PROCESSING SPATIAL DATA ON THE INTERNET

Abdulvahit Torun (*), Barend Köbben, Rob Lemmens
Department of Geoinformatics
International Institute for Aerospace Survey and Earth Sciences (ITC)
Enschede, the Netherlands
{torun, kobben, lemmens}@itc.nl

KEY WORDS: Internet, GIS, Mapping, Spatial data, Visualization

ABSTRACT

With the enabling Information Technology, disseminating every kind of information through the Internet became an enforcing condition for information providers. Now, it is more than being fashion and fun. Serving information on the Internet is a demand from the consumer as a commercial necessity. Internet allows all levels of society to access spatial data and even interactive mapping facilities that are available through the Web. Furthermore, Internet reshapes conventional characteristics of accessing, sharing, dissemination, visualization and analyzing conventions of spatial information handling. With today's information technology support, user requirements determine remote geo-data processing as a need. In this paper a description is given of an application aiming at providing spatial data processing functions on the Internet is implemented using topographic data. In this framework, an architecture with the capability of services such as mapping, analyzing spatial data and dynamical visualization of results through the Web is designed for the users who reach the Internet with an ordinary Web browser. The architecture is implemented as a client/server model by using ESRI Map Objects, Map Objects Internet Map Server (IMS), Visual Basic and HTML.

A Generalization model for Top10 vector data of the Dutch Topography Service (TDN) is built based on spatial data coding structure. Generalization is achieved by thematic feature selection operations considering graphical density in the generalized views. Although the model should not be seen as a substitute for full cartographic generalization, its simplicity makes it applicable. Generalization operations are activated as the user zooms in/out. This operation is named 'dynamic generalization'.

1 INTRODUCTION

Internet became a cheap and easy way of disseminating any type of data. An Internet user with a Browser is provided to access mapping and GIS applications residing in the Web servers. The user is enabled to process spatial data with GIS application modules on the Web without having them on the local machine. Although standard web functionality does not support GIS operations, there are techniques enabling the browser with some simple querying and visualization functionalities (Boehner et all, 1997; Augusto et all, 1997).

2 EXTENDING INTERNET CAPABILITIES TO PROCESS SPATIAL DATA

Internet GIS is a network oriented GIS tool that uses Internet as a major means to access and transmit distributed data and analysis tool modules, and to conduct analysis and visualization (Peng, 1997). However, Internet mapping is a mapping service with some interactive control by the client.

Acquisition, analysis and treatment, interpretation and visualization of geographic data is not only very costly but also requires specific expertise and technical equipment. In order to access remote geographical data and query it, a user friendly and platform independent interface is needed (Boehner et all, 1997).

2.1 Specifications of Internet GIS services

A mapping/GIS application on the Internet should be efficient to respond to client requests in time and effective with more mapping/GIS functions to supply user demand. The goals of a public GIS application on the Internet are set as follows; it should be simple enough for inexperienced users, the simplicity should not mean non-functional and it should support non-specialist users to make decisions. Users from different level of knowledge and experience should be able to select necessary tools for their specific application. The application should enable the advanced users to

(*) Currently at General Command of Mapping, Turkey, atorun@hgk.mil.tr

customize it for their own purpose, finally it should enable the users to communicate with other users in a distributed data and process environment (Stasik et all, 1998).

In several current Internet GIS applications the sequence of processes is as follows. When a request is sent to the Web server that has a link to an Internet GIS application, first query parameters are transferred to a GIS engine, afterwards, the query is executed by the GIS software and the result is summarized into a single image which is then exported and transformed to be visualized in a standard graphic image file as a result page on the WWW interface. Other Internet GIS applications use more sophisticated forms of data transfer.

3 GENERALIZATION IN WEB ENVIRONMENT

One of the challenging problems in mapping is producing accurate maps containing a proper density of information at different scales. Changing the visualization to represent maps at smaller scales is called cartographic generalization. Indeed changing the scale is not the only reason for generalization. Semantic modeling determines the content of generalization (Dettori et all, 1996). Therefore, another reason for easing the information density is particularly emphasizing on some of the information.

3.1 Concepts in Generalization

Generalization, in a GIS environment can be defined as the process which generates a derived data set with more desirable and usually less complex thematic and geometric properties compared to the original data set. Determination of the desirable level of detail of a data set and representing the derived data set with effective communication tools are main processes in generalization. These two processes of generalization are named as 'conceptual generalization or thematic information abstraction-' and 'graphical or cartographic generalization-'. The former should be done before the latter. The user needs are fulfilled within conceptual generalization process. Visualization problems due to graphical limitations by taking into account the target scale are solved within graphical generalization process (Molenaar, 1999; Smaalen, 1996). An object has two components such as thematic and geometric. Conceptual generalization is a process on thematic component of the object. Cartographic generalization mainly affects geometric component.

