

第8课 项目实操答案

题目要求:

- 已给出网址请大家自行分析网站结构, 将下图用红框圈出来的部分爬取到本地并写入excel表格中, 数据取前12条即可。



独家视频 | 习近平：解决问题的宗旨就是为人民服务

30日下午，习近平总书记浙江安吉县考察。

置顶 作者：央视新闻客户端 2020-03-31 12:27

👍(9) 💬(0)

地址: <https://www.shobserver.com/news/list?section=42>

```
1 import requests
2 import openpyxl
3 from bs4 import BeautifulSoup
4
5 wb = openpyxl.Workbook()
6
7 # 利用openpyxl.Workbook()函数创建新的workbook（工作簿）对象，就是创建新的空的
  Excel文件。
8 sheet = wb.active
9
10 # wb.active就是获取这个工作簿的活动表，通常就是第一个工作簿，也就是我们在上面的图片
   中看到的sheet1。
11 sheet.title = 'kaikeba'
12 sheet['A1'] = '标题'
13 sheet['B1'] = '内容'
14 sheet['C1'] = '作者及日期'
15 headers = {
16     'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_3)
   AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.149 Safari/537.36'
17 }
18 res = requests.get('https://www.shobserver.com/news/list?
   section=42', headers=headers)
19 data = res.text
20 # print(data)
21 html = BeautifulSoup(data, 'html')
22 data_ = html.find_all('div', class_='chengshi')
23 # print(len(data_))
24 for i in data_:
25     a = i.find('div', class_='chengshi_wz_h').text
26     res_data = i.find('div', class_='chengshi_wz_m').text
27     zz_data = i.find('div', class_='chengshi_wz_f').text
28     print(a)
```

```
30     print(res_data)
31     print(zz_data)
32     print('====='*10)
33     sheet.append([a,res_data,zz_data])
36 wb.save('new.xlsx')
37 # 保存修改的Excel
38 wb.close()
39 print("数据已经爬取完毕，已经保存到文件中，文件名称为new.xlsx")
```

结果截取展示：

第一观察 | 鉴往知来——跟着总书记走进绿水青山
早在15年前，习近平到这里调研，提出了“绿水青山就是金山银山”，留下了一段佳话。

置顶

作者：新华社 2020-03-31 15:29

独家特稿 | 疫情“大考”下的中小企业发展 习近平始终牵挂于心
一句问候，牵挂的是一线工人的身体健康；一句问候，惦念的是中小企业的生产状况。

置顶

作者：央视网 张媛 2020-03-31 15:16

15年来，习近平不断发展深化这个理念
绿水青山就是金山银山。

置顶

作者：新华网 2020-03-31 15:16