REQUIREMENTS OF A MULTIPLE REPRESENTATION DATABASE FOR TOPOGRAPHICAL DATA WITH EMPHASIS ON INCREMENTAL GENERALIZATION

Tiina Kilpeläinen
Finnish Geodetic Institute
Department of Cartography and Geoinformatics
Ilmalankatu 1 A
FIN-00240 Helsinki
Finland

Abstract

The starting point of this paper is model generalization in a multiple representation database environment. The main issues encountered in constructing multiple representation databases are data abstraction models and data maintenance problems. A data model for multiple representation database is presented and discussed in this paper. One of the advantages of multiple representation databases lies in updating problems. They also provide a control mechanism for updating and data consistency. The requirements of a multiple representation database are discussed, with special emphasis on so called incremental generalization, which is an updating technique for the generalized versions from the database.

1 Introduction

11 What are Multiple Representation Databases

The multiple representation research problem was initialized in the research program of NCGIA, National Center for Geographic Information and Analysis in 1988 [4,8,11,28]. In a couple of research works the same problem is referred to as multiscale databases. During the late 1980s and early 1990s generalization research has been much involved with knowledge-based techniques. Lately, the discussions have centred on the role of model generalization. By model generalization we mean simplification of the abstract and digital model represented by geographical information [31]. Model generalization should be seen as a preprocessing stage for cartographic generalization that includes the constructing of *multiple representation* databases. In this context, awareness of the strong relations of multiple representations to the generalization problem in the database environment in particular has made the multiple representation problem a topical subject of research.

In multiple representation databases several representations of the same geographical object are included in the database. The representations can consist of, for instance, geometrical representations in 2-D or 3-D, but also of conceptual representations in the form of some kind of model. Representations from different time intervals can also be included. In a multiple representation database the relationships between the different representation levels are defined in such a manner that changes to the base level can be propagated automatically to the other levels [8,20]. One of the challenges of the multiple representation problem is to develop a theory to define these connectivities formally.

In our model, real world phenomena are assumed to be represented according to an object model that includes rule processing, as discussed in [17, 18, 19], Figure 1. The object model includes locational as well as nonlocational attribute properties.The object model is not discussed here, as it already been discussed in several other papers and proposed as a suitable alternative for GIS and generalization purposes [5, 14, 18, 19, 22, 23, 27].

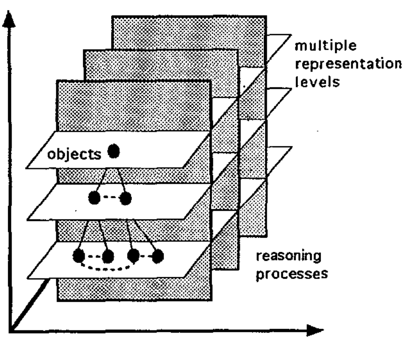

Figure 1. Multiple representation and its two related dimensions.
——— connectivities between levels,
-------- relationships between objects.

Figure 1 illustrates the multiple representation concept. Maintenance of the different versions of generalized versions deduced from the base level is one of the difficulties in a multiple representation database. In Figure 1 the representation levels contain objects at different degrees of details. The connectivities between the levels can be maintained by hierarchical aggregation, the base level being the most accurate and the levels above containing fewer details. In other words; the grade of conceptualization of the objects increases with distance from the base level.

The base level also contains the largest quantity of information. In such it will never in practice be visualized; selection processes are always needed. Objects and/or object groups may have relationships between each other at the same level. The same relationships must be preserved at all the levels, which here means that data consistency will be preserved. An example of such relationships is topological relations.

Here we want to separate the relations between the different objects at one level, that is, relationships, from the relations between the different representations of the same object at different levels, that is, connectivities. Another difference between relationships and connectivities is that, connectivities enable the higher-order representations to be deduced from the lower-level representations. Knowledge of the relationships between objects are defined in object attributes but cannot be derived.

Figure 1 also shows the two related dimensions needed for the multiple representation database: the reasoning dimension and the representation dimension, modified from [29]. The reasoning processes in a representation of the problem between the different representation levels are needed. One example of reasoning is the rules, which may be different for the different levels. For example, when the higher-order representations are derived from the lower-level representations, the rules may be different for each derivative. Reasoning processes are also required for multiple representation data queries. In hypermedia user interfaces it is useful to view the geographical reality in different contexts; vertical and horizontal, [24]. The idea of corresponding vertical and horizontal links are very much similar with the concepts of connectivities between the levels and relationships between the objects.

