**ABOUT Scraping**

**Web scraping** (**web harvesting)** is a computer programming technique of extracting information from sites automatically or perform specific tasks. This is accomplished by either directly implementing the HTP (on which the Web is based), or embedding a web browser.

This technique is basically used in web automation. Web scrapers can also be used as triggers to specific operation on websites etc. There are several languages to write web scrapers. Web scraping focuses more on the transformation of unstructured data into structured data that can be stored and analyzed in a spreadsheet or can be stored on local or an online Database. The technique used to perform this task is using bots or web crawlers which are programed to perform specific tasks according to user requirements.

**User Guide**

**Requirements**

* PC with windows 7 or above / Ubuntu or Linux operating system
* Python2.7
* Test.py
* mail.py
* webdriver.exe for selenium Python

**How to Install Python 2.7 to run the scraper**

* Download Python 2.7 from the link <https://www.python.org/downloads/release/python-2712/>
* Install Python2.7
* Open C: drive
* A folder is created with the name of Python27 copy the path of that folder that would be like ‘C:\Python27’
* Now Right click on This PC on Desktop or on the left side of file explorer
* Click on properties
* Click on ‘Advanced System Settings’ option
* Click on Environmental variables
* You can see two widgets here one is one are ‘user variables’ other is ‘system variables’
* From system variables find variable Path
* Double click on it
* In the variable value textbox paste the path of Python27 i.e ‘C:\Python27’ separated by semi colon ‘;’
* Click on ‘OK’ on every opened window
* Download webdriver.exe for selenium Python from link <http://chromedriver.storage.googleapis.com/2.22/chromedriver_win32.zip>
* Create a folder in C: Drive
* Rename the folder to ‘new’
* Extract the webdriver.exe file to that new folder.
* Open cmd / Command Prompt from run or windows search bar
* Type c: and press enter
* Type cd\ and press enter
* Type cd Python27 and press enter
* Type cd Scripts and press enter
* Type pip install bs4 and press enter (In case of any error type pip install python-bs4 and press enter)
* Type pip install selenium and press enter (In case of any error type pip install python-selenium and press enter)
* Type pip install tqdm and press enter (In case of any error type pip install python-tqdm and press enter)
* Type pip dateutil bs4 and press enter (In case of any error type pip install python-dateutil and press enter)
* Type pip install time and press enter (In case of any error type pip install python-time and press enter)
* Type pip install lxml and press enter (In case of any error type pip install python-lxml and press enter)
* After installing all these packages close the command prompt / cmd.
* Now open C: Drive
* Create a folder
* Rename this folder to ‘Temp’
* Open this folder
* Paste the Test.py & mail.py file to this folder.
* Now you are ready to run the scraper.

**How to Run**

* Open cmd / Command Prompt from run or windows search bar
* Type c: and press enter
* Type cd\ and press enter
* Type cd Temp and press enter
* Type python file.py and press enter

**By: Usama Jamil**