Spatial data

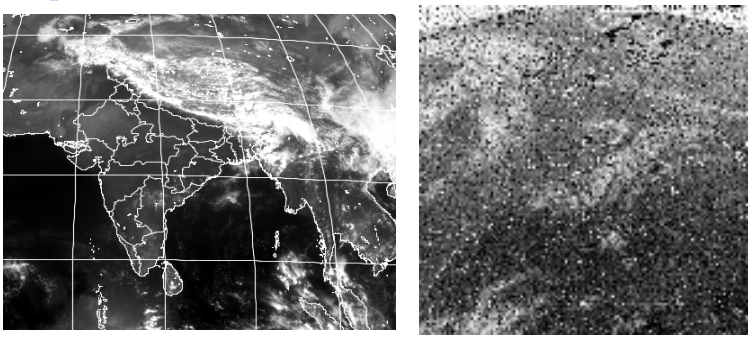
Spatial data is about instances located in a physical space.

Spatial data

- Spatial data has location features.
  - Address, latitude/longitude (explicit)
  - Location-based partitions in databases (implicit)

Spatial data

- Spatial data can be categorized into two types:
  - Raster data (Such data consist of bit maps or pixel maps in two or more dimensions)

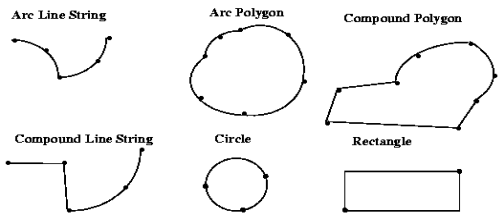


Example of raster image: satellite image of cloud cover, where each pixel store the cloud visibility in a particular area.

3

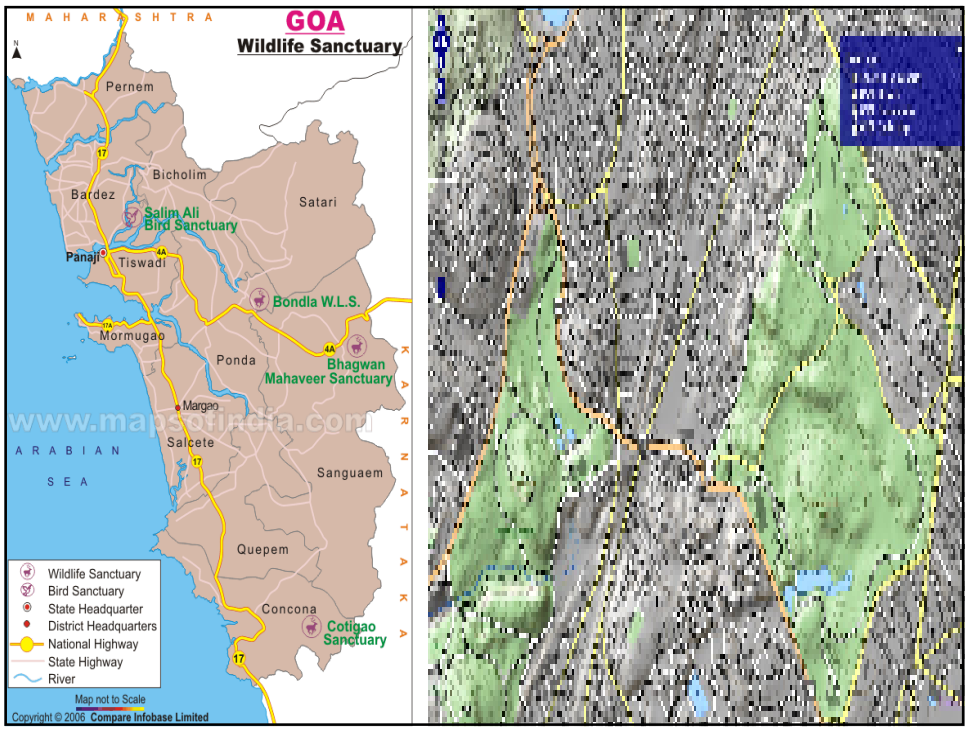
Spatial data

- **Vector data (vector data are constructed from basic geometric objects. such as points, line segments, triangles, polygons in two dimension. cylinders, spheres, cuboids in three dimension.)**



