

爬虫实战——pyquery方法抓取的新闻要闻

使用lxml模块进行快速的 XML 和 HTML 操作。

安装方法:

```
pip install pyquery
```

导入方式:

```
from pyquery import PyQuery
```

使用方法 .text()、.html()

```
html = (
    '<a href="/202309/t20230920_1360717.html" target="_blank"
    title="2023&#x5E74;9&#x6708;20&#x65E5;24&#x65F6;&#x8D77;&#x56FD;&#x5185;&#x6210
    ;&#x54C1;&#x6CB9;&#x4EF7;&#x683C;&#x6309;&#x673A;&#x5236;&#x8C03;&#x6574;">2023
    年9月20日24时起国内成品油价格按机制调整</a><span>2023/09/20</span>'
    '<a href="/202309/t20230907_1360442.html" target="_blank"
    title="&#x56FD;&#x5BB6;&#x53D1;&#x5C55;&#x6539;&#x9769;&#x59D4;&#x6709;&#x5173;
    &#x8D1F;&#x8D23;&#x540C;&#x5FD7;&#x5C31;&#x300A;&#x4E1C;&#x839E;&#x6DF1;&#x5316
    ;&#x4E24;&#x5CB8;&#x521B;&#x65B0;&#x53D1;&#x5C55;&#x5408;&#x4F5C;&#x603B;&#x4F5
    3;&#x65B9;&#x6848;&#x300B;&#x7B54;&#x8BB0;&#x8005;&#x95EE;">国家发展改革委有关负责
    同志就《东莞深化两岸创新发展合作总体方案》答记者问</a><span>2023/09/07</span>'
    '<a href="/202309/t20230906_1360399.html" target="_blank"
    title="&#x56FD;&#x5BB6;&#x53D1;&#x5C55;&#x6539;&#x9769;&#x59D4;&#x53EC;&#x5F00;
    &#x5B66;&#x4E60;&#x8D2F;&#x5F7B;&#x4E60;&#x8FD1;&#x5E73;&#x65B0;&#x65F6;&#x4EE3
    ;&#x4E2D;&#x56FD;&#x7279;&#x8272;&#x793E;&#x4F1A;&#x4E3B;&#x4E49;&#x601D;&#x60F
    3;&#x4E3B;&#x9898;&#x6559;&#x80B2;&#x603B;&#x7ED3;&#x4F1A;&#x8BAE;">国家发展改革
    委召开学习贯彻习近平新时代中国特色社会主义思想主题教育总结会议</a><span>2023/09/06</span>'
    '<a href="/202309/t20230906_1360397.html" target="_blank"
    title="2023&#x5E74;9&#x6708;6&#x65E5;&#x56FD;&#x5185;&#x6210;&#x54C1;&#x6CB9;&#
    x4EF7;&#x683C;&#x4E0D;&#x4F5C;&#x8C03;&#x6574;">2023年9月6日国内成品油价格不作调整
    </a><span>2023/09/06</span>'
)
pq = PyQuery(html)
print(pq.html())
```

