MAINTENANCE OF GEOGRAPHIC STRUCTURE FILES AT THE BUREAU OF THE CENSUS

T. McDowell
D. Meixler
P. Rosenson
B. Davis
Geography Division
Bureau of the Census
Washington, D.C. 20233

ABSTRACT

In 1985 and 1986, the U.S. Bureau of the Census developed an interactive program embodying a system of subroutines to store, review, and modify the geography of a map. The process is based upon the grouping of underlying 2-cells according to their geographic cover. These groups of 2-cells are the minimum intersections of geographic entities and designated as the Geographic Tabulation Unit Base, or GTUB. The theoretical framework for the system was developed and documented in a paper presented at Auto Carto 7 [Meixler and Saalfeld, 1985] before implementation of the system. This paper describes the adopted approach and possible extensions.

Specific routines perform the extraction of boundaries, the windowing of geographic entities, the maintenance and referencing of specified geographic entities, and the comparison of geographic hierarchies. These routines work effectively in an interactive environment and confirm the soundness of the theoretical approach. The structure upon which these routines were implemented did not allow development of the full capabilities of the GTUB routines. Various enhancements and extensions to be incorporated into these routines are discussed.

BACKGROUND

Historically, decennial census data have been collected by enumeration districts (EDs). These districts were defined to honor those geographic boundaries necessary for decennial census tabulation. The EDs were sized by estimated housing counts to provide a reasonable work assignment for an enumerator. For the 1980 census, an effort was made to reduce the geographic coding errors associated with controlling in excess of 300,000 EDs. This effort involved the advance definition and entry into the master geographic file of an explicit record containing unique combinations of geographic tabulation codes. This record was named the Geographic Tabulation Unit Base (GTUB). Later this GTUB record was used as the framework under which one or more EDs were added, depending upon the population of the area. This master geographic file contained no digital map representation, thus the GTUB was the basic representation of the geography associated with any particular area.

With the advent of the Topologically Integrated Geographic

Encoding and Referencing (TIGER) System for the 1990 census, both a specific digital map and geographic entities are integrated into a single file. This change does not lessen the number or complexity of the geographic areas required for census tabulation. The GTUB concept was analyzed and adapted to meet the requirements of this integrated structure. It was recognized as representing the atomic elements of the complete geographic lattice, thus proving useful in high level geographic comparability. As an intersection record for multiple geographic hierarchies, the GTUB reduces the pointer overhead in linking 2-cells to the geographic entities. And the multilist structure allows quick collection of the GTUB refinement of these geographic entities.

Thus the GTUB is useful in conceptually and structurally simplifying the handling of this myriad of geographic areas. As part of the TIGER File representation of multiple geographic areas, the GTUB will serve as the connection between each individual geographic entity and the land area belonging to that entity. As in 1980, each GTUB will contain a specific combination of geographic codes that represents the intersection of various geographic entities. This enables it to link together all two cells sharing this unique set of geography.

THE PROGRAM ENVIRONMENT

The program evolved out of a desire to test ideas and concepts in the routines that create and manipulate GTUBs. All work is done on the Unisys (Sperry) system using the TIGER File and a library of graphics routines. The purpose was to show graphically, using a Tektronix 4115 color terminal, geographic changes to the TIGER File.

A TIGER File is split logically into many different subfiles. A subfile is organized and can be accessed in one of two ways: as a balanced tree that is ordered and accessed by a key, or as a random access logical subfile where each record is identified by its relative position in the subfile. The GTUB records are stored in a random access logical subfile. However, each of these records has a directory record for easier program maintenance.

The screen of the Tektronix terminal is divided into three areas: dialog, instructions or information, and graphics. Graphics displayed during a session are saved on the hard disk attached to the terminal and are automatically reloaded when the file is later processed on that terminal. Since the initial display of the file takes much longer than a load of the graphics from its local disk, a batch process can be used to store all or parts of the graphics before work begins.

