述： | 定时抓取体育咨询和视频信息，去重清洗并结构化存储mongdb，并发送kafka消息，进行新闻推送。  定时抓取直播信息，赛程概要、技术分析，提供视频直播。  抓取的项目包括篮球、足球、网球、排球、桌球、跑步、赛车。  抓取的网站包括腾讯、懂球帝、爱奇艺、第一足球、GOAL、凤凰体育、东方体育、猫扑、一点资讯等数十家网站。  基于gecco框架和redis实现分布式咨询信息采集，使用srping的scheduled定时调度框架，主体信息使用mongdb存储，消息体通过kfaka进行定向推送。 | | | | |  |  | | --- | --- | | 责任描述： | 框架设计、网站分析、爬虫程序 | | | | | |  |  | | --- | --- | |  | **主题数据采集** | | |  |  | | --- | --- | | 项目描述： | 根据提供的关键字，通过百度入口全网爬取相关性新闻。  通过httpclient完成网页的采集和链接分析，通过鱼群算法动态控制爬取的深度和广度，关键字匹配实现文章相关性，通过新闻抽取算法实现新闻内容抽取，文章存储hbase，通过solr提供搜索。 | | | | |  |  | | --- | --- | | 责任描述： | 框架设计、爬虫程序 | | |  |  |  | | --- | --- | |  | **个人征信数据采集** | | |  |  | | --- | --- | | 项目描述： | 提供web接口通过用户授权信息，完成用户征信相关信息采集，基于此完成用户画像分析征信评价。  主要抓取网站运营商（移动电信联通）, 京东、一号店, 学信网, 央行征信 基于dubbo完成爬虫调度，通过httpclient完成登录、跳转、数据采集，数据存储mysql。 | | | | |  |  | | --- | --- | | 责任描述： | 网站分析、框架部署、接口程序、爬虫程序 | | | |  | **垃圾文本分类引擎** | | |  |  | | --- | --- | | 项目描述： | 通过对短信网关、邮件服务器举报上行的短信进行类别识别统计。  使用IK算法分词，词频统计，特征选择，贝叶斯分类。 | | | | |  |  | | --- | --- | | 责任描述： | 可行性分析、技术分析、程序实现 | | | | |  |  | | --- | --- | |  | **购书搜索系统** | | |  |  | | --- | --- | | 项目描述： | 图书网站的信息抓取，清洗存储，为用户提供购书搜索实现 | | | | |  |  | | --- | --- | | 责任描述： | 程序实现 | | | |  | | | | |  |  | | --- | | **教育经历** | | |  |  |  |  |  | | --- | --- | --- | --- | --- | | |  |  | | --- | --- | | 2006/9-2008/7 | **青岛大学软件技术学院** | | 大专|计算机科学与技术 | | | | |