Most of the authors tend to divide generalization operations in six parts. These are 'simplification', 'selection', 'symbolization', 'exaggeration', 'displacement' and 'aggregation' (Robinson et all, 1995; Dettori, 1996). Class hierarchies are used to do conceptual generalization by means of generalization, aggregation and selection. *Class Generalization* is expressed with 'is a' relation such that 'street is a type of road'. *Aggregation* is described as a relation with the meaning 'is part of' such that 'street is part of residential area' and 'building is part of residential area'. Aggregation process is based on spatial relation between objects of aggregated classes. Irrelevant objects or classes are eliminated considering target scale or context of the map within the frame of *selection* operation.

Moving from larger to smaller scales requires *symbolization* process while preserving the topological consistency. Transition from large scale to small scale requires change in accuracy and resolution for graphical limitations. Information overload not considering drawing limitations may lead to generate an illegible map. *Simplification* is expressed as eliminating a number of points of lines to get more smooth shapes. Despite being too small at the generalization scale, some features are preserved due to their semantic importance by means of *exaggeration*. Objects are *displaced* to prevent overlapping considering readability of the final map.

3.2 Generalization in Web Environment

In the Web environment, the user is limited to use the operations defined in advance. The question is 'How the generalization operations can be defined before the application system is designed'. In other words, to what extend the generalization operations fulfill the requirements of the Web users which are mostly application dependent. Therefore, specifications of generalization are defined considering the user group and purpose of the mapping service.

The specifications which help to determine generalization operation are map purpose, target scale, reading conditions, output specifications and device capabilities, accuracy requirements, input data, data model of the spatial database. Graphical limits considering feature classes and default symbology are determined by using these specifications (Brazile, 1998).

In spite of those specifications to design the generalization model, the main aim is efficiency in time in Internet mapping applications. Internet users are not patient to wait for a long time. They may prefer less accurate results in a short time rather than an accurate map taking more time. Perhaps, this is the the main issue to be considered for a Web mapping service having generalization functionality. Time limitation forces to emphasize on conceptual generalization operations which are accomplished on the database rather than cartographic generalization operations, which needs human interaction and requires more time. Although, there are algorithms for simplification, displacement and exaggeration operations where local characteristics of the map and information density are needed to be known.

3.3 Generalization Operations and Multi-scale Approach

Most of the GIS problems in reality are solved by using information in different scales. For instance, vegetation planning requires information about cadastral, soil, climate, ecosystem and administrative units. GIS analysis can be accomplished by introducing each database at different scale levels. Multiscale approaches help to determine the processes handling each information at different scales. Multi-scale database allows the user to represent the same phenomenon of the real world at different levels of precision, accuracy and resolution (Molenaar, 1999). There are methods of building multi-scale database. These are;

Cartographic generalization: generating databases at smaller scales from a large scale database.

Thematic feature selection : features are selected to be represented at smaller scale without considering topological relations in between. The low level operations of this method can be constructed by using database generalization operations.

Integrating different models of the real world : Homologous objects in different representations of the same phenomenon are linked to each other (Devogele et all, 1996).

Last two methods can be named as 'conceptual generalization' in which database generalization operations are handled in a GIS environment. Classification and aggregation hierarchies are considered to be the link between different scale levels in database generalization operations. Thus, by using these links small scale database can be generated from large scale ones. This approach reduces information redundancy and in-consistency between different representation of the world at different scales.

Topographic databases usually have only one representation of the real world at one scale. Multi-scale topographic databases can be derived from a large scale topo-database with the help of database generalization operations. The operations are modeled by using expert knowledge and logic rules.

3.4 Proposed Generalization Operations in Web Environment

The proposed system is an Internet Map Service with generalization functionality. The generalization operation is activated by an intelligent zooming action. From a visualization perspective, the intelligent zooming tool helps the user to visualize necessary classes and features due to generalization model at each scale or scale interval. The generalization operation, which is activated by intelligent zooming is named as *'dynamic generalization'* within the context of this study. The method is designed for dynamic visualization purposes.

Five generalization operations are defined which are applicable in Web environment taking the limitations of the Internet into account. These operations are run on the database. Cartographic generalization process is out of the scope of this study. However no attention is paid for preserving topological consistency. For instance, erasing a feature changes the topology. Interaction between generalization operations and topological relations and metric properties of the objects are analyzed by Dettori et all (1996). The operations are as follows;

Theme selection : Considering the application goals, a subset of feature classes are selected. For instance, an application relevant to roads may not need the classes about vegetation.

Feature selection : Features having a particular geometric or thematic property are selected out of the elements of that feature class.

Changing thematic resolution : It is a generalization operation considering semantic data modeling. Objects of lower level classes are represented as objects of more abstract upper level classes due to class hierarchy or with respect to the values of a set of attributes.

