Global News Topic Tracker

Overview

The Global News Topic Tracker is a Streamlit application that collects trending stories from Google News RSS feeds, organizes them into topic clusters, and produces clear summaries.
Users can explore news by country editions, topics, or keyword searches, and generate concise summaries using either a local summarizer or OpenAI models.

Features

- Country-specific Google News editions (e.g., India, United States, United Kingdom).
- Option to select topics such as World, Business, Technology, Sports, Health, etc.
- Keyword-based search to override default topics.
- Grouping of similar articles into clusters for easier understanding.
- Recency-based scoring to prioritize fresh news.

- Summarization using:

- Local Hugging Face summarizer (runs without internet once installed).
- OpenAI models (requires API key).
- Downloadable summaries in Markdown format.
- Expandable article lists for each topic cluster.

Requirements
- Python 3.9 or later

- Required libraries:

- Streamlit
- Feedparser
- Pandas

- Optional libraries:

- Transformers (for local summarizer)
- OpenAI (for GPT-based summarizer)

An OpenAI API key is needed if the OpenAI summarizer option is used.

How It Works

Input Settings
- Choose a country edition.
- Select a topic or provide a keyword search.
- Set the maximum number of articles to fetch.
- Choose between a local summarizer or OpenAI summarizer.

Fetching News
- Builds the correct Google News RSS feed URL.
- Downloads articles and extracts titles, summaries, sources, links, and publication times.

Clustering
- Headlines are cleaned and normalized.
- Articles are grouped into clusters based on keyword similarity.
- Each cluster is scored using both article count and recency.

Summarization
- Each cluster is summarized into a single clear paragraph.
- The summarization can be performed either locally (using a transformer model) or via OpenAI's API.

Display
- Clusters are displayed with a representative headline.
- Each cluster shows a one-paragraph summary.
- Users can expand clusters to view all related stories with publication times and sources.
- A table view of all raw articles is also available.

Export
- Summaries are combined into a Markdown document.
- A download button allows users to save summaries for later use.

Application Flow
1. User selects country, topic, and summarizer settings.
2. Application fetches articles from Google News RSS feeds.
3. Articles are grouped into clusters of related stories.
4. Each cluster is summarized.
5. Summaries and full articles are displayed in the interface.
6. User can download summaries as a Markdown file.

Example Usage
- Select the United States edition and the Technology topic.
- The application fetches the latest technology-related articles.
- Articles are grouped into clusters such as artificial intelligence, semiconductor industry, or major tech company news.
- Summaries are generated for each cluster.
- User reviews clusters, expands to read full lists of sources, and downloads summaries for later reading.

Future Enhancements
- Use semantic embeddings for more accurate clustering.
- Add visualizations showing topic trends over time.
- Improve multilingual support for non-English editions.
- Provide alternative summarization models for more control over length and style.