

Web Scrapping using Python

-Abithan P

Overview

This project is a simple Python-based news scraper that fetches the latest news articles from specified news websites and saves them into text files. It allows users to choose between fetching news in English or Tamil and from different news sources.

Features

- Fetches news from multiple sources.
- Supports both English and Tamil languages.
- Saves news articles to text files.
- Modular functions for easy maintenance and updates.

Setup Instructions

Prerequisites

- Python 3.x
- Required Python packages: **requests**, **beautifulsoup4**

Installation

Install Required Packages:

```
pip install requests beautifulsoup4
```

Usage Guide

1. Run the Script:

```
C:/Users/Abithan/AppData/Local/Programs/Python/Python311/python.exe  
d:/Inteship/Project.py
```

2. Select Language:

- Enter **1** for Tamil.
- Enter **2** for English.

3. Select News Source:

- If Tamil is selected:
 - Enter **1** for Thinamalar.
 - Enter **2** for Thinathandhi.
- If English is selected:
 - Enter **1** for Indian Express.
 - Enter **2** for Timesnow.

4. The script will fetch the latest news articles from the selected source and save them to a text file with a name corresponding to the source.

Function Details

fetch_news

```
def fetch_news(url, article_class, title_class=None, filename="news.txt", header=None):
```

Fetches news articles from the specified URL and writes them to a file.

Arguments:

- **url (str)**: The URL to fetch news from.
- **article_class (str)**: The class name of the HTML element containing articles.
- **title_class (str)**: The class name of the title element to be removed, if any.
- **filename (str)**: The name of the file to write the news to.
- **header (str)**: The header text to be written to the file.

Indianexpress

```
def indianexpress():
```

Fetches news from the Indian Express website and saves it to **Indian_Express.txt**.

Timesnow

```
def timesnow():
```

Fetches news from the Timesnow website and saves it to **Timesnow_News.txt**.

Dhinamalar

```
def thinamalar():
```

Fetches news from the Thinamalar website and saves it to **Thinamalar_News.txt**.

Thinathandhi

```
def thinathandhi():
```

Fetches news from the Thinathandhi website and saves it to **Thinathandhi_News.txt**.

main

```
def main():
```

The main function that drives the script. It prompts the user to select a language and news source, then calls the appropriate function to fetch and save news.

Example

1. Tamil

2. English

Enter number to select language: 1

1. Thinamalar

2. Thinathandhi

Enter number to select Article: 1

Fetching news from Thinamalar...

File 'Thinamalar_News.txt' saved successfully.