**学习中遇到的困难以及希望杨老师解答的问题：**

1. 如何批量爬取网站的视频

循环、多线程

批量？ 同一个网站上的N多个视频 ， 找URL的规律性，分页

不同网站的N多个视频 :根据不同的网站，制度不同的方案

1. 视频的URL 直接在HmtL中 -》糗事百科
2. 视频采用的是流媒体的形式--》N多个小视频组成一个大视频（优酷）

（爬虫学的是思路，因为网站定期有变化 反爬的问题，反爬虫是在不断升级）

要想爬虫学的好，前端不可少 HTML+CSS+JS+JQuery

想要学好至少三遍 （要想学到精至少五六七八遍）

所有编程语言的结构都是相通

1. Python中的乱码问题

Python.py文件的编码 （针对于Python2和Python3）,如何修改Python文件的编码格式?使用 # coding:utf-8 , 只能是Python文件的第一行

文本文件的编码问题 encoding 关键字参数，open函数中的关键字参数，文件读写时常用

右键新建文本文件的编码格式为UTF-8

使用Python代码新建的文本文件编码格式是什么？默认为GBK

爬虫响应结果的编码问题（响应结果数据是乱码）

|  |
| --- |
| **import** requests  url=**'http://www.baidu.com'** resp=requests.get(url) resp.encoding=**'utf-8'** *# 响应结果的编码格式 #resp.encoding=resp.apparent\_encoding* print(resp.text) |

元组、列表、字典、集合都是Python中的组合数据类型

元组 只能查看元素（不可变序列）， for ..in 遍历

字典和列表用的是最多的

1. 遇到淘宝反爬没有解决办法吗？

登录之后添加cookie再爬取（阿里国际站）

1. 代码的例子需要背过吗？

不需要背例子，但是语法结构必须会，无论多大的项目都离不开语法

类属性如果被重新赋值，是否会影响到类属性的引用？

1. 少儿编程需要学习哪些知识，有没有面试题

|  |
| --- |
|  |

1. 利用pandas如何将excel表格中的某一列移动到指定列
2. 函数的返回值理解的不是很深
3. 递归函数
4. 同一个爬虫程序，有些数据能获取到，有些数据不行，经检查无语法错误，XPath或CSS提取用shell检测都是可以获取的，但就是爬不下来是怎么回事？（需要具体网站）
5. 有些课程不操作就很难深入懂，但是找不到好的例题（同一问题反复练习）
6. 在数据模块安装在PyCharm中如何更快捷操作要准确些呢
7. 类，对象，封装多态继承? 上下文管理器？匿名函数？try except？
8. python中的urllib. parse. quote()方法能解析所有url编码吗？

URL编码基本都是UTF-8

1. 一些网页的隐藏数据怎么找到，比如vip的一些数据怎么找出他在那里

登录之后添加cookie(具体网站还要具体分析)

1. def \_\_init\_\_初始化方法什么时候使用

创建对象时自动调用

1. join函数用法不太理解
2. 希望多讲解类似于递归的问题（算法）

每周三下午 左程云老师算法公开课

直播间地址

https://ke.qq.com/course/399017?taid=11394793049822889



1. 9%-4 = 9-（-4）\*（-3）中商等于-3，为什么？个人猜测是：在Python中，商非整数的话，则为整除向下取整得到的结果 （向下取整）
2. 重点关键字是那些



1. 怎样利用好git hub?

1. 什么是类，宏观微观讲一下，还有为什么要建立类
2. 在使用python给excel创建折线图的时候怎么让列的数据为一条线，而不是以行的数据为一条线。爬虫如果将代理设置为｛“http”：None，“https”：None｝是否有可能会被封ip
3. 感觉目前学习到的一些Python的基础知识与C语言的知识是共通的，但是又不是很确定

（Python解释器的底层是使用 C语言编写的）

除了CPthon，还有JPython

1. python的开发环境有什么用

用于执行Python代码

1. pycharm编译运行时好久打不开，下面显示一直在转

（只能等）

1. 数据结构中的元组、列表与字典有何关联，能否结合运用列表与字典实现动态IP地址讲解？循环语句中的for与while的区别？

动态IP，需要花钱买IP

1. 如何获取不知sheet表序号的sheet表数据

可以使用sheet表的中名称

1. return用的时机不太清楚
2. 格式化｛：^6｝\t｛：^｝…不太明白

100问75

1. 有时候scrapy框架的有的问题总是找不到错误的原因，希望杨老师补充一些scrapy更系统的类
2. 在爬取链家二手房时只显示100页，剩余部分如何爬取

（有多少页就可以爬取多少页）

1. scrapy框架：定义类方法def parse(self, response):

推送时yield scrapy.Requst（class.parse） parse方法不用传参数response？

1. crawl框架的rules使用 定义的如rule正则匹配不要加域名么？ 举例？
2. formRequest是不是等同于request.post发送请求，就多了一步回调？
3. scrapy中间件的概念应用？
4. turtle模块怎么去掉黑箭头（二级第9章）



1. 希望多讲讲正则表达式 （问班主任要肖老师的直播）
2. 希望能写出类似画小猪佩奇的程序
3. 云服务器是什么？是怎样来调用？购买学生云盘是否性价比比较高
4. 爬虫有时候网页爬不到怎么办
5. class的类里面的self和init是什么意思
6. 本地电脑git的用法
7. 面向对象
8. python re模块中 (?P<name>) (?P=name) 及 \g<name> 三者的使用区别
9. 学生管理系统的delete函数编写看不懂
10. Pandas模块中有DataFra/me数据格式与excel格式互换，Excel数据如何转换成DataFrame数据格式
11. 请教一个Python的问题，网站爬虫，我知道一个网站的网址，这个网址没有index，只有视频或者图片资源，当不知道资源名字就无法打开，比如：http://12345.com/aaaaa.mp4 就可以打开电影文件，如果我用爬虫，如何获取该网址的目录和文件，如何爬取该网址下面的所有资源。



https://item.jd.com/13183806.html

