MULTIPLE REPRESENTATIONS
AND
KNOWLEDGE-BASED GENERALIZATION
OF
TOPOGRAPHICAL DATA

Tiina Kilpeläinen

Finnish Geodetic Institute
Ilmalankatu 1 A
SF-00240 Helsinki
Finland

Commission III

ABSTRACT

One of the stumbling blocks to the utilization of fully automated geographical information systems is the lack of consistent well-tested data models providing multiple views of topographical objects in different scales. In order to build up a theoretical basis for such models, a systematic analysis of topographical maps is needed. The first part of this paper reviews the problem area. The second part of the paper describes and discusses the results of an analysis of some Finnish topographical maps. The main changes in geographical relations and objects between scales are described. Predicate calculus is mainly used in the formalization phase and as a result we have started an implementation in Prolog++. This Prolog version also supports object-oriented data representation, which, together with rule-based techniques, seems to offer an interesting way of attacking the problems of multiple representations and knowledge-based generalization.

KEYWORDS:

Artificial Intelligence, Cartography, Conceptualization, Generalization, GIS, Knowledge base, Predicate calculus.

1 INTRODUCTION

1.1 General

Nationwide geographical databases will provide opportunities for the effective utilization of topographic data. Our primary goal will no longer be to produce maps on different scales nor will we have to think solely 'in maps'. Rather, we will collect the information we need from topographical databases for different applications. The possibility of choosing a piece of data, exactly the one we want, will make the traditional maps redundant for purposes they were used for before.

However, the demand of different types of data does not imply that all the data should be saved in the databases. The data not available in the database must be transformed and produced from the existing data. At the same time multiple views and representations of the same data should be possible. The structure of the data stored will have consequences for the data that can be extracted from the databases. The challenge we face is to try find a flexible data model according which to store the data. This data model should even be able to satisfy the demand of multiple representations.

1.2 The Objective of This Work

The work described in this paper is the initial part of a research project on multiple representations and generalization of topographical data currently under way at the Finnish Geodetic Institute. The main objective of this research project is to establish a theoretical basis and principals for treating the problem of automatic generalization and multiple representations in connection with topographic data. A data model to be used in utilization of joint use of geographical data bases is of particular interest. The objective of this paper is to identify the main problems in the domain and to suggest ways of attacking them.

The paper is in two parts. The first part reviews the problem area. Some underlying concepts of generalization and multiple representation are discussed and ongoing research topics are referred to. The second part of the paper describes a case study of some Finnish topographical maps. Predicate calculus is proposed as a formal language for topological rules of consistency, and the implementation of Prolog++ is discussed.

2 REVIEW OF GENERALIZATION AND MULTIPLE REPRESENTATION

2.1 What is Generalization?

Generalization has been regarded as one of the most difficult of the cartographer's tasks. The decision about the characteristics of the map objects or the elimination of nonessential features of the map data required cartographers to be closely familiar with the maps they were making (Robinson et al., 1978). However, the authors insist, that complex though cartographic design may be, communication with the graphic symbolism of maps can be learned. They separate the graphic generalization of the map from the generalization made in the real world in connection with mapping. This aspect seems to have got lost in recent years, and research into generalization has been more concerned with the technical aspects of the generalization process.

Robinson et al. (1978) argue that generalization arises from the need to reduce the spatial relations to be more comprehensible. Both reduction and enlargement are means for increasing comprehensibility. The visual importance of the general compared with the specific is emphasized, and thus the grade of the effectiveness of cartographic communication is increased. The variety of modifications that can and must be made as a result of reduction are called cartographic generalization. Robinson et al. list the elements of generalization as follows: simplification, classification, symbolization and induction. In Bertin (1983) generalization is regarded as the spatial equivalent of simplification. Categorizing is taking place in order to simplify. Note that selection is not considered part of cartographic generalization by Robinson et al. This is because the modification of data is involved in the cartographic generalization. Selection is regarded as the intellectual process of deciding which information will be necessary; it does not demand modification, and it can be carried out irrespective of the map format or scale. The argument of Robinson et al. concerning selection is interesting, because generalization is often done in order to reduce data volume. Robinson's statement appears to be somewhat limited, as generalization done in connection with mapping closely involves selection.

Nyerges (1991), in contrast, underlines the different meanings of the term generalization in cartography and in the database literature. He argues that when cartographic generalization is applied to selection, simplification, classification, induction and symbolization, generalization in connection with databases, is a concept having a more general interpretation than some other concept with a more specific interpretation.

2.2 What is Topographic Generalization and Multiple Representation?

Though most cartographers do not regard cartographic generalization as only graphical, both cartography and generalization have conventionally been limited to display problems. More emphasis should be given to differentiating the graphical part of cartographic generalization from the abstract generalization of digital data done in the database, and these problems should be treated separately. We thus want to put the problem into connection with geographical databases.

