

# Almeera Scraper Documentation

## Overview

The **Almeera Scraper** class is a web scraper designed to extract product details from the Almeera online store's water category. The code performs the following key operations:

1. Fetches the main page for the water category.
2. Identifies and extracts the list of subcategories.
3. Scrapes information about products within each subcategory.
4. Saves the scraped product data in a JSON file.
5. Downloads product images.

## Requirements

- Python 3.x
- BeautifulSoup4
- requests
- joblib
- tqdm
- json
- os
- multiprocessing

Install with **pip install beautifulsoup4 requests joblib tqdm**

## Key Methods

1. **url\_to\_soup\_obj(url)**: Fetches and parses a web page into a BeautifulSoup object.
2. **download\_image(image\_url, image\_path)**: Downloads an image from a given URL to a local path.
3. **sub\_categorise\_extractor(sub\_category\_url, category\_name)**: Extracts subcategory and product information, returning a dictionary.
4. **process\_category(category\_element)**: Processes a category element, extracting category, subcategory, and product data, and saves it as a JSON file.
5. **main()**: Initiates the scraping process, fetching and processing the main page, extracting category and subcategory details, using multi-processing, and saving results in JSON files.

## Usage

1. Create an instance of the **AlmeeraScraper** class with the main page URL and output paths:  
`class_obj = AlmeeraScraper(main_page_url, output_jsons_path, output_images_path)`
2. Start the scraping process by calling the **main()** method:  
`class_obj.main()`

This will initiate the web scraping process, extracting category, subcategory, and product data from the Almeera online store and saving the results in JSON files.