

Weekend Codes

Codes and random stuff I am up to share with you

Emulating a Browser in Python with mechanize

Posted by **Rogério Carvalho Schneider** 16 Aug 2009

It is always useful to know how to quickly instantiate a browser in the command line or inside your python scripts.

Every time I need to automate any task regarding web systems I do use this recipe to emulate a browser in python:

```
import mechanize
import cookielib

# Browser
br = mechanize.Browser()

# Cookie Jar
cj = cookielib.LWPCookieJar()
br.set_cookiejar(cj)

# Browser options
br.set_handle_equiv(True)
br.set_handle_gzip(True)
br.set_handle_redirect(True)
br.set_handle_referer(True)
br.set_handle_robots(False)

# Follows refresh 0 but not hangs on refresh > 0
br.set_handle_refresh(mechanize._http.HTTPRefreshProcessor(), max_time=1)

# Want debugging messages?
#br.set_debug_http(True)
#br.set_debug_redirects(True)
#br.set_debug_responses(True)

# User-Agent (this is cheating, ok?)
br.addheaders = [('User-agent', 'Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.9.0.1) Gecko/2008071615
Fedora/3.0.1-1.fc9 Firefox/3.0.1')]
```

Now you have this *br* object, this is your browser instance. With this its possible to open a page, to inspect or to interact with:

```
# Open some site, let's pick a random one, the first that pops in mind:
r = br.open('http://google.com')
html = r.read()

# Show the source
print html
# or
print br.response().read()

# Show the html title
print br.title()

# Show the response headers
print r.info()
# or
print br.response().info()

# Show the available forms
for f in br.forms():
    print f

# Select the first (index zero) form
br.select_form(nr=0)

# Let's search
br.form['q']='weekend codes'
br.submit()
print br.response().read()

# Looking at some results in link format
for l in br.links(url_regex='stockrt'):
    print l
```

If you are about to access a password protected site (http basic auth):

```
# If the protected site didn't receive the authentication data you would
```

```
# end up with a 410 error in your face
br.add_password('http://safe-site.domain', 'username', 'password')
br.open('http://safe-site.domain')
```

Thanks to the Cookie Jar we've added before, you do not have to bother about session handling for authenticated sites, as in when you are accessing a service that requires a POST (form submit) of user and password. Usually they ask your browser to store a session cookie and expects your browser to contain that same cookie when re-accessing the page. All this, storing and re-sending the session cookies, is done by the Cookie Jar, neat!

You can also manage with browsing history:

```
# Testing presence of link (if the link is not found you would have to
# handle a LinkNotFoundError exception)
br.find_link(text='Weekend codes')

# Actually clicking the link
req = br.click_link(text='Weekend codes')
br.open(req)
print br.response().read()
print br.geturl()

# Back
br.back()
print br.response().read()
print br.geturl()
```

Downloading a file:

```
# Download
f = br.retrieve('http://www.google.com.br/intl/pt-BR_br/images/logo.gif')[0]
print f
fh = open(f)
```

Setting a proxy for your http navigation:

```
# Proxy and user/password
br.set_proxies({"http": "joe:password@myproxy.example.com:3128"})

# Proxy
br.set_proxies({"http": "myproxy.example.com:3128"})
# Proxy password
br.add_proxy_password("joe", "password")
```

But, if you just want to quickly open a webpage, without the fancy features above, just issue that:

```
# Simple open?
import urllib2
print urllib2.urlopen('http://stockrt.github.com').read()

# With password?
import urllib
opener = urllib.FancyURLopener()
print opener.open('http://user:password@stockrt.github.com').read()
```

See more in [Python mechanize site](#) , [mechanize docs](#) and [ClientForm docs](#).

Also, I have made this post to elucidate how to [handle html forms and sessions with python mechanize and BeautifulSoup](#)

Tags: python en-US
Meta: permalink atom




ShareThis ★ Favorites 51 Comments and 0 Reactions

- Previous: [Listing GitHub projects in your blog](#)
- Next: [Sunset in Porto Alegre](#)

Related Posts

- [Django Admin on Google App Engine](#)
- [Traduzindo com pytranslate diretamente do Google](#)
- [Navegando com segurança e sem tarifação em cafés e hotéis](#)
- [handling html forms with python mechanize and BeautifulSoup](#)
- [git branch in your shell prompt](#)

Like  and 6 others liked this.