Topographic generalization is not seen as a reduction of graphic complexity in this context. Rather the generalization of real world entities is emphasized. Several authors prefer to look at geographic generalization than cartographic generalization. Geographic generalization is concerned with the relations between phenomena. Mark (1991) states that the importance of generalization is not only for graphic display but for efficient and appropriate spatial analysis, too. He uses geographic generalization as a concept which aims to preserve the recognizability of geographic features of the real world, and their relations. In this study, topographic generalization is defined as a subclass of geographic generalization with the emphasis on topographical data.

In practice, every map has a size and the dimensionality of spatial entities can be given in scales. The scale of mapping affects what might be presented in the database (Laurini et al. 1992). The statement : "The further away you are the less you see", has began to reflect also into the concepts of the database environment.

Laurini et al. (1992) discuss multiple representation, which refer to the modelling of different forms for different spatial objects or instances of them. In multiple representation particular classes of object are stored in a database in different representations. Different representations have distinct inherent properties in terms of geometric computations, positional error, maintenance of topological consistency and the search for pieces of space.

The multiple representation problem was initialized also in a research program of NCGIA, National Center for Geographic Information and Analysis (Buttenfield et al., 1989, Frank, 1990). A GIS database must be able to represent objects at different levels and to support modification across resolution levels. Starting from object descriptions at each resolution level, the connections between them must be formally described such that changes applied to one can propagate to the others, allowing other resolution levels to be deduced automatically. Multiple representation problems arise during extraction as well as during production of data.

Jones (1991) gives various reasons for storing multiple representations of the same objects in the database. One reason is the relatively limited capabilities of automatic generalization. An aspect discussed by Jones is the extent to which data duplication and data redundancy can be tolerated. By duplication he means that one representation is a simplified version of the other if its geometry is a subset of that of the other version. The smaller scale version may be redundant if an automatic procedure exists for performing the simplification.

Thus, in the present study, multiple representation is considered to be connected to generalization in the database environment. This includes the different data representations that have been stored in the database, as well as the representations that can be deduced from the existing ones. The demand of storing multiple representations of data at different levels will decrease, as the possibilities to deduce the data will increase. Data abstraction models are essentially included in this domain. Resolution, the minimum addressable spatial unit, is related to the scale but it is not the same. In this context the different conceptual resolution levels are referred to, instead of the scales. We thus want to emphasize the utilization of geographical databases for different kinds of applications, not only for such that have been defined beforehand.

2.3 Models for Generalization and Multiple Representations

Several models have been developped for the generalization process. The following examines some of those, developed particularly for the digital environment.

Shea et al. (1989) discuss aspects of digital generalization, asking why, when and how we should generalize. The decomposition of the how aspect is divided into twelve operators: simplification, smoothing, aggregation, amalgamation, merging, collapse, refinement, typification, exaggeration, enhancement, displayment and classification. These operators result in spatial and attribute transformations. The when aspect of the generalization process has three parts: conditions, measures and controls. They argue that the organization of the when and how processes is important if a complete approach for digital generalization is to be achieved.

Brassel and Weibel (1988) make a distinction between statistical and cartographic generalization. Statistical generalization is described as an analytical process having to do with information content reduction in a database under statistical control. Some generalization researchers consider that the Brassel and Weibel model for digital generalization is best suited for the integration of expert systems for generalization tasks (McMaster, 1991). Five processes are identified for digital generalization: structure recognition, process recognition, process modelling, process execution and data display. Structure recognition and identification in geographical neighbourhoods occur at multiple levels of conceptual resolution.The structure recognition problem is the first phase of map generalization.

Bertin (1983) discusses two possible ways of generalizing: conceptual and structural generalization. Though Bertin's model was not especially developed for digital environment, the conceptualization point of view is of particular interest for us. The level of conceptualization implies maintaining the implantation and the planar structure of the phenomenon. In Bertin's model the structural generalization includes both the conceptualization and at the same time simplifying the distribution.

Jones (1991) has presented a concept of deductive knowledge-based system architecture which may provide a suitable basis for building multi-scale geographical databases. The rule base of the deductive system is used to make decisions about the strategies for update and retrieval of the stored objects. His arguments also concern the best way to store the objects to achieve efficient access at different scales. Since multiscale databases may be very large, and objects may occur at widely differing levels of class-generalization hierarchies, data management problems cannot be avoided. Multiresolution data structures should provide rapid access to generalized versions which are geometric subsets, and they therefore prevent data duplication.

Bruegger et al. (1989) propose a formal, dimension-independent approach for building multiple, hierarchically related representations of spatial objects. The same objects are presented in different layers which provide different spatial resolution and degree of detail. The geometry of spatial objects is defined as aggregation of cells on a certain higher level. The layers are interconnected by hierarchical relations between their cells.

