Weather Data Extraction Tool

Overview

This script provides a class-based approach to extracting structured and unstructured weather data from web pages. The extracted data is organized into standardized formats suitable for further analysis or input into other systems.

Key Features

- **Dynamic Web Scraping:** Extracts text and tabular weather data from web pages using predefined XPath queries.
- **Data Structuring:** Converts scraped data into structured formats like JSON and pandas DataFrames.
- **Standardization:** Outputs data in a consistent format, including short reports, detailed descriptions, and tabular weather information.

Approach

1. **Initialization:**
    - The class accepts a weather page URL and a timezone flag.
    - Sets up predefined headers and XPath queries for data extraction.
2. **Data Extraction:**
    - Fetches the web page content using `requests` and parses it with `lxml`.
    - Extracts:
        - Text-based details (e.g., weather descriptions, sunrise, sundown).
        - Tabular data (e.g., temperature, rainfall, wind conditions).
3. **Data Storage:**
    - Organizes extracted data into dictionaries, DataFrames, and JSON files.
    - Provides a standardized JSON output for downstream use.
4. **Saving Data:**
    - Saves extracted data in a raw JSON file and a standardized JSON format for consistent representation.

Usage

```
url = "https://www.wetter.com/schweden/stockholm/SE0ST0012.html"
extractor = WeatherDataExtractor(url)  # Create an instance of the extractor
extractor.save_data_to_json("Abuja.json")  # Save the extracted data to a file
```