输出:

```
'<a href="/202309/t20230920_1360717.html" target="_blank" title="2023&#x5E74;9&#x6708;20&#x65E5;
24&#x65F6;&#x8D77;&#x56FD;&#x5185;&#x6210;&#x54C1;&#x6CB9;&#x4EF7;&#x683C;&#x6309;&#x673A;&#x523
6;&#x8C03;&#x6574;">2023年9月20日24时起国内成品油价格按机制调整</a><span>2023/09/20</span><a href="/.2
02309/t20230907_1360442.html" target="_blank" title="&#x56FD;&#x5BB6;&#x53D1;&#x5C55;&#x6539;&#x9
769;&#x59D4;&#x6709;&#x5173;&#x8D1F;&#x8D23;&#x540C;&#x5FD7;&#x5C31;&#x300A;&#x4E1C;&#x839E;&#x6D
F1;&#x5316;&#x4E24;&#x5CB8;&#x521B;&#x65B0;&#x53D1;&#x5C55;&#x5408;&#x4F5C;&#x603B;&#x4F53;&#x65B
9;&#x6848;&#x300B;&#x7B54;&#x8BB0;&#x8005;&#x95EE;">国家发展改革委有关负责同志就《东莞深化两岸创新发展合作
总体方案》答记者问</a><span>2023/09/07</span><a href="/202309/t20230906_1360399.html" target="_blan
k" title="&#x56FD;&#x5BB6;&#x53D1;&#x5C55;&#x6539;&#x9769;&#x59D4;&#x53EC;&#x5F00;&#x5B66;&#x4E6
0;&#x8D2F;&#x5F7B;&#x4E60;&#x8FD1;&#x5E73;&#x65B0;&#x65F6;&#x4EE3;&#x4E2D;&#x56FD;&#x7279;&#x827
2;&#x793E;&#x4F1A;&#x4E3B;&#x4E49;&#x601D;&#x60F3;&#x4E3B;&#x9898;&#x6559;&#x80B2;&#x603B;&#x7ED
3;&#x4F1A;&#x8BAE;">国家发展改革委召开学习贯彻习近平新时代中国特色社会主义思想主题教育总结会议</a><span>2023/
09/06</span><a href="/.202309/t20230906_1360397.html" target="_blank" title="2023&#x5E74;9&#x670
8;6&#x65E5;&#x56FD;&#x5185;&#x6210;&#x54C1;&#x6CB9;&#x4EF7;&#x683C;&#x4E0D;&#x4F5C;&#x8C03;&#x657
4;">2023年9月6日国内成品油价格不作调整</a><span>2023/09/06</span>'
```

```
print(pq.text())
```

输出:

```
[34]: 1 pq.text()
[34]: '2023年9月20日24时起国内成品油价格按机制调整2023/09/20国家发展改革委有关负责同志就《东莞深化两岸创新发展合作总体方案》答记者问2023/09/07国家发展改革委召开学习贯彻习近平新时代中国特色社会主义思想主题教育总结会议2023/09/062023年9月6日国内成品油价格不作调整2023/09/06'
```

看一下类型:

```
pq[0]
# <Element span at 0x7f65aa0dd5e0>
```

是一个

lxml.etree._Element 类型, 可以用xpath语法解析

检索方式

对pyquery对象('标签'), 即可获得所有标签, 支持css选择器

.items 可以将每个元素转化成pyquery迭代器

- id选择器 #
- 标签选择器 标签
- 类选择器 .
- 选择器分类,
- 后代选择器 空格

```
for i in pq('a').items():
    print(type(i))
'''
<class 'pyquery.pyquery.PyQuery'>
<class 'pyquery.pyquery.PyQuery'>
<class 'pyquery.pyquery.PyQuery'>
<class 'pyquery.pyquery.PyQuery'>
'''
```

元素属性可以使用attr()方法检索。

```
for i in pq('a').items():
    print(i.attr('href'))
'''
./202309/t20230920_1360717.html
./202309/t20230907_1360442.html
./202309/t20230906_1360399.html
./202309/t20230906_1360397.html
'''
```

.remove 删除标签

```
url = 'https://www.ndrc.gov.cn/xwdt/xwfb/'
reponse = requests.get(url, headers=head)
reponse.encoding= 'utf-8'
pq = PyQuery(reponse.text)
pq('ul.u-list')('li')
```

输出

```
# [<li>, <li>, <li>, <li>, <li>, <li.empty>, <li>, <li>, <li>, <li>, <li>,
<li.empty>, <li>, <li>, <li>, <li>, <li>, <li.empty>, <li>, <li>, <li>, <li>,
<li>, <li.empty>, <li>, <li>, <li>, <li>, <li>, <li.empty>]
```

删除 li.empty 标签

[illegible]

要闻抓取

观察新闻信息都在li标签中

☒ Preserve log
 ☒ Disable cache
 No throttling

Filter ☐ Invert ☐ Hide data URLs

All **Fetch/XHR** JS CSS Img Media Font Doc WS Wasm Manifest Other ☐ Has blocked cookies ☐ Blocked 3rd-party requests

