**姓名：王鹿丽**

电话：18611378520

邮箱：1151682534@qq.com

GitHub : https://github.com/cainiaomama520/scrapy\_spiders

王鹿丽 性别：女 生日：1989-5-20 河南大学 2011级 软件工程 专业

|  |  |  |
| --- | --- | --- |
| 目标 |  | **Python爬虫工程师 或 Python相关工作** |
| 技能 |  | 熟悉使用Python 3编写脚本，有较好的Python语言编程基础  熟悉使用Python 3爬取站点数据，熟悉爬虫常用第三方库，如os.path, time, random, optparse, csv, json, openpyxl, requests, lxml, re, Beautifulsoup, pymongo等  熟悉Scrapy爬虫框架，熟练使用之爬取静态和动态网页，GitHub项目记载了平时爬取的一些站点  熟悉HTTP协议，HTML5标签，CSS表达式，熟悉XPATH语法及正则表达式  熟悉Linux bash常用命令，如ssh, chmod, scp, rm, mv, mkdir, touch, grep, cut, diff, sed, gawk, wc, sort等，能编写基础的bash别名、函数、脚本  熟悉使用MongoDB数据库; 熟悉SQL Server 及 CURD（增删改查） SQL语句  熟悉VMware,XenCenter,XenServer,SCVMM企业级虚拟化软件，有简单Docker使用经历 |
| 实践经验 |  | 1. 项目：https://github.com/cainiaomama520/scrapy\_spiders/tree/master/cookie\_login   该项目爬取知乎个人信息，其难点在于如何绕过表单提交实现登录。由于目前验证码越来越复杂，表单提交登录的方式使得爬取站点数据越来越难。  该项目利用第三方库browser\_cookie3对CookiesMiddleware中间件进行改良，实现了一个能使用浏览器Cookie的中间件。  其运用的核心思想是，在构造BrowserCookiesMiddleware对象时，使用 browser\_cookie3提取浏览器中的cookie，然后存储到CookieJar字典self.jars中。   1. 项目：<https://github.com/cainiaomama520/scrapy_spiders/tree/master/jd_splash_example>   该项目爬取京东商城中所有python书籍的名字和价格。其中难点在于利用splash和scrapy\_splash第三方库获取动态网页的数据。  该项目基本思路如下：首先使用scrapy.Request提交请求，提取商品总数。然后根据自己分析出的URL规则构造各个页面的URL。最后使用splashRequest提交请求（除了渲染页面以外，还需要执行一段JavaScript代码——为了加载后30本书)。  此处核心思想在于如何确保JS代码的执行：使用了splash的execute端点将endpoint参数设置为'execute'，通过args参数的lua\_source字段传递我们要执行的lua脚本。由于爬取每个页面时都要执行该脚本，因此使用cache\_args参数将该脚本缓存到splash服务器。lua\_script字符串是自定义的lua脚本，其中的逻辑是: 打开页面——>等待渲染——>执行js触发数据加载(后30本书)——>等待渲染——>返回HTML。 |
| 工作经验 |  | **三星中国研究院 ---- 三星S8语音助手语言专家岗 位，外包于慧博云通有限公司2017-03 – 至今**   1. 语言专家岗位主要负责语音助手可能接收到的各类语音交互指令的语义覆盖。我们按照三星S8手机的常用系统功能（比如设置中的各项内容）分组分模块具体到人，每人对其负责的功能模块做大量的语义覆盖。提交的自然语言覆盖示例将由服务器端验证采纳合并到中央服务器，以使三星语音助手的功能更加完善智能。 2. 该岗位要求熟悉常用Linux命令行，包括scp,rsync等常用跨服务器文件传输命令。同时需要一定的手机端日志分析及定位、排查语音助手功能问题的能力。Python编程能力也在其中为解决各类问题（如处理任务分配文件，提取相关数据）大大提高了我的工作效率。   **Citrix ---- PC服务器端产品测试，*文思海辉技术有限公司(Pactera)*** **2014-****07 -- 2015-08**  1. Citrix XenDesktop是一套桌面虚拟化的解决方案。无论你是什么样的用户（普通用户，商用用户还是自由职业者）无论你使用什么样的设备（手机，Mac,个人电脑等等）它都可以为你提供自己所需要的桌面。  2. 我主要对XenDesktop的部分功能进行测试。首先根据测试用例需求在SCVMM，XenCenter,Vmware上独立搭建测试环境，创建测试需要的虚拟机；然后安装XenDesktop产品，其后对该产品的catalog,delivery group,controller,storefront功能进行测试。  3. 如果发现bug，就写邮件向leader说明详细信息，然后在专门的bug提交网站提交自己发现的bug，跟踪该bug的进度，直到此bug关闭，该项任务完成。 |
| 个人经历 |  | 完成了人生中最重大的事情 之一 —— 结婚生子 ：2015-07 — 2017-02 如今，有了一个将要满2岁的小男孩的我，希望自己用更充沛的精神拥抱生活，爱人和我的家人。他们永远是我的爱的港湾，精神的支持。我们会每半个月聚在一起分享生活。 2011级 -河南大学-软件工程 —— 学习计算*机专业知识，通过大学英语四级考试，多次获得奖学金和优秀大学生称号* 在大学期间主要学习了计算机概论，计算机操作系统，Java，C#, C++编程语言，数据库，高等数学，线性代数，概率论等课程，并 通过了大学英语四级考试。  在校期间，四次获得河南大学奖学金和优秀大学生的称号，并于大三上学期获得优秀党员称号。 |
| 个人简介 |  | 我来自河南省周口市（与安徽省亳州市交界的位置），是一位乐观开朗喜欢结识新朋友的女性。  由于上初三时，偶遭车祸，所以在家休养两年才开始继续上学。因此，我比同时期读书的朋友年龄稍大。而身份证上显示年龄也比实际年龄大两岁，我在大家眼里就成了一个大姐姐。  考虑到年龄因素，**在2015年，**我选择了结婚生子，目前孩子将要两岁了。  对待工作，我始终坚持认真，踏实，负责。此外，生活稳定，钻研好学也是我如今的优势！  无论工作在三星，还是曾经的文思海辉，我都从工作中学到了很多知识：虚拟化技术的了解，Python语言的掌握，以及爬虫的摸索和实践。  每一次翻阅自己记下的笔记，写的简书，GitHub，心中都满是感触！  一路走来，工作促我前进的更远，走的更稳更好——它让我明白要做得更好，就要不断地更加深入地学习，才能拥抱更芬芳的生活！  日常生活中，我喜欢欣赏街舞和唱歌，特别享受唱给我们家小宝宝听！ |