Multimedia (spatial) databases

Principles of Modern Database Systems
2007

Tore Risch
Dept. of information technology
Uppsala University
Sweden
New kinds of data

- Text, documents
- HTML, XML, XML-Schema documents
- bitmaps, raster images
- audio
- video
- maps
- time series
- vector data, geometrical models
Properties of multi-media data objects

- (very) large data items
- more or less complex internal structure
  • E.g. query by humming:
  - Need special data entry equipment (e.g. microphone)
  - Need special result presentation equipment (e.g. loudspeaker)
  • Can be stored as BLOBs
  • Alt. filenames in tables
  • Problem: How to index and query BLOB contents?
Multi-media query

• Query:
  select track from songs where
    contains(content,:mysong,0.9);
  - Need way to enter :mysong and to realize result.
  - Large results
  - Order as top-10 list
  - Similarity matching
  - Contains can be foreign function
  - Indexing desirable!
Multi-media indexing

- Different indexes sensitive to different predicates
  - Hashing: \( x = x_0 \)
  - B-tree: \( x_l < x < x_u \)
  - R-tree: \( x_l < x < x_u \) and \( y_l < y < y_u \)
    2D matching and up
- Proximity queries (nearness)
  - B-tree supports nearest(\( x_0 \))
  - R-tree supports nearest (\( \{x_0,y_0\} \)), overlaps(\( r_1, r_2 \))
  - SS-tree supports within dist(\( \{x_0,y_0\}, d \))
- High dimensionality not efficient (> 8)
Feature indexing

- Extract feature vector from object
  - For example color spectrum, sharpness, pitch
  - not(overlaps(f(x),f(y)))) => not(overlaps(x,y))
- Feature matches => object may match
- Use R-tree to store feature vectors
- Extract feature vector from compared object
- Search objects in database where feature vectors overlap
- Make careful test for each found object (e.g. detailed image analysis)
- May need to limit dimensionality of vector