Name : BIKASHITA KALITA
Paper 604(Unit- 2) GIS

SPATIAL DATA STRUCTURE#####

INTRODUCTION:. Geospatial data is a representative of both spatial and attribute data. For example,. The location information of the river constitutes the spatial data whereas the characteristics information represents attribute data. GIS is used to represent spatial features on the surface of the Earth as map features on a plane surface. Data models are used to represent spatial features in a GIS environment.There are two types of data models namely,vector data model and raster data model and vector data model.

Geospatial data is numerical representation which analyses and describes real world features in GIS. The nature of geospatial database is a dynamic rather than static and allows a range of functions such as organising, storing, processing, analysing and visualising spatial data. Geospatial data depicts there all world in two basic models such as the object-basedmodel and tthe field-based ***model*** as shown below:

1.Object-Based Model: The object is a spatial feature and has some characteristics like spatial boundary, application relevant and feature description (attributes). Spatial objects represent discrete features with well- defined or identifiable boundaries, for example, buildings, parks, forest lands, geomorphological boundaries, soil types, etc. In this model, data can be obtained by field surveying methods (chain-tape, theodolite and total station surveying, GPS/DGPS survey) or laboratory methods (aerial photo interpretation, remote sensing image analysis and onscreen digitisation).
Depending on the nature of the spatial objects we may represent them as graphical elements of points, lines and polygons.

2. Field-Based Model: Spatial phenomena are real world features that vary continuously over space with no specific boundary. Data for spatial phenomena may be organised as fields which are obtained by direct or indirect sources. Source of direct data is from aerial photos, remote sensing imagery, scanning of hard copy maps, and field investigations made at selected sample locations. We can obtain or generate the data by using mathematical functions such as interpolation, sampling or reclassification from lselected sample locations.This approach comes under in direct data source. Forexample,Digital Elevation Model(DEM) can be generated from topographic data such as spot heights and contours that are usually obtained by indirect measurements.

(The Digital Elevation Model (DEM) consists of an array of uniformly spaced elevation data. A DEM is point based but it can be easily converted to raster data by placing each elevation point at the centre of excellence).

.Geospatial data base possess two distinct components such as locations and attributes. Geographical features in the real world are very difficult to capture and may requires large scale data base.Each model tends to fit certain types of data and applications better the anothers.All spatial data models fall into two basic categories: raster and vector data model.

Advantages and disadvantages of vector and raster DataModels:--

The representation of raster and vector models for geospatial data has

two divergent views of the real world as well as data processing and analysis. To solve different geospatial problems oviously the see two models can be used.

Advantages of Raster Data:--

1.data structure issimple

2.good for representing continuous surfaces.

3.location specific data collection.

4.spatial analytical operations arefaster

5.different forms of data are available (satellite images, field data, etc.),and

6. Mathematical modelling and quantitative analysis can be made in easiestway due to the inherent nature of rasterimages.

Disadvantages of Raster Data

1. ata volumes arehuge

2. Poor representation for points, lines andareas

3.cartographic output quality may below

4.difficult to effectively represent linear features (depends on thecell resolution). Hence, the network analysis is difficult toestablish

5.oordinate transformation is difficult which sometimes cause distortion of grid cellshape

6. Suffer from mixed pixel problem and missing or redundant data,and

7. Raster images generally have only one attribute or characteristic value for a feature or object, therefore,limited scope to handle the attribute data.

Advantages of Vector Data

- data structure is morecompact
- data can be represented with goodresolution
- it can clearly describe topology. Hence,good forp proximity and network analysis
- spatiala adjustment of the data is easy with the utilisation of techniques such as rubber sheeting, affine,etc.
- graphic output in small scale as well as large scale gives agood accuracy.
- geographic location of entitiesaccurate
- modernising and generalisation of the entities arepossible
- easy handling of attribute data,and
- coordinate transformation techniques such as linear transformation, similarity transformation and affine transformation could be doneeasily.

Disadvantages of Vector Data

- data structures arecomplex
- overlay analysis is difficult in processing. Often, this inherently limits the functionality for large data sets, e.g., a large number offeatures
- data collection may beexpensive
- high resolution drawing, colouring, shading and also displaying may be time consuming and unbearable

- technology of data preparation is expensive
- representation of spatialv variability is difficult ,and
- spatial nalysis and filtering with in polygons is impossible.

5.1 SPATIAL DATASTRUCTURE

Structures that provide information required for computer to reconstruct spatial data model in digital form are defined as ***spatial data structure.*** Many GIS software have specific capabilities for storing and manipulating attributes data in addition to spatial information. However,basic spatial data structures in GIS are mainly vector and raster.

