XYZ简历

同济大学浙江学院 | 电子信息工程专业 | 本科 | 男 | 3年开发经验

电话：13830004000 邮箱：hellokitty@126.com

·目前状况：

在校学习情况

在校实践经验

培训经历

**求职意向**

* Python爬虫工程师

在校学习情况

在校实践经验

培训经历

**专业技术**

**Python：**

* 熟练运用Python语言
* 熟悉各种网页源代码解析工具，如：Xpath/BeautifulSoup/Re
* 熟练Python爬虫框架Scrapy，Scrapy-Redis，及爬虫技术Urllib/Requests/Selenium
* 熟练运用Web 开发框架Django，了解Flask开发框架
* 熟悉常见的反爬虫策略
* 熟悉网络编程，熟悉HTTP/TCP/UDP等网络协议，熟悉Ajax等开发技术
* 熟悉MySQL数据库
* 熟悉Redis，MangoDB非关系型数据库

**其它：**

* 熟悉HTML和CSS， 对JQuery和Boostrap前端技术有一定了解
* 熟悉版本管理工具Git的使用
* 熟悉Sublime文本编辑工具的使用

在校学习情况

在校实践经验

培训经历

**工作经历**

2015.10 -- 2016.4 | 浙江闪龙科技有限公司 Python开发

2016.5 -- 2018.5 | 浙江驿道网络科技有限公司 Python开发

**项目经历**

**项目名称：各大旅游网站的景点信息抓取系统**

**开发工具：**Scrapy + Fiddler2 + Mysql。

**硬件环境：**

**项目描述：**负责抓取某省份的各大网站上的城市旅游景点信息（标题，评分，评论等）。

**应用技术：**1.分析网页的URL的跳转。

2.负责分析数据爬取过程，编写爬虫文件, 设置User-Agent获取用户头。

3.使用scrapy框架进行爬取网页，使用Re/Xpath进行数据的提取。

**责任描述：**1.网页结构分析，搭建爬虫程序的架构。

2.使用了requests、BeautifulSoup模块进行抓取。

3.将我们所要的文本内容提取出来放入数据库中。

**项目名称：网游、单机和手游游戏相关资讯爬虫**

**开发工具：**Scrapy + Mysql等。

**项目描述：**对大型游戏门户网站如17173, 游侠网, 太平洋游戏网等进行数据爬取分析。

**应用技术：**1.主要使用scrapy框架进行抓取。

2.根据网页数据在items.py中建模。

3.定义爬虫行为，编写对应spider文件。

4.利用Xpath对数据进行抽取,获取所需的各种游戏信息, 新闻, 图片等资讯的链接。

5.编写User-Agent从中随机抽取用户头。

6.运用第三方ip代理池，防止ip被封获取不到数据，影响效率。

**项目职责：**1.负责设计和开发游戏资讯网络爬虫程序。

2.负责分析数据爬取过程，提供技术实现方案。

3.负责分析静态、动态的网页结构、解析网页游戏资讯数据。

4.负责爬虫系统的后期维护。

**项目名称：行业信息动态爬取**

**开发工具：**Python + Scrapy **硬件环境：**

**项目描述：**对不同地区，不同行业动态信息进行爬取，并进行进一步的解析处理，将有效数据存储

**应用技术：**1.主要使用scrapy框架进行抓取，在中间件里写了代理、header等，防止爬虫中断，由于部分网页采取了动态网页js加载，因此采用了Selenium进行了对动态html的处理。

2.根据网页数据在items.py中建模。

3.定义爬虫行为，编写对应spider文件。

4.利用Xpath对数据进行抽取。

5.编写中间件模拟客户端访问，编写User-Agent从中随机抽取用户头。

6.运用第三方ip代理池，防止ip被封获取不到数据，影响效率。

7.编写pipelines文件，将数据存入数据库。

**项目名称：浙江驿道网络科技有限公司内网交流论坛**

**开发工具：**Ubuntu + Python + Django**硬件环境：**

**项目描述：**用于公司内部人员登录、访问、交流。

**应用技术：**1.使用Django ORM（Object Relational Mapping）系统搭建数据模型。

2.使用ManyToManyField 进行多对多的关联。

3.定forms 模块自动为创建常规的表单代码。

4.利用SessionMiddleware 用户处理用户会话。

**项目名称：酷狗音乐歌曲爬取(个人)**

**开发工具：**Python + Request + Fiddler2

**项目描述：**提取酷狗音乐中歌手信息(歌手可自选), 歌曲。

**应用技术：**1.运用fidder抓包工具, 对网页内容进行抓取。

2.分析所要抓取网页的内容, 找到并解析Ajax和json文件。

3.解析后发现, 下载歌曲链接需要hash加密之后的参数, 通过使用Jsonpath库, 获取

参数hash值。

4.运用hash值去请求下一页的Ajax接口,再提取出真正的mp3下载地址。

5.用request获取并下载mp3的二进制。

**项目职责：**负责开发酷狗音乐爬虫。

**自我评价**

有责任心、进取心、吃苦耐劳。有良好的逻辑思维能力及良好的代码风格。