Figure 2 illustrates the dataflows and consisting conditions in a multiple representation system. The connectivities between the levels are formally defined, thus enabling automatic update propagation between the levels. As the connectivities between all the levels are defined, the consistency between the levels is automatically checked. Here, a multiple representation database system also assumes that the different generalized outputs from the different representation levels can be automatically derived. These automatic generalization processes require the knowledge for the graphical representations to be stored in the object attributes in the multiple representation database. The update propagation in Figure 2, can be fulfilled in three ways: 1) from base level to level 1, from level 1 to level 2, and so on; through 2) propagation from base level to each of the levels separately; and 3) by a combination of 1) and 2): both of the propagation strategies are possible.

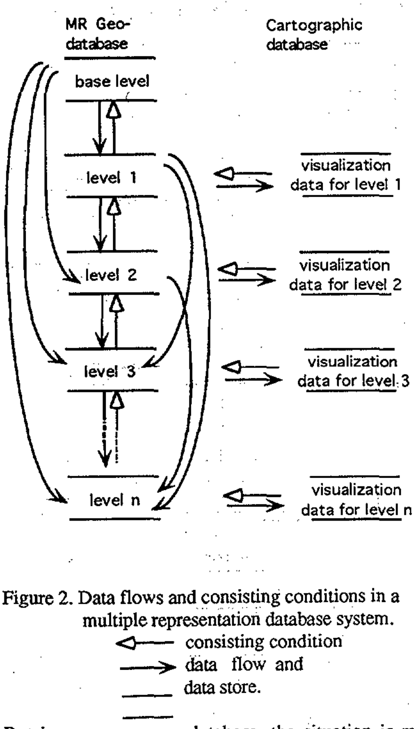

Figure 2. Data flows and consisting conditions in a multiple representation database system.
◁—— consisting condition
——> data flow and
____ data store.

The idea of a multiple representation database system in a fully automatic environment is based on the two-folded concept of 1) model generalization and 2) cartographic database for visualization. Model generalization in the database is a preprocessing stage for cartographic generalization. Separation of the conceptual generalization problems from the problems related to the graphic display is a way of restricting the generalization problems in terms of complexity and quantity [18]. The distinction between model generalization and cartographic generalization is apperate already in the early works of Grünreich [13]. The respective models are there called digital landscape model and digital cartographic model. A similar concept can also be found in the conceptual framework for map generalization, presented by Brassel and Weibel [6].

Mark [24] claims that, in automatic mapping systems, it may be better to view the selection process as direct competition for map space between and among all available features in the database. Therefore, in a cartographic database a small change to cartographic object symbolism accumulates to the symbolisim in the neighbourhood and the result is: a fight for map space (not a fight for geographical space). Since in a geodata base the objects have defined geographical extent they do not have to fight for space for their representations. But in a cartographic database, the situation is more complex. The map symbols for the objects competite in a different way, depending on the actual application. In the visualization phase, the symbols may also be exaggerated and emphazised; the picture on the map becomes an abstraction of the geographical reality.

Model generalization and cartographic generalization can be realized by separate modules in a GIS. The model generalization could be considered to belong to the analysis functions of a GIS. When a system for cartographic visualization is thought as a subsystem of a GIS [2], the cartographic generalization could be carried out by this subsystem.

these cases can be deduced from the multiple representation geodata base (MR) with the aid of graphical representation knowledge stored in the attributes of the objects. Practical projects seem to be emerging for implementing similar principles, as reported in [10].

The essential difference between the updating propagation in case A with an ordinary geodata base and cases B,C,D lies in consistency control. As the connectivities between the levels are formally defined, cases B,C,D with multiple representation database environments provide an automatic control mechanism for data consistency.

3.1 Incremental generalization in updating process

In this section we discuss the role of incremental generalization in an updating process. It is assumed that the new datasets are to be updated at the base level of the multiple representation database. One of the factors in favour of updating at the base level is that it enables us to check that the data are homogeneous at the same time. The concept of updating situations is illustrated in Figure 4, which lists the possible update situations for incremental generalization, modified from [17].

