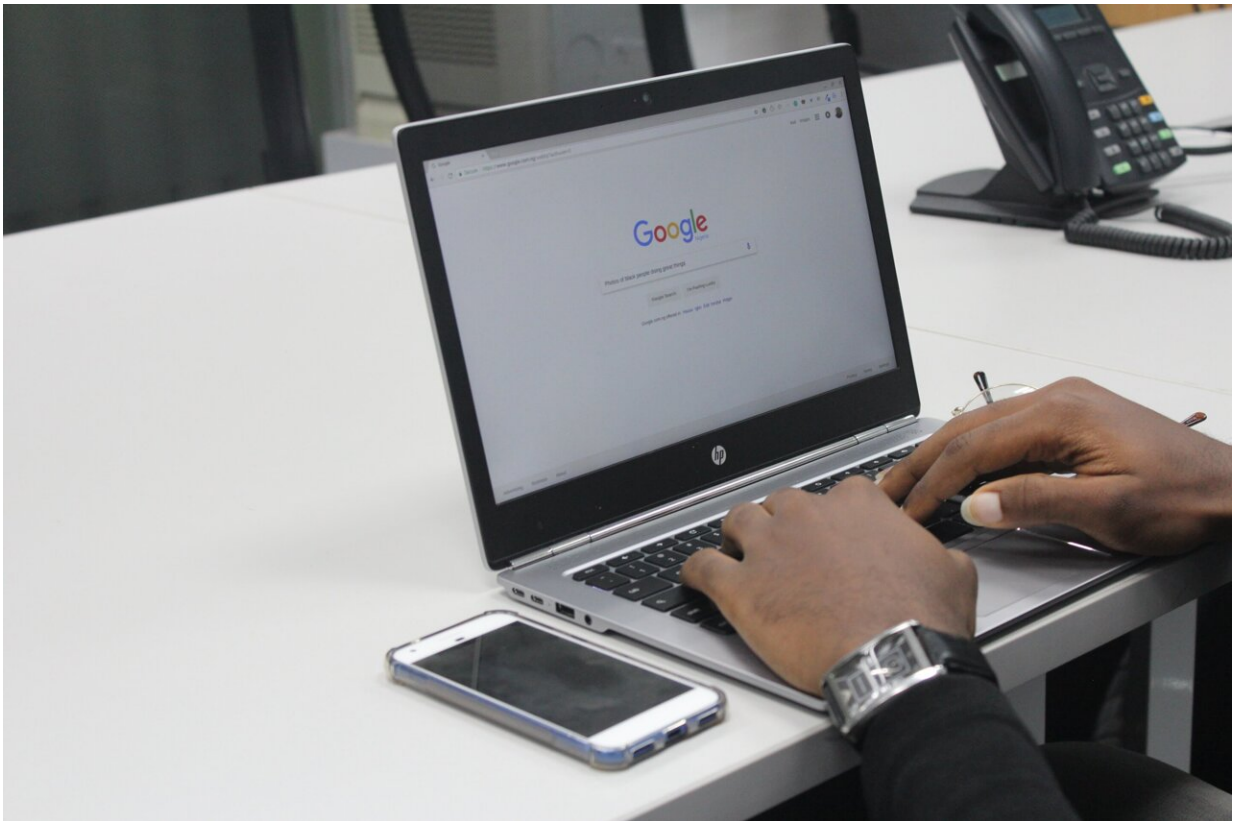


# New technique boosts online medical search results

September 6 2022

---



Credit: Unsplash/CC0 Public Domain

When looking for medical information on the internet, having the precise terminology makes the search fairly straightforward.

But what if the person doing the searching doesn't know the exact [terminology](#), or wants to see what other information may be available without using technical terms? Will internet queries yield any useful results—or worse, will they produce incomplete or downright incorrect information?

A Cornell-led group of researchers has developed a search method that employs [natural language processing](#) and [network analysis](#) to identify terms that are semantically similar to those for cancer screening tests, but in colloquial language.

"If the traditional way of searching for information is by using those official names or concepts, then it will lead to some bias in identifying the content because many people on the internet aren't familiar with official medical vocabularies," said Chau Tong, a postdoctoral associate in the Department of Communication, in the College of Agriculture and Life Sciences.

Tong is lead author of "Search Term Identification Methods for Computational Health Communication: Word Embedding and Network Approach for Health Content on YouTube," which published Aug. 30 in the open-access journal *JMIR Medical Informatics*.

Drew Margolin, associate professor of communication, is the paper's senior author. Also contributing was Jeff Niederdeppe, professor of communication; Teairah Taylor, doctoral student in the field of communication; Andy J. King, associate professor at the University of Utah; Rumi Chunara, associate professor of global public health, computer science and engineering at New York University; and Natalie Dunbar, graduate student at Iowa State University.

This research stemmed from a four-year National Institutes of Health grant that Margolin, Niederdeppe and King received in March 2021 to

work on ways to monitor and evaluate public information and communication disparities regarding screening for colorectal cancer (CRC). Tong is a member of Niederdeppe's research lab.

The disease disproportionately affects African Americans; according to a 2019 study by the American Association for Cancer Research, the overall CRC mortality rate in the U.S. was 14.8 deaths per 100,000 people, but the rate was 20.9 per 100,000 for Black people and 14.7 per 100,000 for white people.

"The question we asked with the grant was, 'Are there messages or aspects of social media that can be used to increase information, increase access, increase screening rates—something that would kind of [have] helped to equal that out?'" Margolin said.

Margolin's group chose YouTube—which more than 80% of Americans use at least sporadically—as the [platform](#) for their study. Starting by searching off "colonoscopy," the group retrieved a set of 250 videos. They then employed word embedding—using neural network modeling to identify words that appear in similar contexts to the main term—to come up with an additional 4,304 related videos.

The group found that colon prep brand names (Miralax, Suprep, Plenu) were often found in user-generated content, where the word "colonoscopy" may not have been used.

"These findings," Tong said, "highlight the value of innovative, data-informed research strategies that can expand the conventional data-collection and analysis pipelines, to cover a range of user-generated health content. This can uncover information disparities that could negatively impact important health equity outcomes."

The group did similar searches using seed terms "FOBT" (fecal occult

blood test, another colon cancer screen), "mammogram" and "pap smear," the latter two being screens for breast and cervical cancer, respectively. They found similar results to the colonoscopy searches, retrieving a range of new videos using words that were semantically close to the seed term.

Margolin said the group's goal is to adapt this technique in platforms other than YouTube, which suggests related relevant videos based on user behavior, making it more likely that a user will find useful content after an initial search.

Margolin thinks computational health researchers should think about this alternative search protocol.

"We don't need to do computational research on YouTube to find out what hospitals have to say about colonoscopy," he said. "The whole purpose of this is to find out what someone who's not certified to talk about colonoscopy will say. For example, a random person is telling you about what happened when they did their 'prep' (for a colonoscopy), but maybe they didn't use the word 'colonoscopy.'"

"They're telling a story," he said. "Now you're getting what social media can reveal."

**More information:** Chau Tong et al, Search Term Identification Methods for Computational Health Communication: Word Embedding and Network Approach for Health Content on YouTube, *JMIR Medical Informatics* (2022). [DOI: 10.2196/37862](https://doi.org/10.2196/37862)

Provided by Cornell University

Citation: New technique boosts online medical search results (2022, September 6) retrieved 26 June 2024 from

<https://medicalxpress.com/news/2022-09-technique-boosts-online-medical-results.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.