

python web scraping request error(mod security)

Asked 1 year, 4 months ago Active 11 months ago Viewed 2k times



3



I am new and I try to grap source code of an Web page for tutorial.I got beautifulsoup install,request install. At first I want to grap the source.I am doing this scraping job from "<https://pythonhow.com/example.html>".I am not doing anything illegal and I think this site also established for this purposes.Here's my code:



1



```
import requests
from bs4 import BeautifulSoup
```

```
r=requests.get("http://pythonhow.com/example.html")
c=r.content
c
```

And i got the mod security error:

```
b'<head><title>Not Acceptable!</title></head><body><h1>Not Acceptable!</h1><p>An
appropriate representation of the requested resource could not be found on this
server. This error was generated by Mod_Security.</p></body></html>'
```

Thanks for all who re dealing with me.Respectly

[python](#) [python-requests](#) [mod-security](#)

[Share](#) [Improve this question](#) [Follow](#)

asked May 23 '20 at 6:55



[özgür Sanli](#)

45 4

3 Answers

Active	Oldest	Votes
--------	--------	-------



12



You can easily fix this issue by providing a user agent to the request. By doing so, the website will think that someone is actually visiting the site using a web browser.

Here is the code that you want to use:



```
import requests
from bs4 import BeautifulSoup
```



```
headers = {
    'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10.12; rv:55.0)
    Gecko/20100101 Firefox/55.0',
```

```
}

r = requests.get("http://pythonhow.com/example.html", headers=headers)
c = r.content

print(c)
```

Which gives you the expected output

```
b'<!DOCTYPE html>\n<html>\n<head>\n<style>\n<div>\n  background-
color:black;\n  color:white;\n  margin:20px;\n  padding:20px;\n}
\n</style>\n</head>\n<body>\n<h1 align="center"> Here are three big cities
</h1>\n<div class="cities">\n<h2>London</h2>\n<p>London is the capital of England
and it's been a British settlement since 2000 years ago. </p>\n</div>\n<div
class="cities">\n<h2>Paris</h2>\n<p>Paris is the capital city of France. It was
declared capital since 508.</p>\n</div>\n<div
class="cities">\n<h2>Tokyo</h2>\n<p>Tokyo is the capital of Japan and one of the
most populated cities in the world.</p>\n</div>\n</body>\n</html>'
```

Share Improve this answer Follow

edited May 23 '20 at 7:23

answered May 23 '20 at 7:08



Siddharth Dushantha

994 8 26



Instead of requests, try using url module

0

```
from urllib.request import urlopen
```



```
page = urlopen("http://pythonhow.com/example.html")
page.read()
```



Share Improve this answer Follow

answered May 23 '20 at 7:08



sushanth

7,544 3 14 24



We just need to pass an argument called headers...

0

```
headers = {
    "User-Agent": "Mozilla/5.0 (X11; Linux x86_64; rv:60.0) Gecko/20100101
Firefox/60.0"
}
page = requests.get(url, headers=headers)
soup = BeautifulSoup(page.content, 'html.parser')
```



Share Improve this answer Follow

answered Nov 1 '20 at 5:30



AgentNirmites

131 9

