

Forum:[Hack Forums](#)

Thread: [\[HELP\] img grabber script](#)

Bestandsextensie - October 26, 2018, 4:00 pm

hi there, i wanted to fill up my folder with some good HighResolution pictures nam sayin so i created this script the thing is on the site theres two links one thumbnail and one full resolution, it grabs both.

so i am asking for a way to not get duplicates and explain it please, feel free to use this script if u dont care about duplicates or use it yourself

becareful it grabs every url that has an image on it ending with .jpg

on the site i targeted the thumbnail = 123.jpg thats can pop up but the real link also redirects to 123.jpg so i get two values

```
Code import requests
from bs4 import BeautifulSoup
from random import randint
```

```
def imageGrabber(url):
    req = requests.get(url)
    soup = BeautifulSoup(req.content, 'html.parser')
    images = soup.findAll('a', href=True)

    for img in images:
        img = img.get('href')
        if img.endswith('.jpg'):
            f = open(str(randint(0,999))+'.jpg','wb')
            img = 'https:'+img
            print(img)
            f.write(requests.get(img).content)
            f.close()
```

```
def main():
    imageGrabber("")
```

```
if __name__ == '__main__':
    main()
```

to use it add an url from a 4chan thread in the imageGrabber function in main. i cant add urls yet so

dontspamthisplease - October 26, 2018, 4:24 pm

Yeah I don't know why are you getting 2 copies, didn't try your code but here is a little improvement I want to make.

```
Code import requests
```

```
from bs4 import BeautifulSoup
# from random import randint ##NOPE!
from itertools import count
```

```
c = count(0)
```

```
def imageGrabber(url):
    req = requests.get(url)
    soup = BeautifulSoup(req.content, 'html.parser')
    images = soup.findAll('a', href=True)

    for img in images:
        img = img.get('href')
        if img.endswith('.jpg'):
            with open('{0}.jpg'.format(next(c)), 'wb') as f:
                img = 'https:'+img
                print(img)
                f.write(requests.get(img).content)
```

```
if __name__ == '__main__':
    imageGrabber("") #Don't need a main function
    Hope it help a little bit!
```

roach - October 26, 2018, 5:01 pm

Here you go buddy no more duplicates:

```
Code import requests
from bs4 import BeautifulSoup
from random import randint
```

```
def imageGrabber(url):
    imgList = []
    req = requests.get(url)
    soup = BeautifulSoup(req.content, 'html.parser')
    images = soup.findAll('a', href=True)

    for img in images:
        img = img.get('href')
        if img.endswith('.jpg'):
            if not img in imgList:
                imgList.append(img)

    return imgList
def saveImages(list):
    for img in list:
        f = open(str(randint(0, 999)) + '.jpg', 'wb')
        img = 'https:' + img
        print(img)
        f.write(requests.get(img).content)
```

```
f.close()
```

```
def main():  
    list = imageGrabber('http://boards.4chan.org/k/')  
    saveImages(list)  
  
if __name__ == '__main__':  
    main()
```

taikkau - *October 29, 2018, 8:25 am*

and enjoy getting your IP banned. Should have tricked any website with giving user-agent and fake browser.

Louis - *November 14, 2018, 6:49 pm*

There is a wonderful 4chan api wrapper for this
<https://github.com/bibanon/BASC-py4chan>