The New York Times | http://nyti.ms/1WDGCrL

TECHNOLOGY

Microsoft Finds Cancer Clues in Search Queries

By JOHN MARKOFF   JUNE 7, 2016

Microsoft scientists have demonstrated that by analyzing large samples of search engine queries they may in some cases be able to identify internet users who are suffering from pancreatic cancer, even before they have received a diagnosis of the disease.

The scientists said they hoped their work could lead to early detection of cancer. Their study was published on Tuesday in The Journal of Oncology Practice by Dr. Eric Horvitz and Dr. Ryen White, the Microsoft researchers, and John Paparrizos, a Columbia University graduate student.

"We asked ourselves, 'If we heard the whispers of people online, would it provide strong evidence or a clue that something's going on?'" Dr. Horvitz said.

The researchers focused on searches conducted on Bing, Microsoft's search engine, that indicated someone had been diagnosed with pancreatic cancer. From there, they worked backward, looking for earlier queries that could have shown that the Bing user was experiencing symptoms before the diagnosis. Those early searches, they believe, can be warning flags.

While five-year survival rates for pancreatic cancer are extremely low, early detection of the disease can prolong life in a very small percentage of cases. The study suggests that early screening can increase the five-year survival rate of pancreatic patients to 5 to 7 percent, from just 3 percent.

The researchers reported that they could identify from 5 to 15 percent of pancreatic cases with false positive rates of as low as one in 100,000. The researchers noted that false positives could lead to raised medical costs or create significant anxiety for people who later found out they were not sick.

The data used by the researchers was anonymized, meaning it did not carry identifying markers like a user name, so the individuals conducting the searches could not be contacted.

A logical next step would be to figure out what to do with that search information. One possibility would be some sort of health service where users could allow their searches to be collected, allowing scientists to monitor for questions that indicate warning flag symptoms.

"The question, 'What might we do? Might there be a Cortana for health some day?'" said Dr. Horvitz, in a reference to the company's speech-oriented online personal assistant software service.

Although the researchers declined to offer specific details, Dr. White is now the chief technology officer of health intelligence in a recently created Health & Wellness division at Microsoft.

They acknowledged that health-related data generated from web search histories was still new territory for the medical profession.

"I think the mainstream medical literature has been resistant to these kinds of studies and this kind of data," Dr. Horvitz said. "We're hoping that this stimulates quite a bit of interesting conversation."

The new research is based on the ability of the Microsoft team to accurately distinguish between web searches that are casual or based on anxiety and those that

are genuine searches for specific medical symptoms by people who are experiencing them, he noted.

Both a computer scientist and a medical doctor by training, Dr. Horvitz said he had been exploring this area in part because of a phone conversation with a close friend who had described symptoms. Based on their conversation, Dr. Horvitz advised him to contact his doctor. He received a diagnosis of pancreatic cancer and died several months later.

The availability of vast sets of behavior data based on individual web queries using the search engines offered by companies like Google and Microsoft has for a number of years been seen as a potential indicator of health-related information.

In 2009, Google published a research paper that explored the potential of early detection of flu epidemics based on statistical analysis of web search logs, though the results of that effort ultimately fell short of what had been hoped.

More recently, Microsoft researchers have had significant success in finding early evidence of adverse drug reactions from patterns observed in web logs. In 2013, they detected unreported side effects of prescription drugs before they were found by the Food and Drug Administration's warning system.

The researchers are exploring evidence related to a range of devastating diseases. They also said that unlike the drug interaction data, which would be of direct value to the F.D.A. as an early alert, it was possible that symptom alert data might be made available as part of a broader online health service that a company like Microsoft might offer.

A version of this article appears in print on June 8, 2016, on page B1 of the New York edition with the headline: Microsoft Sees Cancer Tip-Offs in Searches.