AUTOMATED POLYGON CREATION FROM DIME NETWORK FILES

Jean-Margret Hynes, University of Illinois Chicago

ABSTRACT

There are two ways to organize and store cartographic information, by polygon or areal unit or by network or linear unit. Available cartographic software for the display of geographically associated information requires that maps be viewed as a series of connected polygons. These polygons are expedient for the display of data collected or aggregated by geographic unit such as, for example, population by county or median income by zip code. DIME is a nationwide Census Bureau program to store base map representations of urban areas in machine readable formats. Although this cartographic information is uniformly maintained and available, it is organized as a series of arcs and nodes, i.e. as a network. This network structure is conducive to geocoding address specific data to allow the aggregation of such data to geographic units.

This paper suggests an algorithm based on existing network theory to efficiently create polygon boundaries for geographic units from a large set of such arcs and nodes. The algorithm will enable base map conversion from linear segments to areal units thus allowing one to have a single cartographic information base, e.g. a network, from which particular areal units can be generated as needed. The algorithm isolates boundary segments in the network and arranges them in continuous plottable boundaries.

The individual boundary is verified by testing that all boundary segments of the polygon are used and that all non-boundary segments are contained within the delineated polygon.

KEYWORDS AND PHRASES; polygons, automated cartography, map databases, geographic information systems

Il y a deux manieres d' organiser et de stocker une information cartographique: par unite polygonale ou aerienne, et par unite lineoir ou unite de reseaux. Le software cartographique disponible necessite qui les cartes soint considerees comme une serie de polygones connectees pour aceeder a l' information geographicque contenue. Ces polygones sont bien adaptes - la representation de donnees collectees ou rallemblees par unite geographique comme par exemple, la population par comte ou le revenue moyen par code postal. DIME est un organisme national dont le but est de stocker des cartes relatives aux zones urbaines sous une forme lisible par machine.

Bien que cette information cartographique soit mise a jour et disponible continument, elle est organisee comme une suite d' arcs et de sommets, c'est-a-dire en un reseau.

Cette structure de reseau est utile pour allocier des codes geographiques a des donnees dependant de leur location, and pour permettre l' allociation de telles donnees a des unites geographiques.

Cet article suggere un algorithme base sur la theorie des graphes existante pour determiner a partir du graphe la frontiere du polygone representant l'unite geographique. L'algorithme permet d'avoir une information cartographique unique, par exemple un graphe, a partir duquel des unites aeriennes particulieres peuvent etre crees a la demande. L'algorithme isole des arcs dans la frontiere du reseau et en permet un dellin cintinu par machine.

En peut verifier pour chaque polygone que tous les arcs de la frontiere sont utilises et que les ceux n'en faisent pas partie sont contenus dans l'interieur du polygone delimite.

INTRODUCTION

Visual display of spatial information is a powerful way to communicate a large amount of data in a concise and efficient manner. The viewer assimilates information, makes comparisons, and draws inferences far more quickly from a graphic display than he can by reading an equivalent amount of information organized as numeric tables. For geographic information systems, understood as the storage and retrieval of geographically associated information, a set of locations onto which a set of data may be mapped is required.

Maps organize large quantities of linear or spatially associated data for easy visual retrieval. Collecting, organizing, and maintaining this data has been a clerical function without the clearly defined storage - retrieval methodology of sequential filing cabinet organization. Displaying the data is a graphic artist's function. These artists, known as cartographers, generally organize the information to be stored/ displayed in an esthetically pleasing way. They also maintain or change the information by painstakingly and repetitively copying all the existing information while deleting, altering, or adding information in the process. Computers are particularly skilled at these two tasks: storing large quantities of data and painlessly and accurately copying and updating the data. Computer driven plotters are capable of displaying this machine readable map information in traditional cartographic format for visual map uses.

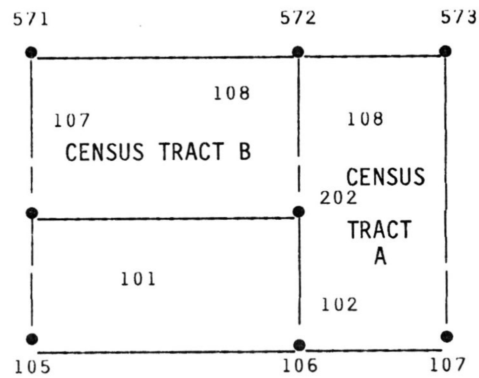
Two software packages exist which have been extensively utilized to produce choropleth maps. Choropleth maps are those that display ranges of values or categories of data within geographic boundaries. They are popular for the cartographic presentation of information, especially to highlight differences between areas. Both packages require that a set of x-y coordinates be prepared to describe the boundaries of the polygons or geographic areas.

