UC SANTA BARBARA

THE Current

October 7, 2020
Sonia Fernandez

Connecting Through Data

COVID-19. The global economy. Food systems. Supply chains. Climate change effects. Social movements. If the past several months have taught us anything, it's that we all are part of vast, complex networks that stretch across space and time. What happens in one area will have impacts elsewhere, though how they manifest can be unclear and are often counter-intuitive. As a result, decision-makers can be left unprepared, reactive and vulnerable.

Geographic information scientist Krzysztof Janowicz, a professor at UC Santa Barbara, knows there's a better way. In most cases, all information required for better data-driven decision making is already out there, he says, but finding the data, extracting them, interlinking them across several domains, and then quickly turning them into relevant, actionable insights has been the challenge.

"Today, data-driven decision making is hindered by the fact that 80% of the entire budget — be it money, person power, or time — of a data science project is spent on finding the data, cleaning them, checking the veracity of the data, then tailoring the data to the analyst's specific study area, and so on," Janowicz said. "By the time you have done all of this, only 20% of your resources are left for the actual analytics. This is hardly sustainable. During our project, we plan to break apart this data acquisition bottleneck."

Armed with a $5 million cooperative agreement from the National Science Foundation's (NSF) Convergence Accelerator, he is working to build an information architecture that incorporates and interprets multiple dimensions of data to do the

heavy lifting for some of our most essential sectors. Dubbed the "KnowWhereGraph," it aligns information from several massive and rapidly changing datasets into an open geo-knowledge graph for use by decision makers in the areas of environmental policy, food security, soil health and humanitarian aid. The KnowWhereGraph project is one of nine Phase II projects that were selected by the NSF Convergence Accelerator out of a Phase I cohort of more than 40 highly competitive teams that were all focused on high-impact societal solutions.

The NSF Convergence Accelerator focuses on delivering tangible solutions that have a nationwide societal impact, and at a faster pace. Using innovation processes such as human-centered design, user discovery and team science, and integration of multidisciplinary partnerships between academia, industry, non-profit, government and other sectors, the Convergence Accelerator makes investments to solve high-risk societal changes through use-inspired convergence research.

"I congratulate Krzyzstof Janowicz on this impressive and well-deserved grant from the National Science Foundation," said Pierre Wiltzius, dean of mathematical, life and physical sciences at UCSB. "As a leading scholar in spatial data science, Prof. Janowicz is uniquely qualified to lead an interdisciplinary team of researchers as they develop a groundbreaking open knowledge graph. This project promises to affect decision-making in both public and private sector and I look forward to its exciting outcomes."

Led by UC Santa Barbara, the highly multidisciplinary team consists of several universities (Kansas State University, Arizona State University, Michigan State University, and the University of Southern California); industry partners (Esri, Oliver Wyman, Hydronos Labs, LLC, IN1OT); Direct Relief as an NGO; and the U.S Geological Survey and the U.S. Department of Agriculture representing the government. The UCSB team consists of the Center for Spatial Studies, the Climate Hazards Center and the National Center for Ecological Analysis and Synthesis. The overall project team is complemented by a growing list of collaborators that contribute data, licenses and expertise, among others. A core deliverable of Phase II will be turning the KnowWhereGraph into a public-private partnership.

Over the next 24 months the Janowicz-led team will work on high-impact deliverables including a prototype and sustainability plan to ensure the solution is impactful beyond NSF support.

Context is King

Years ago, Janowicz said, it was enough for decision makers in some sectors to be concerned only with their local context, within a regional scope of enterprise. “Now, whether you are an individual farmer, a retail company, or a humanitarian relief organization, you act in a global context,” he said. “You have to track commodity prices, tariffs, listen to the pulse of society, be sensitive to the culture of your markets, monitor weather forecasts, and even be able to react quickly to freak events such as pandemics or terrorists.”

What KnowWhereGraph aims to do, according to Janowicz, is to bring together a wealth of highly diverse sources of relevant information to form an open, spatially-explicit knowledge graph — a model that integrates not just different kinds of data, but, importantly, also their relationships, in a way that can be accessed by those for whom the information matters most.

To interlink and be able to query all these data sources requires a “universal” language. “To some degree, such language already exists,” Janowicz said. “It’s called the resource description framework (RDF), and it enables us to describe the world around us in human and machine understandable terms.” These resources can include things like maps, images, tabulated data, and text — all of which are built into a global and decentralized graph.

“RDF triples, which are statements in a subject, predicate, object form, enable us to publish knowledge about the world around us and irrespective of the fact who made these statements, what they are about, or when and where they were made,” he said. “Everybody gets to contribute and connect statements to already existing ones.”

Of course, the value of such a data graph relies heavily on the richness of the data, how current it is, and how the connections between disparate bits of information can become solutions to current problems or predictors of future scenarios. For that, there’s artificial intelligence.

“We’re going to develop AI methods to help decision-makers communicate with our KnowWhereGraph,” Janowicz said. “Essentially, the graph will deliver contextual background information about an analyst’s study area using a process called geo-enrichment. We are particularly interested in graph summarization techniques to find task-relevant triples from a pool of billions of other statements.”

The grant is also closely aligned with Janowicz’s recent NSF RAPID award, in which knowledge graphs technologies are used to compare COVID-19 related forecast data.

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.