

十一、抓取微博所有评论

例子：得到APP用户意见反馈微博

[https://weibo.com/5745019302/DCC6EAQdO?
type=comment#_rnd1529984083763](https://weibo.com/5745019302/DCC6EAQdO?type=comment#_rnd1529984083763)

分析翻页规律

https://weibo.com/5745019302/DCC6EAQdO?type=comment#_rnd1529984083763

https://weibo.com/5745019302/DCC6EAQdO?type=comment#_rnd1529984484944

https://weibo.com/5745019302/DCC6EAQdO?type=comment#_rnd1529984589206



满足2个条件，可用本节的翻页抓取方式

- 1、网页的 URL 无规律
- 2、可以点击“页码”翻页

设置Selector

The image shows a Weibo (微博) page with several comments. A red box highlights a comment by '杨翌 Xeyes' with the text '招点牛人做做技术吧, 跟不上啊'. Below it, another comment by '心有林夕 020' is highlighted with the text '精品课, 可不可以有文字版啊, 我有点听不来语音, 🤔'. A third comment by '北葵向暖之小鱼' is highlighted with the text '希望 App 可以完善, 更新'. A fourth comment by '得一 GO' is highlighted with the text '好想退订凡登读书会。。。不喜欢这个人, 你们又不给试听。。。后悔买了, 浪费节操'.

Below the comments, a red box highlights the selector 'div.list_con div.WB_text' and a button labeled 'Done selecting!'. The number '11' is written below this box.

The bottom part of the image shows the Web Scraper configuration interface. The 'Id' field is set to 'comment' (1). The 'Type' is 'Element click' (2). The 'Selector' is 'div.list_con div.WB_text' (3). The 'Click selector' is 'a.page.S_txt1' (4). The 'Click type' is 'Click once (pagination, tabs)' (5). The 'Click element uniqueness' is 'Unique Text' (6). The 'Multiple' checkbox is checked (7). The 'Delay (ms)' is '2000' (8). The 'Parent Selectors' are '_root' and 'comment' (12). The 'Save selector' button is highlighted (12). The text '见下张图' (See next figure) is written in red, with an arrow pointing to the 'Selector' field.

Web Scraper configuration fields:

- Id: comment
- Type: Element click
- Selector: div.list_con div.WB_text
- Click selector: a.page.S_txt1
- Click type: Click once (pagination, tabs)
- Click element uniqueness: Unique Text
- Multiple: ☒
- Delay (ms): 2000
- Parent Selectors: _root, comment
- Buttons: Save selector, Cancel

设置Selector

4.2

4.3

4.4

4.1

Imtheworld: 请问往期直播在哪里看?
11月14日 20:13

回复 | 赞

为了避免骚扰, 微博智能反垃圾系统会过滤掉部分广告用户。

1 2 3 4 5 6 ... 96 下一页

微博精彩 手机玩微博 认证 & 合作 微博帮助

热门微博 热门话题 扫码下载 更多版本

a.page.S_txt1 S P C Done selecting!

微相册 微游戏

申请认证 链接网站 常见问题
企业微博 广告服务 自助服务
微博标识 广告代理商

私信聊天

Elements Console Sources Network Performance Memory Application Security Audits Web Scraper

Sitemaps Sitemap (comment) Create new sitemap

Id comment

Type Element click

Selector Select Element preview Data preview div.list_con div.WB_text

Click selector Select Element preview a.page.S_txt1

Click type Click once (pagination, tabs)

Click element uniqueness Unique Text

☒ Multiple
☐ Discard initial elements that are available before click

Delay (ms) 2000

Parent Selectors _root
comment

Save selector Cancel

设置Selector

添加子选择器，不再赘述

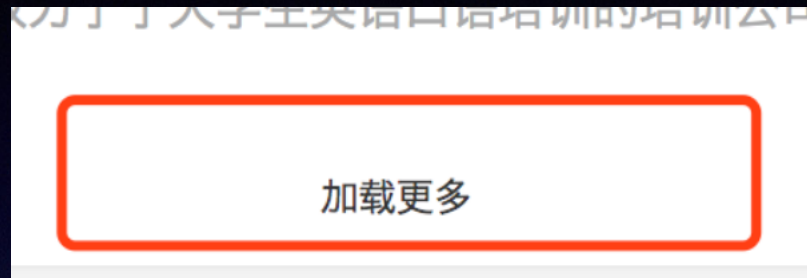
要点解析

Element click也是 Element（结构体）

只是多了一个**鼠标点击翻页的功能**

所以，如果要用它抓取数据，也需要创建一个子选择器

要点解析



Click more

Unique CSS Selector



Click once

Unique Text

数据不完整

网页显示的数字，和具体信息不一致