# *Web scraping*

*Data collection and quality*

***Description:***

The purpose of this lab is to perform data collection using web scraping. Web scraping is one of the data collection methods and is used to capture dynamic data, images and links from online resources such as websites, etc. You will learn to ethically scrape necessary data. The quality of data depends on the source of data which needs to be considered prior scraping. You will learn to scrape textual and numerical information from websites. Task 1 will aim to capture textual data from websites while task 2 aims to capture market data from websites. Task 3 aims to get data from multiple websites which helps in scaling and fault tolerance. Bonus task show you one of the complexity in extracting data.

***Task1:***

1. Choose a topic like: Machine learning, Artificial intelligence, Deep learning, Regression, etc.
2. Search for websites of reliable sources
3. Choose 2 websites and scrape, headline, and textual information from these two websites and store it in a text file

***Task2:***

1. Choose an e-commerce website like mediamarket, elgiganten, etc.
2. Go to search page of any product and scrape name and cost of the product.
3. Search results from one page is enough
4. Store these details in a csv file

***Task3:***

1. Scrape data from two weather websites like timeanddate, wunderground, etc.
2. Handle errors and exceptions
3. Save the data locally in a text file or online database like firebase.

***Task 3***

Go to the website https://nopecha.com/demo pick one from easy to hard difficulty catcha. Develop an algorithm to get past the captcha.

Record and discuss your code and discuss the extracted data for each task. Maximum time of recording is 15 minutes. You can share the video file or YouTube link when submitting the recording. You can use screen recorder software's or zoom application to record your screen.