An Integrated Topologic Database Design for Geographic Information Systems

Terrence Keating, William Phillips, and *Kevin Ingram*
Kork Systems, Inc., 6 State Street, Bangor, ME 04401

ABSTRACT: Several database management system (DBMS) design concepts used by Kork's Geographic Information Management System (KGIS) for the creation, maintenance, and rapid retrieval of geographically referenced data sets are discussed. The DBMS design fosters fast interactive response during input, editing, analysis, and output regardless of the geographic extent.

Important geographic DBMS design concepts include:

(1) Create the DBMS to support three levels of organization: cartographic features (points, lines, and polygons), topologic elements (nodes, edges, and faces), and geographic (spatial) clustering (sorted by size and location). The system software must allow fast access to graphics and associated elements stored on disk from each of these three levels.

(2) Support memory-resident DBMS operations and internal buffering (according to feature type and geographic location) between the disk-resident database and the computer's memory such that the geographic data sets reviewed or modified occupy the smallest possible memory space, can be retrieved in the fewest number of disk accesses, and contain only the necessary data to support the task at hand.

Once these concepts are integrated within a DBMS, an efficient and cost-effective computer graphics approach can be applied to managing geographic information. The procedure is equally appropriate for management of any large two-dimensional data set.

INTRODUCTION

THE DEVELOPMENT of a new and comprehensive geographic information management system with data structures specifically designed to handle geographic features and associated non-graphic attributes is addressed. Topological relationships, inherent in the data structure, are automatically generated to assist in the complex analysis (overlays, proximity, networking) of geo-referenced data sets.

In this development, consideration was given to all the activities which support collection, analysis, and presentation of mapped data, including hardware, training, support, and applications. Design criteria included efficient storage, rapid retrieval of selected geographic regions, an ability to validate and maintain new and existing data, support for complex queries, and simple reformatting to exchange data and associated relationships with other systems.

Many critical limitations exist in using commercially available data base management systems (DBMS) for handling geographic data (Abel and Smith, 1985). Many cannot efficiently handle variable length records to accommodate lines of different lengths (Waugh and Healey, 1985). In relational systems, attributes associated with features are spread between many tables rather than being stored with the feature itself. Features are seldom clustered by location for rapid retrieval. Linkages between features are awkwardly handled and, thus, are expensive to store or generate. Better mechanisms to manipulate and store graphics and associated attributes are necessary to build a responsive system (Frank, 1984).

This paper discusses DBMS design concepts used by Kork's Geographic Information Management System (KGIS) to support the creation, maintenance, and rapid retrieval of geographically referenced data sets, including multiple levels of organization and internal buffering for memory-resident operations. Before concentrating on the DBMS design, a brief overview of the system operation is presented.

KGIS SYSTEM OVERVIEW

The geographic information management system can be thought of as a graphic spreadsheet. Data are entered into the system by direct or scanned digitization, through bulk loading from foreign databases, and by data entry directly between the keyboard and the database. The system contains a rich set of algorithms to construct and edit the topology from cartographic features as each feature or theme (layer) is transacted to the database. Mapped features and attribute data sets may be independently entered and later associated. Map data are displayed on a high resolution color graphics screen as the associated attributes or modeled information sets are reported on a conventional alphanumeric terminal. A Digital Equipment Corporation MicroVAX (VMS) computer, several Tektronix 41xx series color graphics display screens, each with a mouse or tablet, plus the KGIS software comprise a typical multi-workstation configuration.

With the mouse or digitizing cursor, a user can select a specific cartographic feature (point, line, or polygon) visible on the graphics screen. When a feature is selected, it is highlighted on the graphics screen. If attributes have been associated with the mapped feature, these attributes are selectively shown on the adjacent alphanumeric screen. Features similar to the one selected will also be highlighted, if the user requests. If a graphic feature needs attributes assigned to it, they may be entered at that time. As mapped features are modified, their associated attributes are redefined where appropriate.

