**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

--- QUESTION not clear.

1. Which of the following is not a web scraping library in python?

A) selenium B) Beautiful soup

C) Requests C) scrapy

Ans- C

1. Selenium tests \_\_\_\_\_\_\_\_\_\_?

A) Browser based applications B) DOS applications

C) GUI applications D) All of the above

Ans -A)

1. Task of crawling is performed by a complex software which is known as:

A) Scraper B) Crawler

C) Boat D) Spider

Ans -D-

1. 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’]

Ans-B

1. Which of the following is the default parser in Beautiful Soup?

A) html.parser B) html5lib

C) lxml D) lxml-xml

Ans – C

1. 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

Ans -A

1. 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’

Ans - C

1. 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

Ans - C

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

1. Which of the following is(are) tags of HTML?

A) <a> B) <b>

C) <image> D) <href>

Ans – A,B,D

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

1. What is the main difference between a web scraper and a web crawler?

Ans – The main difference between a scraper and web crawler are as follows :-

Scraping is the technique of extracting data from a webpage (structured) . for example we want to extract some product , price information from flipkart , then we will be in need of scrapping , which will be helped by some libraries like selenium , beautiful soup etc.

Web crawler is a technique of find and fetch web links and extracting content in indiscriminate manner ,

For example what google , yahoo does is a web crawler . it is also called spider .

1. What is **‘robots.txt’** file? What is the use of **‘robots.txt’** file?

Ans – Robots.txt is a text file webmasters create to instruct web robots (typically search engine robots) how to crawl pages on their website

The robots. txt file, also known as the robots exclusion protocol or standard, is a text file that tells web robots (most often search engines) which pages on your site to crawl. It also tells web robots which pages not to crawl.

1. What are static and dynamic web pages?

**Ans - Static web pages**  are the pages that remains same for every user through out the task , for example application form for an exam , whereas **dynamic web pages** changes and does not remain same for every user , one basic example is companies portal , which changes as per the position of the user , like manager will have different view and CEO will have different views of the of the webpage.

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

1. Write a python program to check whether a webpage contains a title or not. –is submitted via python notebook file
2. Write a python program to access the search bar and search button on images.google.com. -----is submitted via python notebook file