Scraping Web Content

Web Scraping

Automated extraction of data from websites; website content is first fetched (usually using HTTP) and then parsed to extract specific information.

Web Scraping

Automated extraction of data from websites; website content is first fetched (usually using HTTP) and then parsed to extract specific information.

Web Scraping

Automated extraction of data from websites; website content is first fetched (usually using HTTP) and then parsed to extract specific information.

Web Scraping

Automated extraction of data from websites; website content is first fetched (usually using HTTP) and then parsed to extract specific information.

Web Scraping

Automated extraction of data from websites; website content is first fetched (usually using HTTP) and then parsed to extract specific information.

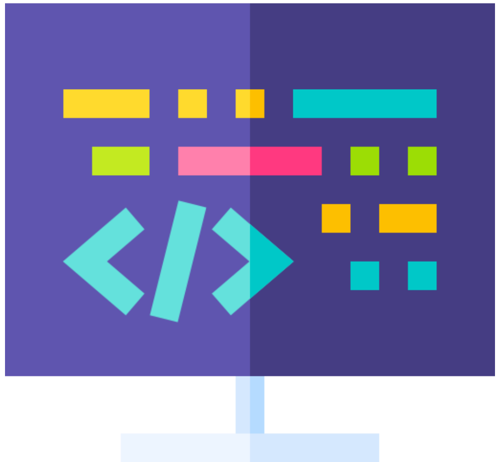
Web Scraping

Automated extraction of data from websites; website content is first fetched (usually using HTTP) and then parsed to extract specific information.

Web Scraping

Automated extraction of data from websites; website content is first fetched (usually using HTTP) and then parsed to extract specific information.

Web Pages

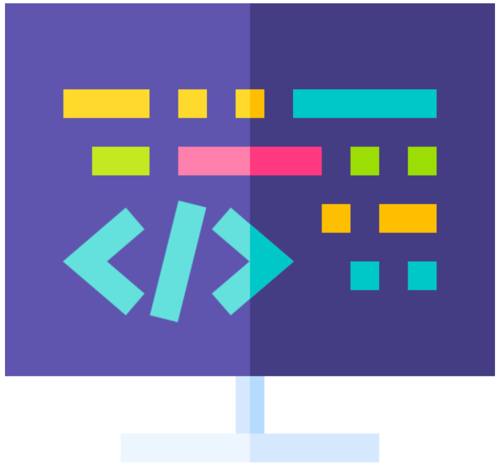


Websites are collections of web pages

Web pages consist of marking e.g. HTML

This markup is understood and rendered by browsers

Fetching and Parsing



The same HTML markup can be accessed (fetched) via HTTP

Possesses an in-built hierarchical structure

Parsers can exploit this structure to extract information