Add New Comment

[Login](#)

Type your comment here.

Showing 51 comments

Sort by popular now

**gmon**

I used your tutorial to close my Facebook account permanently. Thanks for the nice information.

2 years ago 10 Likes

[Like](#) [Reply](#)

**Peter**

How about javascript handling? If cannot, your solution is limited to a small portion of websites.

1 year ago 2 Likes

[Like](#) [Reply](#)

**jestinjoy**

This is awesome

1 year ago 1 Like

[Like](#) [Reply](#)

**Dany**

Very good info, simple and clear, it helped me to resolve a proxy issue I was having. Thanks!

1 month ago

[Like](#) [Reply](#)

**David**

Hi, how to get the result of google and the number of result

4 months ago

[Like](#) [Reply](#)

**Cromulent**

What I have problem with is using Mechanize to loop thru a set of webpages. Doesn't matter what the URLs are. The second time thru the loop my "mechanize.urlopen(url)" call always gets a 400 error.

7 months ago

[Like](#) [Reply](#)

**Gurman**

you're my hero

1 year ago

[Like](#) [Reply](#)

**Regis Github Com**

thx for this code. It just helps me a lot to make my first step in python.
Regards,

1 year ago

[Like](#) [Reply](#)

**Jatin**

Hello



Nice tutorial, i am using mechanize. I want to create a session using mechanize in one file and then use that session for my other python files. I read about saving a session at <http://wwwsearch.sourceforge.n...> but could not figure out what exactly to do to achieve the same. If you could please guide me how to go about this.

1 year ago

[Like](#) [Reply](#)



coach outlet online

We keep talking about spring trends in handbags, clearly there are many. We have brought you stories on [coach outlet online](#) and [Coach Outlet](#), and today I will tell you about another trend; floral patterns. Typically we would cover this Coach bag on [Coach wallets](#), but I actually wanted to feature this on [PurseBlog](#).

1 year ago

[Like](#) [Reply](#)



Odixaltec

i get unknown node name in url when i try and save the form data, after submit and response.read to an html for parsing!

1 year ago

[Like](#) [Reply](#)



books2love

Thanks so much for showing how mechanize should actually be used. Saved me a lot of headaches trying to figure out a great but poorly documented package.

1 year ago

[Like](#) [Reply](#)



SG

Thanks! One question: where you say "this is cheating, ok?" what do you mean? Is adding headers supposed to be optional? I'm just curious, because I can't get anything to work without including that line.

1 year ago

[Like](#) [Reply](#)



Manoj

Headers are supposed to fool the website that it isn't being browsed by a robot but a browser

5 months ago [in reply to SG](#)

[Like](#) [Reply](#)



alvin_mites

Find I keep coming back to this page for quick reference. Thank you.

1 year ago

[Like](#) [Reply](#)



tv8688

Hi, thanks for a very helpful tutorial.
In your tutorial, you show how to download an http link. I tried it and it saves to the temp folder.
How can you specify the destination folder for that?
Thanks.

1 year ago

[Like](#) [Reply](#)



Visame

Hi expert
Could you please take a look at this problem:
<http://stackoverflow.com/quest...>
Thanks a lot!

2 years ago

[Like](#) [Reply](#)



stockrt

Try this one:

<http://aspn.activestate.com/AS...>



[2 years ago](#) [in reply to Visame](#)[Like](#) [Reply](#)**Silasus**

hello!
I have a problem when I open a web, this have a redirect to a subdomain to load an image. When I use Live Http Headers Firefox Add-on I can see this redirect but no using python mechanize.
For example I try to open www.domain.com/login which have a redirect to img.domaing.com/?image=1. I want to see it because I need this link.
Also I have the same problem with cookies, this subdomain set a cookie which I don't see (I see the www.domaing.com/login cookies)

[2 years ago](#)[Like](#) [Reply](#)**Pmreghu**

thank you for posting this code snippet. It helped me develop a plugin for xbmc.

[2 years ago](#)[Like](#) [Reply](#)**Fibinse Xavier**

Great tutorial and quickly bookmarked! THANKS!

