

Forum:[Hack Forums](#)

Thread: [Minecraft Username Scraper - McBans.com Scraper - Threaded Version](#)

Hopeful Goods - March 8, 2015, 7:52 pm

You can edit the server to whatever you want and get different lists.

You could edit it to detect how many pages are on the server.

Enjoy and remember to say thank you!

PHP Code

```
<?php
```

```
import pycurl
```

```
import StringIO
```

```
import sys
```

```
import time
```

```
def Usage():
```

```
    exit("./scrape.py <Pages> <Timeout>\nExample: ./scape.py 50 4")
```

```
if len(sys.argv) < 3:
```

```
    Usage()
```

```
for i in range(1,int(sys.argv[1])):
```

```
    c = pycurl.Curl()
```

```
    b = StringIO.StringIO()
```

```
    c.setopt(pycurl.WRITEFUNCTION, b.write)
```

```
    c.setopt(pycurl.TIMEOUT, int(sys.argv[2]))
```

```
    c.setopt(pycurl.CONNECTTIMEOUT, int(sys.argv[2]))
```

```
    xml_ip = "http://mcbans.com/server/13125/c.beastsmc.com/bans/page/" + str(i) + "/embed"
```

```
    c.setopt(c.URL, xml_ip)
```

```
    c.perform()
```

```
    html = b.getvalue()
```

```
    html_tr = html.split("<tr")
```

```
    for me in range(1,len(html_tr)):
```

```
        html_tr_in = html_tr[me].split(">")
```

```
        done = html_tr_in[11].split("<")
```

```
        print done[0]
```

```
    c = 0
```

Threaded Version

PHP Code

```
<?php
```

```
import pycurl
```

```
import StringIO
```

```
import sys
```

```
import time
```

```
from twisted.internet import reactor, threads
```

```
def Usage():
```

```
    exit("./scrape.py <Pages> <Timeout>\nExample: ./scape.py 50 4")
```

```

def Scrape(x):
    c = pycurl.Curl()
    b = StringIO.StringIO()
    c.setopt(pycurl.WRITEFUNCTION, b.write)
    c.setopt(pycurl.TIMEOUT, int(sys.argv[2]))
    c.setopt(pycurl.CONNECTTIMEOUT, int(sys.argv[2]))
    xml_ip = "http://mcbans.com/server/13125/c.beastsmc.com/bans/page/" + str(x) + "/embed"
    c.setopt(c.URL, xml_ip)
    c.perform()
    html = b.getvalue()
    html_tr = html.split("<tr")

    for me in range(1,len(html_tr)):
        html_tr_in = html_tr[me].split(">")
        done = html_tr_in[11].split("<")
        print done[0]
    c = 0

if len(sys.argv) < 3:
    Usage()
commands = [(Scrape, [1], {})]
for i in range(2,int(sys.argv[1])):
    commands.append((Scrape, [i], {}))
threads.callMultipleInThread(commands)
reactor.run()

```

Can you please read my code and tell me if I could have done anything different.

Hopeful Goods - March 16, 2015, 9:05 pm

Nobody thought this was cool?

This is one of my favorite pieces of code.

The Pizza Slut - March 17, 2015, 3:18 am

This is a really nice piece of code! Thanks for the contribution :)

yoloyoloswag9 - March 17, 2015, 5:58 pm

thank you for this hope it still works

Aurora - March 17, 2015, 6:00 pm

I started learning Python the other day, it's good to see I can read and understand some bits of the code. It looks good to me. :)

Hopeful Goods - March 17, 2015, 7:46 pm

(Mar 17, 2015 06:00 PM).Sinister Wrote: [\[CITE\]](#) I started learning Python the other day, it's good to see I can read and understand some bits of the code. It looks good to me. :)

Yeah, I really like coding. Even just the idea of it is cool.

Like, how many "normal" people would know how to write 34 lines of text then run it and boom get a list of minecraft usernames??

I just like having this skill and being able to apply it to simple things like this.

Keep learning python it is a very awesome language that is easy to get the hang of.

psychadelic - March 18, 2015, 9:29 pm

Awesome man, thanks for the share ?

??Darth Vader - March 20, 2015, 1:54 pm

Few things:

1. By convention, one generally uses the "exit" function to return a completion code as in C/C++ (0 for success, otherwise, any other number), not to print an error message. For error messages, one would generally use "sys.stderr.write()" or even just print.
2. I've never heard of PyCurl. Maybe, to improve cross-platform compatibility, use a more standard module such as Requests, or even one of the builtins such as urllib (IMO, urllib is a bitch, use requests).

Hopeful Goods - March 20, 2015, 2:26 pm

(Mar 20, 2015 01:54 PM)??Darth Vader Wrote: [\[CITE\]](#) Few things:

1. By convention, one generally uses the "exit" function to return a completion code as in C/C++ (0 for success, otherwise, any other number), not to print an error message. For error messages, one would generally use "sys.stderr.write()" or even just print.
2. I've never heard of PyCurl. Maybe, to improve cross-platform compatibility, use a more standard module such as Requests, or even one of the builtins such as urllib (IMO, urllib is a bitch, use requests).

I could use beautifulsoup. Have you ever heard of that?

??Darth Vader - March 20, 2015, 3:00 pm

(Mar 20, 2015 02:26 PM)Hopeful Goods Wrote: [\[CITE\]](#) I could use beautifulsoup. Have you ever heard of that?

Beautifulsoup is intended mainly for parsing HTML and retrieving tag values. I meant you could use requests or urllib for fetching the page, rather than parsing it (a builtin solution you could use for parsing is the xml module).

K.ing - April 9, 2015, 8:54 am

second for using python requests over pycurl.

it shouldn't be too hard to change the code out.

<http://docs.python-requests.org/en/latest/user/quickstart/#make-a-request>

Still a great project!

MR.FREE - April 10, 2015, 11:37 pm

(Mar 16, 2015 09:05 PM)Hopeful Goods Wrote: [\[CITE\]](#) Nobody thought this was cool?

This is one of my favorite pieces of code.

I personally don't play mind craft so it didn't appeal to me, I commend you for not using your standard everyday modules, if that was on purpose. Since everyone else did it, let me do it too. Check out the Queue lib and see if you find it useful, that's assuming there should be a shit ton of usernames to go through.

Thanks for the contribution