

CS-480 Process Book

ACTUALIA

DASHBOARD

newsmaster-dashboard.netlify.app

Prepared by
NewsMaster team

Florian Déjean
Eugène Bergeron
Paul Madelénat

THE IDEA

In an age of overwhelming information, understanding what truly matters globally has become a challenge. Headlines shift by the second, and the sheer volume of news makes it difficult to identify the most important conversations shaping our world.

Our mission is to change that. We are designing a set of **intuitive, interactive visualizations** that allow anyone, regardless of background or expertise, to instantly grasp the most significant global topics and how they are interconnected.

This is not just another news tool. It's a **global pulse**, made visual.

0. A few definitions	4
1. The principle	4
2. The dataset & the pipeline	4
3. The website infrastructure	5
4. The Network.....	5
5. The World Map	6
6. The Temporal Trends	7
7. Telling a Story	8
8. Work distribution	8

0. A few definitions

- **Article:** A news story or report published by a media in their website
- **Provider:** The source or organization that publishes the article (e.g., BBC, CNN, Le Monde).
- **NER (Named Entity Recognition):** A technology that automatically finds and identifies important names in a text, such as people, places, or organizations (e.g., "Joe Biden", "Paris", "Apple Inc.").

1. The principle

Our work is organized into several detailed visualizations (see below), brought together into a single, cohesive "**story**" that helps make sense of current news. One of the most important things to understand is that **every day** at 4 a.m., we automatically add articles from the previous day, and **the visualizations update** and adapt accordingly. On this first page of the story, we highlight why **XXX** is at the forefront of today's news — **XXX** being the most frequently mentioned named entity in the day's articles. More in the **6. Telling A Story** section.

2. The dataset & the pipeline

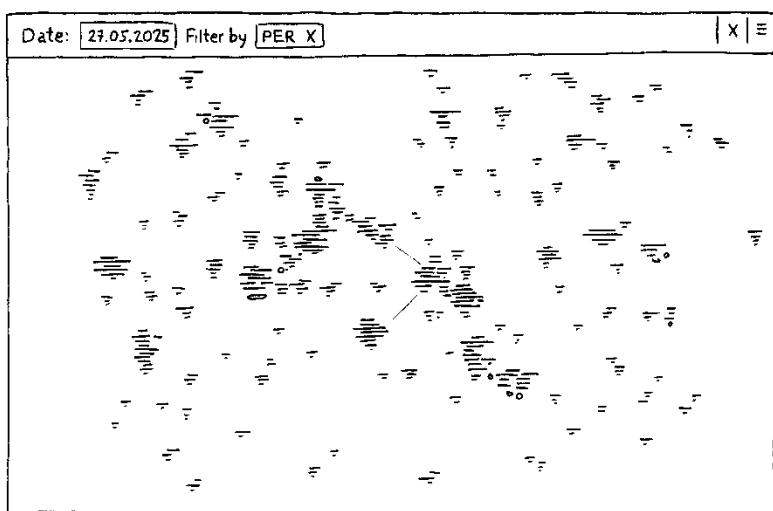
To gather our dataset, we are using a single **GitHub Action**. This action fetches articles from all providers via **RSS feeds** and processes each article daily. Our processing includes translation using Google's service, NER using a Distil BERT, entity normalization, and finally formatting for each visualization to ensure efficiency. This enables each visualization to load a specific, optimized **JSON file**. This pipeline is entirely free. GitHub Action allows for around 1 hour of run-time per day.

The main challenges we encountered are the NER normalization, for example some articles mentioned simply "Trump" instead of full name "Donald Trump". In the scope of the class, we opted for a "sub-set" method to focus our time on visualization. The other challenge was gathering enough providers to ensure world coverage. We use Claude 3.7 using a feedback loop with the objective of having a provider for each country. For feedback, we used a script to check current provider coverage by country and identify dead links. Another script added new providers and verified if the URLs were valid. However, a pitfall of this approach was that Claude was using web search using English term instead of the country language, it managed to reach a near compete worldwide coverage.