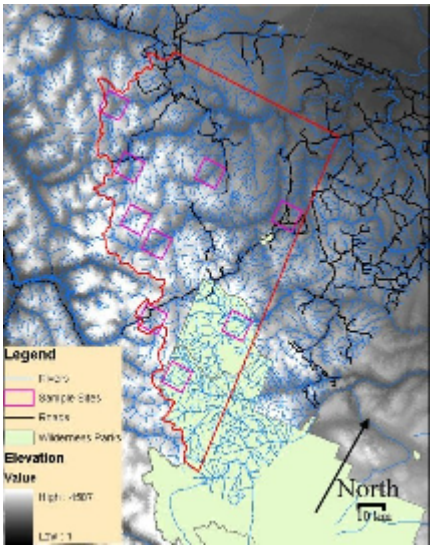
- **Example of vector data : (such as map data. Rivers & roads represented as line segments, states & countries may be represented as polygons)**

4



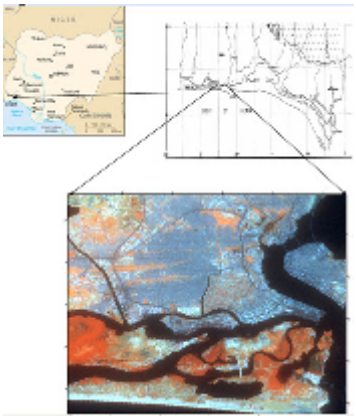
SPATIAL DATABASE

- A spatial database stores a large amount of space-related data.
  - Such as maps data, preprocessed remote sensing or medical imaging data, VLSI chip layout data.



SPATIAL DATABASE

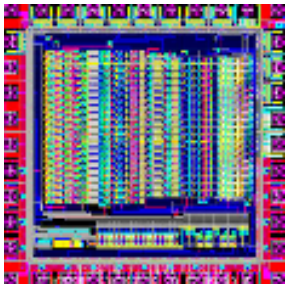
- Example of Spatial database:
  - Cartographic databases (that store maps)
  - Meteorological databases (for weather information)

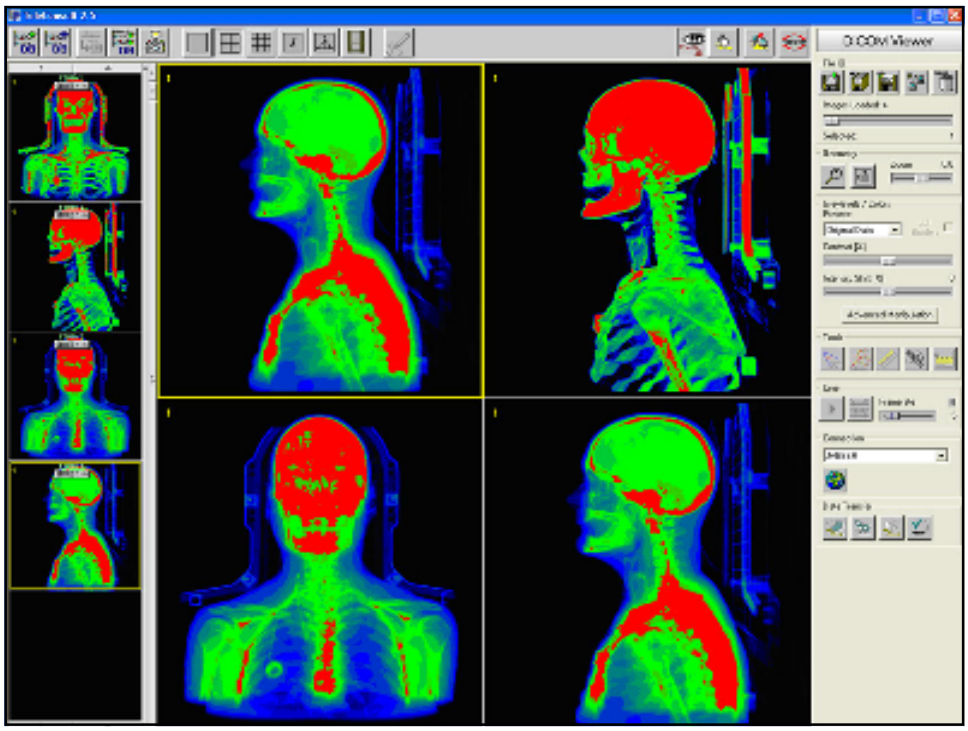


7

Spatial Database

- **A spatial database is a database that is optimized to store and query data that is related to objects in space, including points, lines and polygons**
- **A spatial database stores a large amount of space-related data.**
- **Such as maps data, preprocessed remote sensing or medical imaging data, VLSI chip layout data.**



