UNIT 10 - SPATIAL DATABASES AS MODELS OF REALITY

UNIT 10 - SPATIAL DATABASES AS MODELS OF REALITY

Compiled with assistance from Timothy L. Nyerges, University of Washington

This begins a three unit section covering some basic principles of spatial databases. As these issues are very fundamental, many of them are introduced here but dealt with in much greater detail in later units.

UNIT 10 - SPATIAL DATABASES AS MODELS OF REALITY

Compiled with assistance from Timothy L. Nyerges, University of Washington

A. INTRODUCTION

- the real world is too complex for our immediate and direct understanding
- we create "models" of reality that are intended to have some similarity with selected aspects of the real world
- databases are created from these "models" as a fundamental step in coming to know the nature and status of that reality

Definition

- a spatial database is a collection of spatially referenced data that acts as a model of reality
  - a database is a model of reality in the sense that the database represents a selected set or approximation of phenomena
  - these selected phenomena are deemed important enough to represent in digital form
  - the digital representation might be for some past, present or future time period (or contain some combination of several time periods in an organized fashion)

Standards

- many of the definitions in this Unit have been standardized by the proposed US National Digital Cartographic Standard (DCDSTF, 1988)
  - these standards have been developed to provide a nationally uniform means for portraying and exchanging digital cartographic data
  - these cartographic standards will form part of a larger standard being developed for the digital representation of all earth science information

B. DATABASE CONTENT AND AN ORGANIZATION'S MISSION

Organization mandates

- organizations have mandates to perform certain tasks that carry out their missions
  - mandates are the reasons they exist as organizations
- organizations have different needs for data depending on their mandates and the activities required to carry out these mandates
  - mandates often help identify and define entities of interest, requiring a certain view of the world
  - what might seem at first glance to be the same data need in two different organizations can actually be quite different when we look at a more detailed level
    - e.g. wildlife and forestry departments both need information on vegetation

but the detail needed is different

Database contents

Example: Transportation

- highway data from the different points of view of a natural resources organization and a highway transportation organization
  - a natural resource organization might only need logging roads and the connecting access to state highways
  - the transportation organization's main interest is in characterizing highways used by the public
    - the database might also be used to store detailed highway condition and maintenance information
  - we would expect their need for highway data to be more detailed than would the natural resource organization's

Example: wetlands

- wetlands data from the different points of view of an ecological organization and a taxing authority
  - ecological organization might define wetlands as a natural resource to be preserved and restricted from development
    - that perspective might require considerable detail for describing the area's biology and physical resources
  - a taxing authority might define a wetland to be a "wasteland" and of very little value to society
    - that description might require only the boundary of the "wasteland" in the database

Database design

- in each organization only certain phenomena are important enough to collect and represent in a database
  - the data collection process involves a sampling of geographic reality, to determine the status of that reality (whether past, present or future)

- identifying the phenomena and then choosing an appropriate data representation for them is part of a process called database design

- see Units 11 and 66 for more on database design

C. FUNDAMENTAL DATABASE ELEMENTS

- elements of reality modeled in a GIS database have two identities: 1. the element in reality - entity

  2. the element as it is represented in the database - object

- a third identity that is important in cartographic applications is the symbol that is used to depict the object/entity as a feature on a map or other graphic display

- these definitions and the following concepts are based on those defined by the DCDSTF, 1988 (see references)

  handout - Definition of terms

Entity

- an entity is "a phenomenon of interest in reality that is not further subdivided into phenomena of the same kind"
  - e.g. a city could be considered an entity and subdivided into component parts but these parts would not be called cities, they would be districts, neighborhoods or the like
  - e.g. a forest could be subdivided into smaller forests

Object

- an object is "a digital representation of all or part of an entity"

- the method of digital representation of a phenomenon varies according to scale, purpose and other factors
  - e.g. a city could be represented geographically as a point if the area under consideration were continental in scale
  - the same city could be geographically represented as an area if we are dealing with a geographic database for a state or a county

