**The amateur web scrapper**

So here I began web scraping for the first time.

First success I hit was with this small code snippet:

import re

import requests

from bs4 import BeautifulSoup

from urllib.parse import urlparse

import os

f = open("images\_epithelizing.txt", "w")

res=[]

def download\_google():

url = ' <https://www.google.com/search?q=flowers&sxsrf=ALeKk01iIQFHyA_sDt8bILDwhxcuNjO7iw:1588884074973&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjm5oK3zqLpAhX1j3IEHaCZA_kQ_AUoAXoECA8QAw&biw=1440&bih=740>’

page = requests.get(url).text

soup = BeautifulSoup(page, 'html.parser')

for raw\_img in soup.find\_all('img'):

link = raw\_img.get('src')

res.append(link)

if link:

f.write(link +"\n")

download\_google()

f.close()

This could give only 20 images at a time and the size also seemed to be small.

2. The popular code at geeks for geeks didn’t work for me as apparently there has been some change in the way google search engine has started publishing images.

<https://www.geeksforgeeks.org/how-to-download-google-images-using-python/>

3. I made an unsuccessful attempt to go through the selenium way.

4. This python code didn’t work either

from google\_images\_download import google\_images\_download #importing the library

response = google\_images\_download.googleimagesdownload() #class instantiation

arguments = {"keywords":"flowers","limit":20,"print\_urls":True} #creating list of arguments

paths = response.download(arguments) #passing the arguments to the function

print(paths) #printing absolute paths of the downloaded images

5. This was my fifth attempt to download, where I was able to download at least 135 images , which seemed pretty cool

<https://github.com/teracow/googliser>

bash <(curl -skL git.io/get-googliser)

You might be asked to enter the password, wait for some time . Installation will take time.Just print the search url from the google search engine and download images.

googliser --phrase "puppies" --title 'Puppies!' --number 25 --upper-size 100000 -G

5. The final one was a jackpot. I could download as many as 700 images at a time. The claim is that for 1000 images. The website is

<https://www.pyimagesearch.com/2017/12/04/how-to-create-a-deep-learning-dataset-using-google-images/>

1. Click the download button and download the zip folder.

It is having a .py file and a .js file.

1. Just open the search engine and brows the google search page till the last page, and open the javascript console in chrome. Just paste the .js code in the console and run it. A file named urls.txt will be downloaded.
2. Create a folder where you wish to download the images.
3. Open terminal (It will not work on Jupyter notebook )and go to the path where .py file is present.
4. Run the .py file as follows:

pyhon download\_images.py –-urls **path to the urls.txt** –-output **path to the folder where you wish to download the images**