SYMAP [1] was designed to produce low resolution inexpensive maps on the line printer. CALFORM [1] can draw high resolution, cartographer quality maps on a pen plotter. Those persons who initially attempted to plot the 1970 census data complained that 7/8 of the time to plot choropleth maps was spent developing the x-y coordinates describing the polygons [2].

Much data is collected by areal or geographic units and should remain in that form for analysis. For example, the Census Bureau collects innumerable descriptive statistics about the population and the housing stock by various political units. Marketing companies generate data for zip codes to determine which areas to target. Land use planning requires information on soil type for each buildable parcel. Storing data as attributes of geographic units lends itself to the development of a relational data base. This data base organization allows users to algebraically manipulate the data to answer more complex questions and to generate new information variables which could be mapped. Other papers [3], [8] have suggested the desirability of the relational data base as the underlying structure for a geographic information system with cartographic output. The structural power of a relational database requires data aggregated by and associated with geographic units.

DIME [4], which stands for Dual Independent Map Encoding, is a general purpose geographic base file, a machine-readable map. Each street segment, river segment, or political boundary segment is stored as an identifier such as street name together with related attributes such as, street direction, x-y coordinates of both endpoints, and the right and left side geographic identifiers, such as street address address, school district, or county. As the U.S. Census Bureau uses the DIME file to aggregate data to the various political units, this file is available for all urbanized areas in the country. In particular, it is required to be maintained sufficiently well to enable aggregation of the decennial census. All of the standard political units are included and others are frequently added for any additional local geographic unit, such as police beat. The DIME can be seen as a network, a collection of non-directed arcs con nected to other arcs at the endpoint, called nodes. Each node contains an x and a y coordinate.

A REPRESENTATION OF THE DIME NETWORK



The 200-300 N Apple St. segment has as node endpoints nodes 202 and 572. Its left side geographic identifiers are block 108, census tract B, its right side, block 108, census tract A.

In order to associate a geographic unit or polygon with many different data values, it is desirable to separate the base map (geographic constants) from the data values being displayed. Then, any new set of geographically associated data may be mapped onto any base map describing that geography. Base maps are generally organized by the cartographer as networks of lines but the data displayed is often spatial rather than linear. The cartographer bridges this disparity in view naturally while computer data storage techniques do not. Base map information is often linear, as with street maps, while data is most often areal or polygon oriented.

As an example, the base map may store Block A as four line segments or arcs running between four nodes or corners. Each segment has the attributes of street name, street direction, and areas delimited (Block A to the left, and areas delimited (Block A to the left, B to the right). Nodes have the attribute of location, generally an x and a y coordinate indicating distance from a known point or origin. Information that one may want to display, on the other hand, is likely to refer to Block A as a polygon, or spatial unit, for example, to display the number of children under 2 years on Block A. While current software requires polygons, manual methods for their creation are unsatisfactory. Therefore, another method of generating the coordinates is desired. This paper suggests a fully automatic algorithm designed to take advantage of the available DIME networks.

MAP ORGANIZATIONS AND CONSIDERATIONS

Machine readable maps can have their geographically associated data stored in one of three different organizations. These could be characterized as point, line, or area methodologies.

In the first, the point, or pixel organization, the map space is divided into equal size units. This can best be visualized by overlaying a regular grid of lines on a map showing information. Each unit on the grid is considered a point or pixel. In the machine readable version of the map, each pixel is stored as a location of the pixel and the information content of the map at that point. It is possible to store the location in just one x-y coordinate pair, as either a centroid or a specific corner, since each represents a known amount of space. Resolution is only as precise as the inverse of the pixel size i.e., the smaller the pixels, the higher the resolution. As a result, locations of boundaries are never absolute. On the other hand, gradations of data values between areas can be more clearly represented by this type of data organization. In the pixel methodology - the map does not exist except in concert with its set of data: the spatial concept and the data are inseparable. The pixel organization is the 'natural' organization for maps displayed on a line printer or on some CRTs since each of these devices display a unit of information in each prespecified location. Any cartographic information to be displayed on one of these devices must be translated into pixel organization, each pixel corresponding to one character location.

Line organization of cartographic information is the second major type of storage of map data bases. Each line segment or arc in a map is stored as two x-y coordinates, the end points of the line, and geocodes (geographic identifiers) to represent the meaning of the line or perhaps designate areas bounded by the line. With the line methodology, geographic or spatial data is totally independent of the map data base. This permits differing sets of data to be associated with the space on the defined map without re creating the database as is required with the point methodology. However, associating any spatial data with the network is difficult since the map organization defines lines rather than spaces. The line methodology works well when the data to be represented is naturally associated with the line segment; for example, type of street pavement or street width for a street map. This type of organization is efficient for display on a line plotter or storage type display tube. Some CRTs have hardware for 'line-to-point' conversion, making the line organization suitable for these CRTs also.

