

Scrapy, privoxy and Tor: SocketError: [Errno 61] Connection refused

Asked 7 years, 8 months ago Modified 3 years, 5 months ago Viewed 3k times

I am using Scrapy with Privoxy and Tor. Here is my previous question [Scrapy with Privoxy and Tor: how to renew IP](#), and here is the spider:

5

```
from scrapy.contrib.spiders import CrawlSpider
from scrapy.selector import Selector
from scrapy.http import Request

class YourCrawler(CrawlSpider):
    name = "****"
    start_urls = [
        'https://****.com/listviews/titles.php',
    ]
    allowed_domains = ["****.com"]

    def parse(self, response):
        # go to the urls in the list
        s = Selector(response)
        page_list_urls = s.xpath('///*[
[id="tab7"]/article/header/h2/a/@href').extract()
        for url in page_list_urls:
            yield Request(response.urljoin(url),
callback=self.parse_following_urls, dont_filter=True)

        # Return back and go to bext page in div#paginat ul li.next
a::attr(href) and begin again
        next_page = response.css('ul.pagin li.presente ~ li
a::attr(href)').extract_first()
        if next_page is not None:
            next_page = response.urljoin(next_page)
            yield Request(next_page, callback=self.parse)

        # For the urls in the list, go inside, and in div#main, take the div.ficha
> div.caracteristicas > ul > li
        def parse_following_urls(self, response):
            #Parsing rules go here
            for each_book in response.css('main#main'):
                yield {
                    'editor': each_book.css('header.datos1 > ul > li > h5 >
a::text').extract(),
                }
```

In settings.py I have an user agent rotation and privoxy:

```
DOWNLOADER_MIDDLEWARES = {
    #user agent
    'scrapy.contrib.downloadermiddleware.useragent.UserAgentMiddleware' :
None,
    '****.comm.rotate_useragent.RotateUserAgentMiddleware' :400,
    #privoxy
    'scrapy.downloadermiddlewares.httpproxy.HttpProxyMiddleware': 110,
    '****.middlewares.ProxyMiddleware': 100
}
```

In middlewares.py I added:

```
from stem import Signal
from stem.control import Controller

def _set_new_ip():
    with Controller.from_port(port=9051) as controller:
        controller.authenticate(password='tor_password')
        controller.signal(Signal.NEWNYM)

class ProxyMiddleware(object):
    def process_request(self, request, spider):
        _set_new_ip()
        request.meta['proxy'] = 'http://127.0.0.1:8118'
        spider.log('Proxy : %s' % request.meta['proxy'])
```

If I take out the def _set_new_ip(): method of the class in middlewares.py (and the call to it in class ProxyMiddleware(object): the spider works. But I want the spider to call for a new IP each time, and that's why I added it. The problem is that each time I try to run the spider it returns an error SocketError: [Errno 61] Connection refused , with this traceback:

```
Traceback (most recent call last):
  File "/usr/local/lib/python2.7/site-packages/twisted/internet/defer.py", line
1386, in _inlineCallbacks
    result = g.send(result)
  File "/usr/local/lib/python2.7/site-
packages/scrapy/core/downloader/middleware.py", line 37, in process_request
    response = yield method(request=request, spider=spider)
  File "/Users/nikita/scrapy/****/****/middlewares.py", line 71, in
process_request
```

```
_set_new_ip()
File "/Users/nikita/scrapy/***/***/middlewares.py", line 65, in _set_new_ip
    with Controller.from_port(port=9051) as controller:
File "/usr/local/lib/python2.7/site-packages/stem/control.py", line 998, in
from_port
    control_port = stem.socket.ControlPort(address, port)
File "/usr/local/lib/python2.7/site-packages/stem/socket.py", line 372, in
__init__
    self.connect()
File "/usr/local/lib/python2.7/site-packages/stem/socket.py", line 243, in
connect
    self._socket = self._make_socket()
File "/usr/local/lib/python2.7/site-packages/stem/socket.py", line 401, in
_make_socket
    raise stem.SocketError(exc)
SocketError: [Errno 61] Connection refused
2017-07-11 15:50:28 [scrapy.core.engine] INFO: Closing spider (finished)
```

Maybe the problem is in the port used in `with Controller.from_port(port=9051) as controller:` , but I am not sure. If anybody has an idea that would be fantastic...

EDIT---

Ok, if I go to the browser and go to <http://127.0.0.1:8118/>, it sais:

```
503
This is Privoxy 3.0.26 on localhost (127.0.0.1), port 8118, enabled
Forwarding failure
Privoxy was unable to socks5-forward your request http://127.0.0.1:8118/
through localhost: SOCKS5 request failed

Just try again to see if this is a temporary problem, or check your forwarding
settings and make sure that all forwarding servers are working correctly and
listening where they are supposed to be listening.
```

So maybe it is related to the configuration of SOCKS5... Anyone knows?

python web-scraping scrapy tor

Share Improve this question Follow

edited Jul 11, 2017 at 22:36

asked Jul 11, 2017 at 14:04
user7499416

Look [here](#) about how to connect to Tor using `stem` . – [Tomáš Linhart](#) Jul 11, 2017 at 14:16

Ok, in this web they talk about the `authenticate()` function. In the example they give first they make a `control_socket = stem.socket.ControlPort(port = 9051)` , and after that then `stem.connection.authenticate(control_socket)` . Should I put both of them in the `ProxyMiddleware` class? – user7499416 Jul 11, 2017 at 15:24

Ok, I understand that I have to call the `connect()` function somewhere, but, where? I tried some options but none were succesful... – user7499416 Jul 11, 2017 at 20:49

I have something, update the question. – user7499416 Jul 11, 2017 at 22:34

Are you sure you have Tor running and the Privoxy setup with Tor is correct and working? – [Tomáš Linhart](#) Jul 12, 2017 at 6:07

2 Answers

Sorted by: Highest score (default) ▾

▲

2

▼

🔖

🕒


My guess is either:

- Tor is not running.** To make sure if Tor is running, run `ps` (e.g., `ps -ax | grep tor`) and `netstat` (e.g., for mac: `netstat -an | grep 'your tor portnumber'` . For linux, replace `-an` with `-tulnp`) on terminal to see if Tor is really running.
- You didn't set up the forwarding setting corectly.** Based on the 503 error message, it looks like you didn't set up the forwarding rule correctly (if Tor is running). In the config file of Privoxy, make sure `forward-socks5t / 127.0.0.1:9050` . is uncommented.

Share Improve this answer Follow

edited Sep 18, 2017 at 4:53

answered Sep 18, 2017 at 4:03

 [btaek](#)

572 5 14

▲

0

▼

```
class ProxyMiddleware(object):
    def process_request(self, request, spider):
        def _set_new_ip():
            with Controller.from_port(port=9051) as controller:
                controller.authenticate(password='PASSWORDHERE')
                controller.signal(Signal.NEWNYM)
```




```
request.meta['proxy'] = 'http://127.0.0.1:8118'
spider.log('Proxy : %s' % request.meta['proxy'])
```




Share Improve this answer Follow

edited Oct 1, 2021 at 12:14

 **sseLtaH**

11.3k51634

answered Sep 30, 2021 at 12:03

 **user12354862**

1

As it's currently written, your answer is unclear. Please [edit](#) to add additional details that will help others understand how this addresses the question asked. You can find more information on how to write good answers [in the help center](#). – [Community Bot](#) Sep 30, 2021 at 12:20

