# Specs Extraction Tools

## Approach:(21/Aug/2024)

To create a global code that extracts specifications from any trader's website,we have to build a more generalized and flexible system that can handle different HTML structures. Here’s a high-level approach:

### 1. ****Input Website Link****

* Accept a website link from the user.
* Fetch the HTML content of the page using a library like requests or BeautifulSoup.

### 2. ****Parse HTML and Identify Relevant Sections****

* **Generalized Parsing:** Use a parser like BeautifulSoup to parse the HTML and extract different sections like div, table, ul, ol, etc. You can look for keywords like "specs", "details", "features", or similar indicators to locate the relevant sections.
* **Content Extraction:** For each identified section, extract the content and structure it in a readable format (e.g., key-value pairs for specifications).

### 3. ****Custom Extraction Logic****

* **Pattern Matching:** Implement pattern matching to identify specifications, like looking for tr, th, td tags inside tables, or li tags in lists.
* **Class and ID Usage:** If there are common classes or IDs used across trader websites for specifications, you can target those directly.

### 4. ****Use AI for Content Understanding****

* **Natural Language Processing (NLP):** Use NLP techniques or a pre-trained language model to better understand and extract meaningful specifications from unstructured data.
* **Custom Rules:** Implement custom rules that can be adjusted based on the type of website. For instance, trader websites might have certain recurring patterns in how they display product specs.

### 5. ****Output****

* **Display in Streamlit:** Present the extracted specs in a structured format within the Streamlit app.
* **Download Options:** Provide options to download the extracted data as JSON, Excel, or any other preferred format.

This generalized approach allows you to target various trader websites and extract relevant specifications, even if their HTML structures differ.