☐ Use large request rows ☐ Group by frame

☒ Show overview ☐ Capture screenshots

500000 ms	1000000 ms	1500000 ms	2000000 ms	2500000 ms	3000000 ms	3500000 ms	4000000 ms	4500000 ms
-----------	------------	------------	------------	------------	------------	------------	------------	------------

Name

☐ adv.js
☐ articlePaginationLoad.js
☐ articleSliderLoad.js
☐ ckplayer.js
☐ index.html
☐ load.js
☐ loadHanwebCss.js
☐ page.js?_=1695884779077
☐ sugrec?&prod=pc_his&from=...
☐ unit?parseType=buildstatic&...
☐ unit?webId=8d828e408d904...
☐ video.js
☐ videoLoad.js
☐ webglobal.js?_=1695884779...
☐ webglobal.js?_=1695884779...

×

Headers

Payload

Preview

Response

Initiator

Timing

Cookies

研发出大容量、低成本、长寿命水系液流电池"

```
<ul class="u-list">
  <li>
```

其中跳转链接在a中，标题在正文中，时间在span中，利用PyQuery对其进行解析

```
url = 'https://www.ndrc.gov.cn/xwdt/xwfb/'
reponse = requests.get(url, headers=head)
reponse.encoding = 'utf-8'
pq = PyQuery(reponse.text)
# li.empty 为空行标签, 删除
pq('ul.u-list')('li.empty').remove()
li_info_list = [[url + i('a').attr('href')[1:], i('a').text(),
i('span').text()] for i in pq('ul.u-list')('li').items()]

print(li_info_list)
```

```
'2023/03/16'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202303/t20230303_1350622.html',
'2023年3月3日国内成品油价格不作调整',
'2023/03/03'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202302/t20230228_1350040.html',
'国家发展改革委组织召开推动“十四五”规划102项重大工程实施部际联席会议第二次会议',
'2023/02/28'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202302/t20230227_1349829.html',
'国家发展改革委有关负责同志就推动惠企收费政策落实和整治涉企违规收费情况答记者问',
'2023/02/27'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202302/t20230217_1348864.html',
'2023年2月17日国内成品油价格不作调整',
'2023/02/17'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202302/t20230217_1348858.html',
'首届“中国+中亚五国”产业与投资合作论坛成功举行',
'2023/02/17'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202302/t20230214_1348738.html',
'国家发展改革委负责同志就《新时代洞庭湖生态经济区规划》答记者问',
'2023/02/14']]
```

获取制定时间的li_info_list

```
def get_l_list(url, end_time, page=0):
    # 控制翻页变化的url
    if page != 0:
        crawl_url = url + "_%s" % page + ".html"
    else:
        crawl_url = url + ".html"

    # 获取li_info_list
    reponse = requests.get(crawl_url, headers=head)
    reponse.encoding = "utf-8"
    pq = PyQuery(reponse.text)
    pq("ul.u-list")("li.empty").remove()
    li_info_list = [
        ['https://www.ndrc.gov.cn/xwdt/xwfb' + i("a").attr("href")[1:],
i("a").text(), i("span").text()]
        for i in pq("ul.u-list")("li").items()
    ]

    # 获取当页最早文章发布时间
    publish_time = datetime.strptime(li_info_list[-1][-1], "%Y/%m/%d")

    # 如果时间时间小于截至时间, 停止
```

```

if publish_time < end_time:
    return li_info_list

# 否则翻页, 递归执行
page += 1
li_info_list += get_l_list(url, end_time, page)
return li_info_list

begin_time = datetime.now()
end_time = datetime(2023, 3, 1)
li_list = get_l_list(url, end_time)

```

```

'2023/03/16'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202303/t20230303_1350622.html',
'2023年3月3日国内成品油价格不作调整',
'2023/03/03'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202302/t20230228_1350040.html',
'国家发展改革委组织召开推动“十四五”规划102项重大工程实施部际联席会议第二次会议',
'2023/02/28'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202302/t20230227_1349829.html',
'国家发展改革委有关负责同志就推动惠企收费政策落实和整治涉企违规收费情况答记者问',
'2023/02/27'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202302/t20230217_1348864.html',
'2023年2月17日国内成品油价格不作调整',
'2023/02/17'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202302/t20230217_1348858.html',
'首届“中国+中亚五国”产业与投资合作论坛成功举行',
'2023/02/17'],
['https://www.ndrc.gov.cn/xwdt/xwfb/index/202302/t20230214_1348738.html',
'国家发展改革委负责同志就《新时代洞庭湖生态经济区规划》答记者问',
'2023/02/14']]