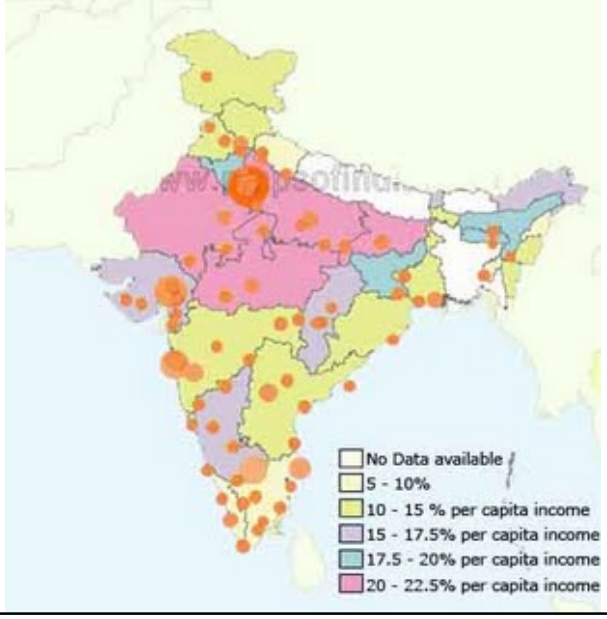


Example of Spatial database:

- **Cartographic databases** (that store maps)
- A database containing x-y coordinates defining a geographical area. When combined with other data (such as any of a wide variety of variables, such as income distribution, age, etc.), a cartographic database can be used to map the distribution of that variable within a geographical region.

10

Per capita income means how much each individual receives, in monetary terms, of the yearly income



11

Meteorological databases (for weather information)

MeteoDB

12

Spatial Database

Special features of spatial databases:

- **That can interpret spatial characteristics.**
- **Special indexing & storage structures to improve performance.**
- **They carry topological & distance information.**
- **Require spatial reasoning, geometric computation, & spatial knowledge representation techniques.**

13

SPATIAL DATABASE

A description of the spatial positions of many types of objects would be needed:

- Some of these objects have **static spatial** characteristics, such as streets, highways, police stations, water pumps, fire stations & hospitals.

14

Spatial Database

Other objects have dynamic spatial characteristics that change over time, such as police vehicles, ambulances, or fire trucks.

Spatial database query

- Accessing spatial data can be more complicated than accessing non-spatial data. There are specialized operations and data structures used to access spatial data.
- Accessing non-spatial data uses the standard comparison operations: **< , > , ≤ , ≥ , ≠ , = .**
- Accessing Spatial data uses the spatial comparators:
  - near, north, south, east, west, contained in, overlap, intersect.

16

Spatial database query

Types of Spatial queries:

- **Range query: (find the objects of a particular type that are within a given spatial area or within a particular distance from a given location.)**

  Example: Find all hospitals within the Bhopal city area.
- **Nearest neighbor query: (find an object of particular type that is closest to a given location)**

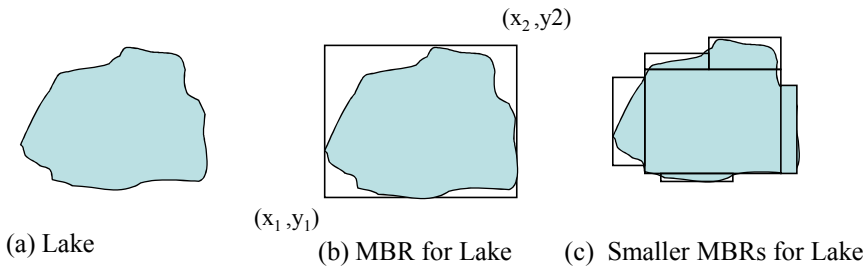
  Example: find the police car that is closest to a particular location(Bhopal Lake).
- **Spatial joins or overlays: (joins the objects of two types based on some spatial condition, such as the objects intersecting or overlapping spatially)**

  Example: find all homes that fall on a major highway, find all homes that are two miles of a lake.