Once the mapped features have the necessary attributes assigned to them, these attributes may be retrieved using relational (Boolean) queries. Queries can combine feature attribute relationships and spatial relationships between features. Queries can be posed either textually (from the keyboard) or graphically (by using a pointer on the graphics system) and the responses can take either textual, graphic (highlighting or plotting), or combined textual and graphic forms. A complete on-line interconnection between attributes and their associated graphics is maintained at all times.

The system is designed for applications requiring the on-line maintenance of geographic information for parcel management (cadastral mapping, registry of deeds), polygon management (soils, forestry, and environmental studies), or linear networking (utility, transportation, military logistics).

MULTIPLE LEVELS OF ORGANIZATION

Geographically referenced data sets are maintained in three levels of organization: cartographic (points, lines, and polygons), topologic (nodes, edges, and faces), and geographic (spatially sorted by size and location). Figures 1 and 2 express the relationships between these various views. The system soft-

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 53, No. 10, October 1987, pp. 1399–1402.

0099-1112/87/5310–1399$02.25/0

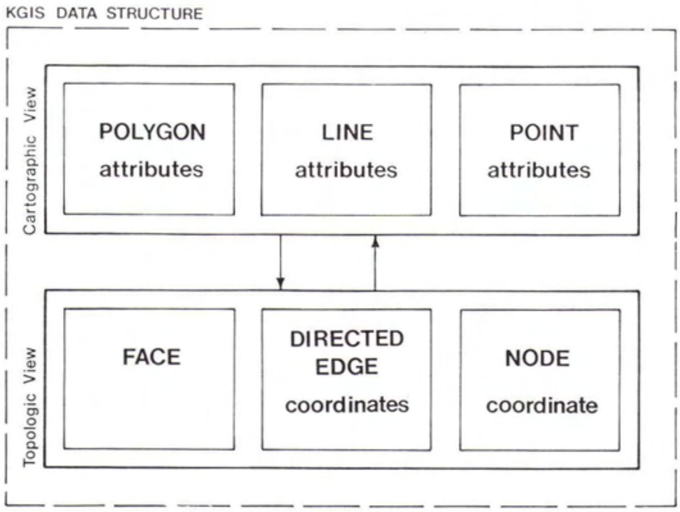

FIG. 1. Internal database organization for the KGIS System. The cartographic and topologic views are represented.

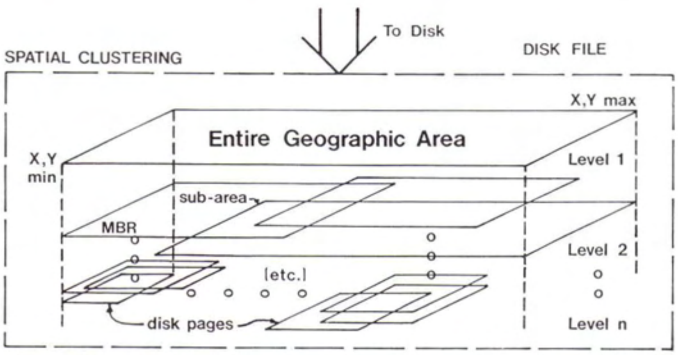

FIG. 2. Clustering of physical data storage according to geographic size and location.

