

Forum:[Hack Forums](#)

Thread: [Help with account scrapper!](#)

Furtivus - May 26, 2014, 11:02 am

Hello. I have been working on a account scrapper for the League of Legends forums. However I can't seem to get the text that's between

Code <td class="author"> text here </td>The code I am using is here:

Code from bs4 import BeautifulSoup

```
import requests
```

```
r = requests.get("http://forums.na.leagueoflegends.com/board/forumdisplay.php?f=2")
```

```
data = r.text
```

```
soup = BeautifulSoup(data)
```

```
print(soup.findAll('td class="author"'))Any help is welcomed.
```

stonerdays - May 27, 2014, 4:28 pm

A lot easier:

Code import urllib2

from bs4 import BeautifulSoup

```
webPage = urllib2.urlopen("http://forums.na.leagueoflegends.com/board/forumdisplay.php?f=2")
```

```
html = webPage.read()
```

```
soup = BeautifulSoup(html)
```

```
for td in soup.findAll('td', ({'class': 'author'})):
```

```
    authorText = td.contents[0]
```

```
    print ("["+] Found author: %s" % authorText)Piece of cake, use Google.
```

TheBeast808 - May 27, 2014, 6:05 pm

For something this simple, you could alternatively use regexes.

Code import urllib2, re

```
webPage = urllib2.urlopen("http://forums.na.leagueoflegends.com/board/forumdisplay.php?f=2")
```

```
html = webPage.read()
```

```
matches = re.findall("<td class=\"author\">.*?</td>", html)
```

```
names = [match.replace("<td class=\"author\">", "").replace("</td>", "") for match in matches]
```

```
print "Authors: " + str(names)
```

The above may also be a bit faster than the above examples using BeautifulSoup since it does not need to parse the html and build an internal structure for it.