5.1.1 Raster DataStructure

Raster or grid data structure refers to the storage of the raster data for data processing and analysis by thecomputer.There are mainly three commonly used data structures such as--------

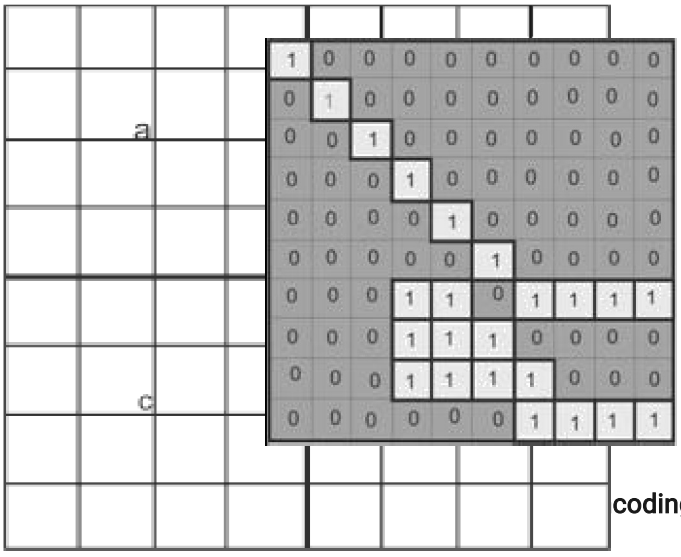
A)cell-by-cell encoding

B) Run -length encoding,and

C)quadtree

A)Cell-By-Cell Encoding Data Structure

This is the simplest raster data structure and is characterised by sub dividing a geographic space into qrid cells.Each pixel or grid cell contains a value.A grid matrix and its cell values for a raster are arranged into a file by row and column. In Figure value 1brepresents the gray cells and 0 has no data.This is called cell-by-cell encoding method. It can also be used for sttorage of data in satellite images. Most of satellite images consist of multispectral bands and each pixel in a satellite image has more than one value. Mainly three formats such as Band Sequential (BSQ), Band InterleavedbyLines(BIL),andBandInterleavedbyPixels(BIP)areusedto storedatainamultiband/multispectralimagery.



coding data

B)Run-Length Encoding Data Structure:--

Run-Length Encoding(RLE) algorithm was developed to handle the problem that a grid often contains are abundant to missing data. When the raster data contains more missing data, the cell-by-cellencoding method cannot be suggested.In RLE method, adjacent cells along a row with the same value are treated as a group

called a*run. If a whole row has only one class*,it is store datas thebclass and the same attributes are kept without change.Instead ofbrepeatedly storing the same value for each cell, the value is store once together with the number of the cells that makes the run. In the Figure , explains the run-length encoding structure of a polygon. In the figure, the starting cell and the end cell of the each row denote the length of group and is generally called as run. RLE data compression method is used in many GIS packages and in standard image formats

Fig. B) Run-length encoding data structure

C)Quadtree Data Structure:--

To compress the data as well as to save the space in original grid, quadtree data structure can be used(Fig.C).A quadtree works by dividing a grid into four quadrants for the available data. The available data quadrant is again split into four half-size quadrants and so on until the individual pixelis reached.The attribute data for all the pixels of the quadrant remains the same even if it is divided.

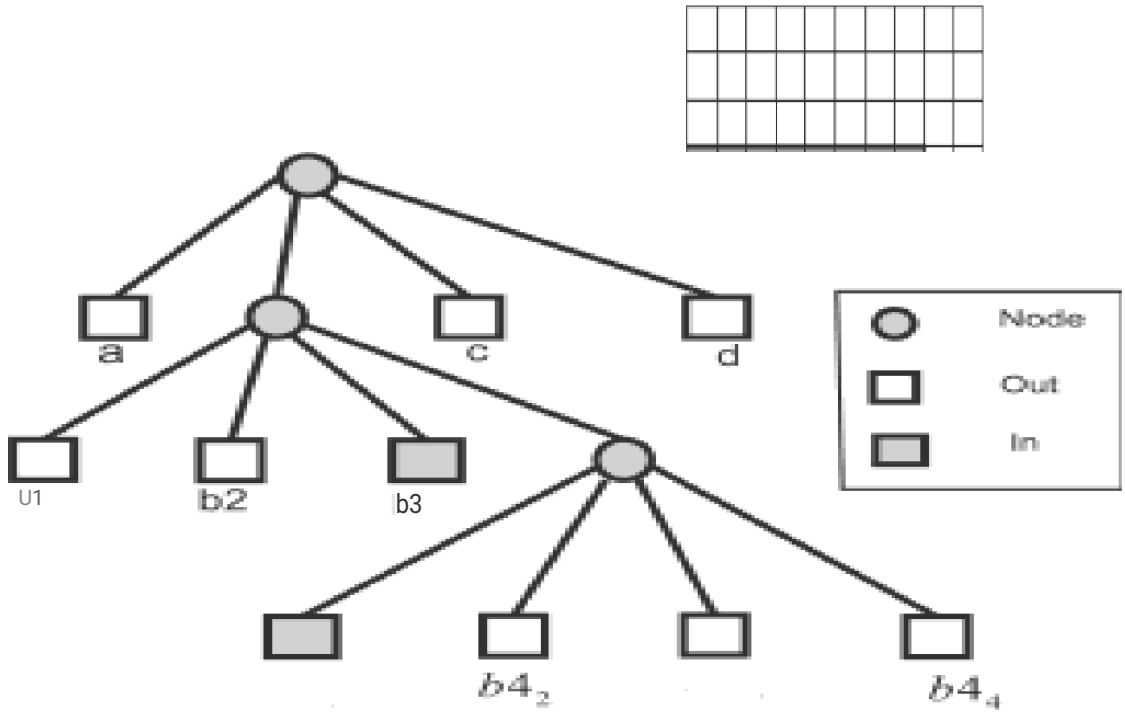


Fig. C: Quadtree data structure

2.Vector DataStructure

Vector data structures are now widely used in GIS and computer cartography.This data structure has an advantage in deriving information from digitisation, and is more exact in representation of complex features such as administration boundaries, land parcels,etc. In GIS, vector files were simply lines and were having only starting and ending points.