The following describes a conceptual model applied to our project for generalization of a GIS database. The basic model was presented by Sarjakoski (1989) and here it is applied to a topographic database. In our model the generalization concept has been divided into three parts: an abstract generalization to be done in the topographic database, then we have the database of modified topographic data and finally, the visual display of the generalized data, Figure 1. It is emphasized that the conceptual generalization should be done before the visualization phase. These parts were often put together, and this makes the problems of generalization even worse. We should, whenever possible, separate the actual generalization problems from the problems related to the graphic display. But at the same time we do not want to disregard the problems connected to visualization. In the generalization phase, additional, often attribute data can be taken advantage of. For example, when storing data on buildings in the database, additional attribute information concerning which blocks the buildings belong to can also be stored. This additional information can then be applied to aggregation problems when, for example buildings should be aggregated into blocks.

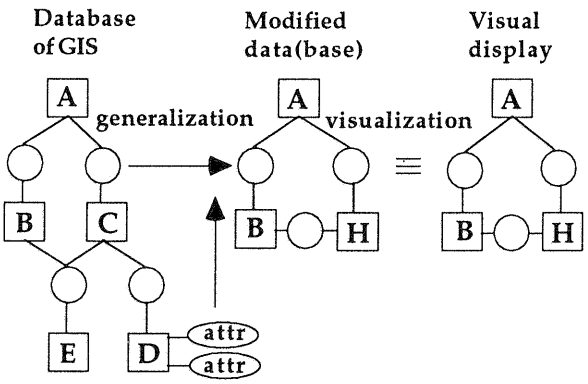

Figure 1. *Conceptual model for generalization of the topographic database (Sarjakoski, 1989). In the notation adopted from Sowa (1984), the rectangles represent entities and the circles relationships.*

2.4 Generalization Research

Three epochs can be identified in automated generalization research (Buttenfield et al. 1991). The first emphasized algorithms for line simplification, the second, algorithm efficiency. The scale dependency was modelled using for example, parametric methods. The third epoch in map generalization research is now in progress, with comprehensive models for formalizing cartographic knowledge, expert systems and knowledge-based methods as on-going research areas. The generalization process has not yet been fully automated, and many researchers wonder whether it would even be possible. This is largely due to the lack of rules and formalism for generalization.

One of the main topics in map generalization today is the establishing of rule-bases for generalization, a task that includes data modelling and representation techniques. This has much in common with other cartographic research areas. In recognizion problems for example, the most critical part of the work, seems to be establishing sound formalism for the problem (Kilpeläinen, 1989).

Generally, establishing rule-bases is a task that seems to be strongly dependent on the application area. In this research project, we therefore try to establish basic rules for topographic data representations that could be applied more extensively to the problem domain.

2.5 Knowledge-based Generalization and Intelligent GIS Databases

Two structures are currently appropriate for storing the knowledge in a rule-based system for conducting generalization operations: rules and parameter tables. Rule-based generalization has been discussed by several authors (e.g Buttenfield et al. ,1991, Muller, 1990).

Production rules in an expert system for generalization have been discussed by Shea (Shea,1991). Production rules can be expressed for example, as logic, network or frames using predicate logic formalism. The advantage of this formalism is the precise and consistent formalism. The disadvantage is that no rule hierarchies exist. Threshold values as presented by some authors, e.g Mark (1991), are important parts of production rules. He discusses the optimal threshold for a given phenomenon, and suggests that common thresholds could be established for groups of rules. Armstrong (1991) discusses the knowledge organization of knowledge-based generalization. Conflicting rules require ordering schemes, which are difficult to develop. Alternative ways are presented using inductive logic and message passing. He says that the IF-THEN structure of rules in a production system is rigid, and maintains that whether the rule is executed or not should depend on how it fares in competition with other rules. Mark (1991) suggests that the importance of the rules can be modelled as a certainty factor. In Shea's proposal (1991) metarules control the priority of generalization rules.

In frame-based systems both declarative and procedural knowledge are maintained. Procedural knowledge is used to guide the selection of appropriate generalization operators and algorithms in a given map context (Amstrong, 1991). Forward-chaining reasoning is a more appropriate choice for generalization since the process is data-driven and situation specific (Shea,1991). The consequences of a starting context are tested by forward chaining . This can mainly be used for what-if reasoning. Backward chaining is interesting for diagnosis, that is, to discover the reasons for the observed situation (Laurini et al. 1992).

Some researchers have started arguing for object-oriented databases for generalization purposes. An object-oriented database environment does seem effective enough to attempt a prototype system (Mark,1991). He proposes that progress will be achieved by attempting to model and generalize real-world objects or features rather than their cartographic representations. An object-oriented approach will probably allow such methods to be implemented and tested. According to Mark a central concern in the object-oriented approach is to identify the specific object classes to be represented, in particular, to find classes of objects with common behaviour. Laurini et al. (1992) also argue that the object-oriented approach is an attempt to improve modelling of the real world. As object-oriented databases were developed as a result of the grafting of several roots, including artificial intelligence techniques, the concept of frames in particular, this appproach seems to be promising for future development of generalization tasks.