17

Spatial data structure

- **Many data structures that have been designed specifically to store or index spatial data.**
  - **Quad tree, R- tree, k-D tree**
- **MBR (Minimum Bounding Rectangle):**
  - **A common technique used to represent a spatial object is by the small rectangle that completely contains that object.**



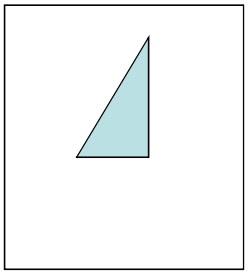
(a) Lake (b) MBR for Lake (c) Smaller MBRs for Lake

MBR Example

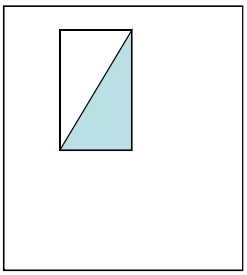
18

Spatial data structure

- **Quad tree:**
  - **A quad tree represents a spatial object by a hierarchical decomposition of the space into quadrants.**



(a) Triangle



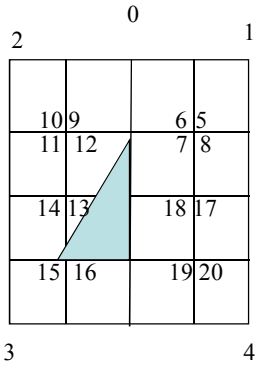
(b) MBR for Triangle

Spatial object example

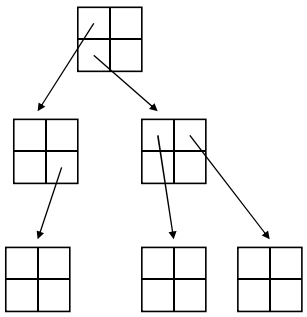
19

Quad tree

- **Quad-trees generally divide each space or subspace into equally sized areas , and proceed with the subdivisions of each subspace to identify the positions of various objects.**



