

Πανεπιστήμιο Πειραιώς, Τμήμα Πληροφορικής

Γεωγραφικά Πληροφοριακά Συστήματα (Geographical Information Systems - GIS)

Χωρικά ΣΔΒΔ – φυσικό επίπεδο
(ευρετήρια, επεξεργασία χωρικών ερωτήσεων)

Γιάννης Θεοδωρίδης, Νίκος Πελέκης, Ηλίας Φρέντζος

Εργαστήριο Πληροφοριακών Συστημάτων (InfoLab)

<http://infolab.cs.unipi.gr/>

Πειραιάς, 7/5/09

Φυσικό μοντέλο σχεδιασμού



- 3 επίπεδα σχεδιασμού ΒΔ
 - **Εννοιολογικό μοντέλο**: υψηλού επιπέδου αφαιρετική περιγραφή
 - **Λογικό μοντέλο**: πραγματοποιήσιμη περιγραφή
 - **Φυσικό μοντέλο**: υλοποίηση
- Τι είναι ένα φυσικό μοντέλο ΒΔ;
 - Πώς θα υλοποιηθεί αυτό που περιγράφεται στο λογικό μοντέλο
 - Με χρήση βασικών στοιχείων, π.χ. hardware H/Y, λειτουργικό σύστημα
 - Με αποτελεσματικό και χωρίς σφάλματα τρόπο

Τι είναι ένα φυσικό μοντέλο ΒΔ;



- Λόγοι μάθησης εννοιών φυσικών μοντέλων
 - Σωστή επιλογή εμπορικών ΣΔΒΔ
 - Μερικά ΣΔΒΔ δεν υποστηρίζουν ευρετήρια πάνω σε χωρικά δεδομένα !
 - Σωστή διαχείριση δυνατοτήτων ΣΔΒΔ με στόχο τον καλύτερο συντονισμό της απόδοσης του (performance tuning)
 - Παράδειγμα: μία ερώτηση εκτελείται αργά,
 - Ίσως χρειάζεται να δημιουργηθεί ένα ευρετήριο
 - Παράδειγμα: δεν μπορεί να φορτωθεί μεγάλος αριθμός εγγραφών
 - Θα μπορούσαν να καταργηθούν τα ευρετήρια πριν τη φόρτωση
 - και να ξαναδημιουργηθούν μετά τη φόρτωση των δεδομένων!

Φυσικό μοντέλο δεδομένων για ΧΣΔΒΔ

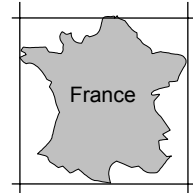


- Είναι το σχεσιακό μοντέλο των ΣΔΒΔ κατάλληλο για τα χωρικά δεδομένα;
 - Τα σχεσιακά ΣΔΒΔ παρέχουν απλούς τύπους τιμών, όπως αριθμοί και συμβολοσειρές
 - Έννοια της ολικής διάταξης
 - τα δένδρα αναζήτησης είναι αποτελεσματικά για αριθμούς
 - Οι έννοιες αυτές δεν έχουν ιδιαίτερο νόημα για τα χωρικά δεδομένα (π.χ. σημεία στο επίπεδο)
- Εναλλακτικές λύσεις:
 - Επαναχρησιμοποίηση εννοιών σχεσιακού μοντέλου
 - Καινοτόμες χωρικές τεχνικές

Υποθέσεις για το φυσικό μοντέλο ΧΣΔΒΔ



- **Χωρικά δεδομένα**
 - Ο αριθμός διαστάσεων του χώρου είναι μικρός π.χ. 2 ή 3
 - Τύποι δεδομένων: συμβατοί με το OpenGIS πρότυπο
 - Προσεγγίσεις αντικειμένων με έκταση (π.χ. γραμμές, πολύγωνα)
- **Minimum (Orthogonal) Bounding Rectangle (MOBR ή MBR)**
 - MBR(obj) είναι το μικρότερο ορθογώνιο παραλληλόγραμμο (παραλληλεπίπεδο κλπ.) που περιβάλλει το αντικείμενο obj
 - Ανάγκη φιλτραρίσματος και εκλέπτυνσης (filter and refinement)
- **Χωρικές λειτουργίες**
 - Λειτουργίες OpenGIS προτύπου
 - π.χ. τοπολογικές, χωρικής ανάλυσης
 - Απλές χωρικές ερωτήσεις – επόμενη διαφάνεια



