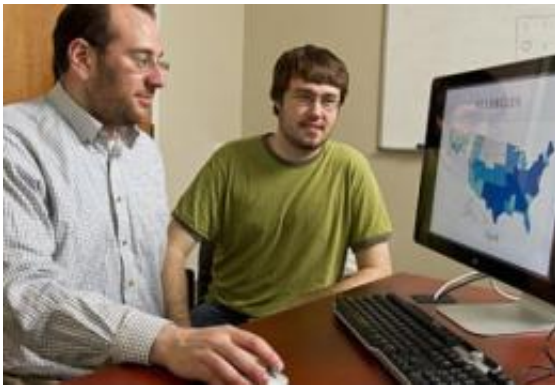


# You are what you tweet: Tracking public health trends with Twitter

July 6 2011

---



Johns Hopkins computer scientists Mark Dredze, left, and Michael J. Paul, found that Twitter posts could yield useful public health information. Credit: Will Kirk/JHU

Twitter allows millions of social media fans to comment in 140 characters or less on just about anything: an actor's outlandish behavior, an earthquake's tragic toll or the great taste of a grilled cheese sandwich.

But by sifting through this busy flood of banter, is it possible to also track important public health trends? Two Johns Hopkins University [computer scientists](#) would respond with a one-word tweet: "Yes!"

Mark Dredze and Michael J. Paul fed 2 billion public tweets posted between May 2009 and October 2010 into computers, then used software to filter out the 1.5 million messages that referred to health

matters. Identities of the tweeters were not collected by Dredze, a researcher at the university's Human Language Technology Center of Excellence and an assistant research professor of computer science, and Paul, a doctoral student.

"Our goal was to find out whether Twitter posts could be a useful source of public health information," Dredze said. "We determined that indeed, they could. In some cases, we probably learned some things that even the tweeters' doctors were not aware of, like which over-the-counter medicines the posters were using to treat their symptoms at home."

By sorting these health-related tweets into electronic "piles," Dredze and Paul uncovered intriguing patterns about allergies, [flu cases](#), insomnia, cancer, obesity, depression, pain and other ailments.

"There have been some narrow studies using Twitter posts, for example, to track the flu," Dredze said. "But to our knowledge, no one has ever used tweets to look at as many health issues as we did."

Dredze and Paul, who also are affiliated with the university's Center for Language and [Speech Processing](#), have discussed some of their results in recent months at computer science conferences. They will present their complete study on July 18 in Barcelona, Spain, at the International Conference on Weblogs and Social Media, sponsored by the Association for the Advancement of Artificial Intelligence.

In addition to finding a range of health ailments in Twitter posts, the researchers were able to record many of the medications that ill tweeters consumed, thanks to posts such as: "Had to pop a Benadryl...allergies are the worst."

Other tweets pointed to misuse of medicine. "We found that some people tweeted that they were taking antibiotics for the flu," Paul said.

"But antibiotics don't work on the flu, which is a virus, and this practice could contribute to the growing antibiotic resistance problems. So these tweets showed us that some serious medical misperceptions exist out there."

Of course, the vast majority of daily tweets have nothing to do with an illness. While a simple approach would be to filter for words that are tied to illness, such as "headache" or "fever," this strategy fails on such tweets as "High price of gas is a headache for my business" or "Got a case of Bieber Fever. Love his new song."

To find the health-related posts among the billions of messages in their original pool, the Johns Hopkins researchers applied a filtering and categorization system they devised. With this tool, computers can be taught to disregard phrases that do not really relate to one's health, even though they contain a word commonly used in a health context.

Once the unrelated tweets were removed, the remaining results provided some surprising findings.

"When we started, I didn't even know if people talked about allergies on Twitter," Paul said. "But we found out that they do. And there was one thing I didn't expect: The system found two different types of allergies: the type that causes sniffing and sneezing and the kind that causes skin rashes and hives."

In about 200,000 of the health-related tweets, the researchers were able to draw on user-provided public information to identify the geographic state from which the message was sent. That allowed them to track some trends by time and place, such as when the allergy and flu seasons peaked in various parts of the country. "We were able to see from the tweets that the allergy season started earlier in the warmer states and later in the Midwest and the Northeast," Dredze said.

Dredze and Paul have already begun talking to public health scientists, including some affiliated with Johns Hopkins, who say that future studies of tweets could uncover even more useful data, not only about posters' medical problems but also about public perceptions concerning illnesses, medications and other health issues.

Still, Dredze and Paul cautioned that trying to take the nation's temperature by analyzing [tweets](#) has its limitations. For one thing, most Twitter users did not comment more than once on their particular ailment, making it tough to track how long the illness lasted and whether it recurred. In addition, most Twitter users tend to be young, which would exclude many senior citizens from a public health study. Also, at the moment, Twitter is dominated by users who are in the United States, making it less useful for research in other countries.

Although social media sites allow users to expose lots of personal information to friends and strangers, Twitter-based research may only reach a certain depth.

"In our study," Paul said, "we could only learn what people were willing to share. We think there's a limit to what people are willing to share on Twitter."

Nevertheless, Dredze says there is still plenty of useful data left to plumb from Twitter posts. "The people I've talked to have felt this is a really interesting research tool," he said, "and they have some great ideas about what they'd like to learn next from Twitter."

Provided by Johns Hopkins University

Citation: You are what you tweet: Tracking public health trends with Twitter (2011, July 6) retrieved 6 May 2024 from

<https://phys.org/news/2011-07-tweet-tracking-health-trends-twitter.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.