# Spatial Databases

M. Andrea Rodríguez-Tastets

Universidad de Concepción,Chile
andrea@udec.cl

I Semester 2008

Outline

Geographic Information refers to data with a location on the Earth's surface.

### Examples

- Census data
- Administrative boundaries of a country, state and county
- Relief data (e.g. from contours)
- Information on transport networks
- A text about a specific place (e.g., textual description of touristics places)
- A photograph or painting of a building
- A satellite image

Geographic information can be used to answer where is a specific entity (physical object) or what is at a particular location on the Earth's surface.


What is it?


lake


Where is it?
lake

Geographic information differs in nature depending on the type of application and the way it is obtained.

### Data From

- Primary data collection in the field. Data is collected from field surveys or remotely collected using remote sensing techniques. An important issue when collecting data is to include delimiting geographic objects and associating thematic attributes.
- Existing maps. It may imply to integrate maps, which could come with different scale and reference system.

## GIS as a Type of Software

- A computer system that allows us to handle information about the location of features or phenomena on the Earth's surface.
- A system that has all the functionality of a conventional DBMS plus much of the functionality of a computer mapping system.

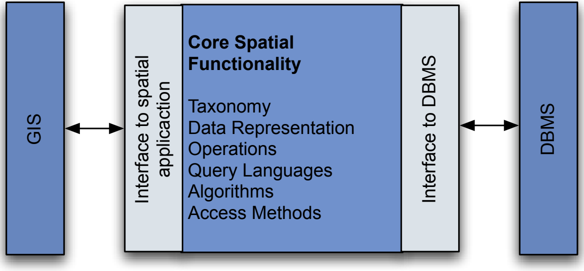
## GIS as a Tool-Kit

- Spatial manipulation (determination of spatial relations, changes in scales and projections, . . . )
- Spatial analysis (quantitative and qualitative spatial analysis)
- Visualization of spatial data (maps, tables, 3D surfaces)

In GIS, themes gather geographical information about a particular topic (e.g., rivers, cities, and countries). They are displayed as maps (e.g., topographic maps, weather maps, soil-type maps).

Until now, GIS has been concentrated on the geometric aspects of spatial representation and manipulation, and how these aspects change over time.

# GIS and DBMS


GIS
Interface to spatial application
Core Spatial Functionality
Taxonomy
Data Representation
Operations
Query Languages
Algorithms
Access Methods
Interface to DBMS
DBMS

## Spatial Databases

- Data of a specific domain, with its own semantics and where particular rules apply.
- Support multiple spatial data models
- Support spatial operations
- Support query languages and query processing over spatial data (indexing and search)
- Should:
  - Contain a good framework to combine geometric with thematic information
  - Be as general as possible and not designed for one particular area of applications
  - Have a formally defined semantics that is closed under set theoretic, geometric and topological operations, and that is defined in terms of finite representations
  - Be independent of a particular DBMS but cooperate with any
  - Use efficient implementation techniques
  - Have a gateway to multimedia and visual interface

## Characteristics of Spatial Databases

- There is no clear separation between what is handled by the SDBMS and an application. For example, is the query "find the shortest path" part of the query language?
- Modeling of a single conceptual object: city, river and road.
- Modeling a collection of objects
  - Coverage-partitions: administrative boundaries
  - Network: transportation networks
  - Aggregation: counties - states -country
  - Terrain digital models

## Geo-Information Requirements

- Geographic information deals with spatial and non-spatial (aspatial) data
- Data are highly structured by the notion of object aggregation
- The existence of user-defined operations that require an extensible underlying model
- The existence of functions that exist at different levels of abstraction (geometric types and maps)

## Types of Users

- Manipulate spatial data
  - Transportation system monitoring
  - Cadastral systems
- Spatial data analysis
  - Analysis of flood risk areas
  - Comparing resonance images
- Final users (high-level users)
  - Navigation systems

| Database | GIS |
| --- | --- |
| Manipulation of spatial data in secondary memory | Motivate SDBMS |
| Spatial semantics embedded in the query language | Clients of SDBMS |
| Spatial Access methods | A large number of spa |
| | SDBMS allows GIS to |

## Objects versus Fields

Two views of the space are: objects (where) and fields (what). They lead us to different abstractions and representations of space.

## Discrete Items: Objects

- Individuals with an identity.
- Individuals whose spatial extend is described by their boundaries.
- They usually have some property values throughout the entire individual.
- Examples:
  - this town
  - the trajectory from home to work
  - the parking lot

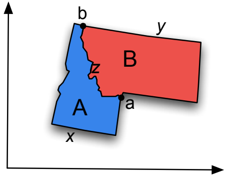
# Objects and Aggregate Objects

Objects are grouped into themes or maps.

