UNIT II SPATIAL AND TEMPORAL DATABASES 9

Active Databases Model – Design and Implementation Issues - Temporal Databases - Temporal Querying - Spatial Databases: Spatial Data Types, Spatial Operators and Queries – Spatial Indexing and Mining – Applications -– Mobile Databases: Location and Handoff Management, Mobile Transaction Models – Deductive Databases - Multimedia Databases.

SPATIAL INDEXING

A spatial index is a specialized indexing structure where the indexing key is the spatial location of objects indexed. The type of searches on a spatial index is the set of spatial queries, for example, range and overlapping queries.

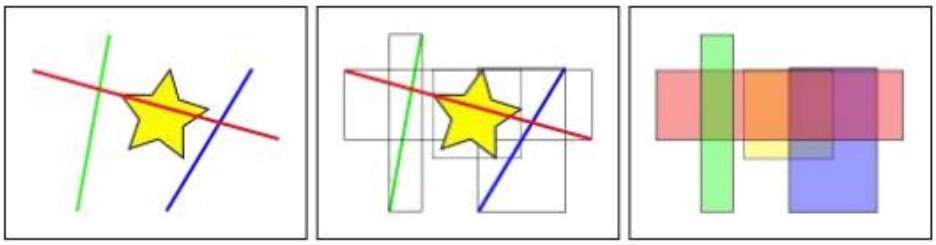
Spatial indexing method divides the space into a manageable number of smaller subspaces, which can be further divided into smaller subspaces and so on. The partitioning continues until the unpartitioned subspace contains the objects that can be stored in a data page. While designing the index structures for spatial databases the storage space must be efficiently utilized and the information retrieval should be fast and easy

A spatial index is a specialized indexing structure where the indexing key is the spatial location of objects indexed. The type of searches on an spatial index is the set of spatial queries, for example, range and overlapping queries

```
CREATE INDEX nyc_census_blocks_geom_idx
ON nyc_census_blocks
USING GIST (geom);
```

How Spatial Indexes Work

Standard database indexes create a hierarchical tree based on the values of the column being indexed. Spatial indexes are a little different – they are unable to index the geometric features themselves and instead index the bounding boxes of the features.

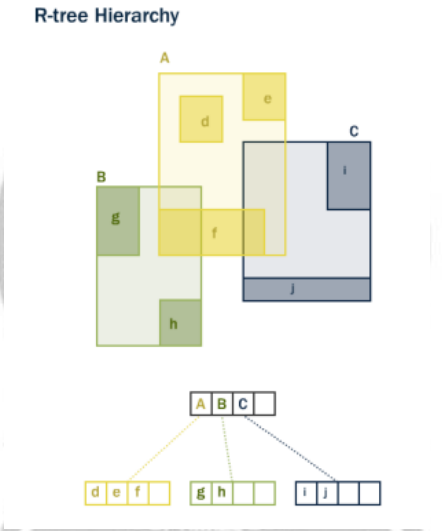


In the figure above, the number of lines that intersect the yellow star is one, the red line. But the bounding boxes of features that intersect the yellow box are two, the red and blue ones.

The way the database efficiently answers the question <what lines intersect the yellow star= is to first answer the question <what boxes intersect the yellow box= using the index (which is very fast) and then do an exact calculation of <what lines intersect the yellow star= only for those features returned by the first test.

For a large table, this <two pass= system of evaluating the approximate index first, then carrying out an exact test can radically reduce the amount of calculations necessary to answer a query.

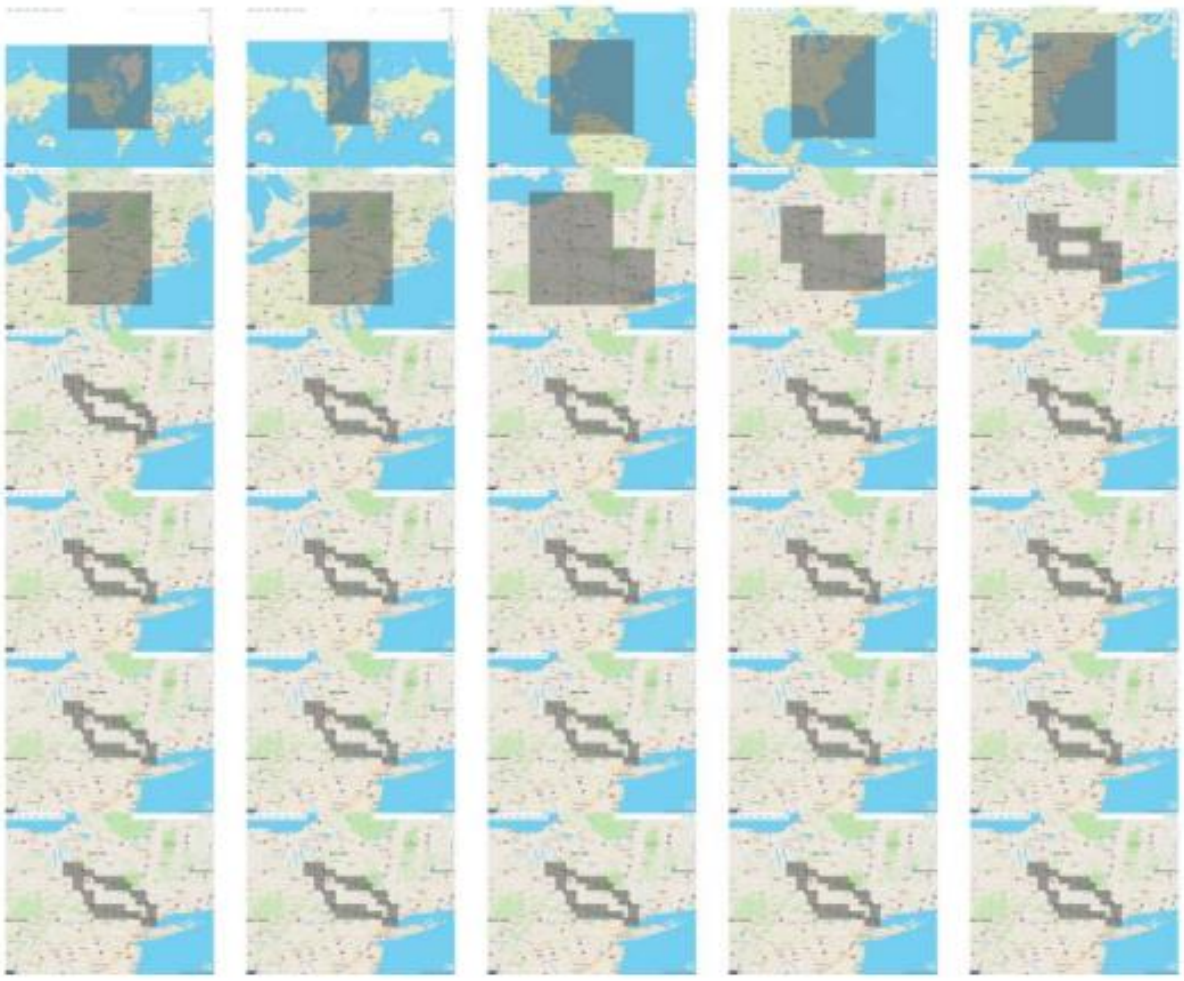
Both PostGIS and Oracle Spatial share the same <R-Tree= 1 spatial index structure. R-Trees break up data into rectangles, and sub-rectangles, and sub-sub rectangles, etc. It is a self-tuning index structure that automatically handles variable data density, differing amounts of object overlap, and object size.



