

第3课 我的第一个小爬虫答案

1、题目要求：

- 请使用requests爬取文章，开掘一座文学理论富矿。并要求使用BeautifulSoup4提取出文本数据，将所有的标签部分去掉。

地址为：<https://baijiahao.baidu.com/s?id=1661382527708632196&wfr=spider&for=pc>

```
1 import requests
2 from bs4 import BeautifulSoup
3 res = requests.get('https://baijiahao.baidu.com/s?
4 id=1661382527708632196&wfr=spider&for=pc')
5 soup = BeautifulSoup(res.text, 'html.parser')
6 items = soup.find_all(class_='article-content')
7 title = items[0].find_all('p')
8 for data in range(8):
9     res = title[data].find('span', class_='bjh-p')
10    print(res.text)
```

结果：

《钱锺书文学思想研究》：罗新河著；中国社会科学出版社出版。

钱锺书是我国知名学者，有人用“精微、会通、自得”概括他的学术研究，有人评价钱锺书学问沟通中西、融汇古今，但更要看到，钱锺书的文学研究具有实践品格，他珍视民族文化独特性，致力于中华优秀传统文化的当代转化，如钱锺书所说：“我们的兴趣和研究是现代的”。结合时代社会需要，对钱锺书的开拓性努力进行探讨和总结，为当下文化发展提供有益参考，是一个有价值的课题。罗新河教授《钱锺书文学思想研究》就围绕这一课题展开。

在长期学术实践中，钱锺书对文学一直有自觉的关注与思考。他深度汲取中西文化养分以丰富自身素养，取得广博精深的学术成果，为20世纪中国文学理论研究留下一笔丰厚遗产。从研究现状来看，钱锺书的文学理论仍是一座富矿，部分命题需要进一步深入探讨。《钱锺书文学思想研究》在现代哲学社会科学逻辑框架下，围绕文学的本质、文本、创作、评论等范畴进行系统化阐释。该书将钱锺书实践性与学理性结合的文学理念予以梳理、整合，把“散点透视”思想火花置入一张“理论地图”之中，使读者把握起来有章可循。这种“整合式”研究彰显了钱锺书理论个性与学术贡献，在前人研究成果基础上进行了丰富和补充。

立足当今，用学术研究成果增强文化自信，是当前学界重要使命。这需要创新研究阐发中华优秀传统文化，处理好文化的“本来”“外来”与“未来”关系问题。

《钱锺书文学思想研究》对这一点的挖掘和阐释尤为突出。在作者看来，钱锺书著作中所体现的“打通”方法论，有效实现了对传统文化的现代阐释、更新与转化，建立了中西文化平等对话的一种范式，扩大了文化阐释空间，释放了传统文化的活力与价值。这一理论阐释，揭示出钱锺书文学观点的开放性、辩证性与时代性，为增强文化自信和增进文化传承，总结了可资借鉴的经验。

该书在研究方法上，将历时溯源与共时扫描相结合，也从空间维度进行跨语际、跨文化的横向比较，力图做到时空交错、视域宏阔。该书从微观到宏观把握钱锺书的文学理论建树，显示出方法上的贴合和创新。作者认为，钱锺书将文学理解为时代心理状态的表现，比较准确把握了文学的本质内涵。他一方面搜索、梳理、总结散落在不同文本中的钱锺书文学观点；另一方面广泛参照各类典籍，将其文学观点放在人类文化的大背景中进行比较；同时结合时代语境，分析钱锺书文学观点产生的条件、目的等。通过视野拓展和理路创新，实现整体把握，构筑起包含文学本质论、文学创作论、文学作品论、文学接受论在内的钱锺书文学观点“地形图”。

《钱锺书文学思想研究》开掘了一座文学理论的富矿，有助于人们理解钱锺书的文学观点，有助于我们汲取理论资源和方式方法，更好实现中华优秀传统文化的创造性转化和创新性发展。

（作者为湖南工业大学教授）

《人民日报》（2020年03月17日 20版）

2、题目要求

- 爬取三国演义每回题目，将第一回到第一百一十九回的题目爬取到本地。

使用技术点引导

- 使用BeautifulSoup4解析器
- requests

地址为：<http://www.shicimingju.com/book/sanguoyanyi.html>

```
1 import requests
2 from bs4 import BeautifulSoup
3 res = requests.get('http://www.shicimingju.com/book/sanguoyanyi.html')
```

```
5 soup = BeautifulSoup(res.text, 'html')
6 items = soup.find_all(class_='book-mulu')
7 title = items[0].find_all('li')
8 for data in range(119):
9     res = title[data].find('a')
10    print(res.text)
```

结果:

第一回·宴桃园豪杰三结义 斩黄巾英雄首立功

第二回·张翼德怒鞭督邮 何国舅谋诛宦竖

第三回·议温明董卓叱丁原 馈金珠李肃说吕布

第四回·废汉帝陈留践位 谋董贼孟德献刀

第五回·发矫诏诸镇应曹公 破关兵三英战吕布

第六回·焚金阙董卓行凶 匿玉玺孙坚背约

第七回·袁绍磐河战公孙 孙坚跨江击刘表