Quick Algorithm of Selecting and Capturing Geographic Information

LI Hongli   XU Haitao   LI Hanrong   YU Jinxing
Tianjin Institute of Hydrographic Surveying and Charting , P.R China
Fax (8622) 28353537

Abstract

From the view of application, this paper discusses the category of geographic information and an accessing method of theoretical data model and data Structure is given.

The main feature of the method is the following. Firstly, the great amount of information is divided into map Tiles moderately according to map scales and different regions. Secondly, map Tile is separated into some coverages with the same structure. Thirdly, coverage is described in spatial data and attribute data which linked together by a common identifier .The spatial data is constituted by control section, index section and data section .The attribute data is presented in database table.

In the process of map digitizing,   coordinate extent value, grid index mark, data pointer and data length are provided in the index section for the rapid selecting and capturing of geographical information .The paper also offers the method of quickly selecting and querying geographical information on the base of above arrangements, and also provide a compound selecting method under both coordinate condition and attribute condition.   It describes in details both the generating method of kinds of checking tag and the course of selecting and capturing geographical objects from figures drawn on the screen.

At last, a instance and its effect applied in the Hydrographic Information System is given.

Introduction

The main objects that GIS (geographic information system) deals with are the coordinate information and attribute information which have spatial topology relations and associated relations. Digital map is the carrier of all those information.

The corresponding object data of those information can be found out and recovered into visual graphics according to user's requirement through information managing and displaying software (including the function of zooming in/out, windowing, layering, panning, inquiring and analyzing, etc).It is accessible that the corresponding object data are found out from the visual graphic and be analyzed .The former processing is named selecting of the objects and the latter is called the capturing of the objects .The speed of selecting and capturing lies on the theoretical data model and data Structure used in a system.

The implement of quick selecting and capturing of geographic information is

very important to a real time GIS .The method we discussed followed is formed in the process of developing Hydrographic Information System and ECDIS. It has a perfect application effect.

1．Theoretical data model and data Structure

1．1 Theoretical data model

Theoretical data modal used in the paper is FULL TOPOLOGY.

It is described in the IHO TRANSFER STANDARD for DIGITAL HYDROGRAPHIC DATA(Special Publication No.57 Edition 3.1) as follows:

A set of nodes, Edges and faces. A planar graph with defined faces. The universe is partitioned into a set of mutually exclusive and collectively exhaustive faces. Isolated nodes may reference their containing faces and edges must reference the faces to their right and left. Point representations are coded as nodes (isolated or connected). Line representations are coded as series of edges and connected nodes. Area representations are coded as faces. Duplication of coincident geometry is prohibited.

The categories of objects include regions which are the aggregation of faces with same characters, compounding objects which are constituted by associated simple objects, annotations as well as nodes, edges and faces .The visual output of geographic information and the relations of kinds of information are showed in the figure below.
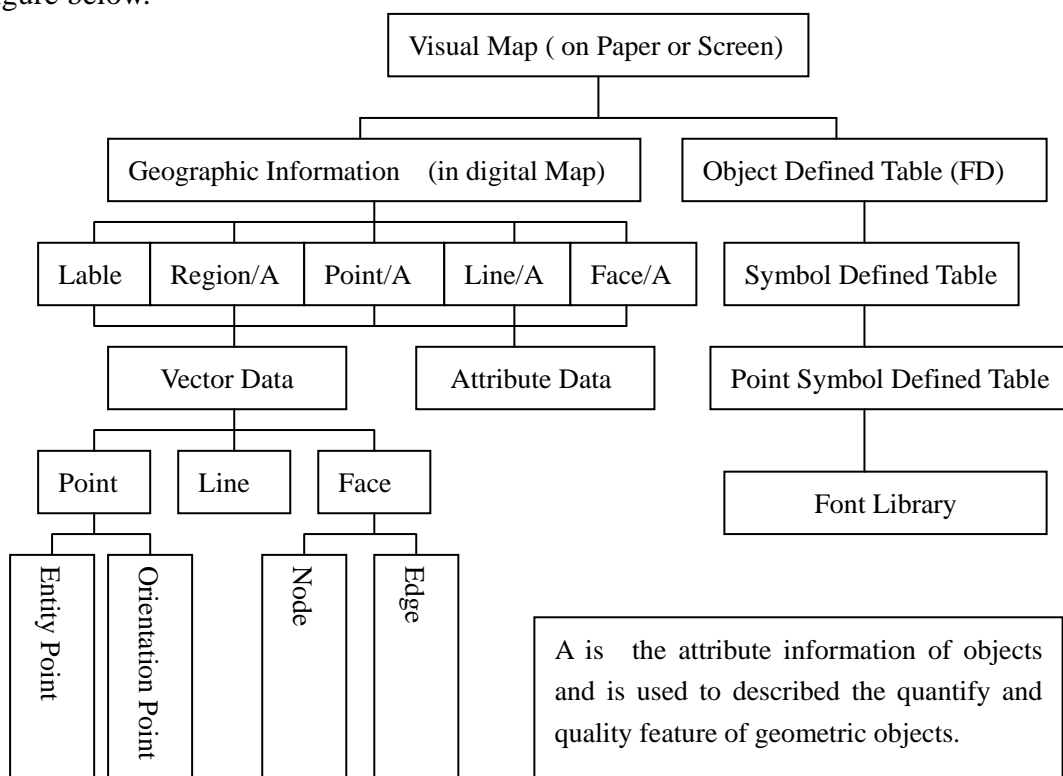
1．2 data structure .