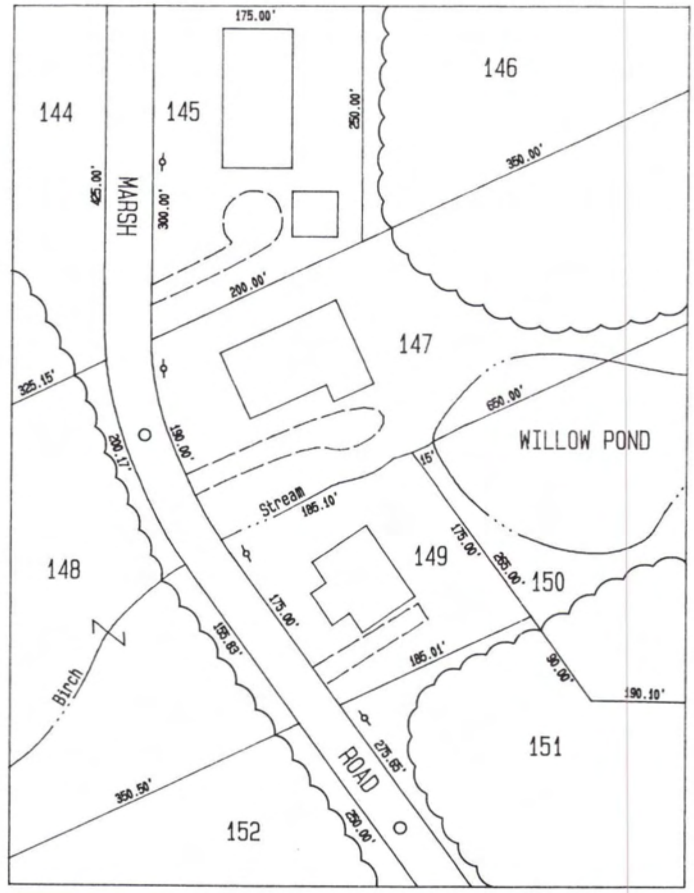

FIG. 3. The Cartographic Features Level contains features which the user can easily identify by name or classification. Examples include houses by street address, roads, streams, lakes, parcels, poles, manholes, sewer lines, woods, etc. Each point, line, and area is described in its entirety.

ware can then allow the user fast access to data with queries expressed in the context of each of these three levels.

THE CARTOGRAPHIC (FEATURES) LEVEL

In KGIS, the user views the world as sets of cartograhpic features, i.e., points, lines, and polygons, to which attributes normally are associated. Figure 3 is an example of the cartographic level. Examples of a "point" at larger map scales include a telephone pole, tree, manhole, or sign, while at smaller map scales examples might include the center of town, a factory location, a parcel centroid, or the location of a famous historic event. A "line" or linear feature might represent an interstate highway, a pipeline, or a stream. A "polygon" could represent a parcel, a stand of trees, a wetland zone, or at larger scales, a building.

Because users initially think in cartographic terms, the cartographic level is especially important. Attributes describing these features (*height* of pole, *name* of highway, or *owner* of a parcel) are easily assigned. This level also is critical in displays because the user can associate information derived from the system with recognizable features.

THE TOPOLOGIC (ELEMENTS) LEVEL

To understand the spatial relationships between cartographic features, the cartographic features may need to be treated in many ways. Rather than pre-determine how features must be treated to answer specific questions (how long, how much, how close, who else, what if), the system will, during data entry, structure all features into what are called topological elements. These system-derived relationships between geographic features allow navigation among features and support the creation and use of associations between features. The system is built with the assumption that complex analysis will be required before meaningful questions can be answered.

Topology provides a means for analyzing the relative spatial relationships among various elements without regard for exact location. The topologic level is composed of zero-cell elements (nodes), one-cell elements (edges), and two-cell elements (faces). Nodes can be isolated or serve as the ends of edges, edges connect two nodes, and faces are individual areas bounded by edges.

Topologic elements are creased from the cartographic features as they are entered into the system. Figure 4 illustrates the topological framework of the cartographic features that are shown in Figure 3. For example, as a line is entered, edges are created. Edges are "directed" with respect to how they are digitized and the direction that is traversed to construct a face. The current edge is terminated, a node created, and a new edge is started wherever the line crosses another line feature. Where an existing topologic edge coincides with a line feature, it defines the feature and no new edge is created. The cartographic line feature then is defined as that series of edges. In general, cartographic point, line, and polygon features are defined respectively by toplogic nodes, edges, and faces.

