**个人简历**

**姓 名： xx** **出生年月：1994.06**

**名 族： 汉**  **联系电话：**

**工作年限：一年**   **邮 箱：**

**职业技能**

熟悉HTTP/HTTPS协议，TCP/IP网络协议

熟悉Python 多线程爬虫及其机制

熟练使用Python urllib、urllib2、requests等网络模块

熟练使用Python lxml、BeautifulSoup、re、json模块进行数据提取

熟悉XPath语法规则和各CSS Selector的使用

了解Tesseract机器图像识别系统，并处理简单的文字验证码

熟练使用Selenium+PhantomJS实施动态HTML抓取

熟练掌握Scrapy框架，以及编写各类中间件

熟练掌握scrapy-redis分布式框架，了解各组件工作机制

 掌握HTML、CSS、jQuery等前端页面的基础制作，了解Django框架；

熟悉常用的数据库，包括MySql/Redis/MongoDB；

 掌握常见的爬虫、反爬虫知识及应对措施；

2016-02-2016.12 重庆亿阳信通 运维

工作内容：负责保障所有工作区域内桌面系统的安全、稳定运行。 负责监控检查系统软硬件运行状态并进行详细的数据化记录；负责监控存储阵列运行状态，确保存储正常稳定；作好值班记录和工作记录，定期汇总并上报给负责人。

2017-03-2017.08 非概率驱动（北京）科技有限公司 python爬虫

工作内容： 参与公司爬虫所需的爬虫系统的开发与维护；

负责分配站点的数据抓取，为公司其它部门提供数据支持；

与数据分析师合作，收集数据采集需求，定向数据的采集与爬去，解析处理，入库等日常工作；

**应聘方向**

|  |  |  |  |
| --- | --- | --- | --- |
| 求职行业： | Python爬虫 | 应聘职位： | Python爬虫工程师 |
| 求职地点： | 北京市 | 薪资要求： | 面议 |

**能力总结**

互联网站、网页、链接的特征挖掘，对指定的多个网站的网页进行网页抓取、数据提取、清洗、入库；

对搜索、匹配、同步等后台守护程序进行性能优化等；

负责系统日常运营资源调度优化及运营指标分析调优，并能实时监控爬虫的进度和警报反馈；

负责特殊网页采集任务的分析及采集方案设计。

搭建分布式爬虫环境。

**自我评价**

性格温和开朗，诚实守信，勤奋稳重，富有幽默感

 学习能力强，对新事物接受能力强，对新技术有着强烈的好奇心

 适应能力强，能很快的融入团队，有良好的团队精神

对待工作认真，责任心强，勤于学习，不断提高自身的能力与综合

**项目经验**

**租房信息采集爬虫**

**项目简介** 这个项目主要是58同城，租房网，搜房网等网站的租房信息的爬取并提供给数据分析部门

**项目职责**

分析网站、网页、链接的特征挖掘，采集任务的分析及采集方案设计，使用scrapy框架爬取数据，使用mango存储数据，准备一个代理ip池和user-agent列表以应对网站的反爬虫机制

Spider Engine负责Spider、ItemPipeline、Downloader、Scheduler中间的通讯，信号、数据传递，获取Spider的URL之后，发送至调度器，请求排列入队，发送处理好的request给引擎，下载器根据中间件的设置下载request请求，spider自行处理下载好的请求，处理完毕之后，有两个结果，既需要跟进的URL和获取到的item数据，item数交给Pipeline处理，URL交给调度器，重复之前的步骤，直到获取到所有需要的数据

此项目是对租房信息的抓取，抓取租房的价格、位置、图片、配置等信息，并做保存，对于爬取的数据使用jupyter notebook进行简单分析。将数据提交给数据分析部门，通过爬取的数据分析当下北京地区的租房现状，快速捕捉市场机遇，实现利益最大化。

**新闻分类资讯分布式爬虫**

**项目简介** 这个项目是对人民网，新浪，腾讯等网站分类新闻资讯爬取的分布式实现。

**项目职责**

采用scrapy-redis分布式框架实现爬虫集群，分布式使用Redis实现。

存储Request请求和指纹集合，并且对各个Slave端爬虫实现集中管理和控制。

利用Redis的高并发和I/O读写来实现高速下载。

采用MongoDB做为本地数据库，将资讯新闻按所属大类、子类以及标题和内容，保存在

MongoDB中。

同时每次下载前会检查请求指纹，防止重复下载，避免资源浪费

**个人项目**

**淘宝网---- 分类店铺、商品信息(使用Selenium+PhantomJS实现动态网页数据的抓取)**

**百**