Reclassification : A new feature class can be created by changing the theme using an attribute of the existing class. Therefore the class hierarchy is destroyed. For instance, administrative boundaries can be reclassified due to values of population density attribute, provided that population density is homogenous in each administrative region.

Aggregation : Features from different classes are combined due to their semantic or geometric properties to create a new feature class (Peng et all, 1996).

4 AN INTERNET MAP SERVICE APPLICATION WITH DYNAMIC GENERALIZATION FUNCTIONALITY

The Internet GIS/mapping configuration implemented has a three-tier model. The Map server side is a PC with Microsoft NT operating system. The Web server is communicating with Map servers over TCP/IP protocol. The Map server has mapping applications to serve them to the clients. The client side is an ordinary Web browser such as Microsoft Internet Explorer or Netscape Navigator.

In spite of complex mapping/GIS functionality demands from the users, they indeed expect the more basic functions such as browsing the map interactively and visualizing within a short time. The other needs for GIS functions like analyzing and exploring the spatial data comes later. Although, there are tools to control visualization and selecting feature classes, generalization is handled automatically referring to an intelligent zooming function.

With the help of computer, a map can be zoomed in even exceeding its original scale. For sure, it is meaningless to display a map at a scale larger then its resolution. For the conventional paper maps, this problem is solved in advance, because the map is already generalized into the new (smaller) scale before it reaches the user. Therefore, the map users never think about to re-generalize or enlarge it. In computer display, problem is completely different than paper maps. If the user zooms out, the map becomes illegible. If the user zooms in lower than original scale, he/she sees unnecessary details, which doesn't give a higher metric accuracy.

4.1 Dynamic Generalization Model and Data Characteristics

Defined generalization operations in Web environment are constituted by using a hybrid approach such that theme selection, feature selection and reclassification operations are handled by querying the Top10 vector database. However, changing thematic resolution and aggregation operations are formed by using visualization tools for only display purposes. Theme selection and reclassification operations are interactively controlled by the user. The user should decide on selecting feature classes relevant to the application. Reclassification operation is accomplished by selecting an attribute and clustering the features with respect to the unique attribute values provided that attribute values are homogenous in each geometric shape. The interface for reclassification operation is given in Figure 1. Feature selection operation handles two issues. These are selection of features considering class hierarchy and graphical density of the features to be displayed at a certain scale with a legible view. Graphical density for each scale level is computed by using data clustering methods explained in sub-chapter 4.4. Class boundaries and class ranges are used to determine the features to be displayed at each scale. Graphical limitations of Web environment are taken into account to determine number of classes and to determine the scale intervals.

The technology used for Web application manipulates only image formats. Therefore the user is not aware of how the view of the database is generated. Because, each time user interacts with the view, a new process is started on the server and the results are sent back to the user. This non-transparent interface between the client and the server –indeed it is a handicap- enabled to do changing thematic resolution and aggregation operations with cartographic modeling tools. The same symbols are assigned to those feature classes to be aggregated or to change thematic resolution. Figure 3 is the user interface to do these operations based on graphical representation.

Dynamic generalization method enables adjusting information density at different scales, selecting appropriate feature classes and visualizing dynamic results. To avoid illegibility at small scales and unnecessary details at large scales, "dynamic generalization" creates appropriate views at different scales. For instance, objects with top priority can be displayed at a very small scale and all the data set can be displayed at the original scale of the database or little larger scale than it. The user is never aware of this transition. Here, application of generalization is based on only 'thematic feature selection' operation, which is one of the methods to create multi-scale database (Devogele et all, 1996). Although, it doesn't satisfy the cartographic generalization rules and topological consistency, simplicity makes it attractive to apply.

4.2 Data Characteristics and TDN Data Coding System

Considering the characteristics of the Top10 database- only the unique data coding system (Tdn_code) helps to re-classify the data into classes. For instance, 'Tdn_code' is the item where every object has a value with respect to class extension rules. A road with certain attributes can be classified into any one of the road classes. So, this crisp data classification is the only issue to classify or reclassify the data into new classes. The hierarchy of importance within the road data and their codes are given in Table 1. The column 'Proposed classification' is a new classification that discards some of the attributes to get a more abstract view of road network in reality. This classification is used for feature selection operation.

The simplest way of doing "dynamic generalization" is applying selection operation considering graphical density on the corresponding view. For the case of *road network*, highways, main roads, secondary roads and streets and cycle paths can be represented at decreasing scale intervals respectively. Selection of scale thresholds, scale ranges and level of information density seems to be not a generic problem for different applications. But there are some other affecting parameters such as technical (screen size, computer power) issues and scale of original spatial database.

4.3 Data Clustering Methods

Two different approaches are experienced in the context of ‘dynamic generalization’. Data clustering methods that are mostly known for histogram generation are used to compute a number of features to be classified at different scale ranges. These methods are ‘equal intervals’ and ‘natural breaks’ which are broadly used in statistical mapping (Kraak et all, 1996).