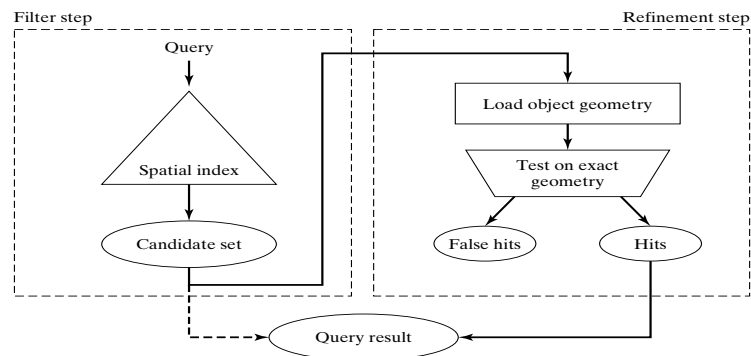
Απλές χωρικές ερωτήσεις και λειτουργίες



- Το φυσικό μοντέλο παρέχει απλές λειτουργίες που χρειάζονται στην επίλυση χωρικών ερωτήσεων!
- Συνηθισμένες ερωτήσεις
 - **Point query**: Βρες όλα τα αντικείμενα που περικλείουν ένα δοθέν σημείο
 - **Range query**: Βρες όλα τα αντικείμενα που κείνται μέσα σε μια (συνήθως ορθογώνια) περιοχή
 - **Nearest neighbor query**: Βρες το πλησιέστερο αντικείμενο σε σχέση με ένα σημείο
 - **Spatial Join** (μεταξύ δύο συνόλων αντικειμένων): Βρες όλα τα ζεύγη αντικειμένων (ένα από κάθε σύνολο) που ικανοποιούν μια χωρική συνθήκη (συνήθως, overlap)

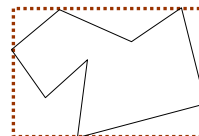
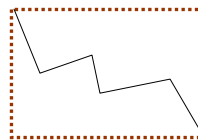
Η διαδικασία «Filter-Refinement»

- Επεξεργασία μίας χωρικής ερώτησης Q
 - **βήμα φιλτραρίσματος** (Filter): βρες ένα σύνολο S που σίγουρα περιέχει και τις απαντήσεις του Q
 - Με χρήση προσεγγιστικών χωρικών τύπων και λειτουργιών
 - **βήμα εκλέπτυνσης** (Refinement): βρες τις ακριβείς απαντήσεις του Q κάνοντας χρήση ενός GIS για την επεξεργασία του S
 - Με χρήση επακριβών χωρικών τύπων και λειτουργιών

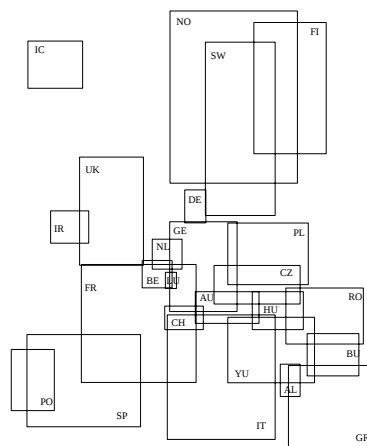


Προσεγγιστικοί χωρικοί τύποι

- Minimum (orthogonal) bounding rectangle (MBR)
 - προσεγγίζει γραμμές, πολύγωνα, ...
- Τα MBRs χρησιμοποιούνται κατά κόρον από χωρικά ευρετήρια, π.χ. R-tree
- Οι αλγόριθμοι για χωρικές λειτουργίες σε MBRs είναι απλοί
- Τα Χωρικά ΣΔΒΔ επεξεργάζονται MBRs στο βήμα filter (για λόγους επίδοσης)