Laurini et al. distinguish in object-oriented data model between attributes (declarative knowledge) and methods (procedural knowledge). The term method refers to an operation on the data, a procedure that can be applied to a class of objects. Both methods and attributes can be inherited. The authors suggest that methods and daemons might be used for checking integrity constraints of the spatial database and propose a two-folded knowledge base: meta rules and expert rules. A metarule is a rule concerning knowledge, for instance, a procedure for selecting other rules. The ability to structure user-defined datatypes is also advantageous. Datatypes such as point, line and area could be used, and combined objects could be structured via inheritance.

The other fields of artificial intelligence and expert systems also provide promising tools for intelligent spatial information systems. There are signs that tools for both object-orientation and logical deduction will be used in the systems for spatial problems. We will have hybrid systems in the future, and thus even the problems of generalization will diminish because we shall have additional knowledge for problem-solving compared with the systems today. Intelligent database systems will encompass several new technologies, including object-oriented programming, expert systems and deductive facilities, hypermedia and interfaces (Parsaye et al. 1989 according to Laurini et al. 1992). It now appears to be feasible to design new spatial information system integrating such technologies. Future intelligent spatial information systems must include object orientation, not only for storing and manipulating data but for modelling the real world more adequately; hypermedia and hypermaps; facilities for logical deduction and geometric computation, that is, reasoning facilities; and person-compatible interfaces (Laurini et al. 1992:621).

2.6 Topology in Cartography

Statements on the importance of production rules including topology are found in the literature on generalization problems. In the following some aspects of topology are discussed.

In terms of cartography, the relevant part of topology, is included in algebraic topology (Hohti, 1987). Cartographic knowledge is divided into four parts: metric knowledge (co-ordinates), topological knowledge (neighbourhood relationships), geometric knowledge (shape) and attribute information (characteristics). Traditionally, area models have been restricted to metric and attribute data. In cartographic databases each area has been defined by the co-ordinates of the surrounding polygon. The area concept has been very important in cartography. With the aid of topological neighbourhood relationships, the problems related to areas can be divided into those concerning co-ordinates and those concerning area codes, (Hohti, 1987). Geometric data connect the topological knowledge to the attribute data.

According to Laurini et al. (1992) spatial concepts may also be measured in both geometric and topological domains. The basis for spatial relationships consists of: metric relations (distance , direction), topological relations (connectivity, orientation (to, from), adjacency, containment) and order relation (inclusion). Laurini et al. give the basic spatial units and particular properties related to them: point, line (length, sinuousity, orientation), area (extent, perimeter, punctured/ indented, connectiviness, overlapping) and volume. Spatial entities may be defined as a combination of more than one spatial unit. If these combinations consist of different types of entity, we have complex objects. If the combination consists of different instances of one type, we have compound objects. These are also kinds of complex objects. Taking only pairwise or binary combinations for point, line or area entities, there are nine binary combinations of spatial units (Laurini et al. 1992:82):

- point-point (Is point near to?)
- point-line (Is on? Is near to?)
- point-area (Is point inside a zone?)
- line-point (Does line (railway) serve town?)
- line-line (Do roads cross rivers?)
- line-area (Is river inside an area? Do river and zone boundary coincide?)
- area-point (Does postal zone encompass schools?)
- area-line (Do zones encompass railways?)
- area-area (Do zones touch? Do zones overlap?)

It is often recommended that topological consistency in geographic databases should be checked. A topologically consistent database is defined in Laurini et al. (1992) as having all spatial entities projected upon a plane surface, with no freestanding features, and having complete topology. Completeness of inclusion means that there are no isolated, non-connected points and that all lines are parts of boundaries of polygons. A problem has been that, owing to errors in digitization, for example, the boundaries between two area objects do not join. Holes and overlaps are often produced. The consistency conditions may be seen as constraints for the integrity of a database of cartographic information.

2.7 Spatial Reasoning

Laurini et al. (1992) claim that the complex data manipulation of spatial knowledge requires reasoning more than numerical computing or data retrieval.

Nyerges(1991) argues that one of the main reasons why map generalization rules do not always succeed is because the process relies on context. The main component for understanding context is meaning. Nyerges argues that bringing the geographical meaning on-line is essential to more efefective results in digital map generalization. Too little information about meaning is the current approach and this contributes to the slow progress in digital map generalization."Meanings are not generated spontaneously, but evolve from what is already known, "(Nyerges 1991). In the same way we do not recognize spatial phenomena as isolated entitites, but we perceive the reality as a whole, constructing our spatial knowledge on phenomena we already know. Learning seems to be a key element for spatial reasoning as for any reasoning.

Laurini et al. (1992) discusses the difference between reasoning and problem-solving methodologies. We reason from our experiences or from knowledge originating from our experiences. For this we do not have the knowledge that would be acquired from empirical evidence from previous experiences, nor do we have any algorithm to perform the task. Two possibilities of spatial reasoning are now emerging: topological and pure geometrical reasoning.