Basically, on the above assumptions, updating propagation to other levels can technically be performed either automatically or interactively, Figure 5. Hence, the first step in our updating process is to decide whether the updating can be done automatically or not. Information is therefore required on whether there are processes for deriving the higher-order representations from the lower-order representations. Since there will certainly be a need for interactive processing in generalization tasks in the future, Figure 5 shows alternative ways for update processing and the role of incremental generalization in the whole process.

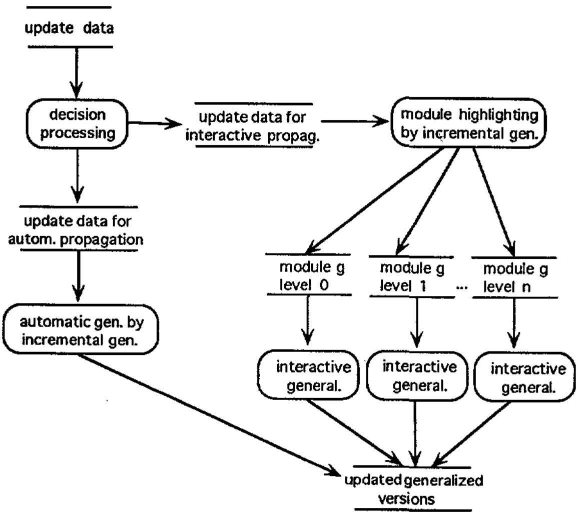

Figure 5. Incremental generalization in a semi-automated environment. The notation is according to SA/SD method [30], where ⟶ data flow, ═══ data store and ▭ data transformation.

In an ordinary generalization environment, after updating the geodata base, we have to re-generalize the new outputs for different applications from the updated base level by generalization processes, as already discussed in [20,21]. In an incremental generalization environment, the generalization process is carried out completely for the whole geodata base only once. After the following update transactions to the geodata base, the old generalized output is also updated in an incremental way, which means that the generalization process is performed only for the modules influenced by the updates. As shown in Figure 5, incremental generalization can be used in two ways. In automatic processing knowledge of the modules where updating is needed is deduced from the update data. The generalized versions can be automatically updated directly in these modules owing to the defined connectivities between the representation levels. In the interactive environment, the incremental generalization can be exploited to highlight the modules that have to be updated, but the updating is then made interactively, module by module. The idea of highlighting the modules could also be exploited to reveal, in interactive working, the modules where possible conflict cases may occur. As discussed earlier and in [20,21], knowledge of the set of conflicting objects has to be defined in the multiple representation database.

The main question in incremental generalization is: what modules are we are talking about? The question is related to the problem of pattern recognition, but can be looked at from another direction, too. We could also say that the modules of the map series consist of separate map sheets. The module would need some overlaps, say 5%, to diminish the data consistency problems at the edges. The other way of seeing the problem is to separate the map sheet into smaller sections with the aid of different kinds of patterns that can be recognized from the sheet, e.g. road networks or real estate divisions. Kidner and Jones [17] state that the least computationally demanding method of determining separateness is that of comparison of the extents of the objects as indicated by minimum bounding rectangles. The order hierarchies of the objects could also be used, as discussed in [18]. The order puts the objects into an importance hierarchy, which defines which of the objects influence each other, and which do not interact. The connectivities between objects on different levels could be described using some object-oriented or semantic data model, as done in [16].

Figure 6 depicts the effects of incremental working. In this case, updating is not done to the base level, but directly to the a module at a higher-order level. Additional update data from outside databases are stored, and the other levels are not involved with the changes. The propagation is done automatically to the higher levels from the updated level owing to the defined connectivities. The generalized outputs in the cartographic database are also updated module by module.

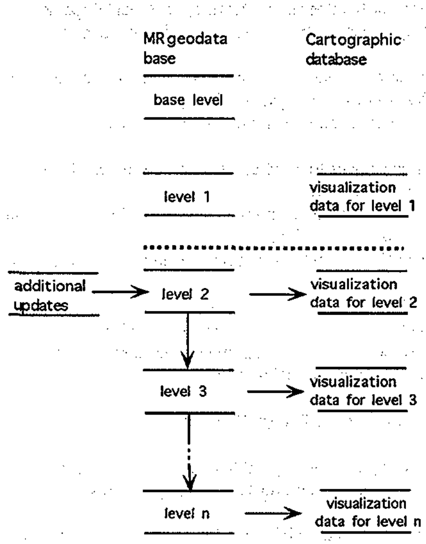

Figure 6. The effect ········ of incremental generalization in an additional updating process.