Data structures like B trees have been designed for efficient insertion and deletion in databases. Spatial indexing is used to look up the values that match the predicate in efficient manner. There are two ways to provide spatial indexing:

i) Dedicated external spatial data structures are added to the system that provide the attributes for spatial databases e.g. a B-tree does for standard attributes, and

ii) spatial objects are mapped into a one-dimensional space so that they can be stored within a standard one dimensional index such as a B-tree.



SPATIAL DATA MINING

Spatial data mining refers to the process of the retrieval of information or patterns that are not explicitly stored in the spatial databases. Spatial data mining methods are used for the better understanding of spatial data, identifying the relationships between spatial data and non- spatial data, query optimization in spatial databases etc.

Statistical Spatial analysis is the most commonly and widely used data mining technique. It assumes that the spatial data are independent which in fact is not true as the spatial data are interrelated with their neighboring objects. Statistical method cannot handle symbolic values and non linear rules and are also very costly in the result computation. Several Machine learning techniques like learning from examples and generalization and specialization are used in spatial data mining.

SPATIAL DATA MINING ARCHITECTURE

Matheus architecture is the most general and widely used architecture in spatial data mining. This architecture is user controlled. All the predefined information about the

objects is stored in the knowledge base which is fetched by the DB interface for query optimization.

The information which is useful for the pattern recognition is decided by the Focus Component and fed as input to the pattern extraction. The output is then monitored and evaluated by the Evaluation module and duplicate values are removed.

All the components interact using the Controller. Geographic data consists of the spatial objects and the non spatial information about these objects (which can be stored in the database as a pointer to the spatial description of the object).

Spatial data is characterized by geometric as well as topological characteristics where geometric characteristics involve the information about length, area, perimeter etc and topological characteristics include the information about neighbours, intersection etc.

SPATIAL DATA MINING METHODS

Various methods have been designed for mining the data related to geometric space like points, polygon, rectangles, network and other complex objects. There are various kinds of rules associated with spatial data mining.

a.) Characteristic Rule: It refers to the general description of object data. Example rule describing the general price range of shops in various geographic regions of a city.

b.) Discriminant Rule: It refers to the properties or features that distinguish one object from another. Example the comparison of the various shop prices in different regions.

c.) Association Rule: It refers to the association of one object with another.

Applications

The following are examples of the kinds of data mining applications that could benefit from including spatial information in their processing:

- **Business prospecting:** Determine if colocation of a business with another franchise (such as colocation of a Pizza Hut restaurant with a Blockbuster video store) might improve its sales.
- **Store prospecting:** Find a good store location that is within 50 miles of a major city and inside a state with no sales tax. (Although 50 miles is probably too far to drive to avoid a sales tax, many customers may live near the edge of the 50-mile radius and thus be near the state with no sales tax.)
- **Hospital prospecting:** Identify the best locations for opening new hospitals based on the population of patients who live in each neighborhood.
- **Hotspot Detection:** Given a set of geospatial points which are related to an activity in a spatial domain, hotspots are the regions that are more active and have higher density of points compared to other regions.

- **Spatial region-based classification or personalization:** Determine if southeastern United States customers in a certain age or income category are more likely to prefer "soft" or "hard" rock music.
- **Automobile insurance:** Given a customer's home or work location, determine if it is in an area with high or low rates of accident claims or auto thefts.
- **Property analysis:** Use colocation rules to find hidden associations between proximity to a highway and either the price of a house or the sales volume of a store.
- **Property assessment:** In assessing the value of a house, examine the values of similar houses in a neighborhood, and derive an estimate based on variations and spatial correlation.