Spatial knowledge is knowledge in which the spatial component is important, especially via locators such as coordinatesor place names (Laurini et al. , 1992:643). In generalization problems, we often suffer from lack of spatial knowledge in its context. Questions that should be asked are, what knowledge do we actually capture, what knowledge do we need and how should we represent this knowledge? Spatial knowledge is often spatially fuzzy. For example, shape and location are not always totally known. Spatial knowledge can be deduced from logical deduction. Therefore we need systems to support techniques of this kind. Laurini states that, on the technical side, encoding spatial knowledge is difficult because logic programming and computational geometry must be combined.

The challenge we face is to describe spatial knowledge so precisely that it can be processed by computer. The level of conceptualization has to give possibilities for deducing new information from the existing one. Topological rules of consistency can be used as means for spatial reasoning. Making use of topological relationships together with attribute and metric knowledge will give us better opportunities for checking spatial consistency.

3. OUR CASE STUDY

3.1 Analysis of Topographical Data

In this work we concentrated on analysing three types of Finnish topological maps: the basic map (scale 1:20000), the topographical map (scale 1:50000) and the GT-roadmap (scale 1:200000). The first two are produced by the National Board of Survey and the road map by the Finnish Map Centre. At the moment, a part of the generalization process from the basic map to the topographical map is done automatically. Generalization of contour lines is almost fully automatic, and a part of the road network generalization is also done automatically. Automatic generalization is not used to produce GT roadmaps.

The first part of analysis examines the main object classes. The objects are divided into point, line and area objects, and the changes between scales are studied. The classes of objects are compared in Table 1.

The main changes in the object types of the maps are made for settlements and buildings. The changes are mostly due to selection and aggregation. The small variations between object types from scale to scale might be due to the narrow distribution of scales.

Maptype / Object class	Basic map 1:20000	Topographic map 1:50000	GT road-map 1:200000
Hydrography			
Lake, seas			
-shoreline	—	—	—
-water surface	□	□	□
Rivers,streams, drains			
>100 m			
-shoreline	—	—	—
-water surface	□	□	□
>20 m			
-shoreline	—	—	—
-water surface	□	□	
5-20	□	—	—
<5m	—	—	—
Roads, railways	—	—	—
Settlement			
-scattered	•	• □	• □
-built-up areas	• □	□	□
Fields			
-figure line	—	—	
-field area	□	□	□
Forests, bogs	□	□	□
Rocky ground	□	□	□
Contours	—	—	—

Table1. The main object classes on some Finnish topographic maps. Point objects are symbolized with •, line objects with —— and area objects with □.

The next part of the analysis deals with the changes in topological relations between objects. The basic types of relations considered are those given by Laurini et al. in 2.6. Some topological relations from the maps analysed are listed. The selection and removal of objects are disregarded, and the changes due to these.

1. Crossing roads will cross in any scale.
2. Islands are inside lakes in any scale.
3. Buildings on an island will be on an island in any scale.
4. Roads and railways do not intersect water if there are no bridges.
5. Buildings are not situated on roads or on water.
6. Neighbourhood relations between area objects will remain.
7. Adjacency between line object and point object will remain; buildings are on the same side of the road in any scale.
8. Roads intersect land areas in any scale.
9. Lakes and rivers are situated in the bottom of valleys.
10. Orientation of the objects will remain.

The main rule is that topology does not change from scale to scale. However there are exceptions, depending on how the generalization is done. If there is not enough space on the map, or when particular objects are emphasized, displacement is done. This will result in changes in topological relations in the form of absence of a certain relation. For example, a field area is adjacent to a forest area. A building in a narrow forest area near the boundary of forest and field may be situated in the field area in the smaller scale. This is due to metric changes in the narrow forest area; there is no space left for the symbol of a building in a smaller scale. A similar situation arises when objects are selected and removed according to certain thresholds. The point object may no longer be inside the same land area. In terms of topology changes are however small, which means that the case when a certain relation will be removed, is an exeption.

In contrast the changes in metric and geometric relations are quite obvious. For example, the extent and shape of lakes or settlements are often changed as a result of exaggeration. The extent of roads is changed, particularly in the GT roadmap, after enhancement according to road classes. Simplification and line smoothing in general also change the shape of the objects. Changes due to selection and removal also result in metric changes as in the case of choosing which buildings are to be represented, and how one building is used instead of many to represent several buildings.

3.2 Formalism for Topographic Data

In this section we discuss one way in which statements such as those given in 3.1 can be described formally. Genesereth et al. (1988) argue that intelligent behaviour depends on the knowledge an entity has about its environment. Much of this knowledge is descriptive and can be expressed in declarative form. Our approach is to describe the topological reality with physical phenomena. The formalization of knowledge in declarative form begins with a conceptualization (Genesereth et al., 1988). This includes the existing objects and their relationships: topological and metric. First, we try to identify such objects and the relationships and use these in a formal language called predicate calculus.

Objects can be concrete or abstract. They can be primitive or compositions.It is up to the knowledge representation task to determine which objects in the world considered are of relevance. The set of objects about which knowledge is expressed is called the universe of discourse (Genesereth et al., 1988). The following example examines especially the object road. In this example, we have applied the presentation given by Genesereth et al. into our problem domain.

