

喜马拉雅-春秋

转载

[weixin_33713503](#) 于 2018-08-17 16:41:00 发布 76 收藏

文章标签: [json](#)

原文链接: <http://www.cnblogs.com/gxsmm/p/9494140.html>

版权

```
import requests
from lxml import etree
from urllib import parse

url = 'https://www.ximalaya.com/lishi/4164479/'

headers = {
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/67.0.3
}

response = requests.get(url, headers=headers)

# with open('chunqiu.html', 'wb') as f:
#     f.write(response.content)

html_str = response.text
# print(html_str)
html_ele = etree.HTML(html_str)
# print(html_ele)
li_list = html_ele.xpath('//ul[@class="d0i2"]/li')
# print(li_list)
for li_ele in li_list:
    # print(li_ele)
    li_href = li_ele.xpath('./div[2]/a/@href')[0]
    # print(li_href)

url_list = parse.urljoin(url, li_href)
# print(url_list)

url_a = url_list.split('/')[-1]
# print(url_a)
url_b = 'https://www.ximalaya.com/revision/play/tracks?trackIds='+ str(url_a)
# print(url_b)

response_a = requests.get(url_b, headers=headers)

json_a_dict = response_a.json()
# print(json_a_dict)

src_a_str = json_a_dict['data']['tracksForAudioPlay'][0]['src']
trackName_a = json_a_dict['data']['tracksForAudioPlay'][0]['trackName']

from urllib import request
request.urlretrieve(src_a_str, trackName_a+'.ma4')
```

