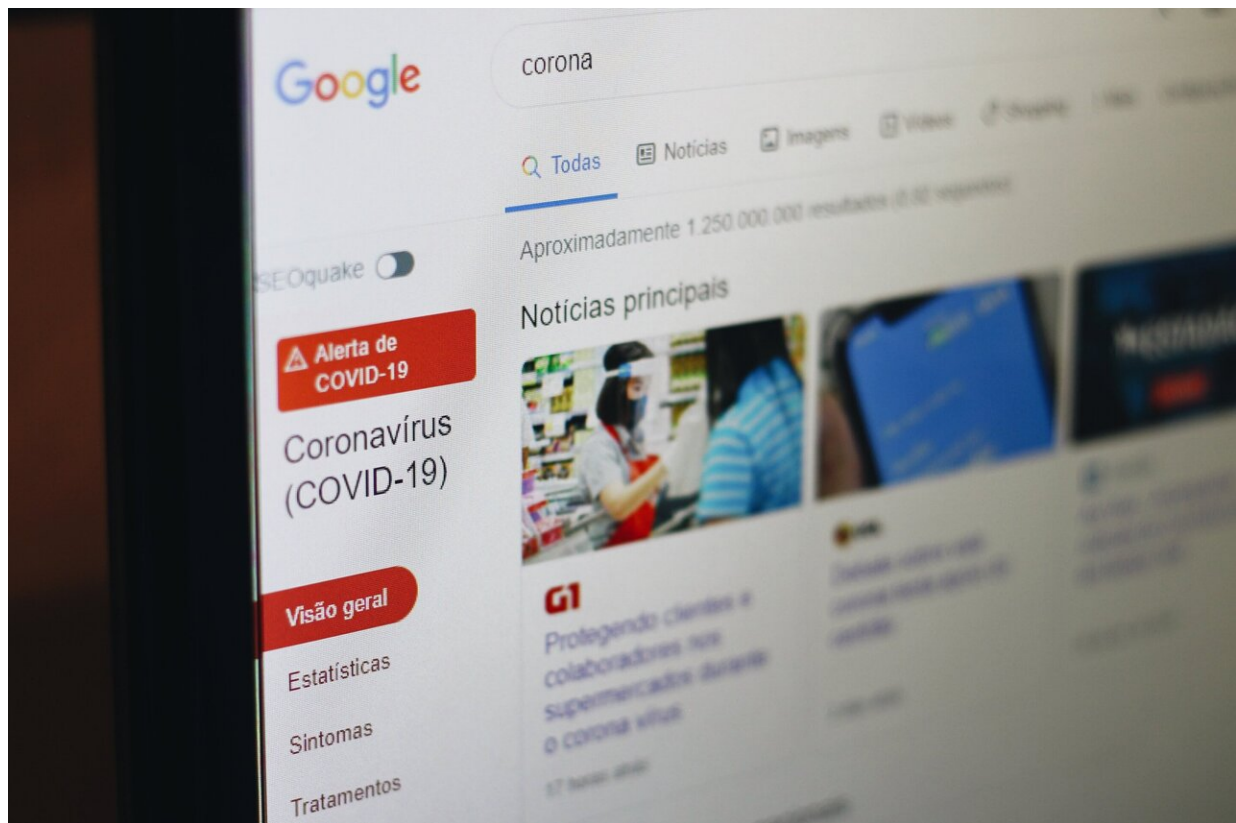


Modular newsgathering makes content aggregation faster, more accurate

April 12 2021, by David Bradley



Credit: Unsplash/CC0 Public Domain

The use of small processing modules can significantly reduce overheads on computing systems with limited resources available to them when large amounts of data must nevertheless be processed. Research by a

team in Greece described in the *International Journal of Web Engineering and Technology* shows how that approach can be used for content aggregation, information extraction, sentiment tagging, and visualization tasks.

Iraklis Varlamis and Dimitrios Michail of the Department of Informatics and Telematics at Harokopio University of Athens and Pavlos Polydoros and Panagiotis Tsantilas of Palo Ltd in Kokkoni, Greece, have demonstrated how this modular approach might function well on the social media and news analytics platform, PaloAnalytics. The team shows how their proposed architecture can easily withstand the pressures of increased content load when an issue goes viral on social media, such as when a major event takes place. The micro-modules that replace the monolithic architecture of conventional data-processing systems can quickly release unused resources when the content load reaches its normal flow.

The researchers point out that even from the early days of primitive web crawlers that became the foundation of search engines and other related tools, it was recognized that distributed processing is the only viable way to taming the vast quantities of textual data being generated even way back then. Today, the scale is almost unimaginable with many petabytes of data to be assimilated, aggregated, processed, indexed, and annotated with meaning. The vast realms of the web and [social media](#) systems offer us a rich seam to be tapped for information and knowledge if the tools can be built to cope with the bits and bytes.

The team's tests so far were based on analysis of 1500 websites, 10000 blogs, forums, hundreds of thousands of public Facebook pages, Instagram, Twitter, and YouTube updates, across six European nations and in six different languages. Their work shows where improvement might be made to build a powerful analytical tool that would be scalable and allow us to soon mine those enormous knowledge seams efficiently

and in an effective way.

More information: Iraklis Varlamis et al. A distributed architecture for large scale news and social media processing, *International Journal of Web Engineering and Technology* (2021). [DOI: 10.1504/IJWET.2020.114029](https://doi.org/10.1504/IJWET.2020.114029)

Provided by Inderscience

Citation: Modular newsgathering makes content aggregation faster, more accurate (2021, April 12) retrieved 18 April 2024 from <https://techxplore.com/news/2021-04-modular-newsgathering-content-aggregation-faster.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.