The mean we adopt is followed:

1.2.1 Be managed by map catalogue database, a wide geographic spatial is divided into some map TILEs with different scales and regions. A map

TILE is controlled by the file whose extension is *.ctl and is separated into some coverages which is characters by geographic categories .The map TILEs control file records Name, Number, scale, scope, projection for a map and also records coverage name, coverage code, coverage number, etc.

1. 2. 2 All the coverages have the same structure and are constituted by vector data and attribute data which are stored separately .The objects stored as vector data are recorded in form of non-fixed length record in the files which have spatial topology and associated relations each other. Attribute data of objects are managed in related data base .The vector and attribute of a object are linked together by record number and object sequence number .

1. 2. 3 Vector data are separated into three parts that are control section, index section and data section. The information stored in control section is the description message of coverage, coordinate extent value, number of objects, address of index section, number of indices, address of data section and data length .The index section allots every objects an fixed length inquiring index that keeps the information of the object such as the sequence number, type, dimension, coordinate extent value, topology relation ,attribute record number , grid index mark, address of data section and data length of the object ,etc. The coordinate message of the object, such as group of coordinates of lines or edges, inner sequence number of boundaries that associated with a face, inner sequence number of child hole to the face, group of coordinates of base lines of annotation, character string of annotation, are all stored in the data section .To Point objects,   all the information of them are stored in index section.

1. 2. 4 Attribute information   are managed by related data base. Every record of the attribute tables correspond to a object and record its quantify and quality features in data items. The code of object category and inner sequence number are dispensable.

1.2.5 The organization method data stored in magnetic disk is followed. A chief catalog represents a region, a child folder represent a map Tile ,a grandson folder represents a coverage in which vector data and attribute data are stored .The relationship of them is showed in the figure below .

文件夹	×	名称	大小	类型	修改时间
□ map		Buldnt. dht	7KB	DHT 文件	99-5-22 9:17
⊞ C10011		Buldnt_1. dbf	1KB	DBF 文件	99-5-22 9:16
□ C10012		Buldntl. dbf	1KB	DBF 文件	99-5-22 9:17
Anvgp					
Arean					
Areap					
Buldn					

2．The generating of index marks

2．1 the classification of index marks

To produce kinds of index marks in advance is the key to the quick selecting and capturing of geographic information .We classify index marks in three types: marks for selecting map Tiles in a region , marks for selecting coverages in a map Tile ,marks for selecting objects in a coverage .

2．2 marks for selecting map cells

Marks for selecting map Tiles are kept in map catalog data base which is a data table . The frame information of a map Tile are recorded as a raw in the table and the parameters of map ,such as map number ,map name ,map scale ,map scope and the path of map stored in disk are described in data items. An example is showed below:

Map_number	Map_name	Map_scale	Map_proj	Area_ll_b	Area_ll_l	Area_ur_b	Area_ur_l	Map_path
10011	黄海北部及渤海	1000000	墨卡托	35.2960	116.59600	40.5960	127.14600	C10011.CTL
10012	青岛、上海至长崎	1000000	墨卡托	30.2960	119.47600	36.2360	130.09600	C10012.CTL
10013	上海、马尾至冲绳岛	1000000	墨卡托	24.5960	119.20360	31.2460	129.40360	C10013.CTL
10014	台湾东、南部诸水道	1000000	墨卡托	18.1460	119.35600	27.4460	126.44600	C10014.CTL
10015	福州至广州	1000000	墨卡托	19.4960	112.54600	26.2460	123.09600	C10015.CTL

When you create a new map Tile, you just need to put the corresponding parameters into map catalog database.

2．3 marks for selecting coverages

Marks for selecting coverages are kept in the parameter control file of a map Tile which is a data table too. The information of a certain coverage is stored in a record, and the name, sequence number, coordinate extent value, number of objects are noted in data items. When you create a new coverage or add a coverage or combine coverages in the map Tile, you just need to put the corresponding parameters into this file.