a) Topological DataStructure

Topologic data structure is often referred to as an intelligent data structure because spatial relationships between geographic features are easily derived when using them.Because it's mportant inundertaking complex data analysis.In a topological data structure, lines cannot overlap without a node whereas lines can overlap without nodes in a non- topological data structure.

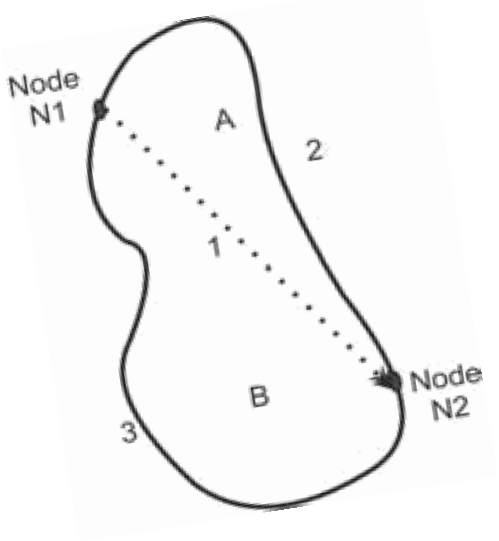
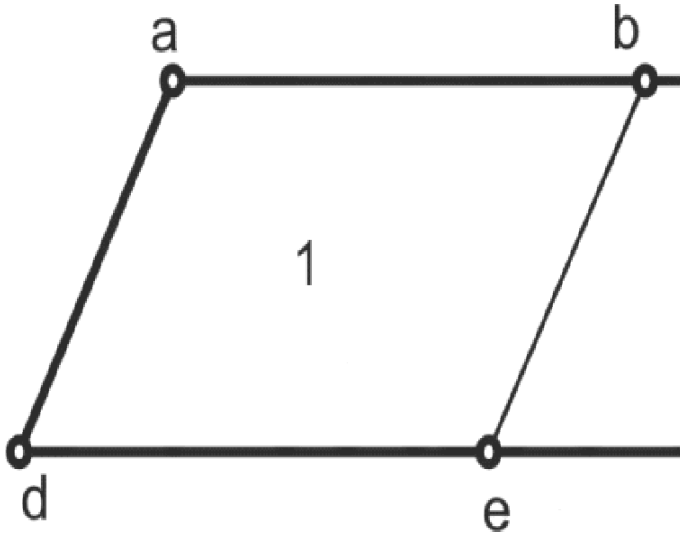


Fig: 2. a)Topological structure of the arc

i) *Topologically Integrated Geographic Encoding and Referencing: (TIGER)*

It is geospatial data created bybUS Bureau of Census as an improvement to the Geographic Base File/Dual Independent Map Encoding (GBF/DIME) data structure. This data

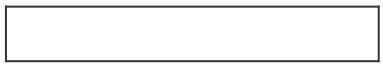


structure or format was used in the 2000 census by US Bureau of the Census. In the TIGER database, points are called 0-cells,lines are 1 -cells, and areas are 2-cell. each1-cell represents a direct line which starts from one point and ending at a another point.The line comprises both sides of the data.Each 2 and 0-cells share of the information of the 1-cells associated with it. The main advantages either the right side or the left side of a street or road.

Fig.2.1 Topology in TIGER database

ii)Coverage DataStructure

Coverage data structure was practised by many GIS companies like ESRI,in their software packages in 1980s to separate GIS from CAD (Computer Aided Design). A coverage data structure is a topology based vector data structure that can be a point, line or polygon coverage. A point is a simple spatial entity which can be represented with topology. The point coverage data structure contains feature identification numbers (ID) and pairs of x,y coordinates, as for example A (2, 4) (. Figure 2.1)Data structure of line coverage is represented in Figure. The starting point of the arc is called from node (F-Node) and where it ends to node (T-Node). The arc-node list represents the x, y coordinates of the nodes and the other points (vertices) that generate each arc.



b)Non-Topological Data Structure:-Vector data structure that is common among GIS software is the Computer Aided Design (CAD) data structure. Drawing Exchange Format (DXF) is used in the CAD package(e.g.,AutoCAD)for transferring of the datafiles.

CONCLUSION:---Data structures are a method of storing data in a systematic way. There are mainly two data structures used in GIS, namely, raster and vector. In raster structure mainly three data structures, i.e. cell-by-cell encoding, run-length encoding and quad tree while in vector, topological and non-topological data structures are used to store the spatial data.