Tdn _code	Tdn_code explanation of features	Proposed classification
02000	Asw (highway)	Highway
02100	Aw-rood (connections)	Highway
02200	H8 (main road > 8 m)	Main road
02300	H > 7 (main road > 7 m)	Main road
02400	H > 4 (main road > 4 m)	Main road
03000	V8 (secondary road > 8 m)	Secondary road
03100	V > 7 (secondary road > 7 m)	Secondary road
03140	V > 4 (secondary road > 4 m)	Secondary road
03300	V > 2 (secondary road > 2 m)	Secondary road
03400	Ow3 (poor road)	Poor road
03410	Gv3 (poor road)	Poor road
03470	Voetgangersgebied	Walking path
03530	Straat (street)	Street
03600	Fietspad (cycle path)	Cycle path
03900	Parkeerterrein	Parking Place

Table 1. Information content and their Id numbers (Tdn_code) and a proposed new classification strategy of the data to be used in ‘dynamic generalization’

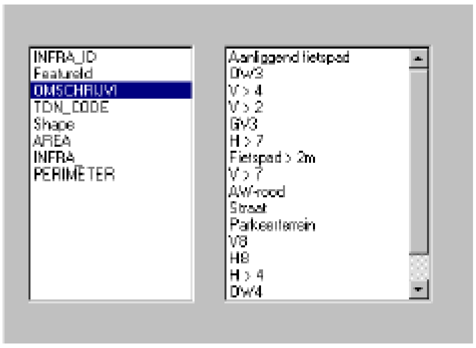

Figure 1. List of items and the unique item values for ‘ValueMapRenderer’ operation

Actually, It should be clearly mentioned that proposed methods are very specific to this particular application and appropriate for road data of TDN Top10 vector database. Thus, they are not generic methods for broad range of spatial databases. Some other more sophisticated data models such as tree structures and searching algorithms for tree structures may be more efficient (Oosterom, 1999). There is no mature technique for dynamic generalization. The clustering methods are;

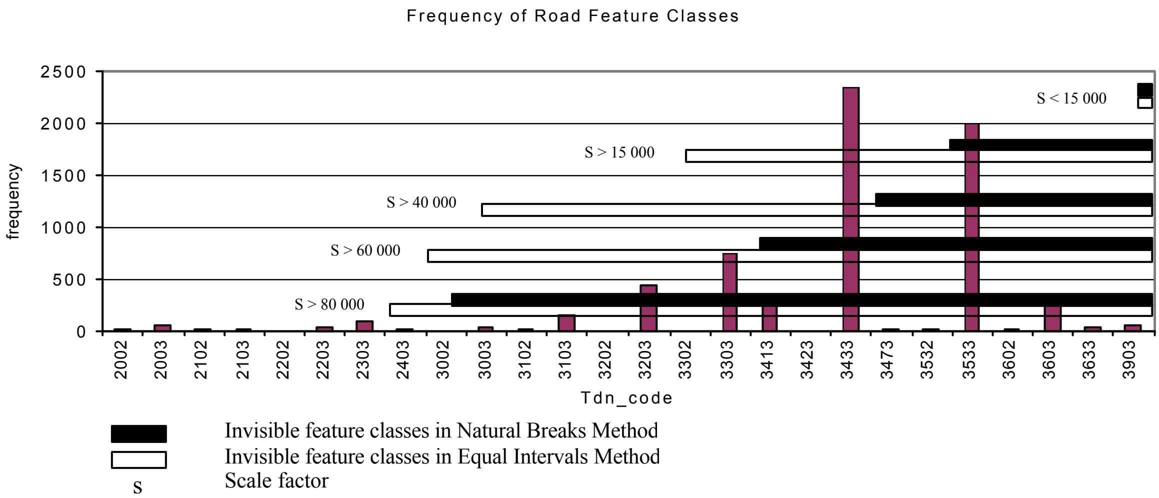

Figure 2. ‘Tdn_code’ (Dutch Topography Service spatial data codes) values, their frequencies and invisible features due to scale interval values.

Equal Intervals Method: Minimum and maximum values in the range of attribute values are taken as boundaries. The range is split into pre-defined number of classes and class boundaries are determined.

Natural Breaks Method: There are as many unique attribute values as the number of different feature types in the domain of item. Therefore, data should accumulate at certain points on the histogram. These points are ‘Tdn_code’ values on the value scale, which can be defined as natural break points. After determining these points class intervals and class boundaries are calculated.

Although, the first is calculated in run time, the latter is defined before compilation. For equal intervals method, calculation of class intervals and class boundaries are easy to cluster data for scale intervals. But, the natural breaks method needs to develop an algorithm. As can be seen from figure 2 and actual map display, for this particular dataset the 'Natural breaks' method appears to have the smoothest transition of visible and invisible features between subsequent scales.

