UC SANTA BARBARA

THE Current

October 1, 2019
Sonia Fernandez

Breaking Data out of the Silos

Our world is teeming with data, all of it just waiting to be placed into the appropriate context. Connecting these enormous bodies of information could, according to UC Santa Barbara geographic information scientist Krzyzstof Janowicz, yield a richer, deeper understanding of the world around us.

“In the previous decades, data has typically been stored in what we call ‘data silos,’ ” Janowicz said. “Data gathered by one entity,” he continued, “is often ‘locked away’ and used for specific purposes, for specific ways of thinking. But what if there was a way to store, connect and provide diverse sets of data that could be useful to the many users who need it and could find creative new ways to use or combine it?”

There is such a way, Janowicz has asserted, and with $1 million in initial funding from the National Science Foundation, he and about 20 colleagues from universities, companies and government agencies across the United States are poised to break data out of their silos. Titled “Spatially Explicit Models, Methods and Services for Open Knowledge Networks,” the project aims to create the connections between vast data sets that can lead to better understanding and more creative solutions to complex emerging problems.

“Even for departments within a single entity, exchanging data has been difficult because one way to talk about things in one data silo is not the same as in another one,” Janowicz said.

Enter the knowledge graph: a combination of technologies, specifications and data cultures for densely interconnecting web-scale data across domains in a human and machine readable and reason-able way. For this project, the main ordering principles to be applied to the interconnected data will be space and time.

Space and time matter not only for the obvious reason that everything happens somewhere and at some time, but because knowing where and when things happen is critical to understanding why and how they happened or will happen. How, for instance, can climate affect politics in areas that rely heavily on agriculture? Is there a link between today's soil health and historic slave trade? Questions like these often take considerable amounts of time and effort to answer, often with work that duplicates previous studies.

"Instead, you can connect your local knowledge repository to global repositories to get a holistic view of your domain or your problem," Janowicz explained, thanks to the increases in computational power and data storage.

It's a huge endeavor. Data can come in many forms, ranging from numerical measurements to images to verbal descriptions. The job of the researchers — who hail from UCSB's Center for Spatial Studies, Earth Research Institute and National Center for Ecological Analysis and Synthesis, as well as Arizona State University, Michigan State University, Kansas State University, U.S. Geological Survey and industry partners such as ESRI, Oliver Wyman, and Princeton Climate Analytics — is to develop artificial intelligence methods of organizing these huge sets of information into formats and relationships that can be read and understood across disciplines, using space and time as ordering principles.

"We would like to develop a knowledge graph together with the partners from other universities, major industry players and government organizations that contains spatial data, and we also want to make methods available for many other knowledge graphs that either use spatial data or want to enrich their data-using spatial data," Janowicz said. He explained that much of this can be done with machine learning models that digest the enormous amounts and various types of data being generated, which can then be organized in graphs that show both the breadth and depth of knowledge of a given topic.

He further explained that the product would be dense, widely accessible knowledge graphs that can not only reach back into history for context, but also widen our

present options and risks and allow us to make informed predictions about things to come. For instance, given the data we already have about local climate, soil health and erosion, what are the chances of having another disastrous debris flow of the type that happened in Montecito, Calif., in 2018, and how should that affect local land-use planning and real estate?

“Currently, there is no way you can query for erosion risks by linking them to extreme event databases,” Janowicz said. “But this should be the most easy thing to do on the planet. These are exactly the kinds of problems that we are tackling.”

The initial grant is for a total of $1 million over nine months, and is part of NSF’s new Convergence Accelerator, which enables research teams to build tools that harness the data revolution and allow people from various sectors — government, academia, industry, nonprofits — to access and use data in an Open Knowledge Network.

Tags
Artificial Intelligence

About UC Santa Barbara

The University of California, Santa Barbara is a leading research institution that also provides a comprehensive liberal arts learning experience. Our academic community of faculty, students, and staff is characterized by a culture of interdisciplinary collaboration that is responsive to the needs of our multicultural and global society. All of this takes place within a living and learning environment like no other, as we draw inspiration from the beauty and resources of our extraordinary location at the edge of the Pacific Ocean.