

Forum:[Hack Forums](#)

Thread: [Simple Proxy Scraper.](#)

Null - April 25, 2017, 7:48 pm

Hey,
I have been messing around with selenium and i made a really simple proxy scraper (credits to [Piero-Bird](#)) that will scrape off gatherproxy.com
you may need to change the chrome driver's path. (get it from here:
<https://chromedriver.storage.googleapis.com/index.html?path=2.29/>)

Here is the source code:

The program basically scrapes off gather proxies and export the proxies in a textfile in a IP:Port format (in the same directory of the script)

[here is the original tutorial by Piero-Bird](#)

LemonTea - April 25, 2017, 7:51 pm

Nice. Very simple and cool

Try using Phantomjs to make it a background process too!

SIM-1866423 - April 26, 2017, 12:28 am

Try this,

PHP Code

```
<?php
import urllib
import re

data = urllib.urlopen("http://gatherproxy.com/proxylist/anonymity/?t=Elite").read()
proxy = re.findall(r"\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}", data)
port = [int(x, 16) for x in re.findall('PROXY_PORT":"(.+?)"', data)]
warehouse = ['{0}:{1}'.format(_[0],_[1]) for _ in zip(proxy, port)]

print(warehouse)
```

PHP Code

```
<?php
['192.241.129.174:3128', '112.82.190.235:8998',
'86.109.100.80:8080', '206.251.227.56:80',
'103.52.252.17:8080', '219.127.253.43:80',
'115.254.104.206:8080', '162.243.76.169:3128',
'159.203.189.158:3128', '104.236.174.88:3128',
'162.243.77.188:3128', '104.131.147.164:3128',
```

```
'54.174.144.206:80', '52.45.121.244:80',  
'185.2.101.31:3128', '186.211.102.57:80',  
'104.224.37.121:3128', '138.197.194.143:3128',  
'119.252.163.66:8080', '93.51.247.104:80',  
'138.197.194.7:3128', '47.52.29.37:3128',  
'104.236.176.142:3128', '36.85.90.8:8080', '23.248.163.235:80']  
[Finished in 1.9s]
```

Imorra - April 26, 2017, 5:56 am

Selenium always works but it's really slow. Use the requests library and XPATHs, it's a lot faster.

Sapience - April 26, 2017, 6:29 am

(Apr 26, 2017 12:28 AM)Mnemosyne Wrote: [\[CITE\]](#) Try this,

```
PHP Code  
<?php  
import urllib  
import re  
  
data = urllib.urlopen("http://gatherproxy.com/proxylist/anonymity/?t=Elite").read()  
proxy = re.findall(r"\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}", data)  
port = [int(x, 16) for x in re.findall('PROXY_PORT": "(.+?)"', data)]  
warehouse = ['{0}:{1}'.format(_[0],_[1]) for _ in zip(proxy, port)]  
  
print(warehouse)
```

```
PHP Code  
<?php  
['192.241.129.174:3128', '112.82.190.235:8998',  
'86.109.100.80:8080', '206.251.227.56:80',  
'103.52.252.17:8080', '219.127.253.43:80',  
'115.254.104.206:8080', '162.243.76.169:3128',  
'159.203.189.158:3128', '104.236.174.88:3128',  
'162.243.77.188:3128', '104.131.147.164:3128',  
'54.174.144.206:80', '52.45.121.244:80',  
'185.2.101.31:3128', '186.211.102.57:80',  
'104.224.37.121:3128', '138.197.194.143:3128',  
'119.252.163.66:8080', '93.51.247.104:80',  
'138.197.194.7:3128', '47.52.29.37:3128',  
'104.236.176.142:3128', '36.85.90.8:8080', '23.248.163.235:80']  
[Finished in 1.9s]
```

Really good simple proxy scraper, nice work!

Personally I prefer having a file output rather than just putting the result in a terminal, so I took your code and "improved" it a bit:

It's still compact, while having a (imo) really useful feature.

?ictor? - April 26, 2017, 4:36 pm

I love Python and great simple code :")