In this application, dynamic generalization is performed by painting (covering) the unnecessary features with the background color. This creates the problem of how the correlation between visible layers should affect generalization, generalization methodology and transition thresholds. With many layers on the display, the unnecessary features painted with background color covers features on the other layers. Therefore there is a loss of display.

4.4 Visualization Modeling

The user interface enables the user to select 'geographic extension', 'layer selection (raster, vector)', 'attribute selection (legend)', 'symbol (area, line and point types) and color selection (web safe colors)', selecting vector (ESRI formats) or raster files in the server, zooming functions, turning on/off layers, changing layer visibility, accessing object's attributes by connecting the database, dynamic generalization of the theme using the original data by considering the magnification level.

Limitations of the Internet from Cartographic perspective are ; communication on the net fully depends on the storage size of the information, the storage size determines the content of information deployed on the Internet, graphics supported by Web Browsers are raster formatted (JPEG, GIF), unless the browser is extended with some plug-ins or applets for vector format, colors used on Internet are not limitless as being on computers (216 colors out of 256 used in web site design are called 'web-safe palette'), a Web map must be small in size to fit heterogeneous screen size of the users. Specifications of the application are given below;

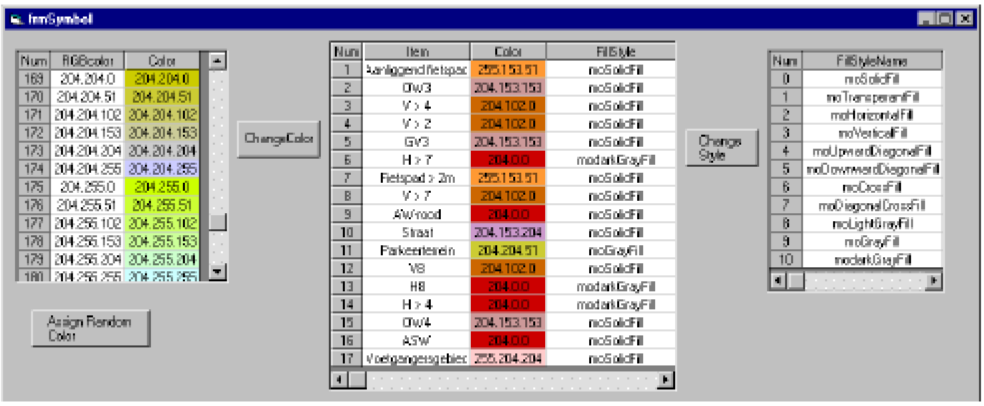

Figure 3. Web Safe colors (left hand side), MapObjects symbol style types (right hand side) and unique item (attribute) values with their assigned color and style parameters (at the middle)

A Web page has been designed with five frames such as frame of descriptive information, frame of control tools, frame of legend information, frame for visualization tools, frame of the map display. First two frames are static, the others are being dynamically created due to user requirements.

4.5 Implementation

Topographic data (transportation, hydrology, vegetation, built-up area, elevation and satellite images) of the Netherlands in neighborhood of Enschede are used in the implementation phase. The display screen (interface) lets the user visualize different layers of the topographic database. Two generalized views of the Top10 database at scales 1:61000 and 1:31000 are given in Figure 4.a and 4.b. At scale 1:61000, only highways and main roads are selected out of the feature space. However, at scale 1:31000 highways, main roads, secondary roads and streets are selected to be displayed. The data is reclassified referring the proposed classification in Table 1. The generalized views are satisfactory, considering the simplicity of the approach and remembering visualization goal on the Internet.

The research is planned as two phases; the first is implementing a mapping/GIS application that handles Top10 vector data with dynamic generalization functionality and the second phase is deploying the application to the Internet as a Web mapping/GIS service. The first phase is accomplished, however the second part is not yet. The application mainly focused on dynamic generalization of the map features and visualization. using the specifications defined above. ESRI

