**学习中遇到的困难以及希望杨老师解答的问题：**

1. python具体用途，如何学好，如何联动

（问题不够具体）

1. 多个类相互调用，里面的初始化参数如self.data等是不是可以相互调用，在不同类中调用修改后，会影响其他类中的self.data的数据。

（要有具体的代码）

1. 代码写了很长，很难找出bug，有没有方便的方法进行检查。

（无）

1. for 循环，while循环

（这两个怎么了？是不会写代码，还是语法不会，问题描述不清楚）

1. 能否介绍自然语言分析的项目

（问题不够具体）

1. 我有一个文件夹里面有400多个excel文件，我想把所有的文件名读取存入一个列表然后根据这个列表依次打开每个文件并提取里面的数据存入一个Excel文件中

**（提供几个Excel文件）**

1. 不懂字典和列表的转换 d=dict(eval(item)) 如果列表中存在重复的元素在写进字典不就报错了么（图一）

（列表里的每一个元素都是一个字典，字典中的Key只有id,name,english,python,java）

1. 讲一下各各数据类型的转换知识和注意事项
2. print(1,1)这句代码执行的结果中，两个1中间的空格能否去掉？

（去不掉）

1. python分词插件在数据采集中的作用是什么，如何嵌入？（图二）（图三）
2. 爬取知乎前三页的数据失败，想请老师看看哪里有问题（图四）（图五）

代码

(没有安装requests模块)

|  |
| --- |
| import requests start\_url="https://www.zhihu.com/api/v4/members/lisanshui1230/articles" headers={     "user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/89.0.4389.82 Safari/537.36"  } offset=0 while True:     print('当前offset的值是:%s,0表示第一页,20表示第二页,40表示第三页' %offset)     params={         "include": "data[\*].comment\_count,suggest\_edit,is\_normal,thumbnail\_extra\_info,thumbnail,can\_comment,comment\_permission,admin\_closed\_comment,content,voteup\_count,created,updated,upvoted\_followees,voting,review\_info,is\_labeled,label\_info;data[\*].vessay\_info;data[\*].author.badge[?(type=best\_answerer)].topics",         "offset": str(offset),         "limit": "20",         "sort\_by": "created"         }     response=requests.get(url=start\_url,params=params,headers=headers)     articles=response.json()     datas=articles['data'] #print(datas)     for article in datas:         title=article['title']         offset=offset+20         if offset>=60:             break         print(title) |

|  |
| --- |
| **import** requests start\_url=**"https://www.zhihu.com/api/v4/members/lisanshui1230/articles"** headers={  **"user-agent"**: **"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/89.0.4389.82 Safari/537.36"** } resp=requests.get(start\_url,headers=headers) print(resp.text) data=resp.json() print(data) |

1. 老师我想请杨老师讲一讲return和print的使用场景和区别方法，另外缩紧可不可以讲一下基础不好的不会用缩紧方法

（函数是否带返回值是吗？）

1. 给定一个字符串 s ，请你找出其中不含有重复字符的 最长子串 的长度。
2. 了解列表生成式else后的执行原理。QualityPointSource = [{"q\_point\_source":"Zp8","q\_point\_nickname":""},{"q\_point\_source":"RoadTest","q\_point\_nickname":"路试"}] qpoint = [[x["q\_point\_source"], x["q\_point\_nickname"] if x["q\_point\_nickname"] else x["q\_point\_source"]] for x in QualityPointSource] print(qpoint) [['Zp8', 'Zp8'], ['RoadTest', '路试']] （图六）
3. 关于id函数的问题，什么时候会内存地址相同，什么时候肯定不同，有没有准确的范围（截图效果在IDLE交互模式下测试）（图七）
4. re.S和re.M模式有什么区别 能否举例说说

（re.S单行匹配，re.M多行匹配）

1. 这部小说有1700多章节 ，普通爬的太耗时 （图8）
2. 能不能用到scrapy框架，下载时间会缩短么，
3. 能不能用多线程来写，怎么实现（整部小说切割成几部分，一个线程下载一部分最后把几个txt整合起来？）
4. a和b方案，哪个好一点（节省时间）
5. 代码

|  |
| --- |
| import requests  from bs4 import BeautifulSoup  import re  import time  # https://www.yqzw2.com/files/article/html/5/5330/1119.html  # https://www.yqzw2.com/files/article/html/5/5330/3190845.html  def parse():  hd={  "User-Agent": "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/86.0.4240.198 Safari/537.36",  "referer": "https://www.yqzw2.com/files/article/html/5/5330/index.html",  "cookie": "PageStyle=#EFF6FC||#000000||||; \_\_guid=26553659.-123641357896290760000.1634467156142.5056; Hm\_lvt\_3d85bd881cef0e5e0f68c531c378f06d=1634467157; monitor\_count=20; Hm\_lpvt\_3d85bd881cef0e5e0f68c531c378f06d=1634471850"  }  url="https://www.yqzw2.com/files/article/html/5/5330/index.html"  resp=requests.get(url,headers=hd)  resp.encoding="gb2312"  bs=BeautifulSoup(resp.text,"html.parser")  div\_lst=bs.find\_all("div",class\_="zjlist4")  lst=[]  # print(div\_lst)  # print("\*\*\*\*\*\*\*\*\*\*\*\*\*")  for item in div\_lst:  title=item.find\_all("a") # [[],[],[]]  # print(title)  # print("------------------")  for i in title:  # print(i)  # if int(re.findall("\d{1,4}",i.text)[0])>=1600:  targe=re.findall("\d+",str(i))  # print(targe[0])  # # for index in i:  t=str(i.text)+"\r\n"+"\r\n"+"\r\n"  # lst.append(t)  # print(t)  html=url.replace("index",targe[0])  # print(html)  resp=requests.get(html,headers=hd)  resp.encoding = "gb2312"  bs = BeautifulSoup(resp.text, "html.parser")  content=bs.find("div",class\_="contentbox")  # print(content.text.strip())  t+=content.text.strip()+"\r\n"+"\r\n"+"\r\n"  # print(t)  lst.append(t)  print("---"\*4)  time.sleep(2)  # print(lst)  write(lst)  def write(lst):  with open("note.txt","w", encoding="utf-8") as file:  for item in lst:  file.write(item)  if \_\_name\_\_ == '\_\_main\_\_':  parse() |

1. 最近做的一个小玩意儿需要用到朗读文本，能否演示一下tts放音模块的使用？比如读取文本了，暂停 变声等功能。不知道有没有相应的现成模块