

Empowering Real-Time Media Research With NewsCatcher API

Tobias Jovall Wessel, Bachelor's in Information science, UIB, tobias.wessel@student.uib.no
Khadiga Seddik, Erik Knudsen, Christoph Trattner

Media Futures

</newscatcher>

INPUT:

```
# Example of advanced search parameters

params = {
  'q': 'technology OR finance', # Search for technology/finance-related articles
  'lang': 'no', # Only Norwegian articles
  'countries': 'NO', # Limit results to Norway
  'sort_by': 'relevancy', # Sort results by relevancy
  'page': 1, # Start at page 1
  'page_size': 100 # Retrieve 100 articles per page
}
```

OUTPUT:

```
Article 16:
Title: Flatt på Wall Street – Trump-aksjen faller
Author: Kristian Aaser
Published Date: 2024-10-30 13:30:54
Content: Etter to kruttsterke dager faller Trump-aksjen tungt.
Karl Biehl
Saken oppdateres ...
Slik ser det ut for de toneangivende indeksene på Wall Street rundt klokken 19:45 norsk tid:
Dow Jones er opp 0,08...
Source URL: https://e24.no/boers-og-finans/i/0ozda3/flatt-paa-wall-street-trump-aksjen-faller
```

Abstract

In the news and media landscape, the timeliness of an article is essential for driving user engagement. Although topical relevance often plays a larger role in attracting readers, recency remains crucial for most news consumers. Media outlets strive to be the first to deliver breaking news, and this demand for up-to-date content became evident in also our work. Previous prototypes and research efforts using mock or outdated articles had received negative feedback, prompting us to seek a more permanent solution to this challenge.

Research question

1. How can we make future prototypes more realistic to further enhance user engagement?
2. How can we access large amounts of real-time news articles from a diverse range of sources worldwide?

Method

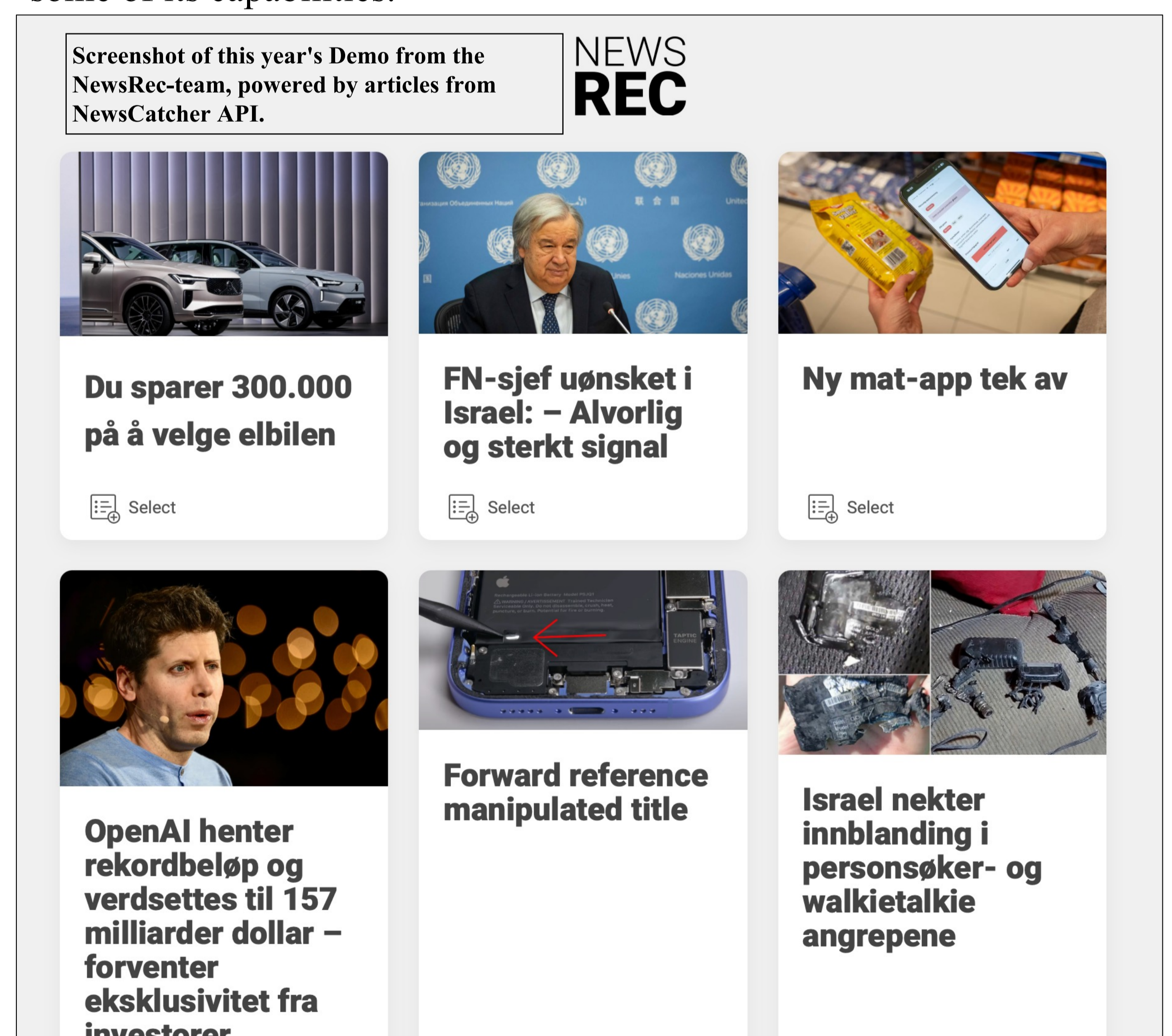
Our research considered various solutions, each with its own challenges. One option was to implement web scraping, but as websites increasingly restrict access to their data, scraping has become more complex and resource-intensive. Fortunately, the NewsCatcher API provides a robust solution, collecting over 100,000 unique news articles daily from 75,000+ sources worldwide, with a latency of less than five minutes. This service includes extensive metadata, such as images, titles, content, dates, authors, and even sentiment analysis.

Current results

Using NewsCatcher API, we have been able to develop a prototype of an automated database on AWS (Amazon Web Services) that fetches and stores up-to-date articles, with old content removed automatically based on expiration settings. Articles are fetched based on parameters tailored to our research requirements, as shown in the example query.

Goal

We anticipate that the NewsCatcher API will be a great asset for future research. Current projects include developing a live-news recommender system adaptable to various research needs, particularly focusing on compliance with the new EU AI Act. The API's custom query parameters, vast data coverage, and low latency significantly enhance our capability to realize this vision. Additionally, NewsCatcher API will be an important resource to the NewsRec project group at MediaFutures, with ongoing research by PhD candidate Khadiga Seddik showcasing some of its capabilities.



PARTNERS



HOST



FUNDED BY

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