Initially, only the outline of the file is displayed on the screen. At this point either the entire map, individual entities, or groups of similar entity types can be displayed. Linear features are color coded according to a subset of the census feature class code. Geographic entities are displayed as colored panels behind the linear features, emphasizing their areal characteristics. Panels can be turned on and off to reduce clutter on the screen. Distinct fill patterns were chosen for each entity type. Thus geographic comparisons may be viewed as overlapping panels. The terminal has a manual zoom feature that can be used to look at areas in more detail. This interactive geographic update program allows a user to review the area of geographic entities, and to change it when necessary. In addition, the user may window on a geographic entity by entering a window command.

THE GTUB RELATED COMMANDS

The program is command oriented, with English-like commands entered from the terminal. Certain commands display a short help screen in the information area. Routines that modify geographic entities alter the GTUBs in the TIGER File directly. Modifications are then extracted from the file and displayed on the terminal for the user to review. The SHOW, WINDOW and CHANGE commands utilize the GTUB routines.

The SHOW command is used to view a geographic entity. The user chooses which geographic entity to display by specifying its geographic codes. The area covered by this geographic entity is constructed by using the GTUBs associated with the entity as building blocks. The boundary is then extracted using information from the 2-cells linked to these GTUBs and used by the panel filling routine.

For each of the GTUBs in the entity, the list linking the GTUB to its underlying group of 2-cells is examined. In this list, each GTUB points to the first 2-cell within the GTUB. Each 2-cell then points to the next 2-cell same GTUB to continue the list of area covered by this GTUB. The boundary of each 2-cell is examined using a similar list that links the 2-cell to the chain of 1-cells that define the polygon.

The boundary file for the geographic entity is created by performing a Boolean add of all 1-cells defining a given 2-cell to the boundary file; in other words, as each 1-cell is examined. If it is not found in the boundary subfile, it is added, otherwise it is deleted. This process will assure a file of distinct 1-cells with all common elements deleted. Repeated application of this procedure on all 2-cells linked to each GTUB in the entity will yield a subfile of boundary 1-cells for that geographic entity. These 1-cells are now ordered and given to the Tektronix hardware which panel fills the area on the screen. During the ordering process, the boundary cycles are determined. If more than one cycle exists, the user is informed that the entity is either discontiguous or contains holes.

The WINDOW command allows the user to fill the graphic screen with a specific geographic entity. The boundary extraction process is done as in the SHOW command. If the entity has not been displayed previously, it is then panel filled. Additionally, the minimum and maximum coordinate of the boundary are found and correctly proportioned for display purposes. They are then sent to the Tektronix terminal which adjusts the graphic display.

The CHANGE command is used to modify the areal coverage for a geographic entity. When the coverage of a given geographic entity changes, the GTUB associated with the affected 2-cell must be modified. Three subcommands have been developed to accomplish such change.

The CHANGE BY command is used in the special case where the boundary of the geographic entity to be inserted or updated encompasses the area of an existing geographic entity in the file. In many cases, only the GTUB records need be referenced. The process entails examining every GTUB associated with the existing entity. If the GTUB already has the proper codes for the new entity then this GTUB is skipped. Otherwise, the appropriate fields on a copy of the GTUB are modified to reflect the requested change. The GTUB directory is searched to see if the modified GTUB already exists. If it does, all 2-cells under the modified GTUB are merged with the list of underlying 2-cells for the existing GTUB. The old GTUB is then deleted from the GTUB directory and the random access subfile. If the modified GTUB does not exist, then the proper fields can be changed on the existing GTUB and a revised directory record created.

The CHANGE AREA command is used to insert a new geographic entity. This command allows the user to delineate an area by chaining the 1-cell boundary around it. Once the boundary is specified, the program retrieves all 2-cells within the area. The GTUB cover for each 2-cell is then adjusted. To change the geography of a given 2-cell, a copy of the current GTUB above that 2-cell is obtained. The 2-cell is then delinked from the list of 2-cell for this GTUB. If the 2-cell is the only remaining 2-cell under this GTUB then, the GTUB and its directory record may be deleted. The copy of the GTUB is then modified to have the requested change and the GTUB directory searched to see if the modified GTUB already exists. If it exists, then the 2-cell is added to the list of underlying 2-cells of this GTUB. If the modified GTUB does not exist then a new GTUB is created and added to the GTUB subfile and directory. Finally the 2-cell is linked to this new GTUB. For efficiency, the current procedure actually handles every affected 2-cell under a single GTUB in one pass.