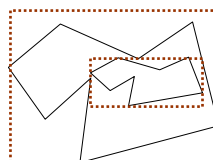
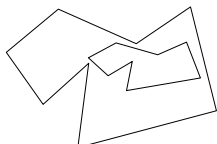
Παράδειγμα εφαρμογής MBR



Παράδειγμα filter – refinement (1)



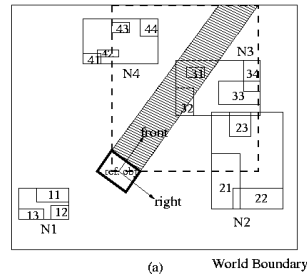
- Οι τοπολογικοί τελεστές του Egenhofer προσεγγίζονται μέσω του τελεστή Overlap
- Παράδειγμα: $\text{inside}(A, B)$
 - η συνθήκη $\text{overlap}(\text{MBR}(A), \text{MBR}(B))$ είναι αναγκαία (αλλά όχι ικανή) για να ισχύει $\text{inside}(A, B)$
 - οπότε στο filter βήμα εκτελείται ο (προσεγγιστικός και πιο φθηνός) τελεστής $\text{overlap}(\text{MBR}(A), \text{MBR}(B))$
 - και στο refinement βήμα εκτελείται ο (ακριβής αλλά και χρονοβόρος) τελεστής $\text{inside}(A, B)$



Παράδειγμα filter – refinement (2)



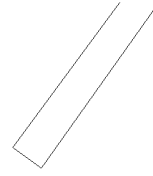
- Ερώτηση:
 - Βρες τα αντικείμενα που βρίσκονται μπροστά από το θεατή V
- Ισοδύναμο με ερώτηση επικάλυψης (overlap query)
 - Η περιοχή κατεύθυνσης είναι ένα πολύγωνο
 - Η απάντηση περιλαμβάνει τα αντικείμενα που έχουν επικάλυψη με το πολύγωνο (front(V))
- Προσεγγιστική ερώτηση
 - Βρες τα αντικείμενα που έχουν επικάλυψη με το MBR(polygon (front (V)))



(a) World Boundary



(b) Range Query



(c) Direction region

Οργάνωση αρχείων & ευρετήρια

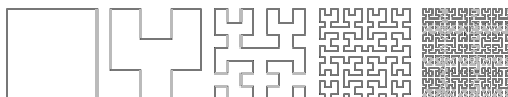


- Επαναχρησιμοποίηση εννοιών σχεσιακού φυσικού μοντέλου – η έννοια της διάταξης (ordering)
 - Οι **καμπύλες διάσχισης χώρου** (space filling curves) ορίζουν μία ολική διάταξη για τα σημεία
 - Αυτή η ολική διάταξη βοηθάει στη χρήση διατεταγμένων αρχείων (ordered files) ή δένδρων αναζήτησης (search trees)
 - π.χ. B-trees
 - Αλλά μπορεί να οδηγήσει σε αναποτελεσματικούς υπολογισμούς!
- Καινοτόμες τεχνικές
 - Χωρικά ευρετήρια, π.χ. δένδρα περιοχών (R-trees), δένδρα τεταρτημορίων (Quadtree), αρχεία πλέγματος (Gridfile)
 - Παρέχουν καλύτερες αποδόσεις στους υπολογισμούς

