

Web scraping is a term used to describe the use of a program or algorithm to extract and process large amounts of data from the web.

We shall see the webscraping processes by means of an example as below:

Firstly, the required modules for webscraping needs to be downloaded.

In the below example beautiful soup is used for extraction and urllib2 for accessing the url.

```
#!/usr/bin/env python
```

```
from bs4 import BeautifulSoup  
import urllib2
```

Secondly, the website needs to be accessed:

```
# Opens the required website  
html_page = urllib2.urlopen("https://www.nytimes.com/search?query=HIV&sort=best")  
soup = BeautifulSoup(html_page)
```

Thirdly the website needs to be inspected for the class and its attributes in order to search for the particular data.

This can be accessed by the following code:

```
# Extract the title of the article  
items = soup.find_all("div", attrs={"class": "css-138we14"})  
for item in items:  
    title= item.parent.h4.text  
    url = item.parent.attrs.get('href', None)  
    print("{} {}".format(title))
```

```
# Extract the body of the article  
subjects = soup.find_all("p", attrs={"class": "css-1dwgixl"})  
for subject in subjects:  
    body= subject.text  
    url1 = subject.parent.attrs.get('href', None)
```

The final step would be to print the title and the body of the article (as per the above example):

```
print("{}:{}".format("Title",title))  
print("{}:{}".format("Subject", body))
```