

第6课 爬取五月天全部歌曲清单答案

1、题目要求

- 爬取猫眼电影信息，将电影名称，主演，上映时间爬取下来

使用技术点引导

- requests
- json
- 分析网址寻找目标网址
- 处理反爬技术-准备请求头

结果图片截取展示：

绿皮书

主演：维果·莫腾森,马赫沙拉·阿里,琳达·卡德里尼

上映时间：2019-03-01

少年的你

主演：周冬雨,易烱千玺,尹昉

上映时间：2019-10-25

```
1 import requests
2 from bs4 import BeautifulSoup
3 url = 'https://maoyan.com/board'
4 #处理反爬技术-准备请求头
5 headers = {
6     'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_3)
7     AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.149 Safari/537.36'
8 }
9
10 res = requests.get(url,headers=headers)
11 html = BeautifulSoup(res.text,'html')
12 data=html.find_all('div',class_='movie-item-info')
13
14 for mov in data:
15     mov_name = mov.find('p',class_='name')
16     print(mov_name.text)
17     mov_title = mov.find('p',class_='star')
18     print(mov_title.text)
19     mov_time = mov.find('p',class_='releasetime')
20     print(mov_time.text)
21     print()
22
```

结果为：

绿皮书

主演：维果·莫腾森,马赫沙拉·阿里,琳达·卡德里尼

上映时间：2019-03-01

少年的你

主演：周冬雨,易烊千玺,尹昉

上映时间：2019-10-25

红海行动

主演：张译,黄景瑜,海清

上映时间：2018-02-16

湄公河行动

主演：张涵予,彭于晏,冯文娟

上映时间：2016-09-30

白蛇：缘起

主演：张喆,杨天翔,唐小喜

上映时间：2019-01-11

流浪地球

主演：吴京,屈楚萧,李光洁

上映时间：2019-02-05

寻龙诀

主演：陈坤,黄渤,舒淇

上映时间：2015-12-18

捉妖记

主演：白百何,井柏然,姜武

上映时间：2015-07-16

滚蛋吧！肿瘤君

主演：白百何,吴彦祖,李媛

上映时间：2015-08-13

一条狗的使命

主演：布丽特·罗伯森,乔什·加德,丹尼斯·奎德

上映时间：2017-03-03

2、题目要求：

- 请自动获取用户输入的关键字完成搜索 360的关键词接口是q 统计出输入关键字后爬取下来数据的数量,通过捕获异常的方法如果爬取失败输出爬取失败字样

使用技术点引导

- requests
- input()
- try...except...

地址：<http://so.com/s>

```
1 import requests
2 keyword = input(":")
3 try:
4     kv = {'q':keyword}
5     r = requests.get('http://so.com/s',params=kv)
6     print(len(r.text))
7 except:
8     print("爬取失败")
```

结果：

```
1
296655
```

3、题目要求：

- 根据已给网址爬取果壳前10页标题和对应的网址信息

使用技术点引导

- requests
- BeautifulSoup
- for循环

地址： <https://www.guokr.com/ask/highlight/?page=1>

```
1 import requests
2 from bs4 import BeautifulSoup
3
4 for page in range(1,11):
5     print('已下是第{}页数据: {}'.format(page))
6     res = requests.get('https://www.guokr.com/ask/highlight/?page=
7     {}'.format(page))
8     html = BeautifulSoup(res.text, 'html')
9     data_ = html.find_all('ul', class_='ask-list-cp')
10    a=data_[0].find_all('div', class_='ask-list-detials')
11    for i in a:
12        res_data= i.find('a')
13        print('标题: ', res_data.text)
14        url = res_data['href']
15        print(url)
16    print()
```