WEB SCRAPING – WORKSHEET 2

## In Q1 to Q10, only one option is correct, Choose the correct option: -

1. Which command method looks through a tag’s descendants and retrieves all descendants that match your filters?
   1. search\_all() B) get\_all()



C) find\_all() D) None of the above

**ANS:** **C) find\_all()**

1. Which operating system among the following is not supported by selenium.
   1. Windows B) Linux

C) Unix D) Solaris

**ANS: C) Unix**

1. Which of these languages is NOT supported by the Selenium RC?
   1. Python B) Java

C) ASP D) C#

**ANS: C) ASP**

1. Which of these commands is used to enter text in text\_boxes.
   1. sentkeys() B) sent\_keys()

C) keys.send() D) send\_keys()

**ANS: D) send\_keys()**

1. Websites fetched by crawler are indexed and kept in huge database, this process is called as:
   1. Optimizing B) Indexing

C) Crawling D) None of them

**ANS: B) Indexing**

1. Web Crawler is also called as:
   1. Search Optimizer B) Web-manager

C) Web-spider D) Linked Directory

**ANS: C) Web-spider**

1. Which selenium command among the following makes WebDriver wait for a certain condition to occur before proceeding further with execution.
   1. verifyTextPresent B) webdriver\_wait

C) WebDriverWait D) none of the above

**ANS: C) WebDriverWait**

1. Which of the following are commonly used expected conditions in Selenium?
   1. title contains B) staleness\_of

C) verify element D) None of the above

**ANS: C) verify element**

1. Where xpath is used?
   1. MS word documents B) MS excel

C) XML D) None of the above

**ANS: C) XML**

1. Which of the following commands is used to find an element in a webpage from the id attribute of its tag .
   1. find\_element\_BYid B) find\_element\_by\_id

C) find\_elements\_by\_id D) None of the above

**ANS: B) find\_element\_by\_id**

## In Q11 to Q13, more than one options are correct, Choose all the correct options:

1. Consider the following line in a HTML code

## <a class="sister" href="<http://example.com/lacie>" id="link2"># Lacie # </a>

Which of the following are attributes of <a> tag here?

* 1. class B) href

C) # Lacie # D) link2

**ANS: A) class B) href D) link2**

1. Which among the following is/are parsers used in BeautifulSoup?
   1. href B) soup\_parser

C) lxml D) None of the above

**ANS: C) lxml**

1. Select the packages you require generally in web-scraping.
   1. Selenium B) BeautifulSoup

C) Requests D) B&C only.

**ANS: A) Selenium** **B) BeautifulSoup C) Requests**

## Q14 and Q15 are programming practice questions. Solve it using JUPYTER NOTEBOOK.

1. Write a python program to scrap 10 images of Rayban Sunglasses from flipkart website and save them in a directory.
2. Write a program to scrap 20 Mouses’ data which includes model name, price from [www.amazon.in](http://www.amazon.in/) whose price

is less than 500 and make a data frame with 2 columns “model\_name”,” price” with the scraped data

JUPYTER NOTEBOOK LINK: