

Forum:[Hack Forums](#)

Thread: [Save only unique entries to text file](#)

Drazard - March 13, 2017, 7:58 am

Hello and thank you for viewing my thread. I am new here to HF and need some advice. Below i have some source code for a script i am trying to write.

The idea is simple, grab the source code from a website, parse ip's from source and dump into a file.

Code while True:

try:

```
req = requests.get(url, cookies=cookies)
```

```
result = req.text
```

```
soup = str(BeautifulSoup(result, "html.parser"))
```

```
# - return URL source code in a string
```

```
parsedLogs = re.findall("\d+\.\d+\.\d+\.\d+"] logged in", soup)
```

```
# - Search Source for IP logs example: "255.255.255.255 logged in"
```

```
f = open("log.txt", "a+")
```

```
HElog = f.read()
```

```
# - Opens and reads the log file
```

```
for ip in parsedLogs:
```

```
    if ip not in HElog:
```

```
        print (ip)
```

```
        f.write(ip+'\n')
```

```
# - Each IP found in the souce code is saved to a log file
```

```
except:
```

```
    print("ERROR... Restarting script")
```

```
# - The script restarts on errorOutput:
```

Spoiler ([Click to View](#))

Code 192.33.160.70

22.95.220.134

192.33.160.70

22.95.220.134

192.33.160.70

22.95.220.134

192.33.160.70

22.95.220.134

192.33.160.70

22.95.220.134

192.33.160.70

22.95.220.134

192.33.160.70

22.95.220.134

192.33.160.70

22.95.220.134

[illegible]

[illegible]

[illegible]

228.203.167.143
84.121.58.161
69.66.31.96
228.203.167.143
84.121.58.161
69.66.31.96
228.203.167.143
84.121.58.161
69.66.31.96
228.203.167.143
84.121.58.161
69.66.31.96
228.203.167.143
84.121.58.161
69.66.31.96
228.203.167.143
228.203.167.143
228.203.167.143
228.203.167.143
84.121.58.161
69.66.31.96
228.203.167.143
228.203.167.143
228.203.167.143
228.203.167.143
84.121.58.161
69.66.31.96
228.203.167.143
228.203.167.143
228.203.167.143
228.203.167.143
84.121.58.161
69.66.31.96
228.203.167.143
228.203.167.143
228.203.167.143
228.203.167.143
84.121.58.161
69.66.31.96
128.26.225.125
128.26.225.125
128.26.225.125
128.26.225.125
128.26.225.125
128.26.225.125
128.26.225.125
128.26.225.125
128.26.225.125

Please note ive only been messing around with python for a short while. but i really cant put it down

it is a great language.

If anyone could outline where i am going wrong with this i would be so great-full.

The output i am looking for is to only print and log one entry of each ip address found, and append new ip's to the log if this makes sense? the script is to run in a loop scanning the page over and over catching new entries.

(???)?(???? - March 13, 2017, 8:12 am

Delete your log file and run it once. Does it look fine?

You're appending (the "a" in your file opener) to the end of the file. You may want it to be "w" which overwrites the file.

Drazard - March 13, 2017, 8:16 am

Sorry, i should explain further.

The script is written to run 24/7 and check the web source for new entries and save the log to a file, appending all new entries.

The ip addresses in the websource disappear after a few seconds and i am trying to harvest them, if this makes sense? so i require the code to be looped and constantly parse the source grabbing it all, sifting through and only adding new ip addresses to the log file.

EDIT:

```
if ip not in HElog:
    print (ip)
    f.write(ip+'\n')
```

This should only write the ip to the text file if its NOT already there. but its not wroking, so i am doing something wrong

EDIT2:

Actually, i understand why it is still duplicating. using "for ip in HElog:" it checks both ip addresses against each in the list and adds the non duplicate, and then goes to the second line and adds the non duplicate again. so my syntax is incorrect for what i am trying to do.

I need it to check all the ip addresses in the log file against the new ones. what is the best way to read an entire file without using too much memory?

