*Webscraping using beautiful soup by python:*

*Requirement:python need to be install(python 3 recommend)*

Step1:

We need to install “pip” by going to get-pip.Then we need to save the file “get-pip.py” and then go to IDLE python and its file location.Then go to script folder where we open cmd.Cmd on script.Then install pip by using the command “python get-pip.py”.

Step2:

In cmd of python:pip install requests

pip install beautifulsoup4

pip install bs4

request.get():

This method is used to request the website for the certain page.200 means ok.

Result=request.get(“url of the website”)  
print(result) [If result is 200 then it is ok to scrap the website.]

200 is the status code for denoting as OK.

#src=result.content:-

The content of the website is saved to the src variable.

Soup=BeautifulSoup(src,’lxml’) :-

The content “src” is passed to the beautifulsoup() method and it format the content as “lxml” format and saves it to soup variables.(not sure about lxml format).

This method helps to create the “soup” as object so that we can access the properties of the object by using object\_name.properties.  
Example: soup.title()

Soup.find\_all(“a”) => find all the “a” tag contents on the website.

IF BeautifulSoup(src,’lxml’) doesnot work then use BeautifulSoup(src,’html.parser’).

Some time “pip install lxml” can worked.

Links=soup.find(“a”) :-returns the first anchor tag on the website

Links=soup.find\_all(“a”) :- returns all the anchor tag on the website.