

Forum:[Hack Forums](#)

Thread: [Simulating http request with Python's requests module.](#)

Daredevil - January 5, 2019, 12:38 pm

Hello. I just recorded HTTP traffic of some website and want to simulate the same process to get datas. When i simulate the same process the website returns 'f*ckoff' as response. Can anyone help me? I will send the details through pm.

Ghxst - January 5, 2019, 4:26 pm

Are you sending valid cookies and headers?

Capture the requests made with your program so you can compare.

Daredevil - January 6, 2019, 2:33 am

(Jan 5, 2019 05:26 PM)Ghxst Wrote: [\[CITE\]](#) Are you sending valid cookies and headers?

Capture the requests made with your program so you can compare.

Yes i am sending valid cookies and headers. I captured the request with Charles, when i repeat the request i am getting same result which is 'f*ck off'.

FriendZone - January 6, 2019, 3:08 am

(Jan 6, 2019 03:33 AM)ZyngaChips Wrote: [\[CITE\]](#) Yes i am sending valid cookies and headers. I captured the request with Charles, when i repeat the request i am getting same result which is 'f*ck off'.

I personally use Beautiful soup which is really good for your purpose, Session keeps any info and you don't need much work for cookies or headers :

```
Code ua = 'Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/41.0.2228.0 Safari/537.36'
s = requests.Session()
s.headers = {'User-Agent': ua}
```

Here a full example for tumblr login+scraping

```
Code #####  HEXECUTOR      #####
##### Hackforums.net #####
#####  Tumblr      #####
#####  jpg,png,gif  #####
#####  DOWNLOADER  #####
```

```
import os, requests, time, re
from urllib.request import urlopen
from bs4 import BeautifulSoup as bs
```

```

def HexeTumblerDownloader():
    username = input("Email/Username: ")
    password = input("Password: ")

    ua = 'Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/41.0.2228.0 Safari/537.36'
    s = requests.Session()
    s.headers = {'User-Agent': ua}

    # form_key
    print('Start...')
    r = s.get('https://www.tumblr.com/login')
    formKey = r.text.split('<input type="hidden" name="form_key" value="')[1]
    formKey = formKey.split('"')[0]

    data = {'user[email]': username, 'user[password]': password, 'user[tos]': 1,
            'context': 'login', 'version': 'STANDARD', 'follow': "",
            'http_referer': 'https://www.tumblr.com/login', 'form_key': formKey}

    # login
    print('Logging...')
    r = s.post('https://www.tumblr.com/login', data=data)

    if 'logged_in' not in s.cookies or s.cookies['logged_in'] is not '1':
        print('Error logging in.')
        return
    else:
        print('Successfully logged in.')

    data = {}
    email = ""
    password = ""
    i=0

    #urls input cycle
    while True:
        page=input("Please paste the URL or Ctrl+z to stop (IT MUST END WITH .com NOT WITH
.com/): ")
        maxpage=input("Scanning max pages: ")
        directory=input("Please write the directory Path e.g. C:\\Users\\NAMEUSER\\Desktop\\ IT
MUST END WITH \\ to save inside the last directory specified (if empty every pic will be saved into
the python folder):")
        pagenumb=1
        while pagenumb <= int(maxpage):
            #the first page doesn't require page number specification (it can cause no contents)
            if pagenumb!=1:
                r = s.get(page+'/page/'+str(pagenumb))
            else:
                r = s.get(page)
        #save jpg png gif
        linkregex = re.compile('[a-z]+:Vv66(?:[^\V]+V)*[^\V]+\.(?:jpg|png|gif)')

```

```

matched = linkregex.findall(r.text)
#sometimes it drops https after the first page so let's use http if the output is empty
if matched == [] :
    r = s.get(page.replace('https','http')+'/page/'+str(pagenumb))
    matched = linkregex.findall(r.text)

for img in matched:
    print('----- Page '+str(pagenumb)+' Pic#'+str(i)+' -----')
    if img.endswith('.jpg'):
        extension='.jpg'
    elif img.endswith('.png'):
        extension='.png'
    else:
        extension='.gif'
    f = urlopen(img)
    with open(directory+'img'+str(i)+extension, 'wb') as w:
        w.write( f.read() )
    i=i+1
    print(img)
    pagenumb=pagenumb+1
return

```

HexeTumblerDownloader()

Daredevil - January 6, 2019, 3:21 am

(Jan 6, 2019 04:08 AM)Hexecutor Wrote: [\[CITE\]](#) I personally use Beautiful soup which is really good for your purpose, Session keeps any info and you don't need much work for cookies or headers :

```

Code ua = 'Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/41.0.2228.0 Safari/537.36'
s = requests.Session()
s.headers = {'User-Agent': ua}