[2 years ago](#)[Like](#) [Reply](#)**Gerard Labadie**

Great article, a typo,

```
# Show the response headers  
print r.info()
```

should be replaced by

```
# Show the response headers  
print br.info()
```

[2 years ago](#)[Like](#) [Reply](#)**stockrt**

Did you at least tried it out? There is no typo. "r" is the response object.
There are more than one way of listing response headers than, one could guess.

Att,
Rogerio Schneider

[2 years ago](#) [in reply to Gerard Labadie](#)[Like](#) [Reply](#)**No**

this is exactly what i was looking for!

thanks a lot! :D

[2 years ago](#)[Like](#) [Reply](#)**Tobias**

Thanks for the nice summary. Python Mechanize documentation isn't the best. It helped me quite well the get started in mechanize.

[2 years ago](#)[Like](#) [Reply](#)**Jim6**

Hey, thanks for this really easy to understand post!

I badly needed something like mechanize, now that my ISP has implemented a bandwidth cap, and the only information on my



current usage is through a series of links on their forum, with 2 nested password fields (what a waste of time!).

My brand new script fetches the page and presents the info snappily. Awesome!

[2 years ago](#)

[Like](#) [Reply](#)



stockrt

Great Jim6 :)

Glad that it was useful to you.

Regards,
Rogério Schneider

[2 years ago](#) [in reply to Jim6](#)

[Like](#) [Reply](#)



omar

I have tried to log into twitter using mechanize. However, the library is not able to identify the forms names, instead, it puts them together in a single form.

I would like to use mechanize cuz the Twitter API wrappers do not offer what I need.

Any ideas?

[2 years ago](#)

[Like](#) [Reply](#)



stockrt

Doesn't this do the trick?

```
br.open('http://twitter.com')
```

```
for f in br.forms():
    print f
```

```
<post (readonly)="" <hiddencontrol(authenticity_token="e8f2cc4ecd1bad183885776de6116e0fe644008f)" application=""
https="" sessions="" twitter.com="" x-www-form-urlencoded="">
<textcontrol(session[username_or_email]=)>
<passwordcontrol(session[password]=)>
<submitcontrol(<none>=Sign in) (readonly)>
<checkboxcontrol(remember_me=[1])>
<hiddencontrol(q=) (readonly)="">
<post (readonly)="" <hiddencontrol(authenticity_token="e8f2cc4ecd1bad183885776de6116e0fe644008f)" application=""
http="" search="" twitter.com="" x-www-form-urlencoded="">
<textcontrol(q=)>
<post (readonly)="" <hiddencontrol(authenticity_token="e8f2cc4ecd1bad183885776de6116e0fe644008f)" application=""
change_locale="" http="" sessions="" twitter.com="" x-www-form-urlencoded="">
<hiddencontrol(lang=en) (readonly)="">>
```

```
br.select_form(nr=0)
....
```

```
Regards,
Rogério Schneider</hiddencontrol(lang=en)></post></textcontrol(q=)></post></hiddencontrol(q=)>
</checkboxcontrol(remember_me=[1])></submitcontrol(<none></passwordcontrol(session[password]=)>
</textcontrol(session[username_or_email]=)></post>
```

[2 years ago](#) [in reply to omar](#)

[Like](#) [Reply](#)



omar

Yeah cool!

Thanks a lot!

[2 years ago](#) [in reply to stockrt](#)

[Like](#) [Reply](#)



Webcammer

Great example code, i was looking for this all day,urllib and mechanize are so terrible documented... the only thing i hate about python is lack of documentation

[2 years ago](#)

[Like](#) [Reply](#)



**stockrt**

Well, Python is well documented. Some modules that may be not as well documented as we would like to :)

Att,
Rogerio Schneider

[2 years ago](#) [in reply to Webcammer](#)

[Like](#) [Reply](#)

**Webcammer**

You are right, python has nice documentation but still its rough when its your first language to grasp everything because most of the time every documentation assumes you already know the basics, now after weeks of studying i feel much better when grasping info :)

[2 years ago](#) [in reply to stockrt](#)

[Like](#) [Reply](#)

**stockrt**

You may also try challenging yourself.
Try <http://pythonchallenge.com>
I have learned a few tricks in that road.

