WEB SCRAPPING TASK

Overview

The provided code is aimed at web scraping product details from the www.acemicromatic.net website. The objective is to extract information about various machining center products. The code is organized into sections, each responsible for specific tasks.

Code Structure

The code is structured into several sections, each serving a distinct purpose:

1. Importing Libraries

This section imports necessary libraries, including requests for making HTTP requests and BeautifulSoup for parsing HTML content.

2. Main Category Scraping

The code starts by downloading the HTML content of the main product category "Milling" from the website.

It uses BeautifulSoup to parse the HTML content.

It extracts a list of machine names from the main category.

3. Subcategory Scraping

For each machine name, the code generates a URL for the corresponding subcategory page.

It downloads and parses the HTML content of each subcategory page.

It extracts a list of machine types within each subcategory.

The code also performs some string manipulation to clean up and standardize the machine type names to be used for the next url.

4. Scraping Product Details

For each machine type, the code generates a URL for the product details page.

It downloads and parses the HTML content of each product details page.

It extracts various product details, such as category, type, and product name and extract X, Y, and Z-axis travel distances

5. Data Storage

The extracted data is stored in a list called main_list. Each item in this list represents a machine type's details.

6. Issues and Improvements

- scrapping incomplete due to host error
- couldn’t save the scraped data to a data frame and export it to excel file/csv file.
- more scope to improvise but couldn’t do due to time constraints.

Thank You

S. Haripriya