Inherent within a database are many-to-many relationships. For example, a parcel boundary may be composed of many topologic edges. Likewise, each of those edges may participate in many cartographic features, such as other parcels, roads, streams, etc. The many-to-many relationships that exist between cartography and topology provide the means for answering

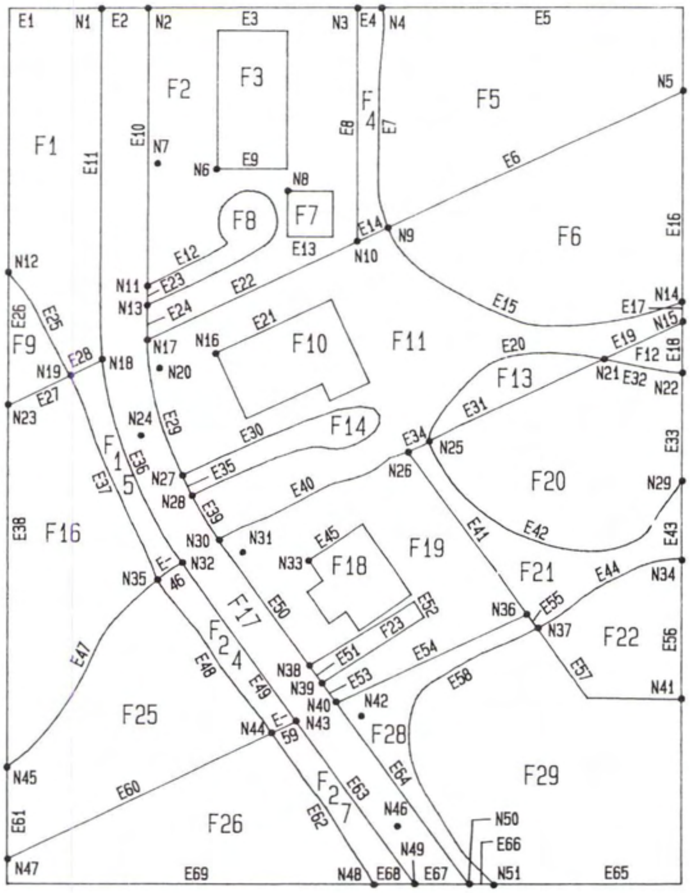

FIG. 4. The Topologic Elements Level remains hidden from the user but links all features to a DBMS navigational framework comprised of nodes (N1, N2, . . ., Nm), edges (El, E2, . . ., En), and faces (F1, F2, . . ., Fp).

questions which involve the spatial relationships between cartographic features such as, "What soil types underlie these parcels?" or "What forest stands are adjacent to this pond?"

In summary, topologic elements are helpful whenever combinations of cartographic features are needed to extract information from collections of data.

The Geographic Features (Spatial) Level

Once each cartographic feature is represented as one or more topological elements (nodes, edges, and faces), the topologic and cartographic combination is clustered geographically. Each feature is allocated to a database storage segment (disk page) which contains other features of (usually) similar size in the same geographic region.

Geographic clustering groups features according to geographic (spatial) location and extent. The purpose of the clustering is to minimize the number of disk accesses to retrieve selected data sets. Clustering occurs at both the topologic and cartographic levels. Figure 2 depicts geographic clustering.

Geographic clustering is performed by first determining the minimum and maximum x, y of coordinates within which the entire topologic element is contained. These coordinates define an important attribute, called the minimum bounding rectangle (MBR), which is stored with each data feature. Each disk page has a comprehensive page MBR as well.

A variety of mathematical algorithms have been devised to sort these MBRs such that all data features within given geographical regions can be rapidly accessed. The system uses a modified quadtree approach where the geographic domain is quartered into successively smaller quadrants and features are stored on disk pages referenced to specific quadrants (Frank, 1983).