Entity types

- similar phenomena to be stored in a database are identified as entity types

- an entity type is any grouping of similar phenomena that should eventually get represented and stored in a uniform way, e.g. roads, rivers, elevations, vegetation
  - provides convenient conceptual framework for describing phenomena at a general level
  - organizational perspective influences this interpretation to a large degree

- precise definitions should be generated for each entity type

  - helps with identifying overlapping categories of information
  - aids in clarifying the content of the database
  - the proposed US National Standard for Digital Cartographic Data Volume 2 (DCDSTF 1988) includes a large number of definitions for entity types

  handout - Sample entity definitions

- the first step in database development is the selection and definition of entity types to be included

    - this is guided by the organization's mandate and purpose of the database
    - this framework can be as important as the actual database because it guides the development

- the second step of database design is to choose an appropriate method of spatial representation for each of the entity types

Spatial object type

- the digital representation of entity types in a spatial database requires the selection of appropriate spatial object types

- the National Standard for Digital Cartographic Databases specifies a basic list of spatial objects and their characteristics

- this classification is based on the following definition of spatial dimensions: 0-D - an object that has a position in space, but no length
        - a point 1-D - an object having a length
        - composed of two or more 0-D objects
        - a line 2-D - an object having a length and width
        - bounded by at least three 1-D line segment objects
        - an area 3-D - an object having a length, width and height/depth
        - bounded by at least four 2-D objects
        - a volume

    overhead - Spatial object types (3 pages)

    handout (cont) - Spatial object types

        - note very specific definitions for line segment, string, link, chain

- spatial objects as representations of reality are dealt with in depth in Unit 11

Object classes

- an object class is the set of objects which represent the set of entities
    - e.g. the set of points representing the set of wells

Attributes

- an attribute is a characteristic of an entity selected for representation

- usually non-spatial
    - though some may be related to the spatial character of the phenomena under study
        - e.g. area, perimeter

Attribute value

- the actual value of the attribute that has been measured (sampled) and stored in the

database

- an entity type is almost always labeled and known by attributes
  - e.g. a road usually has a name and is identified according to its class - e.g. alley, freeway

- attributes values often are conceptually organized in attribute tables which list individual entities in the rows and attributes in the column
  - entries in each cell of the table represent the attribute value of a specific attribute for a specific entity
  - note: attribute table is not an official DCDSTF term

Database model

- is a conceptual description of a database defining entity type and associated attributes
  - each entity type is represented by specific spatial objects

- after the database is constructed, the database model is a view of the database which the system can present to the user
  - other views can be presented, but this one is likely useful because it was important in the conceptual design
    - e.g. the system can model the data in vector form but generate a raster for purposes of display to the user
  - need not be related directly to the way the data are actually stored in the database
    - e.g. census zones may be defined as being represented by polygons, but the program may actually represent the polygon as a series of line segments

- examples of database models can be grouped by application area
  - e.g. transportation applications require different database models than do natural resource applications

Layers

- spatial objects can be grouped into layers, also called overlays, coverages or themes

- one layer may represent a single entity type or a group of conceptually related entity types
  - e.g. a layer may have only stream segments or may have streams, lakes, coastline and swamps
  - options depend on the system as well as the database model
  - some spatial databases have been built by combining all entities into one layer

D. DATABASE DESIGN

- almost all entities of geographic reality have at least a 3-dimensional spatial character, but not all dimensions may be needed
  - e.g. highway pavement actually has a depth which might be important, but is not as important as the width, which is not as important as the length

- representation should be based on the types of manipulations that might be undertaken

- map-scale of the source document is important in constraining the level of detail represented in a database
    - e.g. on a 1:100,000 map individual houses or fields are not visible

Steps in database design

1. Conceptual
    - software and hardware independent
    - describes and defines included entities
    - identifies how entities will be represented in the database
        - i.e. selection of spatial objects - points, lines, areas, raster cells
    - requires decisions about how real-world dimensionality and relationships will be represented
        - these can be based on the processing that will be done on these objects
        - e.g. should a building be represented as an area or a point?
        - e.g. should highway segments be explicitly linked in the database?