```

CSDN @WuRobb

对li标签中每个url进行爬取即可

```

def crawl_li_list(li_list, db_insert_data):
    for li in li_list:
        publish_time = datetime.strptime(li[-1], "%Y/%m/%d")
        if end_time <= publish_time < begin_time:
            a_url = li[0]
            reponse = requests.get(a_url, headers=head)
            reponse.encoding = "utf-8"
            pq = PyQuery(reponse.text)
            db_insert_data.append(
                [
                    # crawl_time
                    datetime.now(),
                    # publish_time
                    publish_time,
                    # 原始网址
                    a_url,
                    # 网站模块
                    pq('div.path')('a:last').text(),
                    # 标题
                    li[1],
                    # 作者或来源
                    pq('div.ly.laiyuantext')('span').text(),
                    # 文章内容
                    pq('div.article_con.article_con_title').text(),
                    # 附件 存储附件地址 media/app03/data/
                    [],

```

```
    ]
    )
db_insert_data = []
crawl_li_list(li_list, db_insert_data)
```

看一下结果

```
[8]: [[datetime.datetime(2023, 9, 28, 17, 56, 23, 477123),
      datetime.datetime(2023, 9, 28, 0, 0),
      'https://www.ndrc.gov.cn/xwdt/xwfb/202309/t20230928_1360986.html',
      '新闻发布',
      '国家发展改革委举行升国旗仪式 庆祝中华人民共和国成立74周年',
      '办公厅',
      '9月28日上午,国家发展改革委举行升国旗仪式,庆祝中华人民共和国成立74周年。委党组书记、主任郑栅洁,在委的党组成员,秘书长、副秘书长出席仪式。\\n“升国旗仪式开始!”上午8时15分,在激昂雄壮的《义勇军进行曲》声中将鲜艳的五星红旗缓缓升起,在场全体人员庄严地向国旗行注目礼,并高唱国歌,表达对伟大祖国的由衷祝福。\\n今年以来,在以习近平同志为核心的党中央坚强领导下,国家发展改革委深入开展学习贯彻习近平新时代中国特色社会主义思想主题教育,不折不扣贯彻落实党中央、国务院决策部署,认真履行宏观管理和经济综合协调部门职责,围绕“政策和研究、规划和计划、投资和项目、改革和创新、综合统筹和综合平衡”五大职能,有力有序开展各项工作,扩内需、提信心、防风险,为推动经济结构持续向优,增长动能持续增强,发展态势持续向好,作出了应有贡献。\\n当前全党全国各族人民迈上了全面建设社会主义现代化国家、全面推进中华民族伟大复兴的新征程,全党同志要更加紧密地团结在以习近平同志为核心的党中央周围,进一步统一思想、坚定信心、振奋精神,在委党组的带领下,以更加饱满的热情、昂扬的斗志、务实的作风,投入到发展改革各项工作中,高质量完成全年经济社会发展目标,为强国建设民族复兴伟业再立新功、再创佳绩!\\n国家发展改革委机关各司局、直属单位和中央纪委国家监委驻国家发展改革委纪检监察组的负责同志和干部职工代表以及驻委武警中队官兵等380余人参加了升国旗仪式。\\n (摄影 高弘杰)',
      []],
      [datetime.datetime(2023, 9, 28, 17, 56, 23, 543104),
      datetime.datetime(2023, 9, 20, 0, 0),
      'https://www.ndrc.gov.cn/xwdt/xwfb/202309/t20230920_1360717.html',
      '新闻发布',
      '2023年9月20日24时起国内成品油价格按机制调整',
      '价格司',
      '根据近期国际市场价变化情况,按照现行成品油价格形成机制,自2023年9月20日24时起,国内汽、柴油价格(标准品,下同)每吨分别提高385元、370元。调整后,各省(区、市)和中心城市汽、柴油最高零售价格见附表。相关价格联动及补贴政策按现行规定执行。\\n中石油、中石化、中海油三大公司及其他原油加工企业要组织好成品油生产和调运,确保市场稳定供应,严格执行国家价格政策。各地相关部门要加大市场监管检查力度,严厉查处不执行国家价格政策的行为,维护正常市场秩序。消费者可通过12315平台举报价格违法行为。\\n附: 各省市区和中心城市汽、柴油最高零售价格',
      []],
```