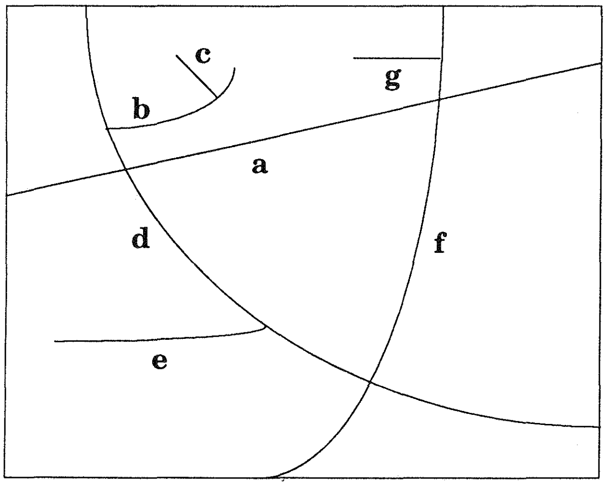

Figure 1. Crossing roads.

We have roads: *{a,b,c,d,e,f,g}* in our universe of discourse. A road is here defined as an arc from a start point to an end point. A function is one kind of interrelationship between the objects in a universe of discourse. The set of functions emphasized in a conceptualization is called the functional basis set. In this case we have the functions *start point* and *end point*. A relation is also a kind of interrelationship between objects in this universe of discourse. The set of relations that we want to emphasize in a conceptualization is called the relational basis set. For example, we have the relation *connect*, which holds between two roads only if the end point of the one is connected to the other road. This could be defined by the following set of tuples:

$$\{\langle b,d\rangle,\langle c,b\rangle,\langle e,d\rangle,\langle g,f\rangle\}$$

The relation *intersect* holds between two roads if, and only if, there is an intersection point between the roads:

$$\{\langle a,d\rangle,\langle a,f\rangle,\langle d,f\rangle\}$$

Besides the relations connect and intersect, we define the relation *on*, that holds between points and roads if the point is immediately on the road. Our conceptualization of the world in Figure 1 is given as a triple of a universe of discourse, a functional basis set for that universe of discourse, and a relational basis set:

$$\langle\ \{a,b,c,d,e,f,g\},\{\ start\ point,\ end\ point\},\ \{connect,intersect\ ,on\}\ \rangle$$

It is important to understand that, no matter how we choose to conceptualize the world, there are other conceptualizations as well (Genesereth et al., 1988). Any conceptualization of the world is accommodated, and we seek those that are useful for our purpose. This raises the question, what is it that makes one conceptualization more appropriate than another for knowledge formalization? One issue mentioned by Genesereth et al. might partly answer this question: the grain size of objects associated with a conceptualization. Too small a grain makes knowledge formalization prohibitively tedious, whereas choosing too large a grain can make it impossible.

The following uses predicate calculus to formalize knowledge as sentences. This seems appropriate to the given conceptualization of the world. One source of expressiveness according to Genesereth et al. is the availability of logical operators that allow us to form complex sentences from simple ones without specifying the truth or falsity of the constituent sentences.

The following sentence defines an object road as an object having a start point and an endpoint:

$$(\ \forall x\ (Road(x) \Rightarrow (\ End\ point(x)\ \Lambda\ Start\ point(x))\ )$$

The relations connect and intersect can be defined in the following way. The first sentence means that if the road x is connected to the road y, the end- or startpoint of the road x are on the road y.

$$(\ \forall\ x,y\ \ (Connect\ (x,y) \Rightarrow On((End\ \ point(x)\ V\ Start\ \ point\ (x)),y)$$

The relation intersect is defined by the sentence which means that if roads x and y intersect there exists a point z that is on both roads x and y.

$$(\forall x,\ y\ \ (Intersect\ (x,\ y) \Rightarrow\ \ (\exists\ z\ (On\ (z,x)\ \ \Lambda\ \ On\ \ (z,y)))$$

This example should not be regarded as a complete description; it is only meant to be an illustration of the expresiveness of predicate calculus.

3.3 Prolog++ for Implementation

We have started to implement the topological constraints of consistency, which are shown in 3.1, in Prolog++ (Vasey et al., 1990), which is based on two technologies: artificial intelligence and object oriented programming systems (OOPS).

Prolog++ inherits the problem solving and database characteristics of Prolog, and also the methodology of an OOPS approach. Compared with Smalltalk, which was one of the first object-oriented languages, Prolog++ combines a declarative language with OOPS, whereas Smalltalk has its origin in procedural OOPS. The following examines some of the features in Prolog++ according to Vasey et al.

