

# python Scrapy framework adding to my code proxy

Asked 2 years ago    Modified 2 years ago    Viewed 85 times



1



I am trying new feature for myself as adding proxy port to my python scraper code.

I took free proxy from this [site](#), and looked for an answer from [SO](#). With help of user @dskrypa I changed in my code meta= {'proxy': '103.42.162.50:8080'}

Now it gives an error which continues all along if I do not stop the code run.

```
File "C:\Users\Admin\AppData\Local\Programs\Python\Python310\lib\site-
packages\scrapy\core\downloader\handlers\http11.py", line 279, in _get_agent
    proxyScheme, proxyNetloc, proxyHost, proxyPort, proxyParams = _parse(proxy)
File "C:\Users\Admin\AppData\Local\Programs\Python\Python310\lib\site-
packages\scrapy\core\downloader\webclient.py", line 39, in _parse
    return _parsed_url_args(parsed)
File "C:\Users\Admin\AppData\Local\Programs\Python\Python310\lib\site-
packages\scrapy\core\downloader\webclient.py", line 20, in _parsed_url_args
    host = to_bytes(parsed.hostname, encoding="ascii")
File "C:\Users\Admin\AppData\Local\Programs\Python\Python310\lib\site-
packages\scrapy\utils\python.py", line 108, in to_bytes
    raise TypeError('to_bytes must receive a str or bytes '
TypeError: to_bytes must receive a str or bytes object, got NoneType
2023-03-12 02:47:32 [scrapy.core.scrapy] ERROR: Error downloading <GET
https://dvlaregistrations.dvla.gov.uk/search/results.html?
search=N11CKY&action=index&pricefrom=0&priceto=&prefixmatches=&currentmatches=&lr
```

Here is my code;

```
import scrapy
from scrapy.crawler import CrawlerProcess
import pandas as pd
import scrapy_xlsx

itemList=[]
class plateScraper(scrapy.Spider):
    name = 'scrapePlate'
    allowed_domains = ['dvlaregistrations.dvla.gov.uk']
    FEED_EXPORTERS = {'xlsx': 'scrapy_xlsx.XlsxItemExporter'}
    custom_settings = {'FEED_EXPORTERS' :FEED_EXPORTERS, 'FEED_FORMAT':
'xlsx', 'FEED_URI': 'output_r00.xlsx', 'LOG_LEVEL': 'INFO', 'DOWNLOAD_DELAY': 0}
    DOWNLOADER_MIDDLEWARES = {
        'scrapy.downloadermiddlewares.httpproxy.HttpProxyMiddleware': 1
    }

    def start_requests(self):
        df=pd.read_excel('data.xlsx')
        columnA_values=df['PLATE']
        for row in columnA_values:
            global plate_num_xlsx
            plate_num_xlsx=row
            base_url
            =f"https://dvlaregistrations.dvla.gov.uk/search/results.html?search=
{plate_num_xlsx}&action=index&pricefrom=0&priceto=&prefixmatches=&currentmatches=
url=base_url
            yield scrapy.Request(url,callback=self.parse, cb_kwargs=
{'plate_num_xlsx': plate_num_xlsx},meta={'proxy': '103.42.162.50:8080'})

    def parse(self, response, plate_num_xlsx=None):
        plate =
        response.xpath('//div[@class="resultsstrip"]/a/text()).extract_first()
        price =
        response.xpath('//div[@class="resultsstrip"]/p/text()).extract_first()

        try:
            a = plate.replace(" ", "").strip()
            if plate_num_xlsx == plate.replace(" ", "").strip():
                item = {"plate": plate_num_xlsx, "price": price.strip()}
                itemList.append(item)
                print(item)
                yield item
            else:
                item = {"plate": plate_num_xlsx, "price": "-"}
                itemList.append(item)
                print(item)
                yield item
        except:
            item = {"plate": plate_num_xlsx, "price": "-"}
            itemList.append(item)
            print(item)
            yield item

process = CrawlerProcess()
process.crawl(plateScraper)
process.start()

import winsound
winsound.Beep(555,333)
```

python

proxy

scrapy

Share

Improve this question

Follow

edited Mar 11, 2023 at 23:02

asked Mar 11, 2023 at 22:30

e

xlmaster

731

2

16

36

1

I'm not familiar with scrapy, but the `meta='103.42.162.50:8080'` argument to `scrapy.Request(...)` is what is causing that error. It expects a different data type for that parameter. From the linked question, it looks like it should be `meta={'proxy': '103.42.162.50:8080'}` – dskrypa Mar 11, 2023 at 22:44

now it gives totally different error. and traceback continues long. Thanks for your assisance by the way) – xlmaster Mar 11, 2023 at 22:59

1 Answer

Sorted by:

Highest score (default)

you should include the protocol in the proxy url:

2

`meta={"proxy": "http://103.42.162.50:8080"}`

Share

Improve this answer

Follow



answered Mar 11, 2023 at 23:07

zaki98

1,106

8

14

Wow. It enormously slowed down my speed, I was guessing it will improve though – xlmaster Mar 11, 2023 at 23:22

1

yeah adding a proxy usually lowers speed because your request has to go to the proxy first before going to it's destination. It's also because these public proxies are low quality so if you want something better consider paid proxy services – zaki98 Mar 11, 2023 at 23:25

1

I would like to pay, but I am not blocked, I needed proxies for better speed. Now, without proxy it runs better, much more accelerated. – xlmaster Mar 11, 2023 at 23:32

Start asking to get answers

Find the answer to your question by asking.

Ask question

Explore related questions

python

proxy

scrapy

See similar questions with these tags.