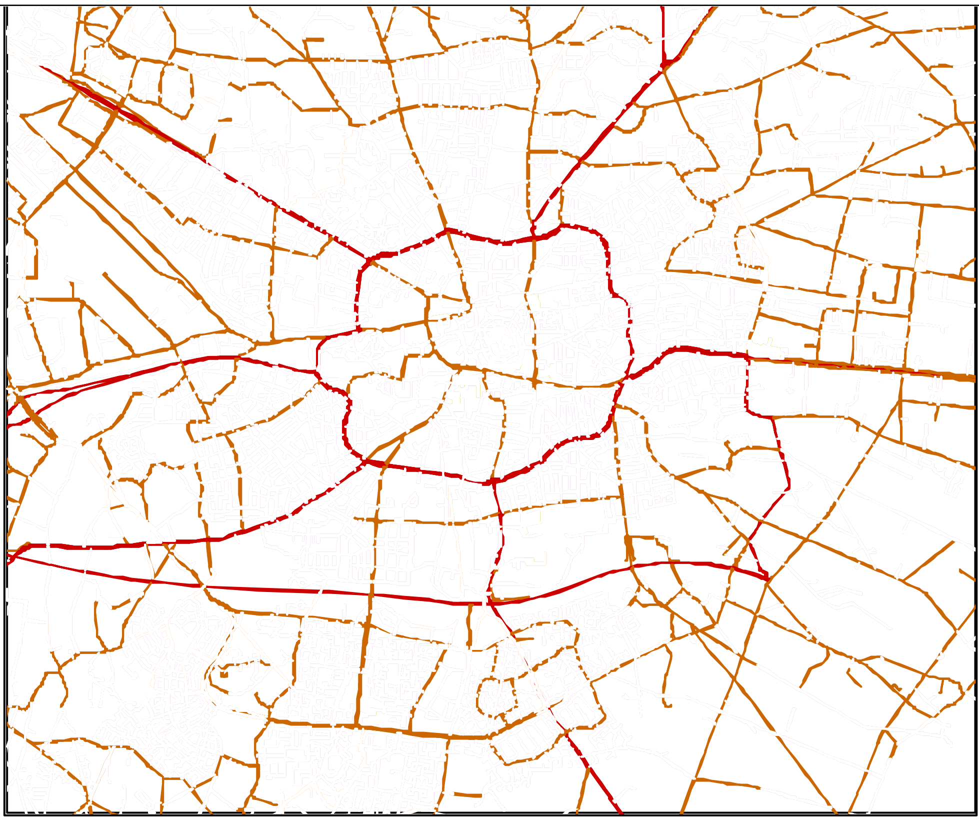

Figure 4a. Generalized Road Data of the city of Enschede at scale 1 : 61 000

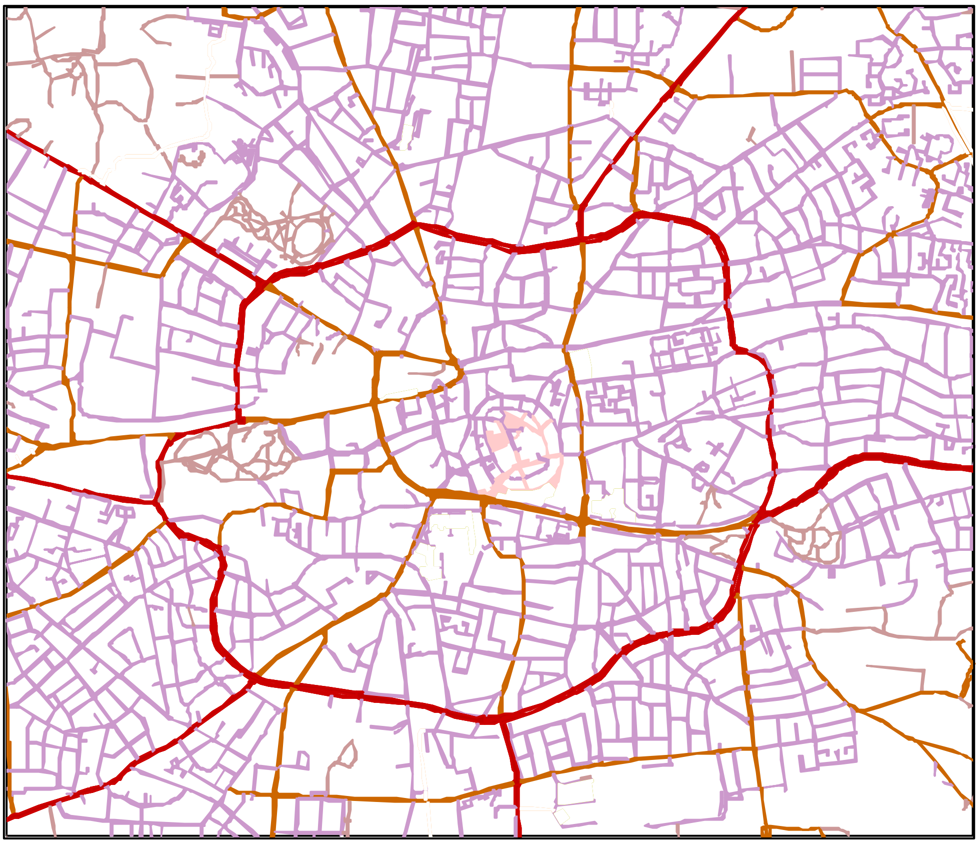

Figure 4b. Generalized Road Data of the city of Enschede at scale 1 : 31 000

MapObjects and Microsoft Visual Basic softwares are used by embedding the first phase into the second phase (Torun, 1999).

