REPRESENTATION OF THE GENERALIZED DATA STRUCTURES FOR MULTI-SCALE GIS

M.O.Govorov
Dept. of Cartography, Siberian State Academy of Geodesy
Plahotnogo 10, Novosibirsk, 630108, Russia

Abstract

This paper considers the urgency of maintenance of the generalized multi-scale or multiresolution data structures (MSDS or MDS) and the principles of its independent implementation. Thus, the different kinds of cartographic generalization (CG) with the point of view of their representations in the multi-scale database (MSDB) are examined and classified. The spatial objects and properties of MSDS are presented. The generalized data structures, which result from different kinds of CG, are implemented in the conceptual model.

1 Introduction

Now the efforts of many researchers are directing to the design of multi-scale GIS (MSGIS) [2, 5-10]. These researchers have chosen different directions to the resolving this task. But the main problem is remained as the problem of formalization and contextual comprehending of automated CG. The formalization of CG implies the searching for axioms and building up the formal language of CG, which is based on those axioms. Last problem is closely connected with former one and usually is discussed during the design MSGIS. MSGIS can be defined as a spatial information system, which supports all function of conventional GIS and additionally maintains MSDS or multi-detailed database structures (MDS). MSGIS must be capable to represent entities of the world at the different levels of details and to support modification across resolution levels [6]. The existing commercial GIS's software, which maintains the spatial models at a fixed scale level, usually contains the following main blocks: 1) analysis management system; 2) graphical management system; 3) data base management system.

There are at least three ways of creation MSGIS. And above mentioned blocks of conventional GIS can play different roles in the construction MSGIS. Also there are different opinions of implementation of these problems. Most researchers mainly have paid attention on the execution of automated CG in the first two blocks [3, 8]. In current years some authors pointed out on the needs of maintenance of a multi-scale or multi-detailed or multiresolution database (MDB) [5-7, 10]. The options for design MSGIS and their advantages and disadvantages are following:

1. The first way is creation of MSGIS, which contains the spatial automatic CG processor and DBMS, which maintains a spatial data model in the single base scale. The smaller scales can be retrieval full automatically at the time of the query. This way is efficient for the capture and the storage of data. But currently however, complete successful design of such systems is doubtful on account of the locational geometry of object in the detailed base scale do not define the geometrical topology of the object in smaller scale. Also such systems can caused the data redundancy during the performing of CG, that is not acceptable for the interactive information system [6].

2. The second option of MSGIS contains CG block, which can maintain the intelligent computer-aided CG design with user interaction [2, 8]. The DBMS maintains, as in the first case, a spatial data model in the single base scale. The CG operations in this case don't create new data in database. The design of intelligent computer-aided generalization systems is recommended by the application of the expert system and knowledge based techniques in order to build software capable of intelligent CG decisions [8]. The disadvantages of these systems are: these systems can be successfully used for computer assisted cartography design, that is for map compilation, but these systems don't guide an application user, who may be introduced not with the techniques of CG; the second option is time delay for real time applications; the data supporting structures are necessary for effective interactive using of the system [2]; the expert systems cannot find precision solution to the ambiguously and unclearly formulated problems of CG [4]; such systems can caused the data redundancy during the performing of CG due to inadequate treatment of scale-related variability within a file [5, 6, etc.].

3. The third way is used the intelligent computer-aided generalization design system with user interaction and DBMS, which maintains the generalized MSDS [5, 6]. In this case, the generalization operation block can be used either for creation, maintenance and update of MSDS's structure or for spatial multiresolution analysis, design user's structures of geographical phenomenon and maps. The third option gives the possibility to avoid same disadvantages of second option. The advantages of maintenance of multiple data base are: reduction of time and labour consuming user interaction for retrieval pre-generalized versions of object and obtaining closely version of appropriate representation accordance to user needs; reduction of processing time for retrieval data at the appropriate resolution by comparison with a SDDB; reduction of data duplication by comparison with the storage of several pre-generalized versions of the object; reduction of the significance of purpose-orientation factor on the process of computer CG; increase of the field of user applications of the system; doing the process of CG more phenomenon-based and impartial due to using whole inner structure of the particular entity, etc. The disadvantages of maintenance of MDB are: the process of realization, maintenance and update of MDB is sophisticated; the assimilation for a user is sufficiently complicated.