Prolog++ incorporates the key concepts generally included in most OOPS environments. The main concepts are encapsulation, class, class hierarchy, inheritance, object, message, method and polymorphism. For generalization task some of these concepts are of particular interest. There are two main aspects to OOPS: organization and communication. Organization takes the form of placing classes within class hierarchies, and communication is carried out through sending messages between the objects of those classes. Polymorphism refers to the ability to use a single name for different methods in different classes. This allows the same message to be sent to several different objects, each having its own set of methods for handling the message.This ability of message sending will be used in our implementation. In generalization task it is conspicuous how modifications of one object will result in modifications of other objects and how different objects will require different kind of modifications though the reason might be the same. However, this argument mostly concerns the metric and geometric changes of objects and are due to the display problem. The message sending concept can also be used to describe the topographical world and its consistency. The messages can thus be used to tell which of the topological constraints will remain or should be removed.

Conventional programming generally concentrates on procedures and procedure calls, while data are often used as arguments to procedures. In OOPS, classes have both procedures, known as methods, and data, known as attributes. Both singular and multiple inheritance are supported. The inheritance mechanism of Prolog++ is based upon the same searching principles as the underlying Prolog system. It is often argued that generalization is data-driven procedure, and therefore there standard Prolog is not always best suited. The paradigm of data-driven programming is expressed in Prolog++ by the use of daemons, which interrupt the main process. Daemons are only activated when certain events occur. It is stated earlier in this article how daemons could be used as metarules for the generalization process. In generalization tasks, we could use this ability as a control mechanism for the decision which rules are to be applied when certain events occur.

The following describes some of the features of Prolog as a logic-based language. Being modelled on predicate calculus, Prolog has a sound theoretical basis. This is one of the reasons why we use Prolog++ in our implementation. In theory, it is always possible for a Prolog program to be proven consistent or inconsistent. Prolog attempts to find a solution by using a built-in inference engine, which automatically infers the solution to a given query using the facts and rules defined in the program. The declarativity of Prolog programs means that the formal definition of a problem may be used as a functional program to solve that problem. In Prolog we declare a given problem domain rather than state the procedural steps for solving the problem. This is useful when we know various aspects of the problem but little about how to find a solution. In our case, when trying to implement the topological constraints this feature of declarativeness is helpful. The formalism inbedded in Prolog itself, can at the same time be used as a formalism for our topological constraints.

One of the advantages of using Prolog as the basis of an object-oriented language is that in a Prolog-based OOPS not only can we dynamically add classes to a class hierarchy, but we can dynamically add classes that have already been established. This means that when a new characteristic is found for an existing object, it can be added to the class structure of that object at run-time. This feature can also give advantages in connection with generalization. For example, aggregation might result in new characteristics for the object: a set of buildings become settlement.

In contrast the introduction of an OOPS methodology to Prolog can be beneficial, for example, as OOPS provides a clear and intuitive structure for programs in the form of class hierarchies. Especially, in large program applications, Prolog has been claimed for bad structure and loose organization. More efficient and easily manageable program structures are archieved by the class hierarchy concept. The ability to define an object hierarchy allows the programmer to partition Prolog programs into dynamic modules which reflect the structure of the problem domain. Prolog++ can be fully integrated with standard Prolog programs.

The next steps for the implementation part in this research project will be to intialize the objects, which messages they should send and to which objects and what would be the individual operations for the objects to be activated. The hierarchical structure of the multiple representation resolution levels will be designed as well the organization of objects. Of particular interest is also, which kind of meta-rules should control the generalization process and its phases.

4 DISCUSSION AND CONCLUSIONS

One of the most difficult problems in topographic generalization is identifying the levels of complexity which should be considered at each conceptual resolution level. Another problem is the appropriateness of the conceptual resolution levels chozen for topographic maps. A greater effort should be made into understand the generalization problem separately from the cartographic context, as a problem of spatial changes varying from one representation resolution to another. Theoretical understanding of the function is essential.

Our intention is to put the problem of generalization into the context of geographical databases. The concept of multiple representation emphasizes generalization aspects in the database environment. Modelling reality, rather than the graphical interpretation and visualization of reality, will give support to the generalization problems. Considering the topological relations between the objects and between the different resolution levels will provide a way of checking the consistency of the database.

But how to model reality? Generalization tasks are often expressed in two dimensional domains, but in modelling reality, our attention focuses on the third dimension, which can be gathered from databases. Perhaps this will be the next epoch for generalization research. Formalization of topographic knowledge is still a difficult problem. It should also be mentioned that formalization can be done in several ways. What we have to find is the way best suited for the generalization task.

This paper has briefly considered one way for formalization, predicate calculus. Although based on a sound theory, it has disadvantages. The expresiveness is not always enough to model real topographical phenomena. It seems that no formal language alone will be powerful enough, which is why hybrid concepts should be used. One point is that generalization tasks will become easier in the future, when additional information on the objects stored in the databases will help reasoning.

In the real world phenomena depend on each other. Therefore in generalization tasks much has to be done iteratively and simultaneously. Stepwise generalization demands an exact schema according to which the generalization steps must be performed. Our suggestion is that we should use the message sending concept of OOPS to solve the problem.