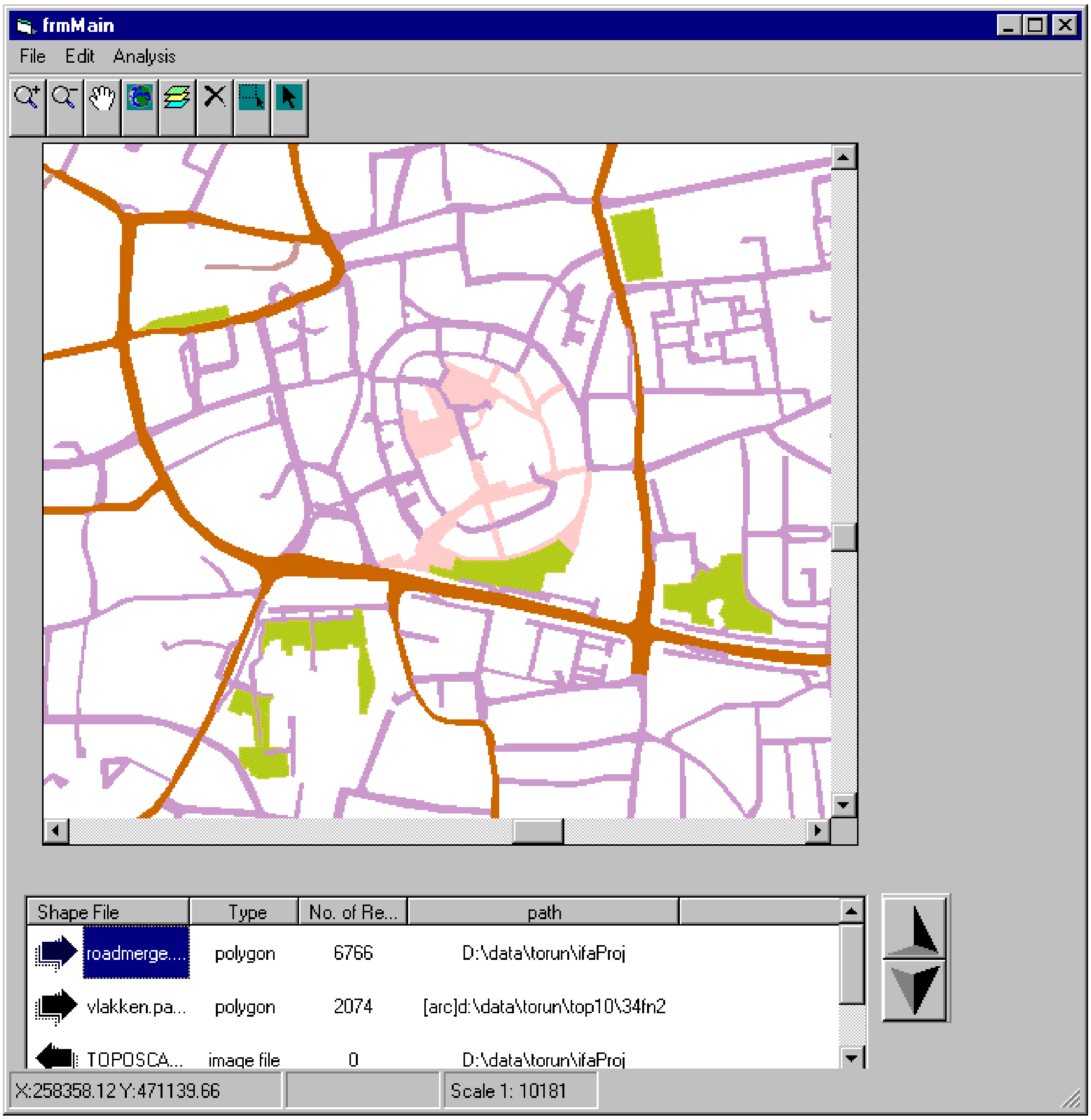

Figure 5. User Interface of the mapping/GIS application.

Although, implementation of deploying the application on the Internet is not yet finished, the basic implementation phases are given in this sub-chapter. This three-tier system provides the user to send a request interactively to Web Server, the request is implemented by Map Objects application, and finally dynamic results (in JPEG or GIF format) are put in an HTML file and sent back the client to visualize. Here, Map Objects IMS acts as an interface between the standard Web Server and Map Objects application and transforming Map Objects application display and mapping to be sent to the client via Web Server (MapObjects Appl.Ref,1996). The Web user interface is designed using HTML, which supports interactive mapping and GIS operations. The Map Objects interface has panning, zooming, querying, activating/disactivating map layers, and other simple onscreen map functions (see figure 5). Results of queries processed on the data are displayed in tabular and graphical format on the user's browser. Although the implementation as a thin client/fat server model in this particular application is not so efficient due to processing all data on the server side, its easy implementation, use of the standard Web Server and flexible programming with Map Objects and Visual Basic make the architecture easy for Web authors.

