

# 抓取北京新发地菜价

```
1 from concurrent.futures import ThreadPoolExecutor
2 import csv
3 import requests
4 from lxml import etree
5
6
7 # 干掉新发地
8 def download(url):
9     resp = requests.get(url)
10    html = etree.HTML(resp.text)
11    table =
    html.xpath("/html/body/div[2]/div[4]/div[1]/table
    ")[0]
12    trs = table.xpath("./tr[position()>1]")
13    for tr in trs:
14        tr_text = tr.xpath("./td/child::text()")
15        tr_text = (item.replace("\\",
        "").replace("/", "")) for item in tr_text)
```

```
16         csv_writer.writerow(tr_text)
17     return url
18
19
20 if __name__ == '__main__':
21     csv_writer = csv.writer(open("菜价.csv",
22 mode="w"))
23     with ThreadPoolExecutor(50) as t:
24         for i in range(1, 100): # 别弄太狠. 不好
25             url_temp =
26                 f"http://www.xinfadi.com.cn/marketanalysis/0/list
27                 /{i}.shtml"
28             t.submit(fn=download, url=url_temp)
29     print("all down!")
```