While this article was being written, we started an implementation in a hybrid language, Prolog++. This kind of language, based on the two paradigms of OOPS and artificial intelligence, may help to handle the problems of generalization tasks. But how long this tool will help the generalization problems remains to be seen. We argue that the main solution to generalization problem, as well as to multiple representation, does not depend which tools we chooze. Taking advantage of the topological constraints of consistency should be put more emphasis on. The best way to solve a problem is to remove it. This should be done in generalization, too. Talking about the abstract generalization in databases, we would say that the generalization problems will become much more easier when the ability to use additional information will be utilized.

When it comes to concepts, techniques and tools, we might say that the main problem for generalization today is neither the techniques nor the tools but the concepts. The problem has returned to its source: how can we model reality?

ACKNOWLEDGEMENTS

I wish to thank Professor Tapani Sarjakoski for his encouragement and valuable comments on this article.

REFERENCES

Armstrong, M.P., 1991, Knowledge Classification and Organization, In: Buttenfield, B., McMaster , R.B., Map Generalization, pp.86-102.

Buttenfield, B., McMaster , R.B., 1991, Map Generalization, 1st edition, Longman Group UK Limited.

Bruegger, B.P., Frank, A.U., 1989, Hierarchies over Topological Data Structures. In: Int Arch. ASPRS/ ACSM Annual Convention, Baltimore, pp. 137-145.

Frank, A.U., 1990, The National Center for Geographic Information and Analysis in the United States of America. In: Int. Arch. FIG XIX, Helsinki, Finland, Commission III, pp. 50-64.

Genesereth, M.R., Nilson, N.J., 1988, Logical Foundations of Artificial Intelligence, 2nd edition, Morgan Kaufmann Publishers, Inc.

Hohti, A., 1987, Kartografinen informaatiojärjestelmä, Helsingin yliopiston maantieteen laitoksen julkaisuja B29, (in Finnish),pp.21-41.

Jones, C.B., Database Architecture for Multi-Scale GIS, 1991 , In: Int. Arch. ACSM-ASPRS, Baltimore , Vol 6,pp. 1-14.

Kilpeläinen, T., Automatic Object Recognition from Vectorized Line Maps, 1989, In: Phogrammetric Reports No 55, Transactions of the Royal Institute of Technology, Department of Photogrammetry, Sweden, 1991.

Laurini R., Thomson D., 1992, Fundamentals of Spatial Information Systems, 1st edition, Academic Press Limited.

Maanmittaushallitus, Karttaosasto / NPK-työryhmä, 20.1.1992,Perusmaastotiedot eli numeerisen peruskartan kohdemalli, luonnos, (in Finnish).

Maanmittaushallitus, Topografikunta, 7.2.1990, Topografisen kartan 1:50000 kuvausohjeet, (in Finnish).

Mark, D.M., 1989, Conceptual Basis for Geographic Line Generalization, In: Int. Arc. AUTO-CARTO 9, Baltimore,pp.68-77.

Mark, D.M., 1991, Object Modelling and Phenomenon-based Generalization, In: Buttenfield, B., McMaster , R.B., Map Generalization, pp.103-118.

McMaster, R.B., 1991, Conceptual Frameworks for Geographical Knowledge, In: Buttenfield, B., McMaster, R.B., Map Generalization, pp.21-39.

Muller, J.-C., 1990, Rule Based Generalization: Potentials and Impediments. In: Int. Arc. 4th International Symposium on Spatial Data Handling, Zurich, Vol 1, pp. 317-334.

National Board of Survey, 1989, Cartographic Devoplepment and GIS Activities in Finland, Finnish Proceedings of the 14th International Cartographic Conference in Budapest, Helsinki.

National Center for Geographic Information and Analysis, NCGIA, Buttenfield, B.P., Delotto, J.S., 1989, Multiple Representations, Scientific Report for the Specialist Meeting, Technical Paper 89-3.

Nyerges, T. L., 1991, Representing Geographical Meaning, In: Buttenfield, B., McMaster , R.B., Map Generalization, pp.59-85.

Parsaye, K., et al., 1989, Intelligent Databases, Object-oriented, Deductive and Hypermedia Technologies, New York, USA:Wiley.

Robinson, A., Sale, R., Morrison, J., 1978, Elements of Cartography, John Wiley & Sons, Inc.

Sarjakoski, T., 1989, Presentation on Knowledge-based Generalization, A seminar on Knowledge-based Systems in Mapping, (unpublished), Helsinki University of Technology.

Shea, K.S., 1991, Design Considerations for an Artificially Intelligent System, In: Buttenfield, B., McMaster , R.B., Map Generalization, pp. 3-20.

Shea, K.S., McMaster, R.B., 1989, Cartographic Generalization in a Digital Environment: When and How to Generalize, In: Int. Arch. AUTO-CARTO 9, Baltimore, pp.56-67.

Sowa, J.F., 1984, Conceptual Structures, Information Processing in Mind and Machine, Addison-Wesley Publishing Company, Inc..

Vasey, P., Spenser, C., Westwood D., Westwood, A.,1990, Prolog++, Version 1.0, Programming Reference Manual, 1st edition, Logic Programming Associates Ltd, England.