In this article attention is paid on the implementation of spatial generalized MDS. For this goal, the CG is examined first of all, because, as it was above mentioned, the understanding of computer CG plays a vital role for successfully automation of this process and also multi-detailed digital model must support the spatial and attribute transformations, which are used in the map generalized design.

2 Generalization in digital environment

A digital CG is the application of abstraction mechanism to the world's entities for the formulation of image of reality by means of distraction and supplement. There are several conceptual models for a CG in digital environment [8]. But these models are far from complete formalized description of automation generalization process. The difficulties for the complete rigorously solution of these problems are caused firstly by the multiformity of the CG. The main factors of CG can be classified as follows [8]: a) phenomenon-based factors, which are caused due to the conceptual nature of modeled entity (the essential of the analyzable phenomenon; the features of territory; the relationship among entities, etc.); b) purpose-orientation factors (user needs and purpose; contents; scale; technology of the compilation map, etc.); c) graphic media and format factors (the visualization purpose, the technology of compilation map; the kinds of cartographic object; the methods of cartographic representation; the rules of map design, etc.); d) computational factors (the effective of information system realization).

CG in GIS environment can be divided on the two stages:

1. The modeling and CG of the real-world objects for the creation of spatial structures in the digital forms.

2. The CG of digital spatial structures for the creation of their cartographic representations.

Above mentioned factors of CG are playing different roles and are having different significant for the CG process on the different stages. The factors a) and d) are more important for the stage 1. The factors b) and c) have influence on the stage 2. The factors a) and b) are more the difficulty to be formalized, former one - due to multivariable of nature of phenomenons, last one - due to coexistence of several user views for the same real-world object and it's cartographic representation. The factors c) and d) are usually programmable. Therefore, for the simplification of the problems of formalization of automation of CG is useful to maximum separate the stages 1 and 2. For this goal in MSGIS must be supported the following demands: 1. The inner multiple representation of entity (its geometry and inner and outer topology) should be embedded in database structure. 2. The date structure and operators should be extensible. 3. The logical independence between multiple data architecture and the generalization processor should be implemented.

Accordance to the definition of CG, as kind of abstraction, and above mentioned supposition about factors of CG the following operation and corresponding CG techniques can be distinguished:

1. Omitting of whole entity or parts of its geometry (selection, simplification, collapse).

2. Systematization (aggregation, concentralization, association, absorption):

2.1. Aggregation of several homogeneous objects to one object on the same level of object's hierarchy.

2.2. Classification:

2.2.1. Concentralization of several homogeneous objects into one object on the higher order object's hierarchical level [1].

2.2.2. Semantic Association [1]:

2.2.2.1. Typification is the process of grouping objects in classes by their types.

2.2.2.2. Classification is the process of grouping objects in classes according to their quantitative properties.

2.3. Absorption of objects into the other inhomogeneous objects.

3. Display operators (exaggeration and underestimation, displacement, smoothing, symbolization).

The techniques of the CG of entity and object, which make up the entity on maps or in a digital database are shown into the Table 1:

Kinds of generalization	Applied to p - points, l - lines, a - areas, at - attribute	Transformation results: g - geometry, s - semantic., t - topology	Level of influence: o - object, e - entity	Affect on the data structure: s - structur., d - display	Affect on the MBR: e - exaggeration, u - underestimation, d - displacement
Selection	p, l, a	t	e	s	-
Simplification	l, at	g	o	s	u
Concentralization	p, l, a	s, g, t	e	s	-
Aggregation	p, l, a	g, t	e	s	-
Absorption	p, l, a	s, g, t	e	s	e
Collapse	l, a	g, t	e	s	u
Exaggeration and underestimation	l, a	g	o	d	u
Displacement	p, l, a	g	o	d	d
Association	at	s	e	s	-
Smoothing	l	g	o	d	u
Symbolization	p	g	e	d	u, d

Table 1: The kinds of CGs and their classification , (where, t - a local neighbourhood topology)

3 The ways of design spatial multiresolution database

The suggestions about creation of MSDB or MDB architecture are described in following articles by Jones, C.B., and Abraham, I. M. (1986), Jones, C.B. (1991), van Oosterom, P. (1991) [5, 6, 10], et al. Mainly these authors paid attention on the two kinds of CG, these are simplification and selection.