The third methodology of organizing a map data base is the area or polygon approach. This is the method required for the common cartographic software packages as seen in figure 2. In this case the polygons are stored as a series of line segments bounding the area. Each polygon represented requires a geocode to identify it and many x-y coordinate pairs to indicate the path of the boundary. Since polygons on a map are generally adjacent to other areas, it can be seen that all the edges in common are stored twice, once for each polygon. The base map will thus require almost twice as much storage as a map stored as line segments, particularly if many geocodes intrinsic to each segment must also be stored. However, polygon organization is ideal for straightforward association with spatial data. Additionally, the data remains independent of the base map. It is the natural organization for the collection and storage of spatial data because in itself a polygon is an areal figure. Polygons are not particularly efficient for any sort of display without transformation as each common edge is drawn twice, often resulting in slivering and overlap.

Maintenance of base map data, most efficient in an interactive mode, must also be considered for the three methodologies. The ideal maintenance system would provide for the section of the base map being altered to be displayed, changes keyed, and the resultant map displayed. In the ODIS [5] system, implemented as a public sector geographic information system, an interactive mode is utilized for inserting segments and changing the alphanumeric labeling portions of existing DIME segments but the resultant map must be viewed viewed as descriptions of the lines rather than as an actual network plot. The actual plot is important for visual feedback because the mistakes made in updating are often very obvious when seen visually. Additionally, improved productivity generally results with interactive file maintenance in contrast with the disjointed time segments necessitated by batch updates.

It can be seen that maintaining a pixel type map would be quite tedious since each pixel would need to be referenced and considered individually. Indeed, pixel type maps are usually generated in batch mode from a set of polygons and their associated data. Resolution problems are often introduced in the polygon to pixel transformation process. To verify such a map, the pixel representation is generated and the data examined with errors in the base map traced back to the polygon data set for correction.

For maintainance of a polygon base map, it is desirable that individual common segments of adjacent polygons be accessible for maintenance as one entity. For this purpose the base map must be organized as a network of segments. Updating of polygon organized base maps is done in batch mode by sorting into segments, editing and resorting/reorganizing back to polygons. Since this is not desirable, having the map base data organized as a network is optimal for maintenance purposes.

Separating the network base map from the data allows for efficient interactive updating of the map with display feedback for verification. Changes can be seen in reference to other neighboring or attached arcs. Although polygons may be interactively edited, they generally contain no information about their connectedness with neighboring polygons. Therefore they are treated more as individual entities and it is more difficult to examine changes in relationship to other parts of the map. Additionally, utilizing the DIME network for which the correction and maintenance procedure is an ongoing federal program [6] would eliminate the need to maintain a separate special purpose database.

POLYGON PRODUCTION

To generate the polygon boundaries utilized by existing software it has been necessary to manually digitize the x and y coordinates for each node of each polygon. This is time consuming, tedious and prone to error. As segments held in common by two adjacent polygons must be digitized twice, the coordinates of the nodes are rarely exactly the same as the digitizer has a higher resolution than the human hand. In addition to being repetitive, the difference in x and y coordinate values for nodes that should be in the same place tend to cause overlap of polygon boundaries and/or space that is not included in either polygon. These problems with manual digitization are covered by Ducker [7].

A significant improvement in this manual method has been suggested by McIntosh [8]. With this method the nodes are each digitized just once and the system generates the second copy of the line segment thus eliminating slivering and overlap. The routine then chains the segments by matching node numbers and displays the resulting polygons for user approval. When the user is satisfied the closed polygon sets are outputted to be used with standard mapping software. If the polygons are not chained correctly the nodes must be redigitized.

In order to take advantage of the already digitized nodes in the DIME, the following algorithm is suggested.

1) Extract from the DIME network all those segments with differing geo-codes left and right where one of the codes names the geographical unit for which polygon boundaries are desired.

2) Generate from each extracted record two boundary segments, one arc for each of the polygons. Each resulting segment is d defined by the names of its two node endpoints and the name of the polygon to which it belongs. The segments or arcs are non-directed.

3) Create a record for each named endpoint containing its node name and its associated x and y coordinates.

4) Sort the endpoints by node name and eliminate duplicates. Since most street intersections involve four or more segments, each node is typically used four or more times. Eliminating the coordinate duplication saves considerable memory space.

5) Sort the boundary segments by polygon name and treat each group of segments potentially forming a polygo as a separate network.

An efficient method for chaining the ordered segments has been suggested by Minieka [9] in his work on algorithms for traversing networks. The cycle method utilized in this system requires an 'even' graph or network. A graph is even if each node is used an even number of times. In the case of a closed polygon this will always be true. The chaining algorithm used is:

1) Starting with the first available segment put the two nodes on a list and mark the segment as used.