2. Logical
    - software specific but hardware independent
    - sets out the logical structure of the database elements, determined by the data base management system used by the software
    - this is discussed in greater detail in Unit 43

3. Physical
    - both hardware and software specific
    - requires consideration of how files will be structured for access from the disk
    - covered in Unit 66

Desirable database characteristics

- database should be:
    - contemporaneous - should contain information of the same vintage for all its measured variables
    - as detailed as necessary for the intended applications
        - the categories of information and subcategories within them should contain all of the data needed to analyze or model the behavior of the resource using conventional methods and models
    - positionally accurate
    - exactly compatible with other information that may be overlain with it
    - internally accurate, portraying the nature of phenomena without error - requires clear definitions of phenomena that are included
    - readily updated on a regular schedule
    - accessible to whoever needs it

Issues in database design

- almost all entities of geographic reality have at least 3-dimensional spatial character, but not all dimensions may be needed
    - e.g. highway pavement has a depth which might be important, but is not as important as the width, which is not as important as the length

- representation should be based on types of manipulations that might be undertaken

- map-scale of the source document is important in constraining the level of detail represented in a database

    - e.g. on a 1:100,000 map individual houses or fields are not visible

REFERENCES

Codd, E. F., 1981. "Data Models in Database Management," ACM SIGMOD Record 11(2):112-114. Explains the nature of data models, their role in constructing databases.

DCDSTF - Digital Cartographic Data Standards Task Force. 1988. "The proposed standard for digital cartographic data," The American Cartographer 15(1). Summary of the major components of the proposed US National Standard.

Robinson, A., R. Sale, J. Morrison, and P. Muehrcke, 1984. The Elements of Cartography, (5th ed.), John Wiley and Sons, New York. Useful survey of cartographic terminology and models.

Unwin D., 1981. Introductory Spatial Analysis, Methuen, London. A spatial analysis perspective on spatial data models.

EXAM AND DISCUSSION QUESTIONS

1. What makes the concept of a spatial database unique relative to other types of databases?

2. Distinguish the construct of an entity from a spatial object.

3. Why are organizational mandates important in database design? Give examples using (a) natural resource data and (b) socio-economic data.

4. What is a database model, and why is it important for designing a database?

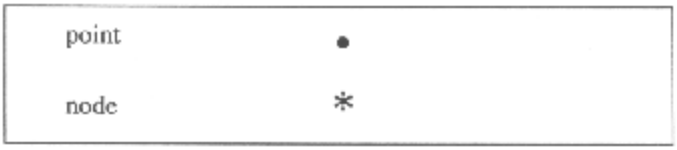
5. Why would a database designer use a chain object rather than a string object for representation of linear features?

6. List and define an example of a spatial object type from each of the 0-D, 1-D, 2-D and 3-D groups of object types.

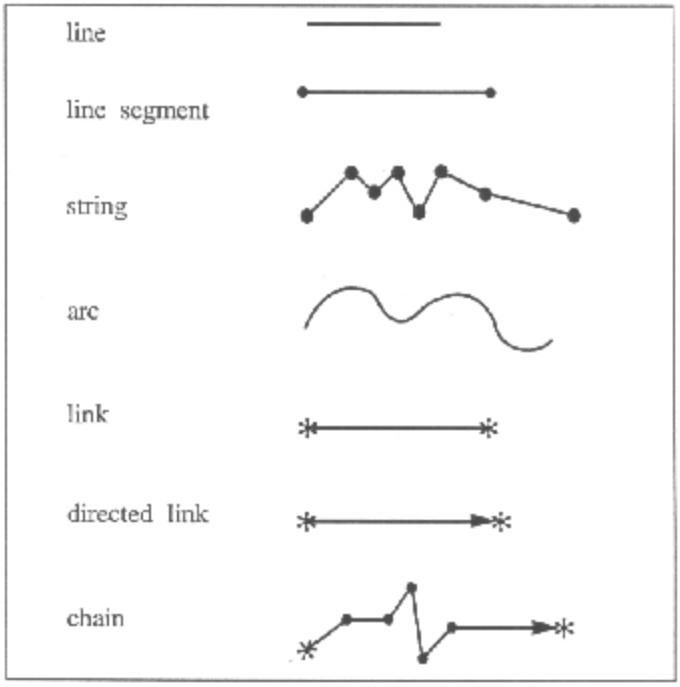
Last Updated: August 30, 1997.