完整代码

```
import json
import os
import re
from datetime import datetime, timedelta

import numpy as np
import pandas as pd
import requests
from bs4 import BeautifulSoup
from pyquery import PyQuery

head = {
    "User-Agent": np.random.choice(
        [
            "Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2228.0 Safari/537.36",
            "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2227.1 Safari/537.36",
            "Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2227.0 Safari/537.36",
            "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2227.0 Safari/537.36",
            "Mozilla/5.0 (Windows NT 6.3; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2226.0 Safari/537.36",
            "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Trident/5.0; chrome/13.0.782.215)",
```



```

        "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Trident/5.0;
chrome/11.0.696.57)",
        "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Trident/5.0)
chrome/10.0.648.205",
        "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Trident/4.0;
GTB7.4; InfoPath.1; SV1; .NET CLR 2.8.52393; WOW64; en-US)",
        "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.0; Trident/5.0;
chrome/11.0.696.57)",
        "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.0; Trident/4.0;
GTB7.4; InfoPath.3; SV1; .NET CLR 3.1.76908; WOW64; en-US)",
        "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_3)
AppleWebKit/537.75.14 (KHTML, like Gecko) Version/7.0.3 Safari/7046A194A",
        "Mozilla/5.0 (iPad; CPU OS 6_0 like Mac OS X) AppleWebKit/536.26
(KHTML, like Gecko) Version/6.0 Mobile/10A5355d Safari/8536.25",
        "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_8) AppleWebKit/537.13+
(KHTML, like Gecko) Version/5.1.7 Safari/534.57.2",
        "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_3)
AppleWebKit/534.55.3 (KHTML, like Gecko) Version/5.1.3 Safari/534.53.10",
        "Mozilla/5.0 (iPad; CPU OS 5_1 like Mac OS X) AppleWebKit/534.46
(KHTML, like Gecko) Version/5.1 Mobile/9B176 Safari/7534.48.3",
    ]
)
}

```

```

def get_l_list(url, end_time, page=0):
    # 控制翻页变化的url
    if page != 0:
        crawl_url = url + "_%s" % page + ".html"
    else:
        crawl_url = url + ".html"

    # 获取li_info_list
    reponse = requests.get(crawl_url, headers=head)
    reponse.encoding = "utf-8"
    pq = PyQuery(reponse.text)
    pq("ul.u-list")("li.empty").remove()
    li_info_list = [
        [
            "https://www.ndrc.gov.cn/xwdt/xwfb" + i("a").attr("href")[1:],
            i("a").text(),
            i("span").text(),
        ]
        for i in pq("ul.u-list")("li").items()
    ]

    # 获取当页最早文章发布时间
    publish_time = datetime.strptime(li_info_list[-1][-1], "%Y/%m/%d")

    # 如果时间时间小于截至时间, 停止
    if publish_time < end_time:
        return li_info_list

    # 否则翻页, 递归执行
    page += 1
    li_info_list += get_l_list(url, end_time, page)
    return li_info_list

