**WEB SCRAPING**

## **ANSWERSHEET – 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?

**B) Web scraping**

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

**C) Requests**

1. Selenium tests \_\_\_\_\_\_\_\_\_\_?
2. **Browser based applications**

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

**B) Crawler**

5.Which of the following commands is used to access name of a tag in Beautiful Soup?

**B) tag.name**

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

**C) lxml**

7.In selenium the webdriver is used to?

**D) to download any content from a webpage**

**8.** In selenium, driver**.**find\_elements\_by\_xpath(‘given xpath’)returns:

**C) the list of all webelements associated with the ‘given xpath’**

**9.** The script **‘**window.scrollBy(0,a) scrolls the webpage by?

**D) ‘a’ number of pixels vertically**

**10**.Which of the following is(are) tags of HTML

**B) <b>**

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

**Ans.** [Web scraping](https://prowebscraper.com/blog/what-is-web-scraping/) is basically extracting data from websites in an automated manner.

* It is automated because it uses bots to scrape the information or content from websites.
* It’s a programmatic analysis of a web page to download information from it.
* Data scraping involves locating data and then extracting it. It does not copy and paste but directly fetches the data in a precise and accurate manner. It does not limit itself to the web; data can be scraped virtually from anywhere it is stored. It does not have to be from the Internet. It is about data and not where it is stored.
* Example of Web Scraping
  + Web scraping would involve scraping specific information from a particular web page or pages.
  + For example, you want to work on price intelligence. You would extract the price of various/specific products from Amazon or any other e-commerce site.
  + This would qualify as web scraping. Likewise, you can extract data and use it for business leads, stock market data, real estate listings.

[**web crawler**](https://en.wikipedia.org/wiki/Web_crawler)

* The term crawling comes from the way a spider would crawl. That’s why a [web crawler](https://en.wikipedia.org/wiki/Web_crawler) is also sometimes called a spider. It’s basically an internet bot that systematically browses (read crawls) the World Wide Web, usually for the purpose of web indexing.
* It is used for indexing the information on the page using bots also known as crawlers.
* It involves looking at a page in its entirety and indexing it, including its last letter and dot on the page, in the quest for information.
* Crawling through every nook and crevice of the World Wide Web, the spider locates and retrieves the information lying in the deeper layers. Web crawlers or bots navigate through heaps of data and information and procure whatever is relevant for your project.
* Example of Web crawling
  + What Google, Yahoo or Binge does is a straightforward example of web scraping.
  + These search engines crawl web pages and use the information for indexing the web pages.

Q12.What is **‘robots.txt’** file? What is the use of **‘robots.txt’** file?

**Ans.** A **robots**. **txt file** tells search engine crawlers which pages or **files** the crawler can or can't request from your site. A search engine bot (like Googlebot) will read the **robots**. **txt file** prior to crawling your site to learn what pages it should deal with. Robots.txt implements the REP ([Robots Exclusion Protocol](https://en.wikipedia.org/wiki/Robots_exclusion_standard)), which allows the Web site administrator to define what parts of the site are off-limits to specific robot user agents. Web administrators can Allow access to their Web content and Disallow access to cgi, private and temporary directories, for example, if they do not want pages in those areas indexed.

Q13.What are static and dynamic web pages?

Ans.

| Sr. No. | Key | Static Web Page | Dynamic Web Page |
| --- | --- | --- | --- |
| 1 | Definition | Static web pages are generally simple HTML written pages which serve as response from browser to server in which all the information and data is static in nature and it does not get changed until someone changed it manually. | On other hand Dynamic webpages are the pages written in some more complex language such as ASP.NET in which data is rendered after some interpretation and capacity to produce distinctive content for different calls. |
| 2 | Complexity | As mentioned in above point as data in static web pages is static and do not require any interpretation before rendering so static web pages are simple in complexity. | Dynamic web pages on other hand does the interpretation process which make data dynamic in nature and due to which dynamic web pages become complex in complexity as compare to static web pages. |
| 3 | Language used | Static web pages are generally written in simpler languages such as HTML, JavaScript, CSS, etc. | On other Dynamic web pages are written in more complex languages such as CGI, AJAX, ASP, ASP.NET, etc. |
| 4 | Rendered Data | For static web pages data do not changes until someone changes it manually and hence data is static in nature. | On other hand for Dynamic web page data is first interoperate at server side and due to which it does not remain same on every call and this makes data dynamic in nature.. |
| 5 | Time | Static web pages due to static data take less time to get load. | While Dynamic web pages due to dynamic data take comparatively more time as compare to static web pages. |
| 6 | Database | In Static web pages generally no involvement of database for data redecoration. | On other hand in case of Dynamic web page database is used for data redecoration. |