

《用 Python 玩转数据》爬虫小实验

1. 请在豆瓣任意找一本图书，抓取它某一页的短评并进行页面解析将短评文字抽取后输出，再对其中的评分进行抽取计算其总分。

参考程序见下一页

【参考代码：bookid 换成自己选择的书名 id】

```
# -*- coding: utf-8 -*-
```

```
"""
```

```
Crawler
```

```
@author: Dazhuang
```

```
"""
```

```
import requests
```

```
from bs4 import BeautifulSoup
```

```
import re
```

```
s = 0
```

```
r = requests.get('https://book.douban.com/subject/bookid/comments/')
```

```
soup = BeautifulSoup(r.text, 'lxml')
```

```
pattern = soup.find_all('span', 'short') # 网站前一版本标签为 “p, 'comment-content' ”
```

```
for item in pattern:
```

```
    print(item.string)
```

```
pattern_s = re.compile('<span class="user-stars allstar(.*) rating"')
```

```
p = re.findall(pattern_s, r.text)
```

```
for star in p:
```

```
    s += int(star)
```

```
print(s)
```