1. Introduction to Spatial Databases

(Everything happens somewhere, sometime)

Overview

Learning Objectives

- Understand basic properties of spatial/geographic data
- Understand basic concepts of Geographic Information Systems (GIS)
- Learn about major spatial information applications
- Learn about components of Spatial Database Management Systems (SDBMS)

Literature

- [RSV02] Chapter 1
- [SC03] Chapter 1
- Ralf Hartmut Güting: *An Introduction to Spatial Database Systems*. The VLDB Journal, 3(4):357-399, 1994. http://www.acm.org/sigmod/vldb/journal/VLDBJ3/P357.pdf
- Shashi Shekhar et al.: *Spatial Databases – Accomplishments and Research Needs.* IEEE Transactions on Knowledge and Data Engineering, 11(1):45-55, 1999. http://www.cs.umn.edu/Research/shashi-group/paper_ps/tkde.ps

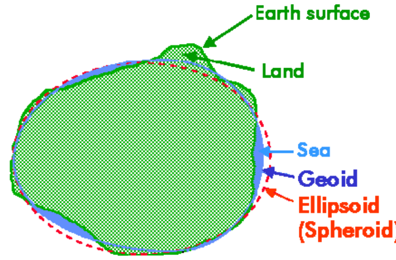
Spatial Data

Spatial Data

- Data with an associated spatial location (with respect to a given reference frame).
- Information about the location and shape of, and relationships among, geographic features, usually stored as coordinates and topology.

Reference Frame

Reference frames are basically the different perspectives of the viewer to describe physical quantities (e.g., position, velocity) while the *coordinate systems* are the different ways to describe physical quantities in these perspectives.



Geographic / Geospatial Data

- Data whose underlying reference frame is the Earth's surface.
- Geospatial information concerns phenomena above, one and below the Earth's surface.

The notions "spatial data" and "geospatial data" are often used interchangeably.

Spatial Data (2)

At the conceptual level, a *geographic object* corresponds to an entity of the real world and has two components:

- A set of alphanumeric attributes, also called *descriptive attributes,*
- A spatial component, which may embody both geometry (e.g., location, shape) and topology (≡ spatial relationships with other objects); the isolated spatial component of a geographic object is called *spatial object*.

Geographic objects can be *atomic* or *complex.*

Spatial data may be obtained from

- *primary data collections* in the field where data is directly collected from field surveys; GPS helps in determining point locations; data is collected remotely through remote sensing techniques.

  Main problem in primary data collection: delimiting geographic objects and associating thematic attributes with objects
- digitized from existing *maps;* note that maps come with a *scale*; also, to represent geographic coordinates, a *reference system* is needed, which involves a choice of a *map projection* (e.g., Mercator, UTM).

What is special about spatial?

Traditional Database Management Systems (DBMSs) are very good in managing non-spatial data, such as strings, numbers, dates. They provide for

- efficient access to and modification of data through declarative query languages and access paths
- transaction management and concurrency control
- persistence and recovery in case of failures
- access control and security mechanisms
- logical and physical data independence through 3-level schema architecture

So why not use a traditional DBMS to manage spatial data?

Spatial data is more complex compared to traditional (business) data. Contributing to the complexity include the

(1) modeling of space, appropriate space/geometric models and data types

(2) complex spatial relationships that need to be managed to speed up query processing,

(3) inadequacy of traditional access paths and (1-dimensional) index structure

Geographic Information Systems (GIS)

Consequence: Techniques and extensions of traditional (relational) DBMSs are necessary to meet requirements of spatial data management.

A GIS is a software package to store, manage, analyze, and visualize spatial data using spatial analysis functions, such as

- *Search* Thematic search, search by region/coordinates, classification
- *Location Analysis* Buffer corridor, overlay
- *Terrain Analysis* Slope/aspect, catchment, drainage network
- *Flow Analysis* Connectivity, shortest path
- *Distribution* Change detection, proximity, nearest neighbor
- *Spatial Analysis/Statistics* Pattern, autocorrelation
- *Measurements* distance, shape, adjacency, direction, perimeter

There are many definitions for GIS....

"A computer system designed to allow users to collect, manage and analyze large volumes of spatially referenced information and associated attribute data."

