Spatial Data Quality 2007

Fuzzy Logic and Predicting Earthqua

The quality of spatial data has always been an important component of mapping and spatial data handling. In Enschede, the Netherlands, it was the subject of the fifth Spatial Data Quality Conference, held June 13 to 15. Hosted by the International Institute for Geo-Information Science and Earth Observation (ITC), it was organized by ITC and the ISPRS commission II, working group 7. The symposium brought together experts from around the world to present the latest developments in spatial data quality.

By Job van Haaften

Michael F. Goodchild

The symposium was a platform for leading international scientists and young researchers to share their experiences in the field of spatial data. Research was reported from all over the world on such topics as 'the quality of meteorological data by earth observation satellites', 'challenges in modelling the uncertainty of GIS-data', 'quality control in large spatial databases' and 'the impact of spatial data quality on decision making in a company'. The following subjects in particular caught my attention.

Granular Computing

The opening keynote speaker, Professor Lofti Zadeh from the University of California, Berkeley in the USA, spoke about fuzzy logic and especially granular computing. Note that it is not the logic that is fuzzy, but it is that the logic is dealing with fuzzy data. Zadeh: "Granular computing is computing with uncertain, imprecise and partially true data. In granular computing, the objects of computation are not the values of variables but the objects are data on the values of variables." Examples include a category or a matter of degree.

"As an illustration, we want to know the age of a woman named Vera. We know she has a son who is about 25 years old and a daughter who is about 35. Furthermore we know that the child-bearing age varies from about 16 to about 42. Given this information we can estimate Vera's age.

Imprecision, uncertainty and partiality of truth are characteristics of the real world. The need for an enhancement of our ability to deal with imprecision is certain to grow in importance. This need motivated the development of granular computing and is driving its progress."

Earthquake Risks in Tehran

From Iran came a good example of fuzzy logic: the estimation of seismic vulnerability in Tehran. Mahmoud Reza Delavar from the University of Tehran explained how they made a hazard map of Tehran which shows the risks of damage by earthquakes. Delavar first pointed out that Iran had 12 major earthquakes in the twentieth century and that more than 70 per cent of Iranian cities, including Tehran, are susceptible to earthquakes. Delavar: "Despite the high level of uncertainty earthquakes pose, they are phenomena which are under special considerations. A number of attempts have been made to predict earthquakes and to mitigate their impact. The prediction of earthquakes is not very successful. However mitigation of the impact is possible through investigation of factors affecting earthquake damage and risk management."

Some of the data used to create the hazard map are the number of floors in buildings, the age of buildings, their construction and the resident population. This information is combined with the knowledge gathered on four major faults, a number of minor faults and the opinions of twelve experts on the risks of earthquakes and their magnitude.

Now it is up to the authorities in Iran to act on the information the hazard map shows. Decision makers and politicians generally tend not to like information on earthquake risks. Delavar himself is positive about it: "Our government paid a lot of money to run these estimates and there are a number of scientific projects concerning earthquakes and their consequences."

Semantics

There were three papers on semantics, a subject not generally thought of in connection with spatial data. But not all data are numeric or digital. Words and sentences also have to be translated for different communities that require reference systems for geospatial information. There is an increasing number of applications that rely on 3D geoinformation including noise mapping, disaster manage-

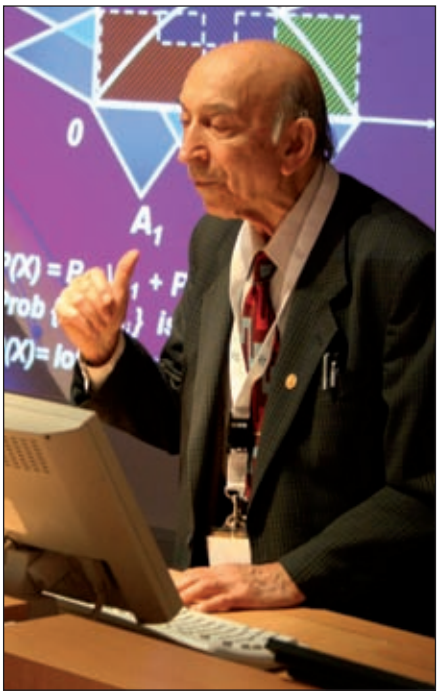
ke Damage

Mahmoud Reza Delavar

ment, architecture and city planning. These applications frequently require complex semantic information from many different sources, often thematically and spatially fragmented. They differ in quality and semantic aspects. Prominent examples involving inconsistencies are 'flying' or 'drowning' houses, where a digital terrain model and 3D models from different sources have been combined. The integration of spatial data with complex object descriptions is a challenge and a change at the same time. The more information is provided by the semantic layer, the fewer ambiguities remain for possible geometric integrations. When a building and a terrain are joined, the lower edge of a polygon marked 'door' needs a surface to step on, either the terrain or a staircase.

Consumers

Semantics is also used for communication to consumers on data and the quality of data. As Anna T. Boin and Gary J. Hunter from the University of Melbourne in Australia put it, spatial databases are more readily accessible to the general public, and every day more people use the World Wide Web to decide whether a dataset is suitable for them. Findings from consumer opinion show that if they use a website to choose data, and if the website provides the information clearly, they are more interested in determining the content of a dataset. In essence, the word 'quality' may not consciously come to mind at all. When using the data and finding mismatches with the real world or another dataset, quality is important. There is therefore a requirement to communicate more explicitly what a dataset contains and to raise awareness of internal quality. This should influence the way datasets are presented on the internet and thus make data more accessible, more understandable and more valuable to the everyday person looking for information.



Lofti Zadeh

Metadata on Quality

One of the five keynote speakers was Michael F. Goodchild from the University of California, Santa Barbara, in the USA. Goodchild: "Worldwide data quality statements are now entrenched in metadata standards. I contrast the needs of the user with the production-control mechanisms of the producer, and argue that metadata standards are producer-centric. To the user the ability of data sets to interoperate is of major concern. The experience of prior users, the accessibility of quality statements are of interest. And the user wants the quality information to be handled easily in local software."

The focus of discussion has been on data quality; however, increasing the sharing of data is prompting a demand for more comprehensive and thus more complex metadata. Goodchild: "Metadata have the potential to exceed data in sheer volume, so it is reasonable to expect that as much effort will be spent documenting data as in compiling them."

Whether data or metadata, numeric or semantic, fuzzy or precise, it is clear that we are and should be concerned about quality.

Job van Haaften (jvanhaaften@geoinformatics.com) is editor of GeoInformatics. For more information about the conference: www.itc.nl/issdq2007