Geographic clustering has been implemented somewhat differently in this system, and is explained as follows:

- When a cartograhpic or topologic feature is first digitized, the MBR is calculated and a feature type is assigned.
- Features are allocated to specific disk pages according to the size and location of their MBRs. Similarly sized (encompassing compatible geographic rectangles) features within the same general geographic region reside on the same disk page.
- Each page contains different types of features but all features within a given page are pertinent to a specific geographic region. The region represented by a page may overlap with those of other pages.
- Many disk pages may be required to store all the features within a given geographic region. Each page has pointers to all other pages within the geographic domain. Thus, all the information about a given geographic location can be found by searching only a limited number of disk pages rather than the entire database.
- Spatial clustering is done during collection. Clusters may be reorganized during input, or clustering may be delayed and performed offline. Features can be directed to pages either higher (into larger page MBR regions) or lower (into smaller page MBR regions) within the indexing hierarchy. This aids in placing features onto pages most likely to contain related feature groups.

Example Using Multiple Levels of Organization

An example is provided to clarify the interactions within a DBMS supporting multiple levels of organization. Assume that the data set includes all interstate highways, all state-maintained highways, all state boundaries, state capitals, and all federal and state owned parcels for the entire United States. Suppose the system is to respond quickly to the request "Interactively display all portions of Interstate 95 within the State of Maine passing through state-owned land." A summary of features within each level would include

Cartographic features:	*Polygons*, of parcels and states which are split into two or more faces by an interstate. *Lines*, representing the (named) interstate highways. *Points*, representing state capitals.
Topologic elements:	*Faces*, defined here as undivided area/portions of parcels and states. *Edges*, defined as undivided boundary segments/portions of faces which have been terminated (noded) at intersections between intersate lines and ownership lines or state border lines. *Nodes*, which are pairs of coordinates either isolated (at a cartographic point rather than at a line crossing) or at the endpoints of edges.
Geographic features:	*MBRs*, representing the Earth coordinates envelope which encompasses each topologic element, or alternately, each cartographic feature.

All three levels are used to generate the graphics display as follows:

- Topologic operations during the collection of the original data sets split linear features into edges wherever lines intersect. Each edge has an MBR assigned to it, and these edges by nature are spread across the continental United States. Likewise, polygon features were split into faces, with MBRs calculated.
- With geographic clustering, the interstate highways (larger feature) are stored in relatively few pages, with one of these pages automatically associated with a geographic location in roughly the Northeast United States. State borders also will cluster into disk pages covering regions, with several pages overlapping the interstate highways page. State-owned properties will be clustered together geographically with the State of Maine's parcels clustered on several pages. Pages associated with smaller features in more local regions are linked together in a hierarchical structure with pages containing features encompassing larger geographic regions.
- To answer the request, the system determines the MBR of the Maine State boundary. One or two disk pages will contain all the Maine state border features and one or two other pages will contain

highway features. All state-owned parcels will occupy several more disk pages.

- All candidate pages are brought from disk into computer memory. All relationships between the features are present on the pages, and the decision-making process to highlight state-owned lands (faces) containing an interstate (edge) begins.
- The cartographic feature known as "Interstate 95" is defined as a series of linked edges. The search begins by identifying the edge associated with Interstate 95 that shares a node with an edge associated with the Maine state border.
- By traversing from edge to edge, and asking whether an adjacent face is part of a polygon which has state-ownership, those edges satisfying the display criteria are identified. Relatively few edges and features within the database are used to complete the search.
- The display is completed by highlighting only those edges meeting the search criteria. Because the intersections of property lines with highway lines were calculated during input, very little additional mathematical processing is required to display the edges.

INTERNAL BUFFERING FOR MEMORY RESIDENT DBMS OPERATIONS

A geographic DBMS design concept which complements multiple levels of organization is the internal buffering of appropriate pages of disk-resident data in the computer's memory, thereby avoiding continuous disk operations. In doing so, system software must protect memory-resident database transactions by locking out other users during the disk update cycle.

An internal two-level buffering strategy is used to keep the most recently used sets of data resident while the earlier data sets (least recently used) are returned to disk. The system buffers first by feature and second by disk page. When a request is made for a particular feature, a search for it is made in the feature buffer. If it is found, the request is satisfied. If it is not found there, the page buffer is searched for the page on which it resides. It the page is in memory, the feature is copied into the feature buffer. If not in the page buffer, the page is retrieved from disk and then the feature is copied into the feature buffer. Both the feature and page buffers employ a least recently used (LRU) replacement algorithm to free up space when they become full.