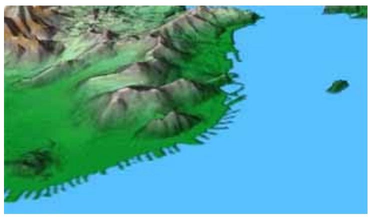
Geospatial Data

GIS are capable of managing a wide variety of spatially referenced data

Satellite imagery

Street maps

Aerial photos



Terrain data

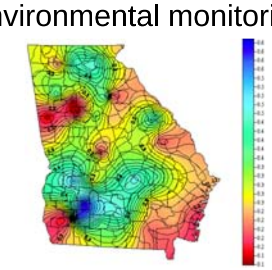
Topographic maps

Classical GIS Applications

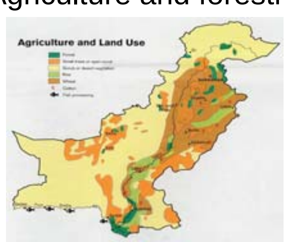
Interest in spatial data and GIS is witnessing a dramatic increase that goes beyond traditional GIS applications; traditional applications include

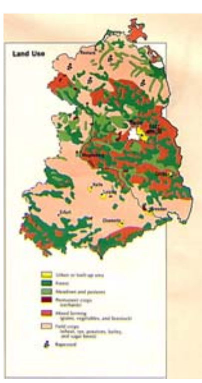
Map making and cartography

Environmental monitoring

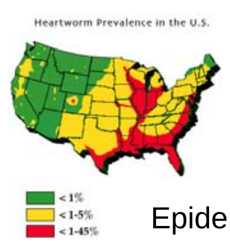


Agriculture and forestry





Land use



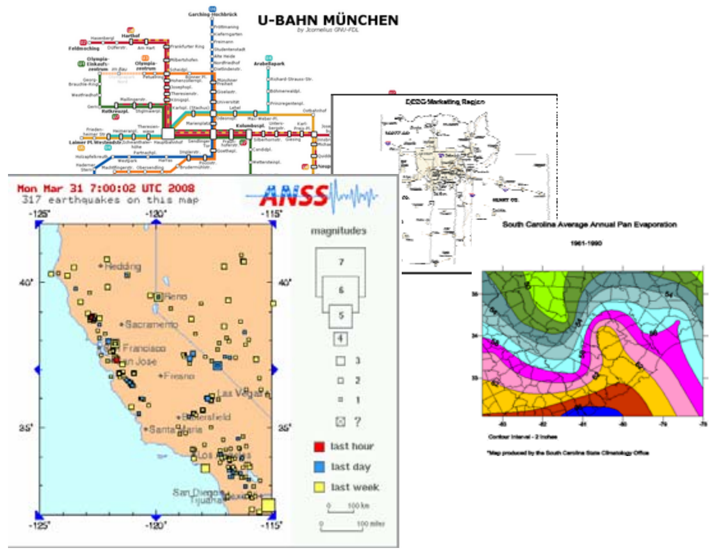
Epidemiology and health

GIS is not only about maps!
But geospatial information is typically visualized using maps.

Classical GIS Applications (2)

Further applications include

- Cadastre (taxation)
- Real Estate
- Economic development
- Marketing
- Navigation
- Climatology
- Utility management
- Transportation networks
- Biodiversity
- Earthquake monitoring
- ...



GIS is fundamentally about solving real-world problems.

Used at various levels, from specific industries and businesses to local governments to federal government to international agencies.

Increase of Spatial Data

- According to unofficial sources, 80% of the data available has a spatial component (just think about addresses, including zip codes and street names...).
- Huge amounts of spatial data is delivered by Earth Observation Systems; e.g., NASA's systems generate several terabytes of satellite image data per day
- Various new services such as MapQuest, Yahoo! Maps, Google Maps, Google Earth, NASA World Wind have spurred public and commercial interest in geospatial data
- GIServices are a rapidly growing form of electronic commerce
  - Route planning
  - Emergency and disaster management and mitigation
  - Crime tracking
  - Simulation of environmental effects
  - Location-based services
- In general, there is a need to integrate spatial information within a wide variety of contexts.

Vocabulary in Geospatial Database Applications

Theme geospatial information corresponding to a particular topic; similar to a relation in the relational model

When a theme is displayed, a user typically sees a *map*, with colors, a particular scale, a legend, and so on.