5 CONCLUSIONS

The Dutch Topography Service data coding conventions help 'dynamic generalization': Spatial data conventions of Dutch Topography Service for road data have a hierarchical system from high priority roads (highway) to low priority roads (poor roads). Although, road data are populated at certain discrete points ('Tdn_code' s), they provided to construct a generalization structure specific to road data of Top10 database. The other data classes have not been studied. Customizing this approach for other databases which has a different scheme and data model may therefor need new modeling issues.

ESRI MapObjects provides a flexible development environment with a visual programming platform (in this application Microsoft Visual Basic). MapObjects provides a set of objects those enable the software developer to implement a mapping and GIS application with basic tools.

Main limitations of the Web for 'dynamic generalization' are the heterogeneous character of user screen size, communication traffic on the Internet and computer power (CPU and RAM) related with technology used for mapping/GIS services. Scale of the database is an important factor to determine the scale intervals and determine the data set to be visualized at that interval. However, the former measures are more or less *de facto* condition, but an efficient generalization algorithm is the main item that needs effort by cartographers and GIS experts.

Only the original data set is used, so it doesn't need duplication of the database for different scales. But it needs immense query operations, which are not optimized actually. The generalization operations work locally without any problem due to communication and time efficiency. But, data transmission band of the Internet, modem capacity and time determine the characteristics of graphic applications on Web. In ordinary Web applications, the server evaluates and manipulates the request of the client and the gathered results are converted into image format to be sent back to the client. Within this application, whenever the user interacts with the map display window, the program checks the need of generalization and if needed, does the operation. So, dense interaction with the map server generates an inefficiency problem, which is the most important one for impatient Internet users.

REFERENCES

Augusto, E.M., Gomes, M.R., Abrantes, G., 1997. Accessing Geographic Data Through WWW, Third Joint European Conf. & Exhibition on Geographical information, Vienne, Austria, pp. 135-144

Boehner, C., Haastrup, P., Reggiori, A., 1997. The GIS-WWW Gateway, Third Joint European Conf. & Exhibition on Geographical information, Vienne, Austria, pp. 125-134

Brazile, F.L., 1998. A Generalization Machine Design that Incorporates Quality Assesment, 8 th Int. Symp. on Spatial Data Handling, Vancouver, Canada, pp. 349-360

Dettori, G., puppo, E., 1996. How Generalization Interacts with the Topological and Metric Structure of Maps, Advances in GIS Research II, Proc. of the Seventh Int. Symp. on Spatial Data Handling, Taylor&Francis, pp. 559-570

Devogele, T., Trevisan, J., Raynal, L., 1996. Building a Multiscale Database with Scale-Transition relationships, Advances in GIS Research II, Proc. of the Seventh Int. Symp. on Spatial Data Handling, Taylor&Francis, pp. 337-351

Kraak, M.J., F.J.Ormeling (1996) : Cartography : Visualization of Spatial Data, Addison-Wesley-Longman Harrow

MapObjects Appl.Ref (1996) : MapObjects, Building Applications with MapObjects, ESRI, http://www.esri.com

Molenaar, M., 1998. An Introduction to the Theory of Spatial Object Modelling for GIS, Taylor&Francis

Oosterom, P., 1999. Spatial Access Methods, GIS : Management Issues & Application, Vol. 2, Second Edition, Editors: P.A. Longley, M.F. Goodchild, D.J.Maguire, D.W. Rhind, J. Wiley & Sons Inc

Peng, W., Templi, K., 1996. An Object-Oriented Design for Automated Database Generalization, Advances in GIS Research II, Proc. of the Seventh Int. Symp. on Spatial Data Handling, Taylor&Francis, pp. 199-213

Peng, Z.R. (1997) : An Assessment of the Development of Internet GIS, ESRI, http://www.esri.com

Robinson, A.H., Morrison, J.L., Muehrcke, P.C., Kimmerling, A.J., Guptill, S.C., 1995. Elements of Cartography, 6 th Ed., John Willey & Sons, pp. 450-473

Smaalen, J.W.N., 1996. A Hierarchic Rule Model for Geographic Information Abstraction, Advances in GIS Research II, Proc. of the Seventh Int. Symp. on Spatial Data Handling, Taylor&Francis, pp. 215-225

Stasik, M., Jankowski P., 1998. GIS Supperted Group Decision Making under Distributed Space and Time Conditions, 8 th Int. Symp. on Spatial Data Handling, Vancouver, Canada, pp. 297-307

Torun, A., 1999. GIS Application on the Internet for Exploring and Dynamic Visualization of a Topograhic Database, Professional Master Degree Final Report, ITC, Enschede, the Netherlands