Regards,
Rogerio Schneider

[2 years ago](#) [in reply to Webcammer](#)

[Like](#) [Reply](#)

**Webcammer**

I heard for that site many times but for some reason those few first examples didnt really make me happy, i guess i am missing something since so many people recommend it. Thanks again for your help, it was really clear and easy to understand for me when i was just starting mechanize and python

[2 years ago](#) [in reply to stockrt](#)

[Like](#) [Reply](#)

**Webcammer**

Btw, maybe you can recommend some good site for code examples, even full scripts but for intermediate beginner level or just beginner level, or anything, thanks again + regards

[2 years ago](#) [in reply to Webcammer](#)

[Like](#) [Reply](#)

**stockrt**

Perhaps you should check this then:

<http://code.activestate.com/re...>

There are good python recipes.

Regards,
Rogério Schneider

[2 years ago](#) [in reply to Webcammer](#)

[Like](#) [Reply](#)

**Webcammer**

Great example code, i was looking for this all day,urllib and mechanize are so terrible documented... the only thing i hate abot python is lack of documentation

[2 years ago](#)

[Like](#) [Reply](#)

**ov3rc10ck**

Not sure exactly were the term emulation comes into play. But its a nice tutorial

[3 years ago](#)

[Like](#) [Reply](#)



**Sébastien Biner**

Thanks for your post. Based on what you wrote I tried to access my "my yahoo" page using my username and password but it doesn't work. Do we have something special to do to access yahoo pages?

Cheers.

[3 years ago](#)

[Like](#) [Reply](#)

**stockrt**

I will think on something :)
You are refering to <http://my.yahoo.com/> right?

[3 years ago](#) [in reply to Sébastien Biner](#)

[Like](#) [Reply](#)

**Sébastien Biner**

Well in fact I wanted to log on the fantasy hockey yahoo website. I just though starting with the my.yahoo.com website would help debugging.
In any case, I managed to get something working with the code presented on <http://matburt.net/?p=81>.
At first I got the same kind of problem as with your code, but I discovered that somehow the urllib2 implementation on my Mac running OSX 10.4 is faulty. I tried with 2 versions of python I had installed.
On a linux machine, the code from maturb worked and yours still did not but with different error message. I think it has something to do with the way both codes deal with cookies. But I am way over my head so I do not really know.

Thanks for the help anyways and keep up the interesting posts.

Cheers.

[3 years ago](#) [in reply to stockrt](#)

[Like](#) [Reply](#)

**rafael81**

nice job~!
I really would want

[3 years ago](#)

[Like](#) [Reply](#)

**stockrt**

rafael81, glad you liked it. What would you want to? I did not catch it... Please explain.

[3 years ago](#) [in reply to rafael81](#)

[Like](#) [Reply](#)

**rafael81**

oops sorry for missed word
I really would want this recipe. :)

[3 years ago](#) [in reply to stockrt](#)

[Like](#) [Reply](#)

**stockrt**

You should see this other too: <http://stockrt.github.com/p/ha...>
Have a good reading!

[3 years ago](#) [in reply to rafael81](#)

[Like](#) [Reply](#)

**Mark**

Hmmm, and if you wanted to log into a site normally? Submit the username and password to the page stated in the post form, then store the results in the cookie?

[3 years ago](#)


[Like](#) [Reply](#)

**stockrt**

Mark, I think I have made this post just for you: <http://stockrt.github.com/p/ha...>
Thank you for the idea :)

[3 years ago](#) [in reply to Mark](#)

[Like](#) [Reply](#)

 **Mark**

Thank a lot, stockrt :) I plan on experimenting with this a lot :D

[3 years ago](#) [in reply to stockrt](#)

[Like](#) [Reply](#)

 **stockrt**

Mark, just let me know if you have any doubts or challenges in this regarding, if I can help, I will, it would be a pleasure.

[3 years ago](#) [in reply to Mark](#)

[Like](#) [Reply](#)

 **stockrt**

Mark, I will update the post with form manipulation tips. Thanks for the question, I will let you know when the post is updated.

[3 years ago](#) [in reply to Mark](#)

[Like](#) [Reply](#)

M [Subscribe by email](#) S [RSS](#)

Reactions



Trackback URL

