PRINCIPLES OF DATA ORGANISATION

Spatial Data

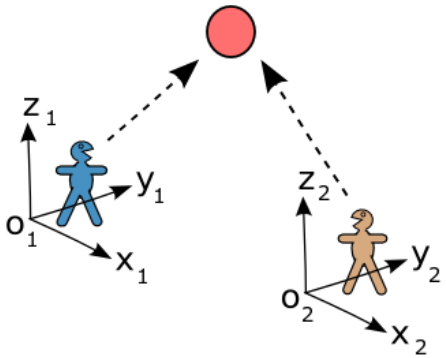
MOTIVATION

- How to search effectively in more than one dimension?
  - One dimension: number, string, …
- Examples:
  - GIS, CAD, 3D modelling, multimedia
  - Longitude and latitude, a vector describing features of a product, images, videos, …
- Queries:
  - List the names of all cinemas within 10 kilometres from the city centre
  - List all rivers crossing at least two states
    - The geometry of a river, the geometry of a state

- Generalisation of algorithms from lower dimension (1D) to higher dimension may seems to be straightforward, but it is usually inefficient

FRAME OF REFERENCE

- How to define a position in a space?
  - Use a reference to a certain point
- Every domain dealing with spatial data needs a spatial frame of reference to anchor the objects to be able to process them
- Examples:
  - GIS : Earth's surface
  - CAD : building's layout
  - Medical imaging : human body
  - Multimedia : vector space, mostly nearest-neighbour queries



SPATIAL DATA

Spatial data

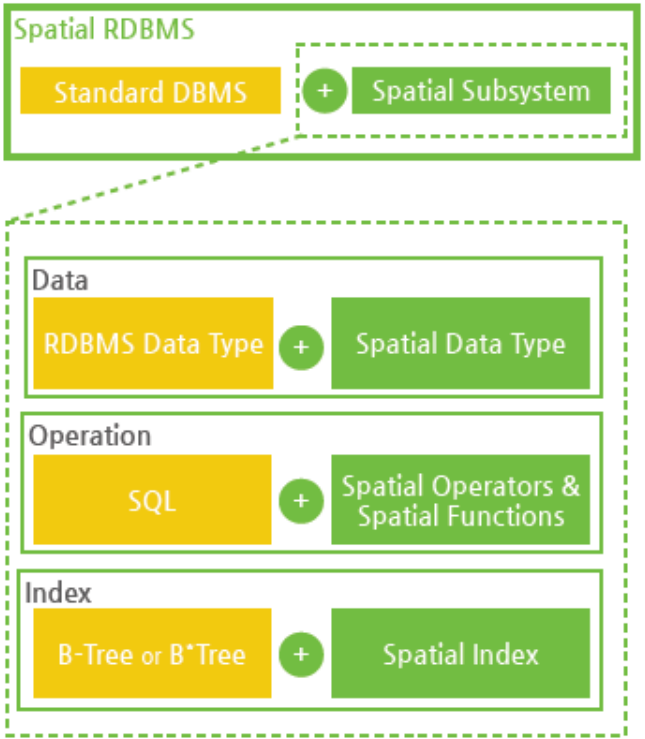
- Data with the assigned location
  - Within a frame of reference
- The spatial component of an object consists of a location and an extent

- They are somewhere and occupy some space

Geographic/geo-spatial data

- Spatial data with assigned location with Earth's surface as the reference frame
- The position moves (continents move), so we need additional information

SPATIAL DATABASE MANAGEMENT SYSTEM



Relational Database Management System (RDBMS)

- Data types, operations (SQL), indices, primary files, …

Spatial Database management System

- Data type – points, lines, polylines, areas, …
- Spatial query language – operators, functions, mostly incorporated into SQL
  - Special keywords
- Spatial indexing – efficient techniques for spatial operations (e.g., spatial join)

We want to connect the worlds

SPATIAL DATA TYPES

- Point
  - n-dimensional points in m-dimensional space
  - E.g., location in the modelled domain, feature vectors
- Line
  - Connected set of line segments
  - E.g., river
- Region (polygon)
  - Point data with spatial extent defined by its boundaries
  - Boundaries are defined by connected line segments (vectors) with common beginning and end
  - E.g., country

SQL EXTENSION

- SQL Multimedia and Application Packages
  - Standard being part of SQL:1999
- ISO SQL Geometry Specification
- Set of user-defined types (UDT)

A spatial database extender for PostgreSQL

PostGIS implementation

Area(POLYGON), Distance(GEOMETRY,GEOMETRY), Contains(GEOMETRY,GEOMETRY), Intersection(GEOMETRY,GEOMETRY), Intersects(GEOMETRY,GEOMETRY), Union(GEOMETRY,GEOMETRY), Buffer(GEOMETRY,double), ConvexHull(GEOMETRY), Perimeter(GEOMETRY), Crosses(GEOMETRY,GEOMETRY), Transform(GEOMETRY,integerSRID), Dimension(GEOMETRY), AsText(GEOMETRY), ST_X(POINT), ST_Y(POINT), NumPoints(GEOMETRY), PointN(GEOMETRY,integer), NumGeometries(GEOMETRY), GeometryN(GEOMETRY,integer), GeometryType(GEOMETRY)

SQL EXTENSION : EXAMPLE

```
INSERT INTO roads (road_id, road_geom, road_name) VALUES
(1, GeomFromText('LINESTRING(19123 24311,19110 23242)', 242),'Jeff Rd.') ;
```

// Insert a road and its coordinates

```
SELECT sum(population) FROM census_tracks
WHERE distance(census_geom, 'POINT(210030 3731201)') < (5 * 1609.344)
```

// How many people live within 5 miles of a certain place?

```
SELECT sum(area(park_geom)) FROM parks, nhoods
WHERE contains(nd_geom,park_geom) AND nhood_name = 'Westside‘
```

// What is the area of municipal parks inside the Westside neighbourhood?

SPATIAL QUERIES/OPERATIONS

Queries

- Containment query
  - Find features (points, lines, polygons) that are within particular polygons
- Region query
  - Find what intersects with given region
- Enclosure query
  - Find what encloses a given region
- Line intersection query
- Adjacency query
- Metric (proximity) queries

Operations

- Clipping
  - Cutting
- Map overlay
  - Union, intersection
- Merge/Aggregation
  - Geometric Union
- Spatial join
  - Does not alter the original geography!
  - Just merges

- Example: intersection between town boundaries and water bodies - water polygons are split at town boundaries

SPATIAL JOIN

Given two sets of spatial objects, spatial join pairs the sets' objects based on a given spatial predicate

- Intersection
- Distance
- … any relation, including a pair of spatial objects

Examples:

- Find all pairs of rivers and cities that intersect
- Identify pairs of objects from two sets which are within a given distance

SPATIAL OBJECTS REPRESENTATION & INDEXING

- Project (serialize) them into 1D space and employ existing single-dimensional methods
  - E.g., B-trees
- Utilize the full spatial information with specialized techniques for spatial management
  - E.g., Quadtree, k-d-tree, R-tree, R+-tree, R*-tree, Hilbert R-tree, X-tree, UB-tree, …

Methods for efficient search in multidimensional data, i.e., spatial queries, should access as few pages as possible.