Multiple representations and multiple resolutions in geographic databases

Christelle Vangenot,
Database Laboratory
Swiss Federal Institute of Technology
1015 Lausanne, Switzerland
Tel :+41 21 693 52 10
firstname.name@epfl.ch

Christine Parent,
University of Lausanne
HEC/INFORGE
1015 Lausanne, Switzerland
Tel :+41 21 692 34 18
christine@lbd.epfl.ch

Stefano Spaccapietra
Database Laboratory
Swiss Federal Institute of Technology
1015 Lausanne, Switzerland
Tel :+41 21 693 52 10
firstname.name@epfl.ch

Interoperability nowadays bridges the gap between information repositories, giving a chance to find complementary information about the same or related facts in various sources that have been independently developed. Unfortunately, semantic interoperability is not easy to achieve, as related knowledge is most likely described in different terms, using different assumptions and different data structures. Reconciling this heterogeneity to build a fully integrated database is known to be a very hard, currently unresolved problem. A simpler step is the identification of related knowledge and the provision of a mechanism that in some way materializes the relationships between different representations of the same fact. Simply stated, DBMS support of multiple representations is a necessary first step to interoperability, but to be really effective it has to be more powerful than the limited functionality currently provided by these systems, e.g., using generalization hierarchies.

One of the most evident examples where multiple representations are needed and are not supported by current technology is in cartographic applications. Map producers need to build maps of the same geographic region at different levels of abstractions (at different scales/resolutions, in technical terms). Typically, they maintain one database per scale, with no interrelationship, hence with no update propagation and no inter-database consistency. Another classical example is given by applications achieving simulation of phenomena or decision support systems. They need several representations because characteristics of phenomena are detected and understood at different levels of description. For instance, to decide about the settlement of a new plant needs an analysis going from the world level, to determine the best country to set up the plant (less charges,…), to the city level, to find the best place (soil composition, proximity…).

A multiple representation GIS would allow storing all representations in a single database, and enforce the consistency of the different representations through appropriate, automatic update propagation. Beyond cartography, multiple representations of geographic data are needed to be able to serve multi-disciplinary user communities, as the same piece of land may support analysis, planning, and forecast activities by city administrations, environmentalists, sociologists, botanists, zoologists, etc.

Actually while the scale concept is perfectly understood and relevant when talking about maps or cartographic databases, it is not relevant anymore when talking about the geographic representation of an object in a database (scale is the ratio between the objects' size in the map and their size in the real world). When focusing on databases, it is more correct to use the term *resolution*, usually defined as the minimum size of an object to be represented. The resolution of information in the database is the resolution used at data acquisition. If different resolutions have been used for the same object, we can talk about multi-resolution objects. Moving among resolutions has a potential impact on the shape of objects (shape may be simplified in less precise resolutions), on their value (because of a corresponding change in semantic resolution according to user-defined hierarchical value domains), as well as on the existence of objects (because of a change in aggregation rules or creation of new aggregates). From the geodata application perspective, multiple resolution is certainly the most urgent multi-representation problem to solve. The potential variety of representations extends over different facets, such as:

- multiple geometries (possibly belonging to different spatial types, like area and point, or area and line) for the same object,
- multiple abstraction levels that make a set of objects coexist with the object(s) that represents their aggregation (whether the aggregation is based on geometric or semantic criteria),
- multiple abstraction levels that result in hierarchical value domains for attributes, and
- multiple representations as for traditional databases.

The consistency of a multi-representation database can only be achieved if the system is capable of interrelating them and dealing with their differences. To this extent, an appropriate data model with advanced concepts is needed.

In this paper, we propose such a set of concepts allowing to describe the same real world object at different levels of spatial and semantic description.

- Specification of several geometries (within the same spatial type or not) for the same real world object.

Attributes having a spatial domain must be able to memorize several values, each of them with their corresponding spatial resolution. For instance, a building could be represented by the area corresponding to its extent at a very precise resolution and by a point giving its position at a less precise resolution.

We propose to stamp spatial and non-spatial attributes with their level of description: several values can then be registered for these attributes as a set of couples (value, level of description). Values of a stamped attribute may belong to different domains. To this extent, spatial attributes are described using spatial abstract data types including some generic types grouping several basic spatial types (point, line, area).

- Description of the validity span of topological relationships.

The explicit description of a topological relationship between two spatial objects is important as it allows to constrain their relative position and to associate to the relationship a semantic value and possibly attributes. Yet, this relationship may also depend on the level of description: we then propose to associate a validity span to topological relationships. This validity span may have different facets. It may be defined on:

- The definition of topological relationships:

For instance, when working at different levels of details in a road database, the adjacency relationship may have several descriptions: “distance between two objects less than 10 meters” at the most precise level of description and “distance between two objects less than 100 meters” at the less precise one.

- The existence of topological relationships:

An object may only have a representation at one level of description if for instance it is below the resolution specifications at the other level of description. Then, its associated relationships are no longer valid and should have a validity span.

- Description of multiple abstraction levels that make a set of objects coexist with the object(s) that represents their aggregation.

For instance, this situation may result from the grouping of objects according to geometric and/or semantic criteria:

- Grouping of areas having the same land use.

- Grouping of objects according to their spatial relationships, like for instance a set of close buildings becoming a built area.

- Grouping of objects according to their semantic and spatial relationships like for instance, a set of Buildings and adjacent Fields belonging to the same farmer grouped into a single object Farm.

To link the two representations, the data model has to support the aggregation concept. Mechanisms of derivation of attributes and of definition of spatial integrity constraints allow to customize the aggregation according to the different criteria. For instance, the geometry of the Farm objects may be derived from their composed buildings and fields. Moreover a spatial integrity constraint may specify that the fields and buildings composing the same farm must belong to the same farmer and that the fields must be adjacent.

- Description of multiple abstraction levels through classification hierarchies for objects and through hierarchical value domains for attributes.

The same entity or attribute may also describe the same real world characteristic according to different levels of semantic description. For instance a building may be seen as a natural history museum at a precise level of description and just as a museum at a less precise one. Or the attribute land use of a plot may be queried at different levels of description: What are the plots growing 'corn', or what are the plots growing 'Cereal'?

We have emphasized the importance of the generalization link with several mechanisms for inheritance (simple, refined, redefined). Thus the objects describing the same real world entity may or may not have the same geometry in the super-class and in the subclass(es). In case of refinement, objects have the same geometry but with different spatial domains. Redefinition allows objects to have different geometries in the super-class and in the subclass. For instance, an object when considered in the superclass Road has a simplified line geometry and when considered in the subclass Motor Road has a detailed area geometry.

- Intelligent zooming through the different levels of description.

The model provides generic methods to manipulate these multiple representations, especially the intelligent zoom that supplies a more precise view of a map: objects are drawn with more precise cartographic representations, small objects, which were invisible at the previous coarser scale, appear, other grouped objects are disaggregated…