The CHANGE BORDER command is used when a boundary update involves annexing a neighboring area. In this case, the user selects a starting point on the existing boundary. From this point, the user chains around the area to be annexed until returning to the existing boundary. All 2-cells contained in this area are determined and then changed using the method described above.

ENHANCEMENTS AND EXTENSIONS

Programs that access or alter the stored relationships between different records in the file do so through standard list handling routines that are part of the I/O management system. The list routines enable programs to add, delete, search, or modify a member of a list. Other capabilities include merging lists or moving between lists that share a common record type. All routines can work with any of the lists defined in the system. These include simple one-way, directory, intertwined, and many-to-many lists. The next release will include teo-way lists and multilists.

The multilist will be used to link the entities and their GTUBs. This list is defined by a generic name and a record type. All entity records reside in the same subfile. They are distinguished by an entity type field. Many entities will be linked to a GTUB, with more than one list threading through the record. Rather than defining each list separately, the routines will determine the correct list and the corresponding pointers by the type of entity that is being manipulated. This will free the programmer from having to be aware of many named lists.

Boundary maintenance has been of questionable efficiency during the first implementation of these routines. It has been proposed that boundaries only be maintained for the GTUBs and that they be extracted for the geographic entities. The GTUB routines can handle the Boolean add of 1-cells to the GTUB boundary subfiles during routine maintenance. Interestingly, the Boolean add should be done for both the GTUB from which the 2-cell is moving and the GTUB to which it is moving. As GTUB boundaries can be determined by the Boolean addition of its component 2-cells, so may the boundary of any geographic entity be determined from its component GTUBs. A technique that is being explored is to store the boundary segments for a GTUB as bit flags in a separate subfile for each GTUB. The GTUB routines would then be optionally responsible for maintaining these boundary subfiles. Most boundary subfiles would only comprise one physical page of data in the system. The boundary for an entity could then be extracted by gathering all the GTUBs contained in that entity and performing an XOR operation on the boundary subfiles. An integral part of the TIGER System is the I/O management system that handles the creation and manipulation of these subfiles. The routines are general purpose and assume nothing concerning type, internal name or number of subfiles within the file. This feature enables an applications program to create and use additional subfiles to store information during processing. Thus, GTUB boundary subfiles may be added easily as part of an adjunct file.

Manual comparability is implemented currently by the graphic overlay of panels. Similarly, coextension and nesting of geographic entities can be inserted into the system using exclusively the CHANGE BY command. In the future it will be useful to edit and test comparability and nesting of geographic entities by reviewing their GTUB relationships. Such high level comparisons need not go to the underlying 2-cells.

CONCLUSION

The GTUB routines have been implemented successfully on a preliminary TIGER file structure. These routines concentrate on the automatic maintenance of the GTUB records. The user appears to change geographic codes for graphically defined areas of the file. The user does not see the addition, or deletion of GTUB records, nor the moving of groups of 2-cells between GTUBs. Because of the file structure, these routines have concentrated on the GTUB maintenance and linkage between the GTUB and its composite 2-cells. The coordination of the GTUBs under the geographic entities will be an integral part of the TIGER File structure for the 1990 census. The list handling capabilities of the I/O management system will provide the maintenance of these GTUBs under the geographic entities.

REFERENCES

Meixler,D. and Saalfeld,A.,1985, "Storing, Retrieving and Maintaining Information on Geographic Structures: A Geographic Tabulation Unit Base (GTUB) Approach", AUTO CARTO 7 PROCEEDINGS, pp 369-376,Washington, D.C.