how to scrape an aspx rendered page using python

Asked 7 years, 1 month ago   Modified 7 years, 1 month ago   Viewed 5k times

▲

1

▼

🔖

🕓

I am scraping an aspx rendered web page Link to Page

the website is .aspx, i chose Selenium , mechanize , urllib , lxml , Beautiful soup , requests . any insights/recommendations on coding the next steps. Also used scrapy.

I have used requests :

```
import requests
from bs4 import BeautifulSoup
request.get(url_to_page)
print request.content
```

it gives

```
<!DOCTYPE html>
<html>
<head><meta charset="utf-8" /><title>
    www.tournamentsoftware.com
</title>
  <style>
    body {
       font: normal 12px/20px Arial, Helvetica, sans-serif;
       color: #505050;
       background: #ccc url(//static.tournamentsoftware.com/images/cw_bg.png)
repeat-x;
    }
    h2 {
       font: bold 16px/16px Arial, Helvetica, sans-serif !important;
       color: #000;
       margin: 4px 0;
    }
    h4 {
       font: bold 13px/13px Arial, Helvetica, sans-serif !important;
       margin: 0 0 -8px 0;
    }
    p {
       font: normal 12px/20px Arial, Helvetica, sans-serif;
       margin: 12px 0;
    }
    p.note {
       font: normal 10px/10px Arial, Helvetica, sans-serif;
       margin: 8px 0 0;
       text-align: center;
       color: #999;
    }
    p.note.error {
       font: bold 13px/20px Arial, Helvetica, sans-serif;
       color: #f00;
    }

    .langtoggle { display:inline; margin-right:6px; }
    .langtoggle.active { display:none; }
    .langmessage { display:none; margin-bottom:20px; }
    .langmessage.active { display:block; }
    input.button {

       margin: 4px 0;
    }
  </style>
</head>
<body>
  <form method="post" action="./default.aspx?
returnurl=%2fsport%2fdraw.aspx%3fid%3dE880C7A5-0A60-4A98-8FF9-
A3B7DD58F3E2%26draw%3d4" id="form1" class="lang1033">
<input type="hidden" name="__VIEWSTATE" id="__VIEWSTATE"
value="p4eGoAC3005ctvGuhkv1w6Nanrs87p7iDcl4Hlk1SNw/cJovTDsJZeq54VdP4JR0injIJb59ok
/>

<input type="hidden" name="__VIEWSTATEGENERATOR" id="__VIEWSTATEGENERATOR"
value="A86F2231" />
    <div id="messagewindow">
      <p class="toggles"><a id="Lang1033" class="lang langtoggle active"
href="#" onclick="switchLang(this)">English</a> </p><div id="divLang1033"
class="langmessage active"><h2>The use of cookies on
www.tournamentsoftware.com</h2><p>We are legally obliged to get your
        elems = document.getElementsByClassName('langmessage');
        for (var i = 0; i < elems.length; i++) {
          elems[i].className = 'langmessage';
        }
        document.getElementById(AThis.id).className = 'langtoggle active';
        document.getElementById('div' + AThis.id).className = 'langmessage
active';
        return false;
      }
      function toggleCookiesHelp(AElmID) {
        document.getElementById(AElmID).style.display = 'block';
```

```
            return false;
        }
        function toggleCookiesHelpByClassName() {
            var elems = document.getElementsByClassName('removecookies');
            for (var i = 0; i < elems.length; i++) {
              elems[i].style.display = 'block';
            }
            elems = document.getElementsByClassName('note');
            for (var i = 0; i < elems.length; i++) {
              elems[i].className = 'note error';
            }
            return false;
        }
        if (storageAvailable()) {
            if (localStorage.getItem('cookiewall')) {
              toggleCookiesHelpByClassName();
            }
            var elems = document.getElementsByClassName('button');
            for (var i = 0; i < elems.length; i++) {
              elems[i].addEventListener('click', function (e) {
                localStorage.setItem('cookiewall', '1');
              });
            }
        }
        function storageAvailable() {
          try {
            var x = '__storage_test__';
            localStorage.setItem(x, x);
            localStorage.removeItem(x);
            return true;
          } catch(e) {
            return false;
          }
        }
      </script>
    </form>
  </body>
</html>
```

Also tried with mechanize, scrapy. All of them are giving this result only. How to scrape those websites. But I am able to see the source code in the browser. Is there any way to scrape those data.

`python` `web-scraping` `python-requests` `mechanize` `scrapy`

Share Improve this question Follow

edited Feb 1, 2018 at 18:47
Brandon Minnick
15.4k   16   81   144

asked Feb 1, 2018 at 16:07
Syed Jafer
306   3   13

What is the issue? Have you tried using any of the library's mentioned above? – Alex P. Feb 1, 2018 at 16:32 ✏️

You need to include the doPostBack arguments in the post (__EVENTTARGET and __EVENTARGUMENT). Mechanize would be a good choice if it were more reliable, I would probably go with requests. – pguardiario Feb 2, 2018 at 2:02

how to add _EVENTTARGET in request ? – Syed Jafer   Feb 2, 2018 at 5:58

2 Answers

Sorted by: Highest score (default) ▾

```
import requests
from bs4 import BeautifulSoup

r_obj = requests.Session()

url = "http://www.tournamentsoftware.com/cookie/default.aspx?
returnurl=%2fdefault.aspx"

fr_soup = r_obj.get(url)


soup = BeautifulSoup(fr_soup.content , "lxml")
#print soup

l = soup.find_all("input",type="hidden")
#print l

data = {
l[0]['name']:l[0]['value'],
l[1]['name']:l[1]['value'],
'btnAccept':'Yes, I accept'}

r_obj.post(url,verify=False,data=data)

url_needed = "http://www.tournamentsoftware.com/sport/draw.aspx?id=E880C7A5-
0A60-4A98-8FF9-A3B7DD58F3E2&draw=4"

final = r_obj.get(url_needed)
```

3

```
#print final.content

soup1 = BeautifulSoup(final.content,"lxml")

detail_tab = soup1.find_all("table")
```

Share   Improve this answer   Follow

answered Feb 3, 2018 at 11:02

Syed Jafer
306   3   13

You need to use a framework which runs the client-side code for you. headless-chrome is one such tool.

-1

Share   Improve this answer   Follow

answered Feb 1, 2018 at 16:11

simbamford
71   1   9

Start asking to get answers

Find the answer to your question by asking.

[Ask question]

Explore related questions

python   web-scraping   python-requests   mechanize   scrapy

See similar questions with these tags.

```
#print final.content

soup1 = BeautifulSoup(final.content,"lxml")

detail_tab = soup1.find_all("table")
```

Share   Improve this answer   Follow

answered Feb 3, 2018 at 11:02

Syed Jafer
306   3   13