Representing Spatial Information

▪ HANAN SAMET

Hanan Samet started working on ways to represent spatial information three decades ago, long before the Internet and sites like Google made interactive maps widely accessible. "At first, everyone told me I shouldn't be working on this stuff," says Samet. But he has always been a contrarian, choosing research directions that were not yet popular, and he also sticks to problems until he works them out. "Eventually they stopped asking me to stop," he says.

"When he started 30 years ago, disk drives were very small, and the concept of putting a map of the entire United States in a car was inconceivable," says Richard Snodgrass, a professor of computer science at the University of Arizona. "Partly because of storage becoming so cheap and partly because of the data structures he developed over the years, we have GPS systems and other devices."

"The theme of my work is sorting in space—or location, location, location," says Samet. One of the fundamental challenges of working with two-dimensional geographic information is sorting the information. Sorting items in one dimension—listing people by height for example—is straightforward, but sorting in two dimensions is much more challenging. "You can't really apply an

Representing Spatial Information

HANAN SAMET

To contact Hanan Samet, go to
www.cs.umd.edu/users/hjs
To view other research in UMIACS,
go to www.umiacs.umd.edu

ordering in more than one dimension," says Samet.

The best way to retrieve geographic data is to use a query language that is graphical, Samet says. He has developed spatial indexing systems, thereby enabling him and his students to devise "spatial spreadsheets" and "spatial browsers" to make large amounts of geographic information accessible and usable. The systems can collect and retrieve vast stores of spatial data and associated non-spatial data quickly, giving users the answers they need in the resolution that they can use.

Samet developed ways to sort two-dimensional information in grids of boxes of varying sizes, with associated indexes identifying the contents of each box of the grid. He was a pioneer in developing quadtrees for sorting spatial information, customizing the sizes of the boxes making up a quadtree to suit the problem at hand. In one variation, he gives every one or two distinct geographical objects in a map their own box. The boxes in a quadtree are denser where objects are more densely packed on a map. "Quadtree methods are simple and easy to implement," says Samet. In fact, his innovations have been used by geographic information systems, or GIS, developers, including Google, the United Parcel Service, and game developers.

Michael Jones, the chief technologist for Google Earth, says, "Quadtrees are a big part of our database processing. They're a very efficient way of organizing information. Google Earth is the world's most widely used graphics application, and it takes advantage of Dr. Samet's work on quadtrees."

Samet's spatial spreadsheets can save the results of different queries and layer them on top of each other. The results for one query can be used as the input for subsequent queries. His system gives the user flexibility in asking questions and deciding how to ask them. For example, spatial spreadsheets are ideal for sorting out geographical pairs, such as the closest monitoring stations to a list of nuclear power plants, the closest distribution centers to a list of Walmart stores, or the closest fire station to a given house.

The spatial browsers that Samet and his students have developed can be used to access information in the spatial spreadsheets. For example, Samet and his students used a spatial browser to reanalyze epidemiologist John Snow's famous 1854 map showing that cholera outbreaks in a region of London clustered around a single water pump. Comparing the locations of homes of people who died of cholera with the locations of water pumps clearly shows that deaths grouped around the suspect pump. This is just one example of how a spatial browser can help researchers test hypotheses and make spatial correlations.

Over his career, Samet has been the author of more than 300 research papers and an encyclopedic book on the subject of spatial indexing. "He has a wonderful expansive perspective on the whole field," says Snodgrass. "He was really ahead of his time, and he's still at the forefront of his field."

Vinton Cerf, a Turing award winner known as one of the founding fathers of the Internet and currently employed as the Chief Internet Evangelist at Google, has tracked Samet's work since serving as his Ph.D. thesis advisor. Samet "is a tireless and comprehensive researcher with energy to spare," says Cerf. "He's persistent and determined. He pioneered the field of spatial information management in the form of quadtrees and their derivatives and is the dean of its school of thinking. His most recent book is a magnum opus on the topic, and the application software he has developed maintains the highest fidelity of representation of spatial relationships."

Of the many popular mapping tools that people now use everyday, Samet says, "There's inevitably an engine under the hood, and that engine is a sorting mechanism. We've devoted our work to building a good engine."

— Profile written by Karin Jegalian

University of Maryland Institute for Advanced Computer Studies
2119 A.V. Williams Building • University of Maryland • College Park, MD 20742

UNIVERSITY OF MARYLAND