Because the geographic clustering minimizes the number of disk pages needed to contain all the data within a region, most DBMS operations will access the required data sets in memory.

The system is designed to perform memory-resident database transactions, to have the results held in memory, and to automatically post the transactions when the user updates the disk. Memory-resident transactions have the advantage that temporary or "what if" perspectives can be tested, displayed, and discarded with no subsequent disk access. Only verified transactions are posted back to disk.

CONCLUSIONS

A comprehensive redesign of DBMS strategies specific to geographic information management requires a fresh look at design standards and operations.

Two important design concepts are integrated into one comprehensive data structure to improve the analysis capabilities for large geographic data sets. First, multiple levels of organization are created for each feature. The user perceives the cartographic features level. The DBMS creates and then navigates between features using the topologic level. Complex analysis of dissected data sets is possible from this level as well. Geographic clustering is used to optimize the storage and retrieval of data for rapid analysis and display, regardless of the extent of the database.

Second, memory-resident DBMS operations with internal buffering greatly improve processing. This is most efficient when geographic clustering is used to minimize those pages carrying all the pertinent data within a selected region. It allows for a responsive system which can handle extensive world-wide and intensive local data sets.

REFERENCES

Abel, D., and J. Smith, 1985. A Relational GIS Database Accomodating Independent Partitioning of the Region, *Proceedings of the International Symposium on Spatial Data Handling*, Seattle, Washington.

Birtwistle, G., *et al.*, 1973. *Simula Begin*, Auerbach, Philadelphia.

Frank, A., 1983. *Storage Methods for Space Related Data: The Field Tree, Spatial Algorithms for Processing Land Data with a Microcomputer.*, Boston Lincoln Institute for Land Policy.

———, 1984. Requirements for Database Systems Suitable to Manage Large Spatial Databases, *Proceedings of the International Symposium on Spatial Data Handling*, Zurich, Switzerland.

Waugh, T., and R. Healey, 1985. The Geoview Design: A Relational Database Approach to Geographical Data Handling, *Proceedings of the International Symposium On Spatial Data Handling*, Seattle, Washington.

Xerox Learning Research Group, 1981. The Small Talk - 80 system, *Byte Magazine*.

(Received 11 July 1986; revised and accepted 20 April 1987)

Forthcoming Articles

Abdalla Elsadig Ali, Geometric Accuracy Testing of Orbital Radar Imagery.
Michael H. Brill and *James R. Williamson*, Three-Dimensional Reconstruction from Three-Point Perspective Imagery.
J.H. Everitt and *A.J. Richardson*, Canopy Reflectance of Seven Rangeland Plant Species with Variable Leaf Pubescence.
M.F. Gross, M.A. Hardisky, V. Klemas, and *P.L. Wolf*, Quantification of Biomass of the Marsh Grass *Spartina Alterniflora* Loisel Using Landsat Thematic Mapper Imagery.
Kurt Kubik, A Note on Photogrammetric Block Adjustment with Additional Parameters.
Robert D. Lamb, H.E. McGarrah, and *J.D. Eick*, Close-Range Photogrammetry with Computer Interface in Dental Research.
Jerry C. Ritchie and *Charles M. Cooper*, Comparison of Landsat MSS Pixel Array Sizes for Estimating Water Quality.
Johnny Skorve and *Peter Vincent*, Some Models of Ice Melt on High Level Lakes in Southwest Norway.
Christopher Stohr, W.J. Su, P.B. DuMontelle, and *R.A. Griffen*, Remote Sensing Investigations at a Hazardous-Waste Landfill.
P.J. Wise and *J.C. Trinder*, Assessment of SIR-B for Topographic Mapping.
R. Yazdani, N. Christou, and *E. Derenyi*, Potential Applications of Digital Image Analysis Systems for Displaying Satellite Altimetry Data.