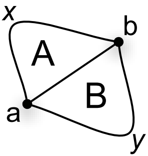
$$
\begin{aligned}
\mathit{theme} &= \{\mathit{geographic\_object}\} \\
\mathit{geographic\_object} &= (\mathit{thematic\_description}, \mathit{spatial\_description}) // \mathit{Atomic} \\
&\quad |(\mathit{thematic\_description}, \{\mathit{geographic\_object}\}) // \mathit{Complex}
\end{aligned}
$$

# Objects: Spatial Description

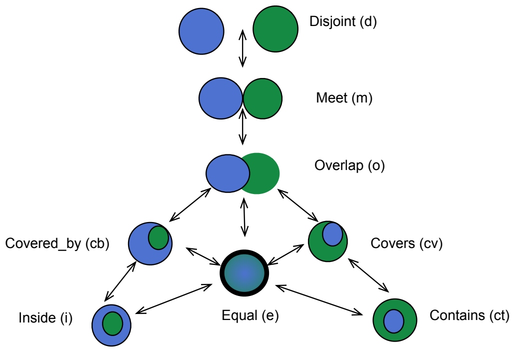
Geometry (location and shape)




Topology



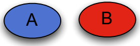
# Spatial Relations: Topological Relations




Topological relations are invariant under rotation, translation and scaling.

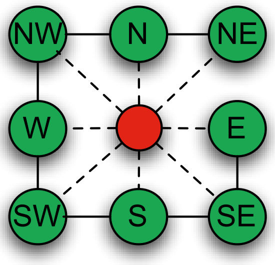
## A Model of Topological Relations

The *9-intersection* model defines topological relations between regions in terms of the set-intersection of objects' boundaries, interiors, and exteriors. For example, in the following we have the 9-intersection definition of *disjoint*.

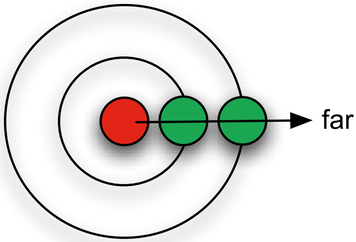



| | $A^\circ$ | $\delta A$ | $A^-$ |
|---|---|---|---|
| $B^\circ$ | $\emptyset$ | $\emptyset$ | $\neg\emptyset$ |
| $\delta B$ | $\emptyset$ | $\emptyset$ | $\neg\emptyset$ |
| $B^-$ | $\neg\emptyset$ | $\neg\emptyset$ | $\neg\emptyset$ |

# Spatial Relations: Direction Relations



# Spatial Relations: Distance Relations



## Distributions: **Fields**

- Properties values change across space.
- Continuos space and property values.
- Values are determined by sampling.
- Require interpolation to derive values at locations without observations.
- Examples:
  - elevation over the Andes Mountains
  - temperature on the coast
  - the speed of my trip from home to work

## Field Properties

- Position (space $\mathcal{S}$ and time $\mathcal{T}$)
- Value of type $A_\alpha$ associated with a position
- Spatio-temporal function as a model of a field (interpolation function):

$$F : \mathcal{S} \times \mathcal{T} \rightarrow A_\alpha$$

## Quantitative versus Qualitative

While quantitative spatial descriptions deal with data we can draw clear boundaries as on maps, qualitative spatial descriptions deal with data whose boundaries are unclear or unknown, but still with data we can apply some kind of spatial analysis.

# Quantitative

- Size limited in terms of absolute units.
- Number of classes depending on granularity.
- Limited vocabulary.
- Infinite amount of valid sentences.
- Spatial configurations captured very precisely.
- Descriptions are typically complete.
- Bottom up.

# Qualitative

- Distinctions in terms of relations instead of absolute values.
- Limited vocabulary, but large vocabulary.
- Only necessary distinctions.
- Variable resolution.
- Finite amount of valid sentences.
- Vagueness.
- Descriptions are typically incomplete.
- Top down.

## Qualitative Spatial Reasoning

- Inference of **missing information**
  **IF** Chile is west of Argentina and
  Eastern island is part of Chile **THEN**
  Eastern island is west of Argentina

- Evaluation of **consistency**
  **IF** A is west B and
  Eastern island is part of Chile **THEN**
  Eastern island must west of Argentina

## Static versus Dynamic Phenomena

- Changes over geometry (location/shape). So, we have geometries that are valid at a particular time instant or time interval.
- Changes may occur discretely.
- Changes may occur continuously.
- In addition to spatial relations, we have now temporal and spatio-temporal relations.
- We may be interesting in historical data or in predicting future data.

## Spatial Data Models

- Represent geographical information in the *n*-dimensional real space $\mathbb{R}^n$ (an infinite or non-enumerate set of points).
- They use (linear) approximations to represent objects' boundaries or discrete approximations for objects and fields.
- They have an important influence on the operations (embedded in the model or user-defined operations).
- We would expect that they were closed under operations.