2．4 marks for selecting objects

The index section in the coverage files allots every objects a fixed length inquiring index that keeps the information of the object such as the sequence number, type, dimension, coordinate extent value, topology relation, attribute record number , grid index, address and length of a object in the data section ,etc .We classify the objects into seven categories in geometry : Point ,Node ,Line ,Edge ,Area ,Region and Annotation . The index structure of every object is same in data length and coordinate mark for selecting objects, but also has a little difference .An example of index structure of Edge programmed in C++ is given :

```
typedef struct  target_EDGE
{        short          seq ;       // Sequence number. If seq<0 then feature has deleted.
         Short          type;       // Object Type = 5
         Short          dim;        // Dimension
         short          pn ;      // Points Number
         Cpoint         P_min;    // left_down point ( x, y)
         short          Left_Area ; // left area's sequance
         Cpoint         P_max;     // right_up point ( x, y )
         short          Right_Area ;   // right area's sequance
```

```
        Cpoint          l_s   ;          // start point of   a edge
        short           stangle;         // the angle from start point
        Cpoint          l_e   ;          // end point of   a edge
        short           enangle;         // the angle from end point
        unsigned long   data_pt;             // data pointer
        unsigned short  data_length;         // data length
        unsigned long    GridIndex;  // GridIndex of feature
        short    attr_id ;               // The record number in Attribute database
        short    stnd_seq;               // the first point's   Node seq
        short    ennd_seq;               // the end point's   Node seq
}        EDGE;
```

In the above definition , P_min and P_max(coordinate extent value), , GridIndex (grid index mark) are the indispensable   mark in selecting objects in interspace location. Cpoint has a two dimensions structure .Seq, type, dim,pn, data_pt, data_length and attr_id are the common feature of all kinds of objects.Others are differ in kinds of objects and mainly described topology relations and associated relations of objects .As a region is presented , the sequence number of a object may be quickly found out in the condition of coordinate extent value, grid index mark and data records .The process of implement would be discuss in the next section .Now we will briefly discuss the generating arithmetic of coordinate extent value and grid index mark.

a. coordinate extent value

To the Point object with single point location, coordinate maximum value equals to coordinate minimum value and the coordinate of   the   Point object.

To the Line, Edge and Text Lable objects with multi-point location, coordinate extent values equal to the maximum and minimum value of the group of points which constitute the base lines.

To the face object, firstly you will find out coordinate extent values of the Edges which structure the topology relation with the face object, then constitute those values into groups, at last find out the coordinate extent value of the face from the group.

To the region object, coordinate extent values of every face in the region will be found out and turned into group after the region has been formed, then calculate the coordinate extent value from the group.

b. grid index mark

GridIndex , the grid index mark , is a N bit data storage .It points out the possibly position of a object in a map cell and serves for primary selecting .A rectangle map cell is departed into M*N cells and the length and width of every cell is :

$$Dx=Width/M$$
$$Dy=Height/N$$

Dx and Dy are called the minimum selecting steps. Every cell is represented by a Bit .If the object is located in the Cell, the corresponding value of Bit is 1,if not, the value is 0.In favour to calculate in computer, we stipulate:

$$M*N+L=8T（Bits）$$

M、N、L、T are integers，T is the number of bytes，M/N approximates 1.

If a GridIndex has 64 cells ， then M=8,N=8,L=0,T=8.

The more cells are partitioned, the more exact selecting will do. But the exactness is companied with adding cost of time and storing capacity .It is more suitable to use 32 or 64 Bits as an object's grid index mark generally.
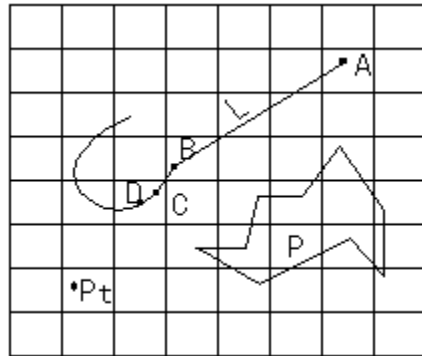
The figure above shows that Line L, Point Pt ,Face P lie in a 8*8 Bits grid index .The generating method is explained follow:

To Point objects, $K=Y/Dy*M+X/Dx$

(X,Y) is the location coordinate of point in the map . Dx,Dy are the length and width of a map cell. M is the number of cells in vertical direction .K is the corresponding index mark byte of that point .The left_down corner of the map is set as (0, 0)．For example, Point Pt is located in the place 9 and the value of the ninth byte is 1.

To Line objects, two cases will be taken account of:

When neighbor points lie in the same cell, like point C and D, the method equals to that of points and overlap points are not recorded.

When neighbor points cross several cells ,like point A and B, new points will be inserted in the line AB so as to make continued cells. We regard Min(Dx,Dy)/3 as increase step .

