

# Web scraping: Automating button click [closed]

Asked 10 years, 2 months ago   Modified 10 years, 2 months ago   Viewed 7k times

▲

6

▼

**Closed.** This question is seeking recommendations for software libraries, tutorials, tools, books, or other off-site resources. It does not meet [Stack Overflow guidelines](#). It is not currently accepting answers.

💡 We don't allow questions seeking recommendations for software libraries, tutorials, tools, books, or other off-site resources. You can edit the question so it can be answered with facts and citations.

Closed 10 years ago.

Improve this question

I am trying to scrape data off a website using [Scrapy](#), a python framework. I can get the data from the website using the spiders but the problem occurs when I try to navigate through the website.

According to [this](#) post [Scrapy](#) does not handle Javascript well.

Also, as stated in the accepted answer, I cannot use mechanize or lxml. It suggests using a combination of [Selenium](#) and [Scrapy](#) .

Function of the button:

I am browsing through offers on a website. The function of the button is to show more offers. SO on clicking it, it calls a javascript function which loads the results.

I also was looking at [CasperJS](#) and [PhantomJS](#) . Will they work?

I just need to automate the clicking of a button. How do I go about this?

javascriptpython selenium web-scraping scrapy

Share   Improve this question   Follow

edited May 23, 2017 at 12:26

Community Bot

11

asked Jan 7, 2015 at 5:56

praxmon

5,1312376124

- Really depends on the button. Can you share the details? – [alecxe](#) Jan 7, 2015 at 5:57
- If you use Selenium, the Javascript will execute in an otherwise normal browser. You can certainly automate simple button clicks with only Selenium IDE or WebDriver. – [BadZen](#) Jan 7, 2015 at 5:59

1 Answer

Sorted by: Highest score (default) ▼

▲

4

▼

First of all, yes – you can use [PhantomJS \\_ghostdriver](#) with python. It is built-in to [python-selenium](#) :

```
pip install selenium
```

Demo:

```
>>> from selenium import webdriver
>>> driver = webdriver.PhantomJS()
>>> driver.get('https://stackoverflow.com/questions/27813251')
>>> driver.title
u'javascript – Web scraping: Automating button click – Stack Overflow'
```

There are also several other threads that provide examples of "scrapy+selenium" spiders:

- [selenium with scrapy for dynamic page](#)
- [Scraping with Scrapy and Selenium](#)
- [seleniumcrawler](#)

Also there is a [scrapy-webdriver module](#) that can probably help with it too.

Using scrapy with selenium would give you a huge overhead and slow things down dramatically even with a headless PhantomJS browser.

There is a huge chance you can mimic that "show more offers" button click by simulating the underlying request going to get the data you need. Use browser developer tools to explore what kind of request is fired and use [scrapy.http.Request](#) for simulation inside the spider.


Share Improve this answer Follow

edited May 23, 2017 at 12:34

Community Bot

11

answered Jan 7, 2015 at 6:04

alecxe

474k1271.1k1.2k

Thank you for this answer. I have started to work on something using our answer as the base. I wanted more clarity on some details here. In the example you have illustrated that I can get the page title using PhantomJS . So, what's the use of scrapy here? I basically want to know the difference between their usage. Could you explain? – [praxmon](#) Jan 13, 2015 at 5:08

2 @PrakharMohanSrivastava the key thing is that scrapy is not a browser and doesn't have a javascript engine built-in. A lot of sites use javascript to construct their pages - this javascript code is executed in the browser - browser follows the `script` links, loads additional js files, executes the code, changes the DOM - does a lot of things. For these sites, it is easier to use a real browser to construct the page, as you would see it in the browser developer tools. Then, you can feed the resulting `.page_source` to scrapy for processing. Hope that makes things a bit more clear. – [alecxe](#) Jan 13, 2015 at 5:13

Start asking to get answers

Find the answer to your question by asking.

Ask question

Explore related questions

- javascript
- python
- selenium
- web-scraping
- scrapy

See similar questions with these tags.