2) Look through the rest of the unused segments for a segment that uses the last node on the list; when it is found add the other end-point's node to the list and mark the segment used.

3) If the last node on the list is not the same as the first repeat at step 2.

4) If all the segments are used, the boundary is complete, otherwise save this subgraph and start again at step 1.

5) When all segments are used, splice the subgraphs together at the repeated points.

This cycle algorithm requires that the boundary not be disjoint and that there be no extraneous or extra segments. Due to the complex nature of some geographical coding schemes, including the application for which this algorithm was developed, not all of the segments may be boundary segments but may be segments within the polygon. Additionally, some political boundaries describe disjoint areas. An acceptable polygon is defined to be one that either uses all the boundary segments, or one for which all the unused segments are inside the boundary described by the polygon. If the cycle algorithm fails to produce such an acceptable polygon then the algorithm is tried with different starting segments. All possible branches from a node or vertex of the network are held in a stack so that all possibilities may be tried without repetition or omission. If a segment does not yield an acceptable polygon when used as a starting point, it must not be on the polygon boundary, and it is flagged to be omitted from future searches.

Vectors drawn from A and C, points inside the polygon, cross the boundary an odd number of times; B, outside, crosses an even number. Segment AC intersects the boundary and is thus not contained although both end-points are contained.

All potential polygon boundaries are checked to see that all unused segements are within the described polygon. Jacobson [10] gives an algorithm for deciding if this is the case. Each endpoint or node is tested to see if it is inside the polgon by determining if a vector drawn from the point intersects the polygon boundaries an odd number of times. If so, then the point is inside. If both end-points are found to be inside the polygon, then the line defined by the end-points is tested for intersection with the segments in the potential boundary. If the line segment being tested does not intersect any boundary segment, the segment is contained within the polygon. If all the unused segments are contained within the polygon, the polygon is considered complete and correct.

When a polygon is accepted the system outputs its boundary in CALFORM standard node chain format. Nodes that are used are flagged. After all network sets have been processed into polygon boundaries, the flagged nodes are outputted also in CALFORM format, to provide the required x and y coordinates for the plotting software.

SUMMARY

While the algorithm discussed above adequately provides coordinate boundaries of geographic polygons for use with mapping software, there are two extensions that might be considered. The alphanumeric labeling package utilized by the mapping software could be generated by the system. After each polygon is bounded, coordinates could be computed to place the name of the polygon in the middle of the polygon. The polygon names and their placement coordinates would be outputted to form the labeling package.

Occasionally the algorithm cannot produce a closed polygon from the network inputted. This occurs due to an error introduced into the DIME when additional local and non-standard geocodes are added to the street segments in the file. During this extension, some of the street segments may be split creating two new node endpoints. This is done because the boundary being geocoded runs down being geocoded runs down an alley rather than a street center. Each end of the former segment belongs in a different polygon. This division implies a new arc in the network, that is, the alley and two new nodes. But these segments are rarely coded and integrated into the DIME by the persons adding the geocodes, as they have no use in geocoding. This results in missing segments and/or node names in the boundary when the algorithm attempts to generate a closed polygon. A solution that could generate these missing segments when required would increase the power of the algorithm.

REFERENCES

[1] SYMAP and CALFORM are program products of the Laboratory for Computer Graphics and Spacial Analysis at Harvard University.

[2] Maxfield, 'Computer Mapping of Census Aggregated Data', URISA, 1968: Urban and Regional Information Systems, Federal Activities and Specialized Programs, p

[3] Alsberg, P. A., 'The Management of a Large Data Base in IRIS', Journal of Chemical Information and Computer Science, Vol. 15 #1, Feb. 1975.

[4] The DIME Geocoding System, Census Use Study Report No. 4, U.S. Census Bureau, Washington D.C., 1970.

[5] Kinzie, Steve, 'ODIS: An Online Dime Implementation System' URISA 1976: Information System as Services to Citizens, pp. 504-515.

[6] 'Geographic Base File Correction-Update-Extension', Dept. of Commerce Publication GEO-202, U.S. Census Bureau, Washington D.C., 1973.

[7] Ducker, Kenneth, 'Cartographic Data Structures' Computer Graphics Vol. 10 #1, July 1976 pp. 167-172.

[8] McIntosh, John, 'The Interactive Digitizing of Polygons' Computer Graphics Vol. 12 #3, August 1978 pp. 60-63.

[9] Minieka, Edward, Optimization Algorithms for Networks and Graphs Marcel Dekker, Inc., New York, 1978.

[10] Jacobson, J. D., 'Geometric Relationships for Retrieval of Geographic Information', IBM Systems Journal, Vol. 7 #3 & 4, 1968 pp. 331-341.