HTML-to-JSON-Blocks-Library

The HTML to JSON Blocks Library is designed to convert HTML content into a structured JSON block format, making it useful for systems that need to transform web content into processable formats. It is highly customizable, allowing users to define their own image conversion methods through the use of abstract classes.

Key Features

- Converts common HTML tags like paragraphs, headers, lists, and links into structured JSON blocks.
- Offers an abstract class for users to customize image transformation (`ImageTransformer`).
- Modular design that allows for extending and adapting the base functionality.
- Ideal for integration into content management systems or applications that process web content.

JSON Block Structure

- **Headers:** Converted into blocks with title levels (`h1`-`h6`).
- **Paragraphs:** Processes inline content such as text, bold, italics, and links.
- **Lists:** Supports both ordered and unordered lists.
- **Images:** Provides an image block with detailed metadata, customizable via the abstract `ImageTransformer` method.

Technical Description

- The library is built using Python and leverages BeautifulSoup for HTML parsing.
- Key modules and features:
    - **HTML to JSON Conversion:** Core functionality to convert HTML elements into JSON blocks.
    - **Image Transformation:** Abstract class `ImageTransformer` for custom image processing.
    - **Modular Design:** Easily extendable to add new HTML tag processing or customize existing ones.
- The system generates structured JSON blocks that can be easily integrated into other systems for further processing or storage.
- Unit tests are included to ensure the reliability and correctness of the library.

How to Use

1. Install the library from PyPI:

```
pip install html-to-json-blocks
```

2. Alternatively, you can clone the repository and install the dependencies:

```
git clone https://github.com/NivekTakedown/html-to-json-blocks.git
cd html-to-json-blocks
pip install -r requirements.txt
pip install .
```

3. Create an instance of the converter:

```python
from html_to_json_blocks.core import HtmlToJsonConverter
from html_to_json_blocks.transformers.image_transformer import
ImageTransformer

class CustomImageTransformer(ImageTransformer):
    def transform_image(self, img_node, images_info):
        # Your custom logic for transforming images
        pass

converter =
HtmlToJsonConverter(image_transformer=CustomImageTransformer())
```

4. Use the converter to transform HTML into JSON blocks:

```python
from bs4 import BeautifulSoup

html_content = "<p>Example HTML content</p>"
soup = BeautifulSoup(html_content, 'html.parser')
images_info = []  # Image information if needed
json_blocks = converter.convert(soup, images_info)
print(json_blocks)
```

Repository

[View the repository](#)

PyPI Package

The HTML to JSON Blocks Library is also available as a package on PyPI, making it easy to install and integrate into your projects.

[View the PyPI package](#)