

# 稀土掘金

目标网站: <https://juejin.cn/>

第一种: 先获取所有的元素, 循环遍历

第二种: 边滑动, 边加载元素, 边获取内容

需求: 获取前50以上的文章, 保存到txt文档当中

翻页是滚轮滑动的, 加载元素, 加载50个元素以上

1. 初始化方法
2. 解析文章
3. 滑动
4. 保存
5. 主方法

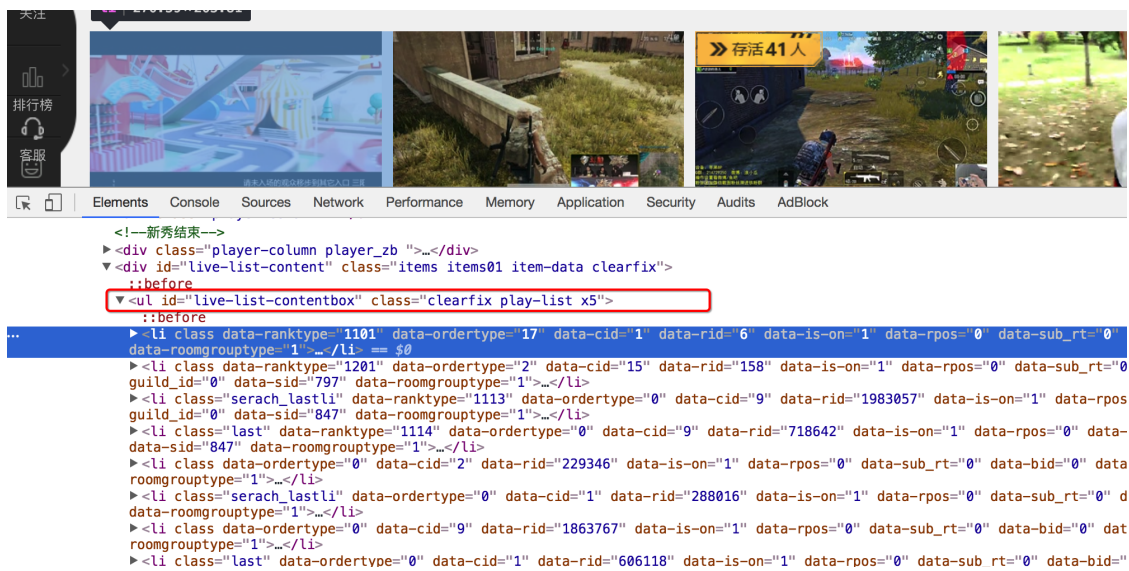
# 斗鱼爬虫

爬取斗鱼直播平台的所有房间信息: <https://www.douyu.com/directory/all>

**思路分析:**

1. 数据的位置

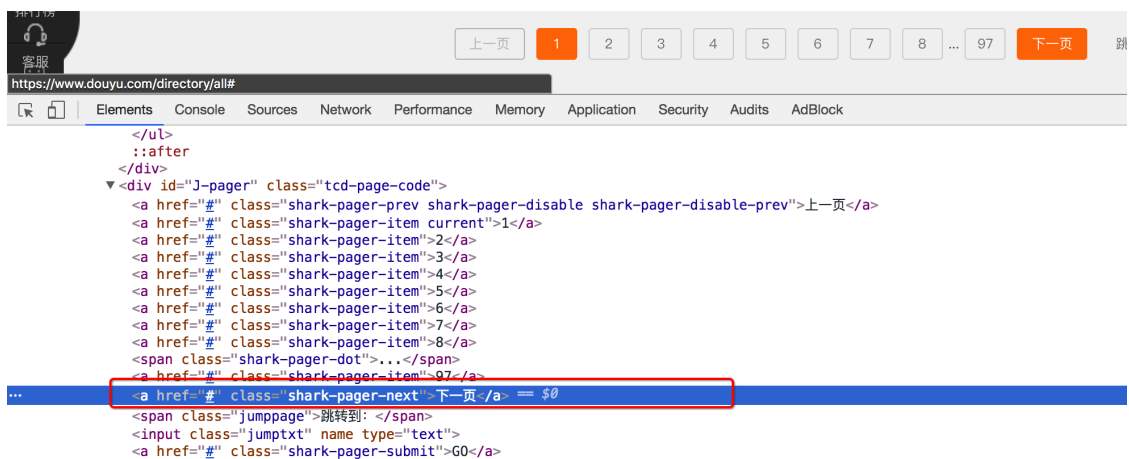
每个房间的数据都在 `id=live-list-contentbox` 的ul中