3. The website infrastructure

We opted to use a static site rendering technology to avoid the need for backend infrastructure management. The data pipeline we implemented places a new static JSON file in the `/public/data/` directory every day (alongside additional JSON files depending on the specific visualizations). These files are then directly loaded by the client's browser. This approach eliminates the need for a database or server-side logic, which significantly reduces costs and simplifies maintenance. We host via Netlify, but it's compatible with any static website provider. We'll migrate to Cloudflare Pages at the end of the project.

4. The Network



same articles. Users can filter the network by entity type through a dropdown menu, allowing for focused analysis of specific categories.

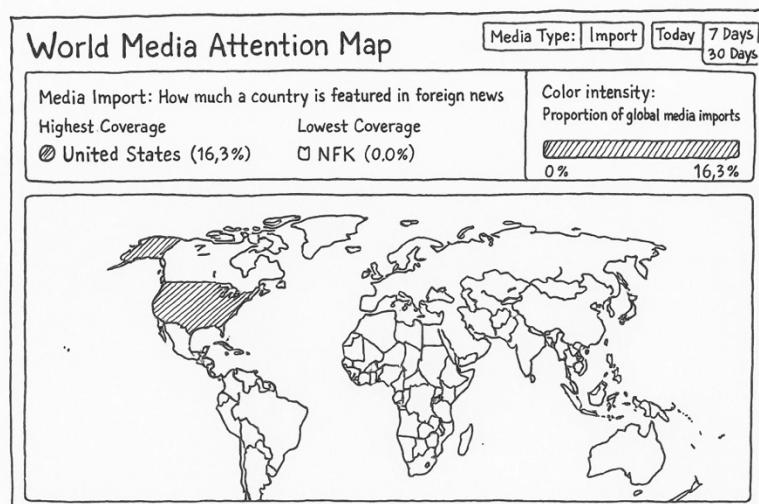
Originally, the idea was to create a 2D representation. However, due to the overwhelming volume of data, the network quickly became unreadable. Introducing a third dimension made more sense, as it allowed for clearer visualization of the differences and similarities between named entities (NERs).

This decision brought several significant challenges. First, it became crucial to simplify the user experience as much as possible, since this type of 3D visualization is inherently more complex to grasp. The goal was to ensure that users could still navigate and interpret the network easily despite the added dimensionality. From a technical standpoint, we also had to learn and implement WebGL, the 3D rendering engine, alongside React Graph, which added another layer of complexity. Finally, one

The **Article Keyword Network** visualization offers an interactive exploration of relationships among named entities, such as people, organizations, and locations, co-mentioned in news articles. Using a force-directed **3D graph** layout, each **node represents an entity**, while the connecting **edges indicate their co-occurrence within the same articles**. Users can filter the network by entity type through a dropdown menu, allowing for focused analysis of specific categories.

major hurdle was managing the heavy computational load of the visualization, especially when loading all articles at once. To address this, we allowed users to select specific data providers, offering a curated default list while giving them full control over what content to load, thus optimizing performance and usability.

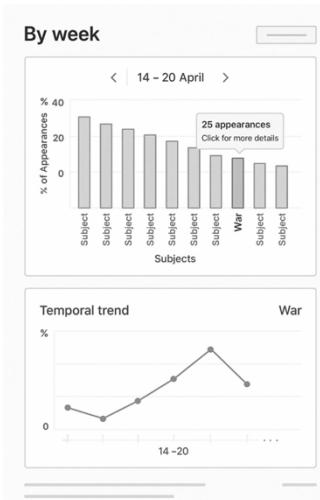
5. The World Map



the color gradient for proportional country size, but oversized countries ballooned even more, and the look was anything but elegant (although it was a fun experiment). To clear up the import-export confusion, we switched to simpler labels: "featured" and "covered", and let users click on a country to see who features or covers it. Finally, we added two charts beneath the map to add context: one shows which Entity the selected country talks about, and the other tracks how its featured-versus-covered balance changes over time.

