个人简历

Life is too short , we need Python

姓 名 :xxx

民 族：汉

电 话：

邮 箱：

学 历：本科 | 英语六级

年 龄：26

期望薪资：12000

工作年限：Python开发两年

求职意向：Python工程师

毕业院校：

婚姻状况：未婚

基本属性

**Python技能**：

**·** 熟练掌握python的基本语法，对面向对象思想有一定的了解

**·** 了解Python垃圾回收机制及其原理

**·** 对系统编程和网络编程有一定的认识

**爬虫技能**：

**·** 熟悉HTTP/HTTPS协议，TCP/IP网络协议  
**·** 掌握常见的爬虫、反爬虫知识及应对措施  
**·** 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  
**·** 掌握Scrapy框架，以及编写各类中间件  
**·** 了解scrapy-redis分布式框架，了解各组件工作机制

**·** 熟悉fiddler抓包工具的使用，能够获取到动态生成的页面

**web技能**：

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

**数据库技能**：

**·** 熟练使用MySQL数据库，了解MongoDB，Redis的相关操作

**其他技能**：

**·** 熟悉Linux开发环境，熟练掌握常用命令行的使用

**·** 了解分布式管理控制系统Git,并掌握常用命令

**·** 了解Numpy, Pandas，matplotlib等的数据分析工具  
**·** 英语六级，具有一定的英文文档阅读能力和翻译能力，能进行日常英语交流

技能清单

工作经历

**2015/06 -- 2017/08**

###### 西安亚森通信股份有限公司 | python工程师

**工作描述：**  
 该公司为外包公司，任职期间根据公司业务需要被外派到其他公司进行项目开发，主要工作包括：  
 1.负责电子商务网站后端开发；  
 2.按照项目计划，按时提交高质量的代码，完成开发任务；  
 3.参与爬虫系统的架构设计与开发，完成采集任务、多线程爬虫。

**2014 / 08 - 2015 / 04**

2016-08至2017-08 天路网络科技有限公司 Python开发工程师

* 负责公司业务系统的设计及改进；
* 负责公司外派项目开发工作；
* 负责公司部分项目后台数据支持及系统维护。

**2014 / 08 - 2015 / 05**

**西安云动力科技有限公司 | 前端工程师**

**工作描述：**

1.负责静态页面的设计；

2.开发技术：html+css+javascript。

项目经验

**新闻分类资讯分布式爬虫**

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

**责任描述：** 1.采用scrapy-redis分布式框架实现爬虫集群，分布式使用Redis实现  
 2.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  
 3.利用Redis的高并发和I/O读写来实现高速下载  
 4.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  
 5.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

**网易云音乐（个人项目）**

**责任描述：**  
 1.找到start\_url；导入selenium的webdriver包  
 2.发送get请求，获得响应  
 3.利用find\_elements\_by\_xpath来获取数据  
 4.对某一首歌曲的所有评论进行点赞

**豆瓣电影分类排行榜（个人项目）**

**责任描述：**

1.分析网页url 地址，获悉该网页是动态加载生成的

2.由抓包获得需要的请求参数，并进行分析

3.发送请求，并获取到每部电影的名字、主演和评分

**有道翻译/百度翻译（个人项目）**

**责任描述：**

1.分析是get请求还是post请求，获得url

2.得知是post请求时，通过抓包获得请求参数

3.发送请求并且通过用户输入的指令进行翻译

**万表官网的爬取（个人项目）**

**项目描述：**

通过对万表官网的爬取，可以知道查到这个平台目前各个品牌的手表的型号、参数、销量，折扣以及价格。

**责任描述：**

1.采用 requests实现爬取

2.通过get 获取页面的内容

3.用 xpath 提取想要的节点，分别为手表的品牌、型号、价格、折扣、销量。以字典形式进行存储

4.分析下一页的链接规则，运用xpath去提取下一页的链接，直到获取不到页面。

5.把爬取的内容存入本地磁盘，也可存入数据库。