```

Here a full example for tumblr login+scraping

```

Code #####  HEXECUTOR  #####
##### Hackforums.net #####
#####  Tumblr  #####
#####  jpg,png,gif  #####
#####  DOWNLOADER  #####

```

```

import os, requests, time, re
from urllib.request import urlopen
from bs4 import BeautifulSoup as bs

```

```

def HexeTumblerDownloader():
    username = input("Email/Username: ")

```

```

password = input("Password: ")

ua = 'Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/41.0.2228.0 Safari/537.36'
s = requests.Session()
s.headers = {'User-Agent': ua}

# form_key
print('Start...')
r = s.get('https://www.tumblr.com/login')
formKey = r.text.split('<input type="hidden" name="form_key" value="')[1]
formKey = formKey.split('"')[0]

data = {'user[email]': username, 'user[password]': password, 'user[tos]': 1,
        'context': 'login', 'version': 'STANDARD', 'follow': "",
        'http_referer': 'https://www.tumblr.com/login', 'form_key': formKey}

# login
print('Logging...')
r = s.post('https://www.tumblr.com/login', data=data)

if 'logged_in' not in s.cookies or s.cookies['logged_in'] is not '1':
    print('Error logging in.')
    return
else:
    print('Successfully logged in.')

data = {}
email = ""
password = ""
i=0

#urls input cycle
while True:
    page=input("Please paste the URL or Ctrl+z to stop (IT MUST END WITH .com NOT WITH
.com/): ")
    maxpage=input("Scanning max pages: ")
    directory=input("Please write the directory Path e.g. C:\\Users\\NAMEUSER\\Desktop\\ IT
MUST END WITH \\ to save inside the last directory specified (if empty every pic will be saved into
the python folder):")
    pagenumb=1
    while pagenumb <= int(maxpage):
        #the first page doesn't require page number specification (it can cause no contents)
        if pagenumb!=1:
            r = s.get(page+'/page/'+str(pagenumb))
        else:
            r = s.get(page)
#save jpg png gif
linkregex = re.compile('([a-z]+:V66(?:[^\V]+V)*[^\V]+\.(?:jpg|png|gif))')
matched = linkregex.findall(r.text)
#sometimes it drops https after the first page so let's use http if the output is empty

```

```

if matched == [] :
    r = s.get(page.replace('https','http')+'/page/'+str(pagenumb))
    matched = linkregex.findall(r.text)

for img in matched:
    print('----- Page '+str(pagenumb)+' Pic#'+str(i)+' -----')
    if img.endswith('jpg'):
        extension='.jpg'
    elif img.endswith('png'):
        extension='.png'
    else:
        extension='.gif'
    f = urlopen(img)
    with open(directory+'img'+str(i)+extension, 'wb') as w:
        w.write( f.read() )
    i=i+1
    print(img)
    pagenumb=pagenumb+1
return

```

HexeTumblerDownloader()

Thank you very much. Tried using session but still the same result.

Infamia - January 6, 2019, 7:46 am

some websites have anti-bot protections, if you're sure it's the exact same request then you're probably retarded and don't know to change certain session cookies or session data; otherwise, if there is no session data, then it's probably a protection mechanism of the website itself.

Daredevil - January 6, 2019, 8:17 am

(Jan 6, 2019 08:46 AM)Infamia Wrote: [\[CITE\]](#) some websites have anti-bot protections, if you're sure it's the exact same request then you're probably retarded and don't know to change certain session cookies or session data; otherwise, if there is no session data, then it's probably a protection mechanism of the website itself.

I think they have a protection mechanism. I am sending exact same request. Will send you the details please give it a look.

SIM-1866423 - January 7, 2019, 7:12 am

I can try to help if you can send me details.

Daredevil - January 7, 2019, 1:23 pm

I used selenium and i just realized if i set the captcha result to g-recaptcha-response element the website is working as i wanted. But i want to get the result directly from the API endpoint using POST request.

```

data = {
    'action': 'imh_hat',
    'captcha': 'XXXX',
    'username': 'test'
}

```

Can you guys please help me? I can't figure what is the problem.

Ghxst - January 9, 2019, 9:24 am

When you're simply repeating the requests captured in Charles and getting the same return, it's mostly likely not a python issue.

Servers often include parameters in payloads, headers or cookies that are meant to stop repeat-attacks. This can be a nonce for example, or a csrf token.

One way to get an idea of the variables and constants you're working with is to capture a specific request legitimately multiple times and see what variables differ from each of the requests.

For example the user-agent is not likely to differ between the same 3 legitimately made requests, however there's a good chance there's a token in another header that might.

Can you share one of the requests?