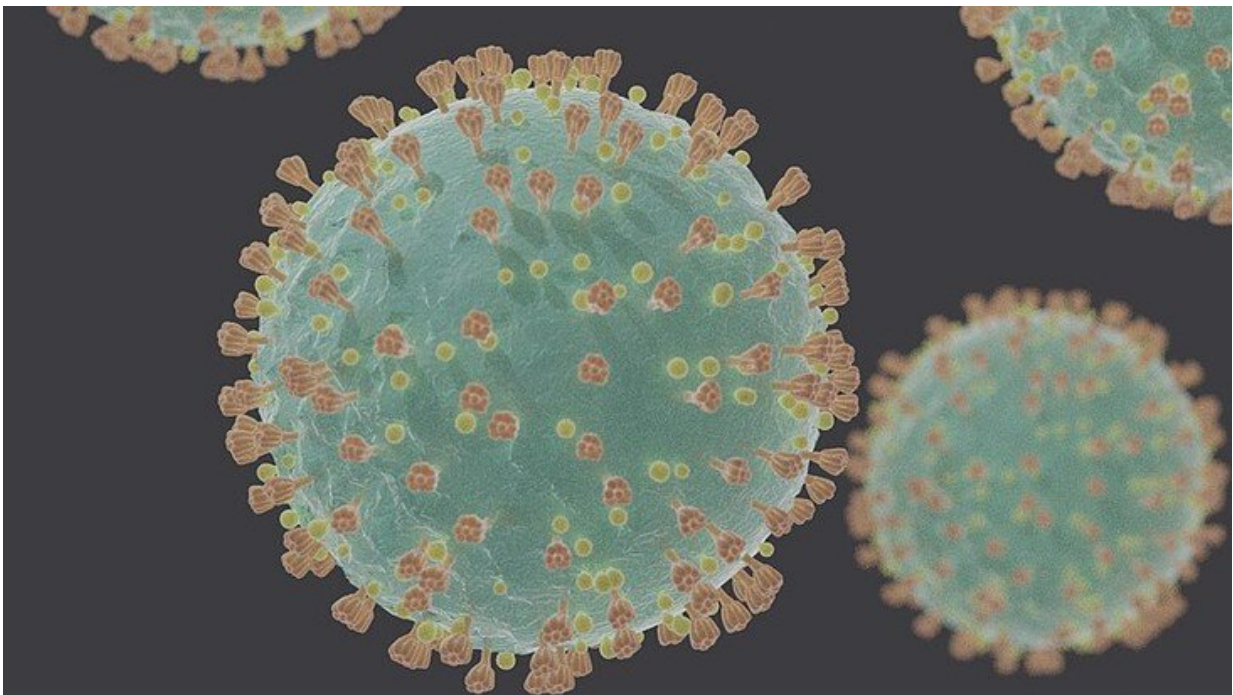# Data scientists say tools may offer COVID-19 researchers faster, more in-depth answers to queries

April 9 2020, by Matt Swayne

Penn State date scientists and students are helping develop technology tools that will connect researchers to vital new evidence-based insights about COVID-19. Credit: Wikimedia

Teams of Penn State data scientists and students, including ones led by the College of Information Sciences and Technology (IST) and the

Applied Research Laboratory (ARL), recently developed tools that they hope will help researchers better understand—and potentially stop—the spread of the novel coronavirus, or COVID-19.

The IST team created a search engine, COVIDSeer, that sorts through the COVID-19 Open Research Dataset—called CORD-19—a free resource of over 45,000 scholarly articles, including over 33,000 with full text, that are focused on COVID-19, said C. Lee Giles, David Reese Professor of Information Sciences and Technology, Penn State, and an Institute for Computational and Data Sciences associate.

"When the dataset was first announced, we immediately thought a worthwhile project to pursue would be to create a search engine because any effort focused on the virus could be useful," said Giles. "We did this in a week. Now we are going after more aspects of the data to better visualize it and make it available. While the site already has a search engine, we wanted to see if we could build one that might have improved performance."

He hopes the tool will provide researchers with quick access to needed peer-reviewed publications on the virus that could help them advance their research.

Compared to other datasets, Giles said that this set is relatively small and, therefore, easier to index. The data is drawn from research sites, such as bioRxiv and MedRxiv. Plans are to update this database each week.

According to Giles, the search engine already has about 100 users—and he expects that number to grow. The search engine is listed in the AllenAI list of search engines and other teams are also integrating the search engine into their own applications.

ARL's COVID-Explorer

In another example of how Penn State scientists are developing tools for COVID researchers, ARL's [COVIDExplorer](link) lets researchers explore data with a level of nuance and specificity not available with existing search engines like Google Scholar.

The tool brings together a suite of machine learning techniques to identify natural "topics" through the linguistic patterns within the documents' title, abstract and body text. The interactive visualization enables rapid querying, contextual [search](link) and critical understanding.

Robbie Fraleigh, assistant research professor at ARL and product design lead on the COVIDExplorer, said, "My work is always about giving context and clarity to the key questions facing the scientific community. Getting to deploy this skill set in response to the crisis we are now facing has been incredibly rewarding."

The team is actively receiving feedback from leading infectious disease researchers to further refine and optimize the COVIDExplorer.

The IST-led team includes: Giles; undergraduate students Jason Chhay, Shaurya Rohatgi, Arjun Menon and Zeba Karishma, all from Penn State; Jian Wu, assistant professor of computer science, Old Dominion University; and Cornelia Caragea, associate professor of computer science, University of Illinois, Chicago.

The ARL team includes: Fraleigh; Chris Griffin, associate research professor at Applied Research Laboratory and the Department of Mathematics; Brady Bickel, research and development engineer; and Kurt Vandegrift, assistant research professor in biology and the Center for Infectious Disease Dynamics.

In the future, the teams plan to continue to add new features and is reaching out to build more partnerships in the battle against COVID-19.

Provided by Pennsylvania State University