WEB SCRAPING
WORKSHEET – 1

In Q1 to Q9, only one option is correct, Choose the correct option:

1. Which of the following extracts information from user generated content?
   A) Java script tagging
   B) Web scraping
   C) A/B testing
   D) MROCs
2. Which of the following is not a web scraping library in python?
   A) selenium
   B) Beautiful soup
   C) Requests
   C) scrapy
3. Selenium tests _____?
   A) Browser based applications
   B) DOS applications
   C) GUI applications
   D) All of the above
4. Task of crawling is performed by a complex software which is known as:
   A) Scraper
   B) Crawler
   C) Boat
   D) Spider
5. Which of the following commands is used to access name of a tag in Beautiful Soup?
   A) tag.attrs
   B) tag.name
   C) tag,id
   C) tag['id']
6. Which of the following is the default parser in Beautiful Soup?
   A) html.parser
   B) html5lib
   C) lxml
   D) lxml-xml
7. In selenium the webdriver is used to?
   A) design a test using selenese
   B) test a web application on firefox only
   C) execute tests on HtmlUnit browser
   D) to download any content from a webpage
8. In selenium, driver.**find_elements_by_xpath('given xpath')** returns:
   A) the first webelement associated with the 'given xpath'
   B) the url of first webelement associated with the 'given xpath'
   C) the list of all webelements associated with the 'given xpath'
   D) all the attributes of the first webelement associated with the 'given xpath'
9. The script **'window.scrollBy(0,a)** scrolls the webpage by?
   A) **'a'** number of horizontal spaces
   B) **'a'** number of lines
   C) **'a'** number of pixels horizontally
   D) **'a'** number of pixels vertically

In Q10, more than one options are correct, Choose all the correct options:

10. Which of the following is(are) tags of HTML?
    A) <a>
    B) <b>
    C) <image>
    D) <href>

Q10 to Q13 are subjective answer type questions, Answer them briefly.

11. What is the main difference between a web scraper and a web crawler?
12. What is **'robots.txt'** file? What is the use of **'robots.txt'** file?
13. What are static and dynamic web pages?

Q14 and Q15 are programming practice questions. Solve it using JUPYTER NOTEBOOK and paste the solution in your answer sheets.

14. Write a python program to check whether a webpage contains a title or not.
15. Write a python program to access the search bar and search button on images.google.com.