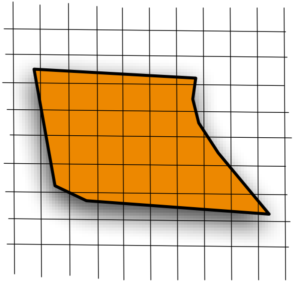
## Three Broad Types of Spatial Data Models

- Tessellation or partitions of space
- Vector
- Linear constraints

In theory, each of these types of data models can represent objects and fields.

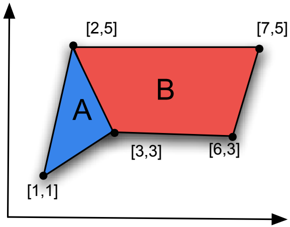
## Objects: Tessellation

A regular (irregular) tessellation divides the space into regular (irregular) cells (pixels). An object is represented by the set of cells that contain it.



## Objects: Spaghetti Model (Vector)

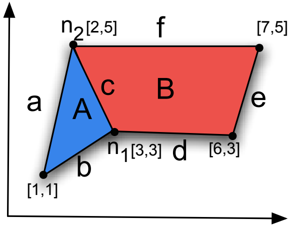
$$\begin{aligned} \textit{point} &: [x : \textit{real}, y : \textit{real}] \\ \textit{polyline} &: < \textit{point} > \\ \textit{polygon} &: < \textit{point} > \\ \textit{regions} &: \{\textit{polygon}\} \end{aligned}$$




$A :< [1,1], [3,3], [2,5], [1,1] >$
$B :< [3,3], [6,3], [7,5], [2,5], [3,3] >$

# Objects: Network Model (Vector)

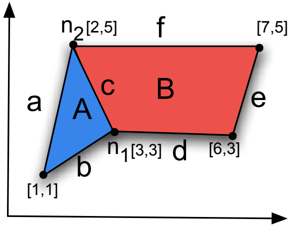
$$\begin{aligned} point &: [x : real, y : real] \\ node &: [point, < arc >] \\ arc &: [node, node, < point >] \\ polygon &: < point > \\ regions &: \{polygon\} \end{aligned}$$




$A :< a, b, c >$
$B :< c, d, e, f >$
$c : [n_1, n_2, < [3, 3], [2, 5] >]$
$n_1 : [[3, 3], < b, d, c >]$
$n_2 : [[2, 5], < a, c, f >]$

# Objects: Topological Model (Vector)

$$\begin{aligned} point &: [x : real, y : real] \\ node &: [point, < arc >] \\ arc &: [node, node, < point >] \\ polygon &: < arc > \\ regions &: \{polygon\} \end{aligned}$$



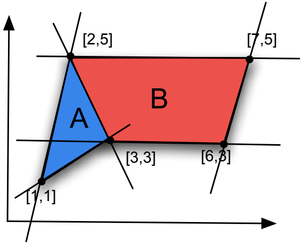

$A :< a, b, c >$
$B :< c, d, e, f >$
$c : [n_1, n_2, A, B, < [3, 3], [2, 5] >]$
$n_1 : [[3, 3], < b, d, c >]$
$n_2 : [[2, 5], < a, c, f >]$

## Objects: Linear Constraint Model

Boundaries are represented by linear constraints of the form

$$\sum_{i=1}^{p} a_i x_i \Theta a_0,$$

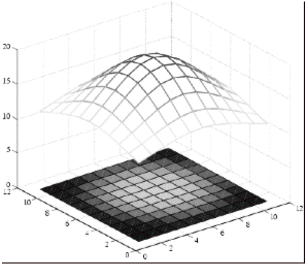
with $\Theta$ a predicate $=$, $\leq$, or $\geq$, $x_i$ a variable, and $a_i$ a constant.




$$
\begin{aligned}
A &: y \geq x \wedge y - 4x + 3 \leq 0 \wedge \\
&\quad y + 2x - 9 \leq 0 \\
B &: y \geq 3 \wedge y \leq 7 \wedge \\
&\quad y - 2x + 9 \geq 0 \wedge 2x + y - 9 \geq 0
\end{aligned}
$$

## Fields: Tessellation

A regular or irregular tessellation divides the space into cells (pixels). The field-based data is still represented as a function from space to a range; however, the function domain in not an infinite set of points but a finite set of pixels or cells.



## Fields: Vector

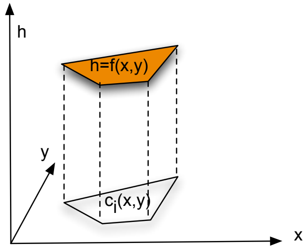
An example of a vector model for fields is a TIN (Triangulated Irregular Network). A TIN is based on a triangular partition of 2D space. No assumption is made about the distribution of the points. The elevation (property) value is recorded at each vertex and inferred at any other point $P$ by linear interpolation of the three vertices of the triangle that contains $P$.

