

目标：使用 lxml、bs、re 或 PyQuery 解析，爬取新浪某个大 V 的微博，mblog 包含以下信息：created_at(发布日期)、reposts_count(转发数目)、comments_count(评论数目)、attitudes_count(点赞数目)，text(微博正文)

代码：

```
import re
import requests
import csv
from urllib.parse import urlencode

base_url='http://m.weibo.cn/api/container/getIndex?'
headers={
    'Host':'m.weibo.cn',
    'Refer':'https://m.weibo.cn/u/1739046981?uid=1739046981&luicode=10000011&lfid=100103type%3D1%26q%3D%E6%9D%8E%E8%8D%A3%E6%B5%A9',
    'User-Agent':'Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:66.0) Gecko/20100101 Firefox/66.0',
    'X-Requested-With':'XMLHttpRequest',
}

def get_page(page):
    params={
        'type':'uid',
        'value':'1739046981',
        'containerid':'1076031739046981',
        'page':page,
    }
    url=base_url+urlencode(params)
    response=requests.get(url,headers=headers)
    if response.status_code == 200:
        return response.json()

dic = []
for i in range(1,10):
    json = get_page(i)
    cards = json['data']['cards']
    for j in range(len(cards)):
        mblog = cards[j].get('mblog')
        if mblog == None:
            continue
        created_at = mblog['created_at']
        reposts_count = mblog['reposts_count']
        comments_count = mblog['comments_count']
```

```

attitudes_count = mblog['attitudes_count']
text = mblog['text']
Text = re.sub('<.*?>', '', text)
info = {'时间':created_at,'转发数':reposts_count,'评论数':comments_count,'
点赞数':attitudes_count,'内容':Text}
dic.append(info)
print('*****')
print('时间: '+created_at)
print(f'转发: {reposts_count}')
print(f'评论: {comments_count}')
print(f'点赞: {attitudes_count}')
print('text: '+Text)

with open('2016110308_何丹阳_实验 4_1.csv','w',newline='',encoding='utf-8-sig')
as f:
    writer = csv.DictWriter(f, fieldnames=['时间','转发数','评论数','点赞数','内容'])
    writer.writeheader()
    writer.writerows(dic)

```

截图：

```

时间: 2018-08-30
转发: 10320
评论: 4283
点赞: 83757
text: 别人的想法 只是个想法
*****
时间: 2018-08-30
转发: 10139
评论: 2373
点赞: 46909
text: 都是这么想 这么想 这么想可以看起来不伟大 却要善良要尝过时间给过的挫败 才能够变得坦然《贫穷或富有》已在各平台上线 欢迎收听
*****
时间: 2018-08-22
转发: 5246
评论: 8533
点赞: 65032
text: 你猜我热不热
*****
时间: 2018-08-16
转发: 10092
评论: 6424
点赞: 67954
text: 彩虹屁什么意思
*****
时间: 2018-08-15
转发: 1086
评论: 667
点赞: 19118
text: 恭喜杉哥
*****
时间: 2018-08-14
转发: 11138
评论: 2476
点赞: 32046
text: 我爱❤️杜康

```