

The Web scraper API provides service to scrape data from websites and retrieve information about images, posts, and sentiment analysis.

Base URL

- The base URL for the API is: ``http://localhost:8000``.

Endpoints

1) Scrape a Website

Endpoint URL

- URL: ``/scrape``
- HTTP Method: **GET**
- Full URL: ``http://localhost:8000/scrape?url=<url>``

Request

- **Parameters:**
 - ``url`` (query parameter): The URL of the website to be scraped.

Response

- **Status codes:**
 - 200 OK: Successful response.
 - 400 Bad Request: Invalid request or URL format.
 - 500 Internal Server Error: Server error during scraping
- **Response Body:**
 - A JSON array containing data for each image on the web page. Each object includes the following information:
 - ``title``: Title associated with the post.
 - ``short_description``: Short description of the post.
 - ``image``: URL of the image.

- **`href`**: URL of the associated post.
- **`date`**: Date associated with the post.
- **`long_description`**: Longer description of the post.
- **`word_count`**: Word count of the post.
- **`sentiment`**: Sentiment analysis result of the text.

Error Responses

- If the request is invalid or there is an error during scraping, the API will respond with a JSON error message containing details.

2) Error Handling

- The API provides error responses with appropriate status codes and error messages to help you identify and handle issues during scraping.

3) Rate Limiting

- The API does not impose rate limiting, but it is important to be respectful when scraping websites to avoid overloading their servers.

Run the application

Prerequisites

Before running the application, make sure you have the required dependencies installed. These include Node.js, Puppeteer, Express.js, CORS middleware and Natural.

Step-by-Step Guide

Clone repository: <https://github.com/CsergezanIstvan/WebScraper.git>

Setup and run server

1. Navigate to the repository folder and inside `Webscraper` to `Back-end`
2. Open a terminal to this location and run: `npm install`
3. To start the server execute command: `npm run start`
 - a. You should see the message: “ Listening on port 8000” indicating that the server is running.

Access the Web Application

1. Navigate to the repository folder and inside `Webscraper` to `Front-end`
2. Open `index.html` from `public` folder.

Using the Application

1. In the web application, you will see an input field labeled “Search”.
2. Enter the URL of the website you want to scrape in the input field.
3. Click the “Search” button/icon to initiate the web scraping process.

View the Results

1. The website will make a request to the `/scrape` endpoint of the local server.
2. The result of the web scraping process will be displayed in the “Results” section of the web application.

Interpreting the Results

The scraped data will be presented in a structured JSON format, including details about images, posts and sentiment analysis.