Scrapy-Splash running Tor Proxy

function main(splash)

Asked 8 years, 1 month ago Modified 7 years, 11 months ago Viewed 991 times



I try to build proxy settings of splash. Here I assign Tor or Polipo port address to set_proxy, but it doesn't work. I get 504 error:







```
local host = "localhost"
        local port = 8123
        --local type = "SOCKS5"
        splash:on_request(function (request)
           request:set_proxy{host, port}
    end)
    splash:go(splash.args.url)
    splash:wait(0.5)
    local image = assert(splash:png{render_all=true})
    return {png=image}
end
url = 'https://www.torproject.org/'
```

In polipo.config(9150 is tor port):

```
socksParentProxy = localhost:9150
diskCacheRoot=""
#socksProxyType = socks5
```

In settings.py:

```
HTTP_PROXY = 'http://127.0.0.1:8123'
DOWNLOADER_MIDDLEWARES = {
    'thefork.middlewares.RandomUserAgentMiddleware': 400,
    'thefork.middlewares.ProxyMiddleware': 410,
```

How can I fix this, or is there an easier way to use splash with Tor?

```
scrapy scrapy-splash
```

Share Improve this question Follow



Sorted by:

茶

asked Feb 16, 2017 at 23:03



辮

※

Experiencing the same issue trying to get splash working with other proxy servers. I think it is a problem with scrapy splash and not splash itself. - eusid Mar 29, 2017 at 16:31

1 Answer



The problem I believe is in the middleware you have installed. Try setting the proxy manually on every request. This is how I solved my issue. It tries to proxy every request and when you send a request to splash it attempts to go to your localhost through the proxy and fails.



Share Improve this answer Follow



*

answered Mar 30, 2017 at 3:03



Highest score (default)



Can you provide more details on how you implemented this solution? – Kurt Peek Jul 13, 2017 at 11:12

The solution is to not use Splash. Splash in my opinion is very specific use cases. Nobody wants to write a LUA script to do some basic logic in their spider. I only use Splash when Javascript is necessary to be rendered and can't be gotten another way. Also, only use Splash when you dont need to interact with web elements but need the reduced footprint of Splash. Other than that use Scrapy and Selenium. - eusid Nov 21, 2018 at 9:29

Start asking to get answers

Find the answer to your question by asking.

Explore related questions

scrapy scrapy-splash

See similar questions with these tags.

\$

Ask question



辮

袾