There are at list three ways for the design of MDB for MSGIS. These ways of multi-scale database organization and their advantages and disadvantages are following:

1. The MDB can be organized as a storage of several pre-generalized versions of the entity. Advantages are [6]: the realization, update and maintenance of MSDB are simple; the efficient access is existing to the different scale representations. Disadvantages are: a storage of data is expense due to data duplication; the database structure maintains not the whole model of geographic entity, but its slices in the different scale levels.

2. The MDB can be organized as a storage of single MDS of the objects, which are contained the entity, with possibility to recover a cartographic representation at a required level of scale or resolution. Advantages are: a data duplication is avoiding by comparison with the first option; a data redundancy is avoiding by comparison with the single detailed database (SDDB); a data duplication for maintenance of different resolution of entity is sufficient; a complete or multiresolution model of entity is stored, but not its scale slices. Disadvantages are: the realization, update and maintenance of integrity are complicated.

3. The MDB can be organized as a storage of a SDDB and additional (auxiliary) information, which supports generalization operators. Advantages are same as for a SDDB. Disadvantages are: very

detailed information for resolved conflicts and for supporting a real-time CG of single detailed representation is necessary to store; the definition of information, which is necessary to store for the rigorously supporting of generalization process, is uncertainty.

In this paper the attention is focused in particular on the maintenance of MDS. The following problems should be defined on the stage of design of MDS: 1. Type of a GIS architecture. 2. Method of implementation of the spatial object's index. 3. Methods of maintenance of integrity of database. 4. Methods of maintenance topological object relationships. 5. Techniques of CG, which should be maintained into a storage structure and, which should be executed on the stage of visualization. 6. The object-layer debate. 7. The tiling and edge matching problems.

There are three different types of a GIS architecture, which has following characteristics [9]: 1) separate subsystems for storing and retrieving spatial data and thematic information, which is stored in relational DBMS; 2) a single system for storing and retrieving the spatial and thematic data in the pure relational DBMS; 3) an extensible or object-oriented DBMS, which can support more attribute types and access methods. This paper outlines only the study of geo-relational implementation for spatial MSDS. Such model of data structure, of course, has same disadvantages by comparison with an object-oriented model, but now even the implementation of conventional last one for GIS has been on the stage of investigation. The further research of this topic is supposing to do for an object-oriented implementation. It can be considered what a geo-relational implementation of MSDB is the first step in the study of multiresolution data structure.

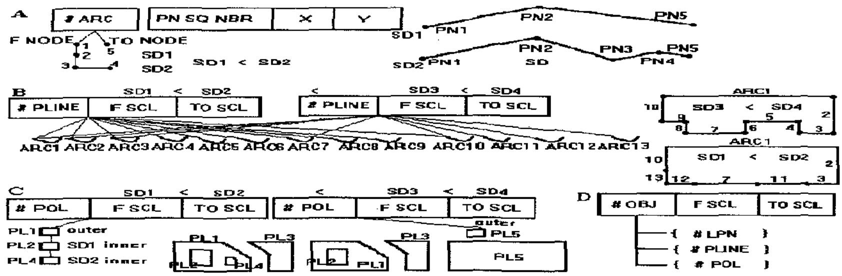

Figure 1: The conceptual schema of multi-scale spatial features: **A.** A representation of arc feature; **B.** A representation of polylinear feature; **C.** A representation of polygonal feature; **D.** A representation of object, (where, # - an ID; F - from; TO - to; { } - a subset of elements; SCL - a scale; OBJ - an object; X, Y - rectangular coordinates; SD - a scale denominator; PN - a point; NBR - a number; OR - an oriented; R - right; L - left; PLINE or PL - a polyline; POL - a polygon; LPN - a label point; SP PR - a spatial primitive; SQ - a sequence)

