**WEB-SCRAPING (**WORKSHEET**)**

1. (B)

2. (C)

3. (D)

4. (D)

5. (B)

6. (A)

7. (D)

8. (A)

9. (D)

10. (A, B)

11. **web scraper:** it is a tool which is used to extract the required data from any websites and saving that data to our local machine in the form of XML, SQL etc. It sends the request to the target page and after getting respond, it extracts the response and downloads the required data.

**web crawler:** it is a tool which is used to indexing the webpages by visiting the website. For extracting more information, it visits each and every page and analyse all the web pages. I t is being used in Google, Yahoo like many search engines to give search results to the user.

12. **robots.txt:** It is a text file which included in the source files of many websites for crawling.

**Format:** User-agent: [user-agent name]Disallow: [URL string not to be crawled]

Within a robots.txt file, each set of user-agent directives is separated by a line break. For any given website, we can view this file by typing the full URL of the home-page and then adding /robots.txt

Basically, it tells search engine crawler to which pages or files the crawler can request or not from your site. For this purpose, these files use a various protocol like Sitemaps protocol, Robots Exclusion protocol.

It can be used to manage crawl traffic and also prevents image, video, and audio files from appearing in Google. search results.

13. **Static webpages:** The static webpages are written in languages such as css,html,javascript etc. In static web pages, pages will remain same until someone changes it manually. After receiving request for a web page, server sends the response to the client without doing any additional process.so it takes less time for loading and very simple in terms of complexity. Static web pages neither contain any application program and nor use any database.

**Dynamic webpages:** The dynamic Web Pages are written in languages such as CGI, ASP, ASP.NET, etc. In dynamic web pages, contents of pages are different for different users. So, it takes more time for loading as compared to static web page. Dynamic web pages contain application program for different purposes and also use database.

14. # Python program to check whether a webpage contains a title or not

from urllib.request import urlopen

from urllib.error import HTTPError

from bs4 import BeautifulSoup

def getTitle(url):

try:

link = urlopen(url)

except HTTPError as e:

return None

try:

BS = BeautifulSoup(link.read(), "lxml")

title = BS.body.h1

except AttributeError as e:

return None

return title

title = getTitle(url)

if title == None:

return "Title could not be found"

else:

return title

print(getTitle("https://www.facebook.com/"))

print(getTitle("http://www.naukri.com/"))

15. #Python program to access search bar and search button

from selenium import webdriver

from selenium.common.exceptions import StaleElementReferenceException

driver=webdriver.Chrome('C:\\Users\\hp\\Downloads\\chromedriver\_win32\\chromedriver.exe')

my\_page=driver.get('https://images.google.com/?gws\_rd=ssl')

search\_bar=driver.find\_elements\_by\_xpath("//div[@class='RNNXgb']")

search\_button=driver.find\_elements\_by\_xpath("//span[@class='z1asCe MZy1Rb']//svg")