

What is Scraping?

In this lesson, an introduction to scraping is provided.

We'll cover the following ^

- Scraping?
- How is it done?

Scraping?

The extraction of data from a website is known as *scraping*. A scraper makes a request to a web page for access either using the *HTTP* protocol or through a browser. The following steps are performed in receiving data from a web page.

- The scraper makes a *GET* request to a web page.
- In response, the structure of the web page, which contains all information in the *HTML DOM* structure is returned.
- The scraper then parses the *HTML* structure to find the information that is required.
- The information is then fetched and stored.

Any kind of data can be scraped from a web page, whether it be a text, video, audio, or image.

How is it done?

The main part of scraping is to fetch the data from the *HTML DOM* structure after receiving the response from the web site. The request and response part can easily be done using built-in tools, but the efficient retrieving of data depends on the skill of the programmer.

Web scraping can be thought of like a spider crawling over each and every corner

of an *HTML* structure and retrieving information when needed.

In the next lesson, the tool used for scraping is discussed.