

python网络爬虫五个小实例

实例一

爬取京东商品信息

```
1 #导入requests库
2 import requests
3 url="https://item.jd.com/6946605.html"
4 #异常处理
5 try:
6     r=requests.get(url)
7     #如果发送了错误请求，通过raise_for_status()抛出异常
8     r.raise_for_status()
9     #编码处理
10    r.encoding=r.apparent_encoding
11    #输出字符串的前1000个元素
12    print(r.text[:1000])
13 except:
14    print("爬取失败")
```

实例二

爬取亚马逊商品信息

```
1 import requests
2 url="https://www.amazon.cn/gp/product/B07D51Y714"
3 try:
4     kv={'user-agent':'Mozilla/5.0'}
5     r=requests.get(url,headers kv)
6     r.raise_for_status()
7     r.encoding=r.apparent_encoding
8     print(r.text[1000:2000])
9 except:
10    print("爬取失败")
11
```

实例三

百度搜索

```
1 import requests
2 keyword="Python"
3 try:
4     kv={'wd':keyword}
5     r=requests.get("http://www.baidu.com/s",params=kv)
6     print(r.request.url)
7     r.raise_for_status()
8     print(len(r,text))
9 except:
10     print("爬取失败")
11
```

实例四

爬取单张网络图片

```
1 import requests
2 import os
3 url="http://img0.dili360.com/ga/M00/4A/77/wKgBzFsFM-
4 2ADQ9iAC7rBKt1uIE377.tub.jpg"
5 root="E:/pics//"
6 path=root +url.split('/')[-1]
7 try:
8     #判断在本地是否有该文件 如果没有创建该文件
9     if not os.path.exists(root):
10         os.mkdir(root)
11     if not os.path.exists(path):
12         r=requests.get(url)
13         #以二进制只写的方式打开文件，生成文件对象f
14         with open(path,'wb') as f:
15             f.write(r.content)
16             f.close()
17             print("文件保存成功")
18     else:
19         print("文件已存在")
20 except:
21     print("爬取失败")
```

实例五

IP地址查询

```
1 import requests
2 url="http://www.ip138.com/ips138.asp?ip="
3 try:
4     r=requests.get(url+'202.204.80.112')
5     #raise_for_status() 方法是用来检测异常的
6     r.raise_for_status()
7     #编码操作 避免出现乱码问题
8     r.encoding=r.apparent_encoding
9     print(r.text[-500:])
10 except:
11     print("爬取失败")
```