Constructing a conceptual schema for all themes of interest is typically the first step in developing a GIS application.

theme = {geographic object} // set of geographic objects

geographic object = (description, spatial object) // atomic object

| (description, {geographic-object}) // complex object

A theme thus is a set of homogeneous geographic objects (objects having the same structure or type) and is represented according to the GIS logical data model, also called *geographic model.*

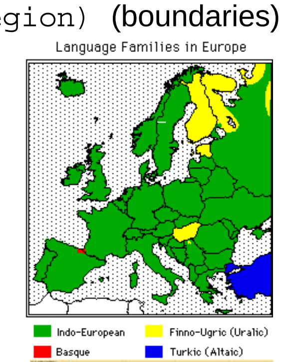
Representation of geometry and topology requires *spatial data model,* which can include basic types such as point (0-D), line (1-D), and region (2-D).

Geospatial Data Manipulation

Assume two-theme schema

`Countries(name, capital, population, geo:region)` (boundaries)

`Languages(language, geo:region)` (colored map)



Simple operations on themes:

- Theme projection: *theme* x $\{A_1, ..., A_n\} \rightarrow$ *theme*
  e.g., projection onto population
- Theme selection: *theme* x $p(A_i) \rightarrow$ *theme*
  e.g., countries with population > 30M
- Theme union: *theme* x *theme* $\rightarrow$ *theme*
  themes must have the same schema
- Theme overlay: *theme* x *theme* $\rightarrow$ *theme*
  creates a new theme; similar to spatial join
  objects are joined if geometry intersects;
  very common operation in GIS applications

Geospatial Data Manipulation (2)

Geometric selection:

- Windowing (window query): one obtains another theme that only includes those objects that overlap the given (usually rectangular) area.
- Point query: retrieves all objects whose geometry includes given point
- Clipping: extracts the portion of a theme located within a given area

Further theme operations include

- Operations using a metric: for example, distance between two points or regions
- Topological operations: adjacent objects of a given object; topological queries on networks, e.g., reachability

Other typical GIS operations include interpolation/extrapolation, location, allocation, location/allocation combination.

Operations can be formally described using a *theme algebra*.

DBMS Approaches to GIS

- A GIS needs to manage both alphanumerical and geographic data.
- Early GIS were built directly on top of proprietary file systems.
- In principle, geographic data can be modeled using relational model.

  For example, a boundary can be described using a contour, which is a sequence of lines, each line described by a start and end point; contours, lines and points are managed in separate relations.

Inappropriate for handling geospatial data, because

- *Data independence* principle is violated; one has to know the internal structure of geospatial objects to formulate queries
- *Poor performance*; a considerable amount of tuples is required to represent simple geometric objects.
- *Lack of user-friendliness*, one has to manipulate points, and queries are complex
- It is difficult to define *new data types*
- Impossible to express certain geometric computations, such as adjacency test, point query, or window query

Spatial Database Management Systems

Loosely coupled approaches to realizing a GIS use relational DBMS and a separate component for spatial data management

⇒ Drawbacks: Heterogeneous data models; partial loss of DBMS functionality; ....

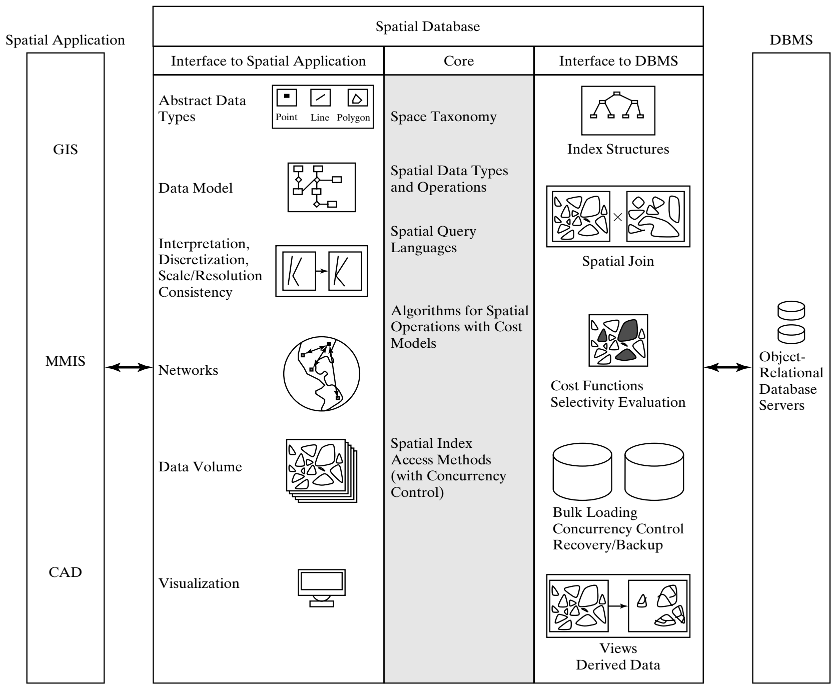
A Spatial Database Management System (SDBMS) is a software package that