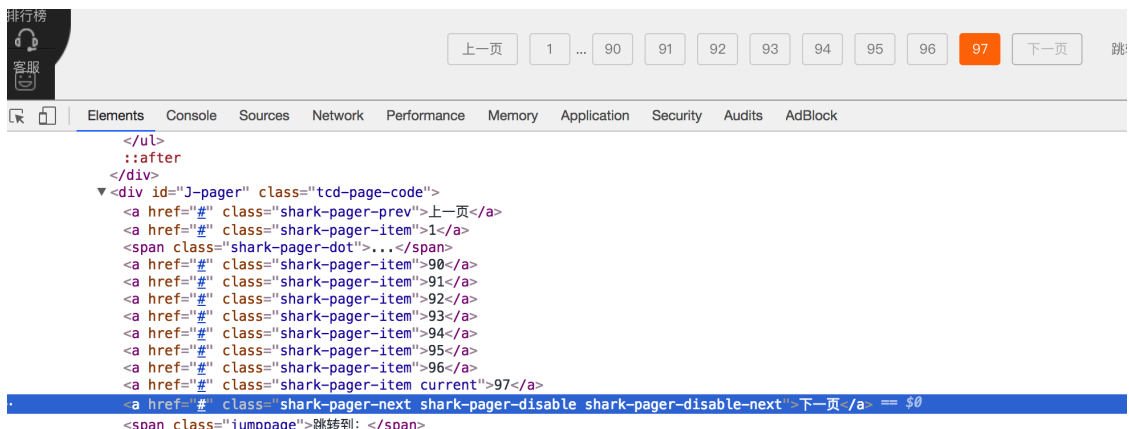
## 2. 实现翻页和控制程序结束

selenium中通过点击就可以实现翻页，观察存在下一页和不存在下一页时候标签的变化

存在下一页时：



不存在下一页时：



## 2. 作业

完成网易云音乐的所有分类下的所有的播放列表，包括：播放列表的标题和url地址

思路分析：

1. 确定程序入口的位置 首先可以定位到所有的大分类和所有的小分类

获取所有的小分类的地址后，遍历请求获取播放列表页的首页



2. 确定首页数据的位置和url地址

数据不是在url地址对应的响应中，而是在另一个url地址中，可以使用selenium，也可以使用requests来获取

重逢是始料未及 | 回...  
by 宠着长岛冰茶吗 ☆

华语说唱里的Trap...  
by 薛定谔的缝纫机 ☆

五一小长假养生秘笈  
by 神威有呆毛

岛·乐队集·台北天空...  
by 鲑鱼先森 ☆

music.163.com  
default\_album.jpg  
core.css?175ec07fc5be1e8aa7405f74aefe20f1  
pt\_frame.css?c8073c80262733bea3646d00c742944a  
core.js?78976c135030c06788fdd7aa9d1569933  
pt\_frame\_index.js?76d8d6ef37086b164baea69c05def0130  
topbar.png?bec8dd4512e808fba4b93843478ac83b  
playbar.png?bf71d7cc8c4ca8afd13856963e9f0fb  
iconall.png?d0b092412ae6f79b94baf81715e552a9  
statbar.png?78b51c3501dcb2d17f03013386edb22e8  
loading.gif?8d7a059b64d7ca828c49c553ea1ab2d5  
?cat=%E5%8D%8E%E8%AF%AD

General  
Request URL: http://music.163.com/discover/playlist/?cat=%E5%8D%8E%E8%AF%AD  
Request Method: GET  
Status Code: 200 OK  
Remote Address: 127.0.0.1:1086  
Referrer Policy: no-referrer-when-downgrade  
Response Headers (14)  
Request Headers  
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng  
Accept-Encoding: gzip, deflate  
Accept-Language: zh-CN,zh;q=0.9,en;q=0.8,ru;q=0.7,zh-TW;q=0.6  
Connection: keep-alive  
Cookie: JSESSIONID=WYYW=q%5CpDuj93miDv9nPmq3qfsfotobBPahD0EaY5530APRcm758SpG550uZ%  
igeH2K3R04R1Zc02E40F%5Ca41jaWjhae0rT5AHTDfY9vN4G%2B3AotF8BIRQ5CFewm%2Fo2x%3A1524%  
a07fd7,1524997467174; \_ntes\_nuid=a9f9602dd35f18458900941a38a07fd7; \_\_utma=94650624  
4; \_\_utmz=04650624\_1524007460\_1\_1\_\_utmcsr=(direct)|utmccn=(direct)|utmcmd=(none);

### 3. 实现翻页和程序停止的判断

存在下一页的情况：

关于网易 | 客户服务 | 服务条款 | 网站导航 | 意见反馈

关于网易 | 客户服务 | 服务条款 | 网站导航 | 意见反馈

网易公司版权所有©1997-2018 杭州乐读科技有限公司运营：浙网文[2015] 0415-135号  
违法和不良信息举报电话：0571-89853516 举报邮箱：cloudmusicservice@163.com

独立音乐人 音乐

Elements Console Sources Network Performance Memory Application Security Audits AdBlock

<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=0” class=“zpgi”>1</a>  
<span class=“zdot”>...</span>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1015” class=“zpgi”>30</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1050” class=“zpgi”>31</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1085” class=“zpgi”>32</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1120” class=“zpgi”>33</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1155” class=“zpgi”>34</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1190” class=“zpgi”>35</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1225” class=“zpgi”>36</a>  
<a href=“javascript:void(0)” class=“zpgi js-selected”>37</a>  
<a href=“javascript:void(0)” class=“zbtn znxt js-disabled”>下一页/a = \$0

不存在下一页的情况

关于网易 | 客户服务 | 服务条款 | 网站导航 | 意见反馈

关于网易 | 客户服务 | 服务条款 | 网站导航 | 意见反馈

网易公司版权所有©1997-2018 杭州乐读科技有限公司运营：浙网文[2015] 0415-135号  
违法和不良信息举报电话：0571-89853516 举报邮箱：cloudmusicservice@163.com

独立音乐人 音乐

Elements Console Sources Network Performance Memory Application Security Audits AdBlock

<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=0” class=“zpgi”>1</a>  
<span class=“zdot”>...</span>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1015” class=“zpgi”>30</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1050” class=“zpgi”>31</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1085” class=“zpgi”>32</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1120” class=“zpgi”>33</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1155” class=“zpgi”>34</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1190” class=“zpgi”>35</a>  
<a href=“/discover/playlist/?order=hot&cat=%E5%8D%8E%E8%AF%AD&limit=35&offset=1225” class=“zpgi”>36</a>  
<a href=“javascript:void(0)” class=“zpgi js-selected”>37</a>  
<a href=“javascript:void(0)” class=“zbtn znxt js-disabled”>下一页/a = \$0

地址: <http://music.163.com/#/discover/playlist>