(a) Representing triangle with quadrants



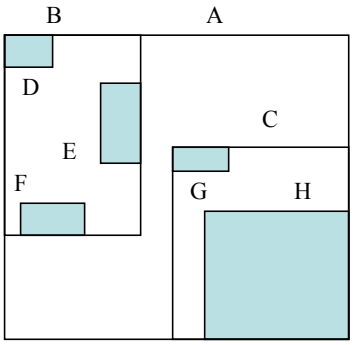
(b) Quad tree

Quad tree Example

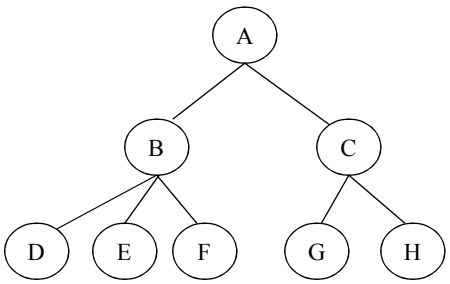
20

R-trees

- **R tree:**
  - **One approach to indexing spatial data represented as MBRs.**



(a) Partitioning with MBRs



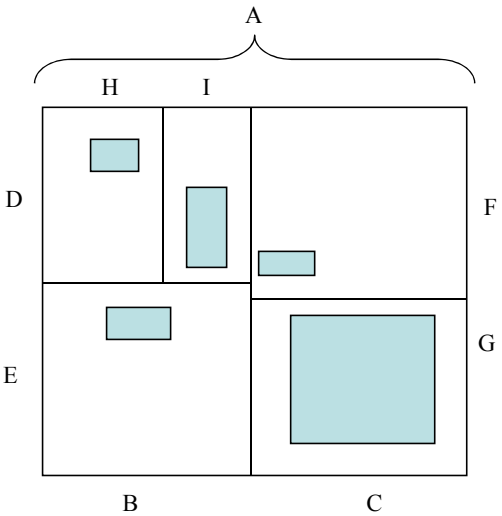
(b) R-Tree

R-Tree Example

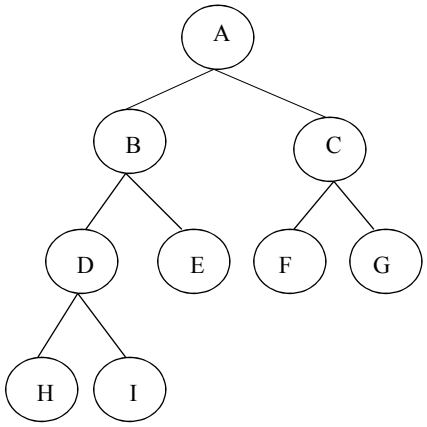
21

k-D tree

- **A k-D tree was designed to index multi-attribute data**



(a) Divide & Conquer partitioning



(b) k-D tree

K-D tree example

22

R Tree & K-D Tree

•Nodes in *k*d-trees represent separating planes, whereas nodes in R-trees represent bounding boxes.
•*k*d-trees partition the whole of space into regions whereas R-trees only partition the subset of space containing the points of interest.
• *k*d-trees represent a disjoint partition (points belong to only one region) whereas the regions in an R-tree may overlap.

23

Pattern

- What is a Pattern?
  - A frequent arrangement, configuration, composition, regularity
  - A rule, law, method, design, description
  - A major direction, trend, prediction
- What is not a pattern?
  - Random, haphazard, chance, accidental, unexpected
  - Without definite direction, trend, rule, method, design, aim, purpose
  - Accidental - without design,
  - Casual - absence of pre-arrangement, relatively unimportant
  - Fortuitous - What occurs without known cause

24

Spatial Data Mining

Defining Spatial Data Mining

- Search for spatial patterns
- **Non-trivial search** - as "automated" as possible—reduce human effort
- **Interesting, useful** and **unexpected** spatial pattern

25

Spatial Data mining

- **Spatial data mining refers to the extraction of knowledge, spatial relationships, or other interesting patterns not explicitly stored in spatial databases.**
- **Such mining demands an integration of data mining with spatial database technologies.**
- **Used for understanding spatial data, discovering spatial relationships and relationships between spatial & Non-spatial data, Constructing spatial knowledge bases and optimizing spatial queries.**

26

What is NOT Spatial Data Mining?

- Simple Querying of Spatial Data
  - Find neighbors of India given names and boundaries of all countries
  - Find shortest path from Bhopal to Indore
  - Search space is not large (not exponential)
- Testing a hypothesis via a primary data analysis
  - Ex. Female chimpanzee territories are smaller than male territories
  - Search space is not large !
- Uninteresting or obvious patterns in spatial data
  - Heavy rainfall in Minneapolis is correlated with heavy rainfall in St. Paul, Given that the two cities are 10 miles apart.
  - Common knowledge: Nearby places have similar rainfall
- Mining of non-spatial data
  - Laptop sales and Digital Camera sales are correlated in evenings
  - GPS product buyers are of 3 kinds:
    - outdoors enthusiasts, farmers, technology enthusiasts

27

Spatial Data Mining: Actors

- Domain Expert -
  - Identifies SDM goals, spatial dataset,
  - Describe domain knowledge, e.g. well-known patterns, e.g. correlates
  - Validation of new patterns
- Data Mining Analyst
  - Helps identify pattern families, SDM techniques to be used
  - Explain the SDM outputs to Domain Expert
- Joint effort
  - Feature selection
  - Selection of patterns for further exploration

28

Spatial data-mining primitives

- Equals:
  - A equal B if all points in the two objects are in common.
- Covered by or inside or contained in:
  - A is contained in B if all points in A are in B. there may be points in B that are not in A..
- Covers or Contains:
  - A contains B if and only if B is contained in A.

29

Spatial rules

- Spatial rules can be generated that describe the relationship between Spatial object and structure of spatial objects.
- Three types of rules:
  - Spatial characteristic rules describe the data.

    Example: In Bhopal the average family income is Rs. 50,000.
  - Spatial discriminant rules describe the differences between different classes of the data. They describe the features that differentiate the different classes.

    Example: In Bhopal the average family income is Rs. 50,000, while In Indore the average family income is Rs. 75,000.
  - Spatial association rules are implications of one set of data by another.

    Example: In Bhopal the average family income for families living near great lake is Rs. 90,000.

    is_a(X, “school”) ^ close_to(X, “sports_center”) ➔ close_to(X, “park”) [0.5% , 80%]

30

Spatial Data mining

Spatial data mining used in following Applications:

- **Medical imaging, Cancer clusters to investigate environment health hazards**
- **Crime hotspots for planning police patrol routes**
- **Navigation, Traffic control, Environmental studies,**
- **Remote Sensing**

31

Reference

Data Mining
Introduction and Advance Topics
Margaret H.Dunham

32