**WEB SCRAPING**

## **WORKSHEET – 1**

1. B
2. C
3. A
4. B
5. B
6. C
7. D
8. C
9. D
10. A,B,C
11. **Web scraping**: Web scraping is basically used for extracting information from websites using automated scripts. It does not copy and paste but fetches data in a precise and accurate manner.

**Web crawling**: It is also called as spider. It is used for indexing the information on the page using bots known as crawlers. Web crawlers navigate through heaps of data and procure whatever Is relavent.

1. A robots.txt file is a set of instructions for [bots](https://www.cloudflare.com/learning/bots/what-is-a-bot/). It is primarily used to manage crawler traffic to your site and to avoid overloading your site with requests.

A web crawler bot will follow the most specific set of instructions in the robots.txt file. If there are contradictory commands in the file, the bot will follow the more granular command.

1. **Static web pages**: In static web pages when a server receives a request for a web page, then the server sends the response to the client without doing any additional process then these web pages are seen through a web browser. In static web pages Pages will remain the same until someone changes it manually. Static web pages does not contain any application program and does not interact with any database.

Static web pages require less work and cost in designing them.

**Dynamic web pages:** In dynamic web pages the content of pages is different for different visitors. It takes more time to load than the static web page. Dynamic web pages are used where the information is changed frequently(like stock prices, weather information, etc). In dynamic web pages, database is used. Dynamic web pages require comparatively more work and cost in designing them.

1. import requests

from bs4 import BeautifulSoup

url="https://towardsdatascience.com/an-introduction-to-web-scraping-with-python-a2601e8619e5"

req = requests.get(url)

soup = BeautifulSoup(req.text, "html")

if (soup.title!=None):

print("This site contains Title")

print ("The title of the page is: %s" % soup.title.text)

else:

print("This site dones not contain any title")

1. from selenium import webdriver

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

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

word = "webscraping"

text\_box = driver.find\_element\_by\_xpath('//input[@title="Search"]')

text\_box.send\_keys(word)

driver.find\_element\_by\_class\_name("Tg7LZd").click()