WEB SCRAPING – WORKSHEET 4

In Q1 to Q14 have one or more than one correct options, Choose all the correct options:

1. Which of the following functions can be used to get an element from webpage when we know the Name attribute of the element?
   A) get_by_name() B) get_element_by_name()
   C) find_element_by_name() D) None of the above

   ANS: D) None of the above

2. Which of the following functions can be used when you want to locate an element by tag name?
   A) get_elements_by_tagid() B) get_element_by_tagsid()
   C) find_element_by_tag_name() D) All of the above

   ANS: C) find_element_by_tag_name()

3. In what type of Waits, a WebDriver waits for a certain condition to occur before proceeding further with execution.
   A) Implicit wait B) Explicit wait
   C) Both of them D) None of them

   ANS: B) Explicit wait

4. Which of the following is an expected condition in selenium (python)?
   A) title_is B) visibility_of
   C) staleness_of D) All of the above

   ANS: A) title_is C) staleness_of

5. Which of the following is a disadvantage of html5lib parser in beautiful soup?
   A) External C dependency B) Very Slow
   C) External Pyhton Dependency D) all of the above

   ANS: B) Very Slow

6. What are the advantages of using Scrapy over Selenium for web-scraping?
   A) For large data Scrapy is faster than selenium
   B) It supports javascript better than Selenium
   C) Scrapy is better than Selenium for simple projects
   D) All of the above

   ANS: A) For large data Scrapy is faster than selenium

7. Which of the following is (are) true regarding Scrapy?
   A) spiders are classes which define how a certain site will be scrapped.
   B) spiders are the place where you define the custom behaviour for crawling.
   C) None of them
   D) both A & B

   ANS:.

8. Full form of HTML:
   A) Hyper Text Markup Link
   B) Hyper Text Mark language
   C) Hyper Text Markup Language
   D) Hyper Text Mining Link

   ANS: C) Hyper Text Markup Language

9. Which among the following is the correct syntax for parsing a html page?
   A) soup=BeautifulSoup(html_doc, html)
   B) soup=BeautifulSoup(html_doc,'html.parser')
   C) soup=BeautifulSoup(html_doc,'html_parser')
   D) none of the above

   ANS: B) soup=BeautifulSoup(html_doc,'html.parser')

10. Which among the following is not a valid parser in BeautifulSoup?
    A) "lxml"
    B) "html.parser"
    C) "lxml-xml"
    D) "html-xml"

    ANS: D) "html-xml"

11. Which of the following functions is used to go to the next element in the page?
    A) findNext_all()
    B) Find_all()
    C) find_next()
    D) None of the above

    ANS: C) find_next()

12. Which of the following functions are used to iterate over an element's siblings that precede it in the tree?
    A) find_previous_siblings()
    B) Get_prev_sibs()
    C) get_siblings()
    D) None of the above

    ANS: A) find_previous_siblings()

13. Which of the following is an argument used in find_all() which tells Beautiful Soup to stop gathering results after it's found a certain number.
    A) stop_at
    B) stop_before
    C) limit
    D) None of the above

    ANS: C) limit

14. How would you set the recursive argument in find_all() if you want Beautiful Soup to consider only the direct children.
    A) recursive = True
    B) recursive = False
    C) recursive argument has no effect
    D) None of the above

    ANS: B) recursive = False

Q15 is subjective answer type question, Answer it briefly.

15. What is the difference between find() and find_all() in Beautiful Soup?

find()	**find_all()**
The find() method get the first tag of the HTML object which satisfy the condition given.	The find_all() method scans the entire HTML object and get all descendants matching the query.
Syntax; soup.find(tag, attribute, text, keywords,recursive) Where, tag – it shows the name of tag to be searched. attribute- it is the attribute value in the tag name.it can be multiple. text- it find the list of tags based on text content of the tag	Syntax; soup.find_all(tag, attribute, recursive, limit, text, keywords)