Goodies - March 13, 2017, 12:37 pm

- 1) use a set to make a collection unique before writing to file.
- 2) use a pythonic storage mechanism like shelves and store a set in there.

Depending on how you want it stored...

Drazard - March 13, 2017, 3:42 pm

i have no idea what im doing. this isnt working.

Isnt there simple way to read a file and check the whole file for a string??

```
Code # - Opens and reads the log file
stuff = set()
```

```
for ip in parsedLogs:
    stuff.add(ip)
    print(stuff)
```

Code Python 3.5.3rc1 (v3.5.3rc1:de530d7f21c0, Jan 2 2017, 06:29:54) [MSC

```
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
{'238.65.255.57', '193.171.10.214', '229.4.158.187'}
```

(???)?(???? - March 13, 2017, 5:03 pm

Code with open("file.txt", "r") as f:

```
file_contents = f.read()
```

```
for foo in bar:  
    if foo not in file_contents:  
        result.append(foo)  
Something like that?
```

Drazard - March 13, 2017, 10:03 pm

(Mar 13, 2017 05:03 PM)(°°°)(°°°° Wrote: [\[CITE\]](#) Code with open("file.txt", "r") as f:
file_contents = f.read()

```
for foo in bar:  
    if foo not in file_contents:  
        result.append(foo)  
Something like that?
```

```
Code req = requests.get(url, cookies=cookies)  
result = req.text  
soup = str(BeautifulSoup(result, "html.parser"))  
# - return URL source code in a string  
  
parsedLogs = re.findall("\d+\.\d+\.\d+\.\d+", soup)  
# - Search Source for IP logs example: "255.255.255.255 logged in"  
ips = str(parsedLogs)  
# - Opens and reads the log file  
with open("log.txt", "r") as f:  
    file_contents = f.read()  
  
for ip in parsedLogs:  
    if ip not in file_contents:  
        parsedLogs.append(ip)  
        print (parsedLogs)  
        f = open("log.txt", "a+")  
        f.write(ip+'\n')
```

Returns the same:

Spoiler ([Click to View](#))Code 184.114.61.160

```
184.114.61.160  
184.114.61.160  
184.112.2.155  
184.114.61.160  
184.114.61.160  
184.114.61.160  
184.112.2.155  
184.114.61.160  
184.114.61.160  
184.114.61.160  
184.112.2.155  
184.114.61.160  
184.114.61.160  
184.114.61.160  
184.112.2.155  
184.114.61.160
```

184.114.61.160
184.114.61.160
184.112.2.155
184.114.61.160
184.114.61.160
184.114.61.160
184.112.2.155
184.114.61.160
184.114.61.160
184.114.61.160
184.112.2.155
184.114.61.160
184.114.61.160
184.114.61.160
184.112.2.155
184.114.61.160
184.114.61.160

EDIT:

I apologise, i have changed the code slightly after re reading all your suggestions and come to a reasonably good outcome.

Code with open("log.txt", "r") as f:

```
file_contents = f.read()
```

```
for ip in parsedLogs:
```

```
    if ip not in file_contents:
```

```
        parsedLogs.append(ip)
```

```
        #print (parsedLogs)
```

```
        f = open("log.txt", "w+")
```

```
        f.write(file_contents+ip+'\n')
```

Its not perfect, and any tweaking advice would be great but it does work.

as it is now i have changed the file to re-writable and added that it writes the contents of the file + the new entries, effectively adding only the new ones. so far it seems successful :D

Thank you for all your help.

Goodies - March 14, 2017, 11:07 am

Did you not do what I said? Put the data into a set first. This will make it unique.

Drazard - March 14, 2017, 9:32 pm

(Mar 14, 2017 11:07 AM)Goodies Wrote: [\[CITE\]](#) Did you not do what I said? Put the data into a set first. This will make it unique.

I tried your method using sets, but not enough ws explained on how to use the set module, besides i worked out how to do what i wanted using text files.

:D