```

```

def crawl_li_list(li_list, db_insert_data):
    for li in li_list:
        publish_time = datetime.strptime(li[-1], "%Y/%m/%d")
        if end_time <= publish_time < begin_time:
            a_url = li[0]
            reponse = requests.get(a_url, headers=head)
            reponse.encoding = "utf-8"
            pq = PyQuery(reponse.text)
            db_insert_data.append(
                [
                    # crawl_time
                    datetime.now(),
                    # publish_time
                    publish_time,
                    # 原始网址
                    a_url,
                    # 网站模块
                    pq("div.path")("a:last").text(),
                    # 标题
                    li[1],
                    # 作者或来源
                    pq("div.ly.laiyuantext")("span").text(),
                    # 文章内容
                    pq("div.article_con.article_con_title").text(),
                    # 附件 存储附件地址 media/app03/data/
                    [],
                ]
            )

url_list = [
    "https://www.ndrc.gov.cn/xwdt/xwfb/index",
    "https://www.ndrc.gov.cn/xwdt/tzgg/index",
    "https://www.ndrc.gov.cn/xwdt/dt/wlddt/index",
]

begin_time = datetime.now()
end_time = datetime(2023, 9, 1)

def spider():
    db_insert_data = []
    for url in url_list:
        li_list = get_l_list(url, end_time)
        crawl_li_list(li_list, db_insert_data)
    return db_insert_data

spider()

```



```
[7]: [[datetime.datetime(2023, 9, 28, 18, 1, 3, 237739),
datetime.datetime(2023, 9, 28, 0, 0),
'https://www.ndrc.gov.cn/xwdt/xwfb/202309/t20230928_1360986.html',
'新闻发布',
'国家发展改革委举行升国旗仪式 庆祝中华人民共和国成立74周年',
'办公厅',
'9月28日上午,国家发展改革委举行升国旗仪式,庆祝中华人民共和国成立74周年。委党组书记、主任郑栅洁,在委的委党组成员,秘书长、副秘书长出席仪式。\\n“升国旗仪式开始!”上午8时15分,在激昂雄壮的《义勇军进行曲》声中将鲜艳的五星红旗缓缓升起,在场全体人员庄严地向国旗行注目礼,并高唱国歌,表达对伟大祖国的由衷祝福。\\n今年以来,在以习近平同志为核心的党中央坚强领导下,国家发展改革委深入开展学习贯彻习近平新时代中国特色社会主义思想主题教育,不折不扣贯彻落实党中央、国务院决策部署,认真履行宏观管理和经济综合协调部门职责,围绕“政策和研究、规划和计划、投资和项目、改革和创新、综合统筹和综合平衡”五大职能,有力有序开展各项工作,扩内需、提信心、防风险,为推动经济结构持续向优,增长动能持续增强,发展态势持续向好,作出了应有贡献。\\n当前全党全国各族人民迈上了全面建设社会主义现代化国家、全面推进中华民族伟大复兴的新征程,全委同志要更加紧密地团结在以习近平同志为核心的党中央周围,进一步统一思想、坚定信心、振奋精神,在委党组的带领下,以更加饱满的热情、昂扬的斗志、务实的作风,投入到发展改革各项工作中,高质量完成全年经济社会发展目标,为强国建设民族复兴伟业再立新功、再创佳绩!\\n国家发展改革委机关各司局、直属单位和中央纪委国家监委驻国家发展改革委纪检监察组的负责同志和干部职工代表以及驻委武警中队官兵等380余人参加了升国旗仪式。(摄影 高弘杰)',
[]],
[datetime.datetime(2023, 9, 28, 18, 1, 3, 306093),
datetime.datetime(2023, 9, 20, 0, 0),
'https://www.ndrc.gov.cn/xwdt/xwfb/202309/t20230920_1360717.html',
'新闻发布',
'2023年9月20日24时起国内成品油价格按机制调整',
'价格司',
'根据近期国际市场油价变化情况,按照现行成品油价格形成机制,自2023年9月20日24时起,国内汽、柴油价格(标准品,下同)每吨分别提高385元、370元。调整后,各省(区、市)和中心城市汽、柴油最高零售价格见附表。相关价格联动及补贴政策按现行规定执行。\\n中石油、中石化、中海油三大公司及其他原油加工企业要组织好成品油生产和调运,确保市场稳定供应,严格执行国家价格政策。各地相关部门要加大市场监督管理力度,严厉查处不执行国家价格政策的行为,维护正常市场秩序。消费者可通过12315平台举报价格违法行为。\\n附:各省区市和中心城市汽、柴油最高零售价格',
[]],
[datetime.datetime(2023, 9, 28, 18, 1, 3, 361639),
```