Selenium Web Scraping Tutorial

Web scraping allows you to extract data from websites. The process is automatic in which the HTML is processed to extract data that can be manipulated and converted to the format of your liking for retrieval and or analysis. The process is commonly used for data mining.

What is Selenium?

Selenium is an automation tool for web browsers. It is primarily used for testing of websites allowing you to test it before you put it live. It gives you the chance to perform the following tasks on the website:

- Click buttons
- Enter information within the website, forms
- Search for information on the website

It is a tool that has been used for scraping website. But you must note that if you scrape a website too often, you risk the chance of having your IP banned from the website so approach with caution.

How to scrape with Selenium?

In order to scrape websites with Selenium you will need [Python](#), either Python3.x. or Python2.x. Once you have that downloaded you will need the following driver and package

Selenium package – allows you to interact with website from Python

Chrome Driver – a platform to perform and launch tasks on browser

Virtualenv – helps create an isolated Python environment

1. In Python, you need to create a new project. You can create a file and name it setup.py and within it type in selenium as dependency.
2. Then open the command line and you will need to create a virtual environment by typing the following command: $ virtualenv webscraping_example
3. You will now need to run the dependency on virtualenv, you can do this by typing th

following command in the terminal: $(webscraping_example) pip install -r setup.py

4. Now going back to the folder in Python, create another file and you can name it, webscraping_example.py. Once done, you need to add the following code snippets

   1. from selenium import webdriver
   2. from selenium.webdriver.common.by import By
   3. from selenium.webdriver.support.ui import WebDriverWait
   4. from selenium.webdriver.support import expected_conditions as EC
   5. from selenium.common.exceptions import TimeoutException

5. You then need to put Chrome in Incognito mode, this is done in the webdriver by adding the incognito argument:

   1. option = webdriver.ChromeOptions()
   2. option.add_argument(" — incognito")

6. You will then create a new instance with this code: browser = webdriver.Chrome(executable_path='/Library/Application Support/Google/chromedriver', chrome_options=option)
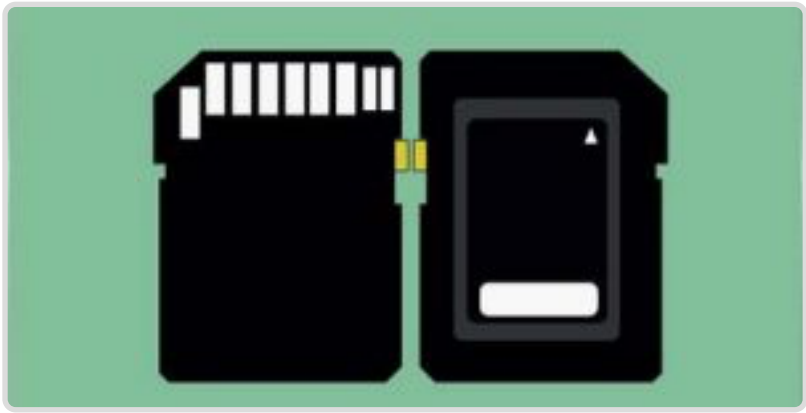
7. You can now start making request you pass in the website url you want to scrape.

8. You may need to create a user account with Github to do this but that is an easy process.

9. You are now ready to scrape the data from the website.

Paulo	Updated January 9, 2018	Web scraping

Data Mining Jobs for Pros

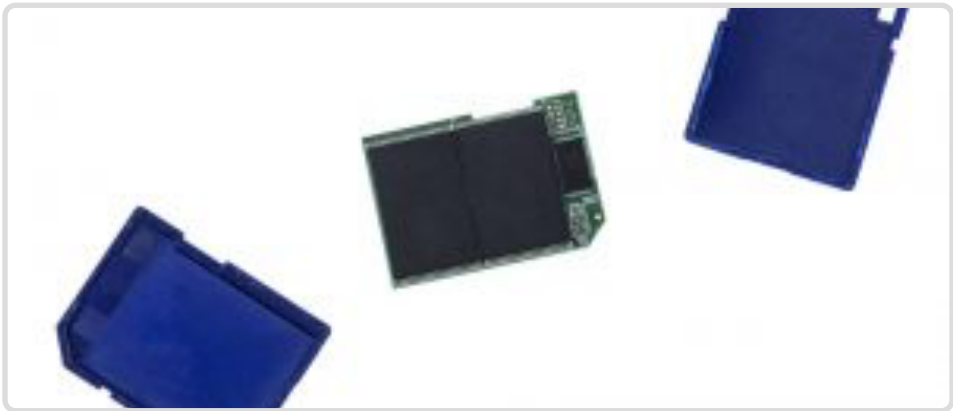**CleanMyPc Review 2017**

What to do if my SD Card Won't Read

How to Download Pictures from Google Drive

How to Repair Damaged SD Card in Android

How to Download Entire Facebook Photo Albums of Friends or Pages

Extract Data from Multiple Web Pages into Excel using import.io

Difference between Data Mining and KDD

Leave a Reply:

Name*

Email*

Website

SUBMIT

Thrive Themes | Powered by WordPress