Our **World Map** gave us a few headaches. Once we'd wired the schema into a web graph and started gathering feedback, two problems jumped out. First, the notions of "import" and "export" felt difficult to grasp. Second, tiny nations disappeared in the visualization, no matter how bold the contrast. We experimented with swapping

6. The Temporal Trends



This sketch was our initial concept for visualizing temporal trends, created prior to implementation. Following the first prototype, we identified two key issues:

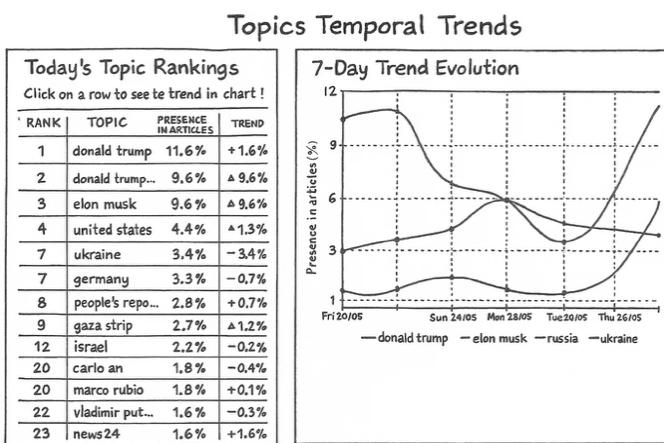
- 1. Inconsistent Time Scale:** The original week-based time scale did not align with the overall narrative and structure of the website. A day-based scale was more appropriate for consistency and granularity.
- 2. Limited Comparability:** While the bar plot provided a basic view, it was not well-suited for analyzing how topics evolved over time and comparing them effectively.

	LAST	CHANGE	CHANGE %	VOLUME	AVERAGE VOL	MARKET CAP
AAPL	215.61	0.14	0.10%	14.543M	6.986M	\$746.1B
COMP INDEX	6134.59	28.58	0.47%	528.3M	1.37M	na
TTWO	73.25	1.26	1.75%	8.837M	77.411M	\$7.5B
FL	50.52	-0.22	-0.45%	6.19M	496.4K	\$6.8B
MSGN	22.15	-0.25	-1.12%	35.22M	1.647M	\$1.7B

To address these challenges, we drew inspiration from **stock market charts** and tables, treating topics as stocks. This approach enabled a more effective temporal visualization and facilitated clearer comparisons across topics.

Final Version

The final version of the visualization consists of two parts:



- The Table:** it uses the native table component of HTML; the CSS was refined using AI to achieve a visual identity to match the artistic direction of the website. Most of the work was put into the pipeline to compute the table's data. This was done using basic JS, then all rendered using React.

- **The Chart:** rendered using *recharts* (based on React and D3), it is an interactive plot that listens to selected components of the table to add or remove curves from the chart. We chose to display only seven days for clarity, while providing an informative view of the relationship between the topics over time.

7. Telling a Story

Our mentioned visualizations were great at providing an overview of our data, but we decided to make a dedicated page to combine their strengths in the Home page. This page focuses on the most important entity in today's news.

First, we feature the world map showing which countries speak the most about this entity, next we temporal trend give us a good idea of how popular this entity is, by how much and how long it's been here. Finally, the network established related relationships and their weights.

8. Work distribution

- **Florian Déjean**
 - World Map visualization
 - Data collection and processing
 - Homepage design
- **Eugène Bergeron**
 - Temporal trends visualization
 - Data processing for trends
 - Homepage implementation
- **Paul Madelénat**
 - Website infrastructure
 - 3d Network visualization
 - 2d Network visualization in Homepage