

题目： 目标网址： https://cd.58.com/danche/?utm_source=market&spm=u-2d2yxv86y3v43nkddh1.BDPCPZ_LOGO&PGTID=0d100000-0006-65cf-8c8e-dfd8022dc585&ClickID=31 目标内容：爬取二手摩托信息，包括题目、内容、地址、链接、价格等内容。 任务要求：使用 XPath 完成，结果保存到 CSV 文件。

代码:

```
import lxml.html
```

import requests

```
import csv
```

```
import re
```

```
url = "https://cd.58.com/danche/?utm_source=market&spm=u-2d2yxv86y3v43nkddh1.BDPCPZ_LOGO&PGTID=0d100000-0006-65cf-8c8e-dfd8022dc585&ClickID=31"
```

```
html1 = requests.get(url).content.decode()
```

[illegible]

```
html3 = re.sub('\t\t',",",html2)
```

```
html4 = re.sub('\r\n\t',",html3)
```

```
html5 = re.sub('\t',"",html4)
```

[illegible]

```
selector = lxml.html.fromstring(html)
```

#获取题目

```
name list = selector.xpath('//td[@class="t"]/a/text()')
```

#获取内容

```
content list = selector.xpath('//div[@class="item-desc"]/text()')
```

#获取地址

```
add list = selector.xpath('//a[@class="c 666"][1]/text()')
```

```
#获取 url
```

```
url_list = selector.xpath('//td[@class="t"]/a/@href')
```

#获取价格

```
price_list = selector.xpath('//b[@class="pri"]/text()')
```

#字典

```
dic = []
```

```
for i in range(3,len(name_list)):
```

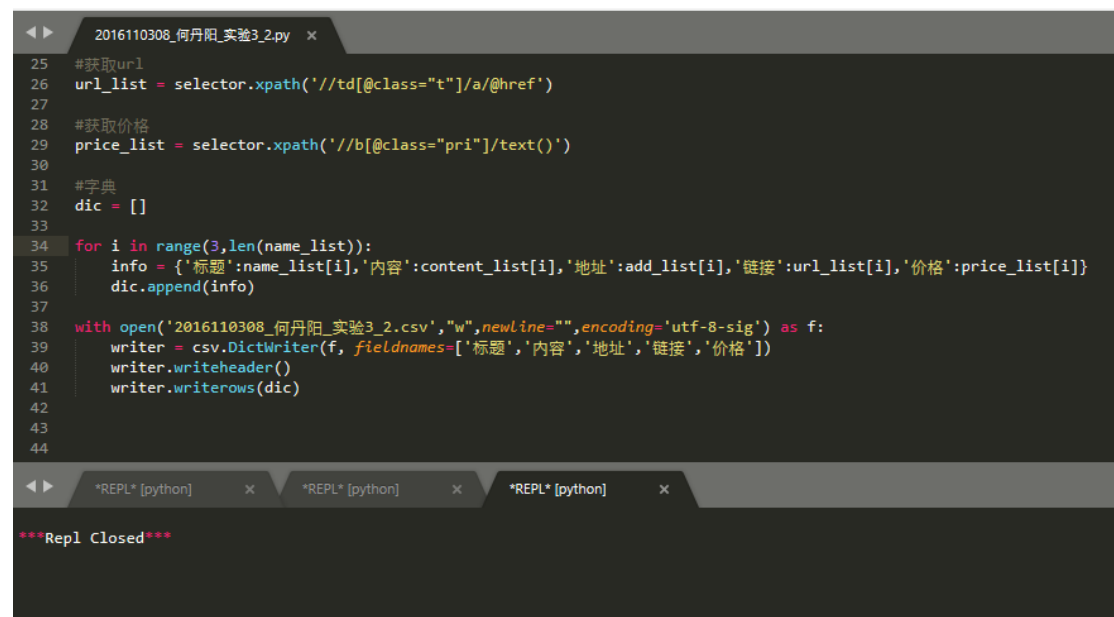
```
info = {'标题':name_list[i], '内容':content_list[i], '地址':add_list[i], '链接':url_list[i], '价格':price_list[i]}
```

```
dic.append(info)
```

```
with open('2016110308_何丹阳_实验 3_2.csv','w',newline="",encoding='utf-8-sig')  
as f:
```

```
    writer = csv.DictWriter(f, fieldnames=['标题','内容','地址','链接','价格'])  
    writer.writeheader()  
    writer.writerows(dic)
```

截图：



```
25 #获取url  
26 url_list = selector.xpath('//td[@class="t"]/a/@href')  
27  
28 #获取价格  
29 price_list = selector.xpath('//b[@class="pri"]/text()')  
30  
31 #字典  
32 dic = []  
33  
34 for i in range(3,len(name_list)):  
35     info = {'标题':name_list[i],'内容':content_list[i],'地址':add_list[i],'链接':url_list[i],'价格':price_list[i]}  
36     dic.append(info)  
37  
38 with open('2016110308_何丹阳_实验3_2.csv','w',newline="",encoding='utf-8-sig') as f:  
39     writer = csv.DictWriter(f, fieldnames=['标题','内容','地址','链接','价格'])  
40     writer.writeheader()  
41     writer.writerows(dic)  
42  
43  
44
```

Repl Closed