- can work with an underlying DBMS
- supports spatial data models, spatial abstract data types (ADTs) and a query language that supports operations on ADTs (in addition to traditional SQL-like operations)
- supports spatial indexing, efficient algorithms for processing spatial operations, and domain specific rules for query optimization
- supports efficient storage, querying, and sharing of large spatial datasets
- Examples include ESRI's SDE, Oracle Spatial

In general, SDBMS (1) integrates representation and manipulation of geometric data with traditional data at the logical level, and (2) provides efficient support at the physical level to store and process the data.

The kernel of most GIS is a Spatial Database Management System

"Three Layer" Architecture

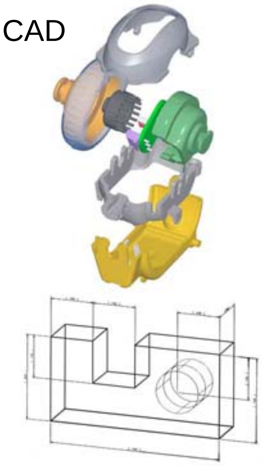


[taken from [Shekhar, Chawla 03], Figure 1.6]

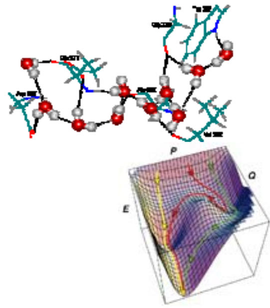
SDBMS

SDBMS can be used by applications other than GIS

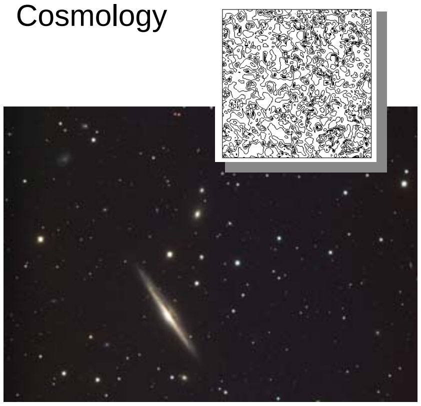
CAD



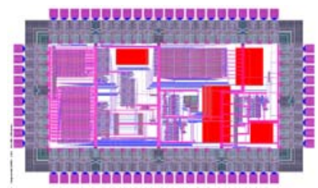
Genomics



Cosmology

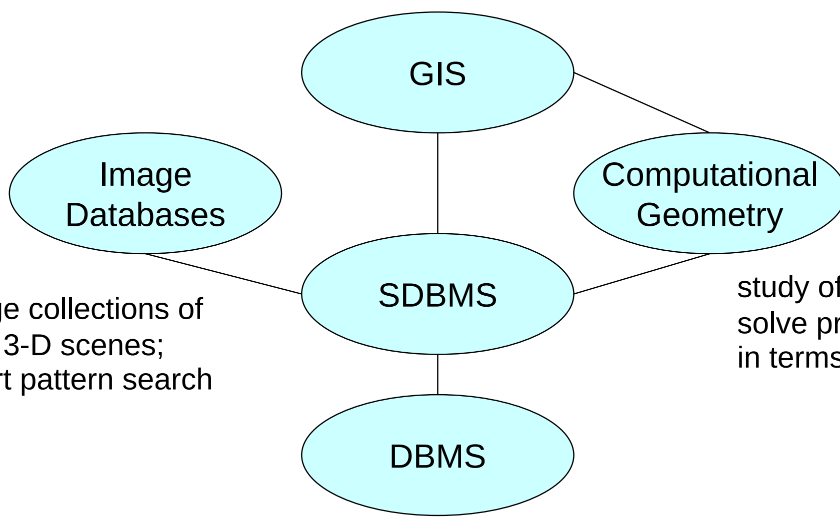


VLSI Design



and many other applications....

SDBMS in Context



Manage collections of 2-D or 3-D scenes; support pattern search

study of algorithms to solve problems stated in terms of geometry