**Assignment 02: Scrape the Simplilearn Website Resource Page**

**Web Scraping of Simplilearn Website Resource page**

**DESCRIPTION**

**What to:**

Scrape the Simplilearn website resource page and perform the following tasks:

1. View and print the Simplilearn web page content in a well formatted format
2. View the head and title
3. Print all the href links present in the Simplilearn web page
4. Search and print the resource headers of the Simplilearn web page
5. Search resource topics
6. View the article names and navigate through them

**Simplilearn website URL:** <http://www.simplilearn.com/resources>

**Assessment:**

(Note: For code refer the notebook attached in pdf format)

**1: View and print the Simplilearn webpage content in a good format:**

* Import the BeautifulSoup and urlopen libraries.
* Provide the URL for web scraping.
* Scrape the web page document.
* Create the soup object and parse the webpage using an html parser.
* Close the webpage object.
* View the soup contents.
* Use prettify () method to view the well-formatted html output.

**2: View the title:**

* View the title of the scraped document using title method.

**3: Print all the href links present in the Simplilearn webpage:**

* Find all href links using findAll ().

**4: Search and print the resource headers of the Simplilearn webpage:**

* View the entire headers name.
* The div information is obtained from the web page (view page source).

**5: Search resource topics:**

* View all the resource topics.
* The class information is obtained from the web page (view page source)

**6: View the article names and navigate through them:**

* Create article object for finding the list (Hint: the class information is obtained.
* View the type of the article object using type ().
* Find all the article names using findAll (‘p’).
* Find next sibling (next article name) using next\_sibling.
* Find previous sibling (previous article name) using previous sibling.