UNIT 10 IMAGES

Basic 0-Dimensional Object Types



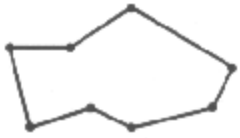
Basic 1-Dimensional Object Types



Basic 1-D Spatial Object Types (Continued)

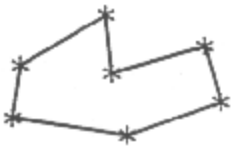
ring

1) ring created from string(s).

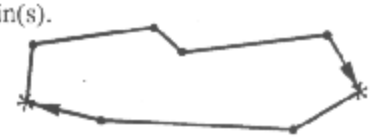


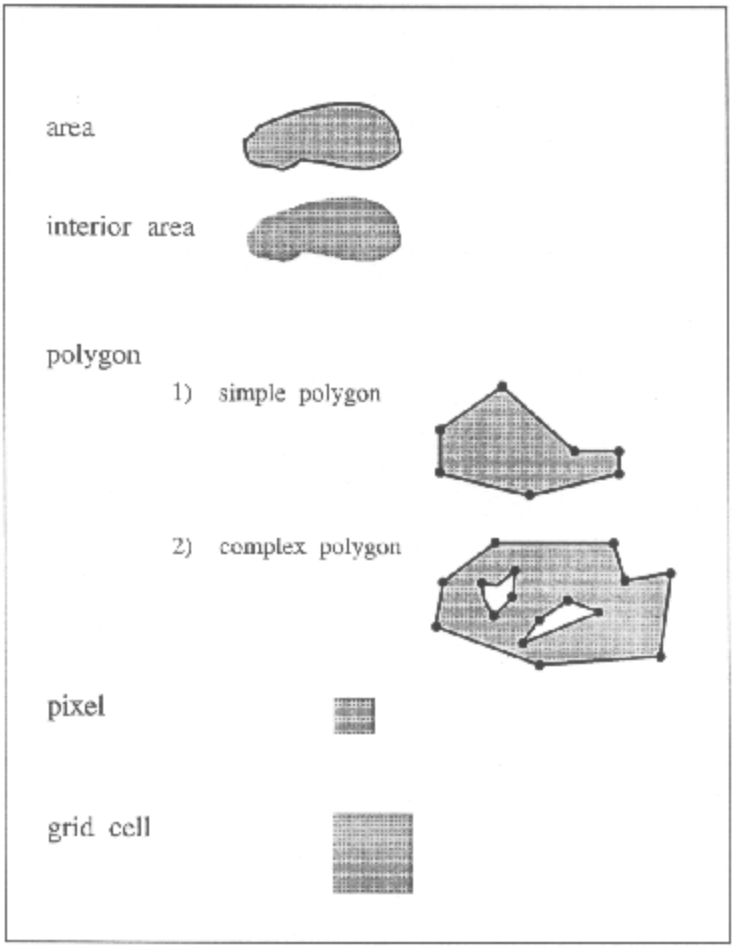
2) ring created from arc(s).

3) ring created from links.



4) ring created from chain(s).




area
interior area
polygon
1) simple polygon
2) complex polygon
pixel
grid cell