## Fields: Linear constraints

We can use linear constraints for representing the equivalent of a TIN with a constraint of the form:

$$c_i(x, y) \wedge h = f(x, y)$$

where $c_i(x, y)$ denotes a conjunction of inequalities on $(x, y)$ representing the triangle o region associated with the property value defined by the equality constraint $h = f(x, y)$ among variables $h$, $x$ and $y$, with $f$ being a linear function of $x$ and $y$.



# Beyond Classical Fields: A framework for Spatio-Temporal Phenomena

A model based on a theory that

- enables prediction
- is verifiable through new observations
- can be improved with higher resolution measurements

## A continuous spatial field

$$\frac{\dfrac{sinxy + \frac{1}{2}cos2x + \frac{1}{3}sin3y + \frac{1}{4}cos4(x+y)}{1 + \left|sin5y + \frac{1}{2}cos6x + \frac{1}{3}sin7y + \frac{1}{4}cos8x\right|}}{1 + \left|\dfrac{(sin10xy + \frac{1}{2}cos10x + \frac{1}{3}sin12y + \frac{1}{4}cos13(x+y))}{1 + \left|sin14y + \frac{1}{2}cos15x + \frac{1}{3}sin16y + \frac{1}{4}cos17x\right|}\right|}$$

## A continuous change over time

$$\frac{\dfrac{sinxy + \frac{1}{2}costx + \frac{1}{3}sinty + \frac{1}{4}cos4(x+y)}{1 + \left|sin5y + \frac{1}{2}cos6x + \frac{1}{3}sin7y + \frac{1}{4}cos8x\right|}}{1 + \left|\dfrac{(sin10xy + \frac{1}{2}cos10x + \frac{1}{3}sin12y + \frac{1}{4}cos13(x+y))}{t + \left|sin14y + \frac{1}{2}cos15x + \frac{1}{3}sin16y + \frac{1}{4}cos17x\right|}\right|}$$

## Geometric Integrity Constraints

Geometric integrity constraints are defined according to the spatial data model to ensure spatial data consistency. They may also be a requirement for geometric algorithms.

Typical geometric integrity constraints are:

- Tessellation: for every point $p$, $p$ is in only one cell, and every cell has a corresponding property value.
- Polyline: with the exception of the extreme point, each line segment ends when it starts the next line segment.
- Polygon: is a closed polyline.
- Simple polyline: two polylines only intersect at the extreme point of a line segment.

## Topological Integrity Constraints

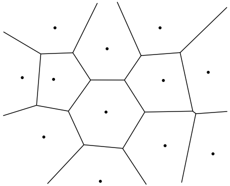
Without concerning the semantics of physical objects, I also distinguish constraints that are not related to the spatial data model, but to the manipulation of data. I refer to these constraints as topological constraints, which are particularly useful among multiple representations of the same physical objects or when there are continuous changes of objects' geometry.

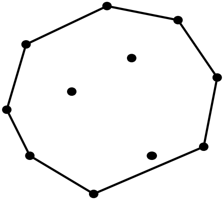
## Data Manipulation: Unary Operations

- Unary operations with Boolean result:
  $op : spatial \longrightarrow Boolean$:
  - convex test
  - connected test
- Unary operations with scalar results
  $op : spatial \longrightarrow int$:
  - length
  - area
  - perimeter
- Unary operations with spatial result:
  $op : spatial \longrightarrow spatial$:
  - topological transformations (e.g., rotation, translation and scaling)
  - transformation to another object of different dimension (e.g., boundary)
  - object extraction (e.g., minimal bounding rectangle, centroide, buffering)

# Data Manipulation: *N*- ary Operations

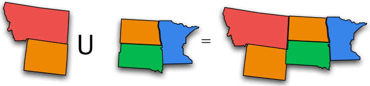
- *N*-ary operations with spatial result:
  $op : spatial* \rightarrow spatial$:
  - Voronoi
  - Convex hull



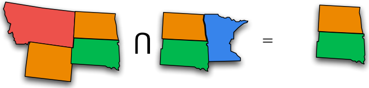


# Data Manipulation: Binary Operations

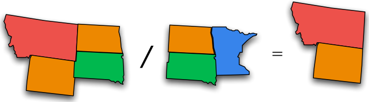
- Binary operations with spatial result:
  $op : spatial \times spatial \longrightarrow spatial$:
  - union



  - intersection



  - difference



## Data Manipulation: Binary Operations

- Binary operations with Boolean result:
  $op : spatial \times spatial \longrightarrow Boolean$:
  - topological predicates
  - direction predicates
  - metric predicates
- Binary operations with scalar result:
  $op : spatial \times spatial \longrightarrow int$:
  - distance