To Face objects, GridIndex of Edges need to be got first, then the GridIndex of perimeter of the face by means of "OR" every bit .The cells included in the second GridIndex decide the GridIndex of the face. The filling arithmetic is followed:

(a) scan the bits horizontally, find couple "1" ,and turn the bit inside into "1".

(b) scan the bits vertically, find couple "1" ,and turn the bit inside into "1".

(c) "AND" the results of (a) and (b),the result of "AND" is the GridIndex of the face

If there are holes in the face, the GridIndex inside the hole should "XOR" the GridIndex of the face and holes are excluded out the face.

3.   the process of quickly selecting and capturing objects

The quickly selecting and capturing of geographic information is based on the theory discussed in above text. Selecting marks are created in the process of digitizing objects .We expatiate the process in the follows.

3．1 Selecting according to position

Selecting refers to finding out objects accord to the two conditions, quality and position. Quality feature refers to type, coverage, dimension and attribute value (as object character code) in attribute database, all that have been mentioned in the text. Position condition selecting follow such steps. Firstly one or more corresponding maps would be chosen out from map catalog database. Secondly the coverage is made certain according to the corresponding geography catalog of the object. Thirdly common selecting is made in the coverage files by grid index mark and coordinate extent value .At last selecting is persecuted by data records.

The flow of single object selecting in a coverage is as follows:

1.To get the GridIndex of the window on certain condition.
2."AND" the GridIndex of object and the GridIndex of the window by bits, the result is not "0" means the object may be inside the window, otherwise means the object could not inside the window.
3.To exclude the objects outside window by means of coordinate extent value.

Objects as Line and Edges need to be careful selecting by data records. The objects chosen after the above prosecution lie in the window absolutely.

From the above steps, selecting speed is accelerated.

The flow of selecting objects in a coverage is as follows:

1.An area (pointer is *Seq_Ptr) of storing sequence numbers is built and a counter (Seq_Num)is used to count .
2.The GridIndex of window given in a map is calculated. Single object selecting program is prosecuted one by one by sequence number from begin to end. The sequence number of the object which fulfill the requirement is recorded in the storing area and the counter is used to count the number .The result of the selecting is applied in displaying and inquiring.

This method of selecting object has advantage in its exclusion character .In mostly cases, the operation which is only related with the index of object is more efficient in a real-time system.

3．2 Selecting according to attribute and position

The attribute of object are recorded in the attribute database and linked the coordinate of the object with attribute record number and the sequence of the object. Compound selecting is necessary. Firstly we use the method discussed in 3.1 and find out sequences inside the window. Then use those sequences to find out the attribute of the objects and get the sequences of the objects which can fulfill the condition. At last make careful selecting through data records and get the conclusive results. The attribute condition should fulfill SQL rules and the operation should be done by database programs , as ADO of Microsoft.

3．3 the capturing process

Capturing refers to finding out some objects around a point in the screen and search out their attributes and sequence .The ideal speed of capturing objects in a mature GIS is that information is gotten immediately as the point be   discriminated.

The objects present as   visible maps   in   screen and have their own sequence in ems memory. We collect the sequences of the objects and pay no attention to the

index .

The searching window does not equal to the whole screen window and is just a circle area which is named discriminate window .The discriminate window is small and take the discriminated point as center of the circle whose radius is usually 3mm.

The following is the capturing process:

Firstly a point is chosen from screen by mouse and the discriminate window and its GridIndex are gained .By the use of discriminate window and its GridIndex,the objects that inside the screen window but outside the discriminate window are put aside .Then the remained objects are chosen by the data records and the sequences of the objects that fulfill the condition are collected .

The objects which lie in the discriminate window are always more than one .It is necessary that they are lightened one by one to be insured by user .The probability of capturing the objects depends on the choosing of the discriminated point .To point object ,the discriminated point should be nearby. To Line object, the discriminated point should be in the area that regard the Line object as center line with width of double discriminated distance. To text object, the discriminated point should be in the area based on the baseline instead of the character string. To face object, the discriminated point should be near the edges of it or inside it.

4．Project sample and effect

According to above thoughts, we manufacture several systems (one of which is Hydrographic Information System) by means of the softwares of MICROSOFT 's VISUALL C++ and VISUALL FOXPRO .When those systems are operated in the computer which has 128MB EMS memory and PII 450 CPU ,it   cost no more than several seconds to find out several special objects from the digital chart data base which include hundreds charts and have 1GB data capacity .To display a chart with more than score coverages only cost two or three seconds .To capture a object from the graphic arrives at the grade of real –time .To search out the information with the compounding method depended on locations and attributes only spend a few seconds.

The arithmetic we introduce above has good value in application.

Reference

1. IHO TRANSFER STANDARD for DIGITAL HYDROGRAPHIC DATA
（Edition 3.1 - November 2000）
2. Hydrographic Surveying and Charting, 1995-4
3. Programming Visual C++ 6.0, Microsoft Press.