The spatial primitives, object and their properties, which have been used for design geo-relational model of MSDS are: 1) a point feature is a one-dimensional spatial object with coordinates and a unique identifier (ID) within the database or a map. 2) an arc is an oriented sequence of ordered vertices, where the beginning of the line may have a special start node, and the end a special end node. Also an arc can be a part of zero or more polygons, and therefore has a left and right polygon identifiers. 3) a node is a junction of one or more arcs. 4) a label point is used for representation of point features and for assign ID to polygon. 5) a polyline consists of one or more arcs. A polyline is an oriented sequence of arcs. 6) a polygon consists of one outer and zero or more inner polylines. 7) an object is a sequence of spatial features, which make up a representation of the real world entity on a map. 8) the cartographic entity or it's object can be existing on maps of intermediate adjacent scales. 9) during of simplification is going to selection of vertices of the arcs, but the nodes of these arcs are saving. 10) the following characteristic of line (arc) cartographic simplification is taken into consideration that points (vertices), selected for small scale representations, are always a subset of those used in larger scale representations. 11) the arc, which is a part of nonzero polygon, can't be selected. 12) the displacement, exaggeration, underestimation, smoothing and symbolization don't give rise of the topology modification, these operations are pure graphical. 13) the results of selections, concentralization, aggregation, absorption, collapse and association must be conducted in different relational tables for maintenance of integrity of database. The conceptual model of spatial features is shown on the figure 1.

For this scheme has been built logical model, which has been implemented as shell software under a commercial geo-relational database management system. The amount of storage in the multi-scale geo-relational database is significantly reduced due to integration multi-scale storage of vertices in the line multi-scale trees [6], which minimized vertex duplication and also using same records for arcs, polylines and polygons into the different scales of entity representation. Although this method introduces additional data in the form of sequence number of vertex, scale denominators, pointers on the files, the total amount of storage is reduced significantly by comparison with the storage of several pre-generalized versions of entities.

Conclusion

The logical model of multi-scale representations for the different kinds of generalized techniques; the possibility of application of hierarchical methods for the cases of a simplification and selection; the spatial access mechanism by means of the overlapping region's scheme; the questions of integrity of database; the supported demands and the contents of the MSDB will be present as poster on the conference and will be published elsewhere. Further research efforts will be concerned with developing an object-oriented approach for multi-detailed representation.

Acknowledgment

The author should like to express their appreciation to the China Postdoctoral Science Foundation and LIESMARS, WTUSM and professors Wu Hehai, Du Daosheng, who have provided financial, technical and consultative support for the research.

References

[1] Alaev,E.B., Social-economic geography (conceptyal-terminologic dictionary).Moscow,1983,350p.

[2] Aasgaard, R., 1992, Real time Cartographic generalization, a requirement for efficient presentation of geographic data. Proc. of the 5th Int. Symp. on Spatial Data Handling, pp. 113-122.

[3] Beard, M.K., 1986, How to survive on a single detailed database. Proceedings, Auto-Carto 8, ACSM-ASPRS, pp. 211-220.

[4] Bruegger, B.P., and Muller, J.-C., 1992, Mechanisms of geometric abstraction. Proce. of the 5th Int. Symp. on Spatial Data Handling, Chareston, South Carolina, IGU Com. on GIS, pp.123-133.

[5] Jones, C.B., 1991, Database architecture for multi-scale GIS. Proceedings, Auto-Carto 10, Baltimore, ACSM-ASPRS, pp. 15-31.

[6] Jones, C.B., and Abraham, I. M., 1986, Design consideration for a scale-independent cartographic database. Proceeding of the Second Int. Symposium on Spatial Data Handling, Seattle, pp. 384-398.

[7] Kilpelainen, T., and Sarjakoski, T., 1993, Knowledge-based methods and multiple representation as means of on-line generalization. Proc. of 16th Int. Cart. Conf., Cologne / Koln, ICA, pp.211-220.

[8] Map generalization: making rules for knowledge representation. Edited by Buttenfield, B.P., and McMaster, R.B., 1991, 1-st edition, Longman Group UK Limited.

[9] Vijlbrief, T., and van Oosterom, P., 1992, The GEO++ System: an extensible GIS. Proc. of the 5th Int. Symp. on Spatial Data Handling, Chareston, South Carolina, IGU Com. on GIS, pp. 40-50.

[10] van Oosterom, P., 1991, The Reactive-tree - a storage structure for a seamless, scaleless geographic database. Proceedings, Auto-Carto 10, Baltimore, ACSM-ASPRS, pp. 393-407.