Διατεταγμένα αρχεία



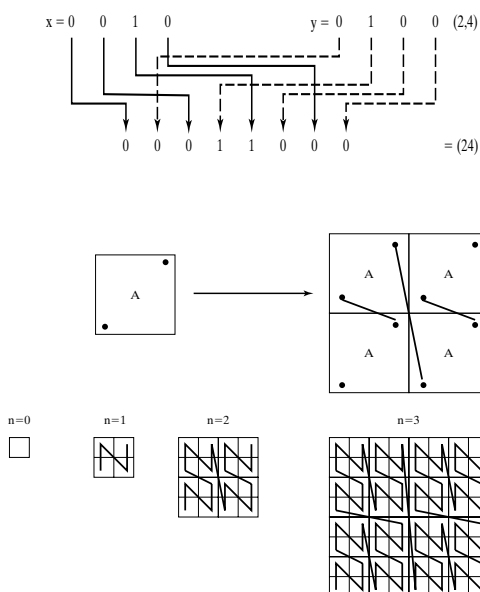
- Γενικά ένα διατεταγμένο αρχείο ομαδοποιεί τις εγγραφές
 - που προσπελαύνονται από πολλές ερωτήσεις
 - σε κοινούς τομείς δίσκου
 - για τη μείωση του I/O κόστους για επιλεγμένες ερωτήσεις
- Πρόβλημα:
 - Η διάταξη δεν έχει κάποιο φυσικό νόημα στα χωρικά δεδομένα
- Πιθανή λύση:
 - Ομαδοποίηση των εγγραφών με χρήση μιας «καμπύλης διάσχισης χώρου» (space filling curve)
 - Παραδείγματα: Z-curve, Hilbert-curve



Διάσχιση Z



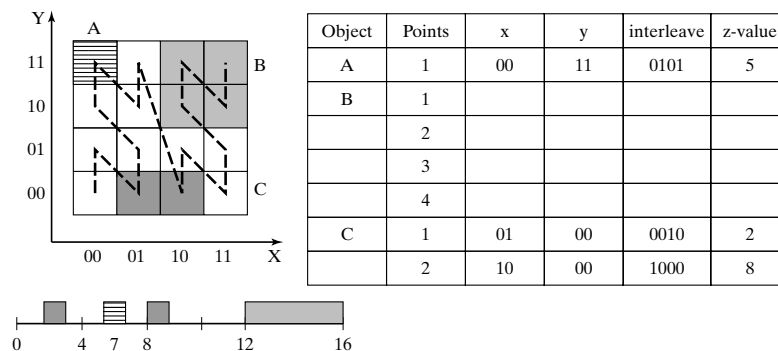
- Τι είναι μία Z-curve?
 - Μία καμπύλη διάσχισης χώρου
 - Παραγόμενη από διατεταγμένα (interleaved) bits
 - x, y συντεταγμένες
 - Εναλλακτική μέθοδος δημιουργίας: συνδέοντας σημεία με z-order
 - Μοιάζει με N ή Z



Παράδειγμα Z-τιμών



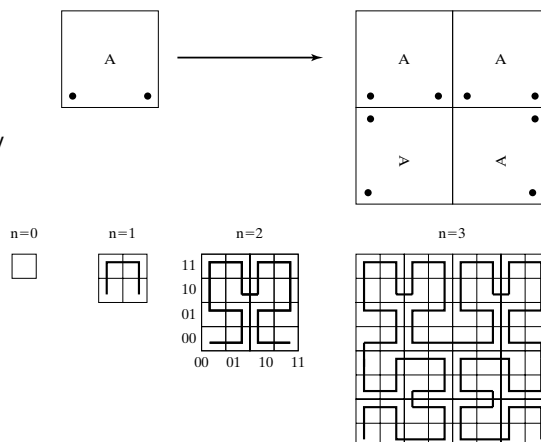
- Το αριστερό μέρος δείχνει ένα χάρτη με τα χωρικά αντικείμενα A, B, C
- Το δεξί και κάτω αριστερό μέρος Z-τιμές στα A, B και C
- Παρατηρήστε ότι το C παίρνει z-values 2 και 8, οι οποίες δεν είναι κοντά
- Άσκηση: Υπολογίστε τις z-τιμές για το B



Διάσχιση Hilbert



- Τι είναι μία Hilbert curve?
 - Μία καμπύλη διάσχισης χώρου
 - Πιο πολύπλοκη στη δημιουργία της από τη Z
 - Λόγω περιστροφών (στα 2 κάτω τεταρτημόρια)

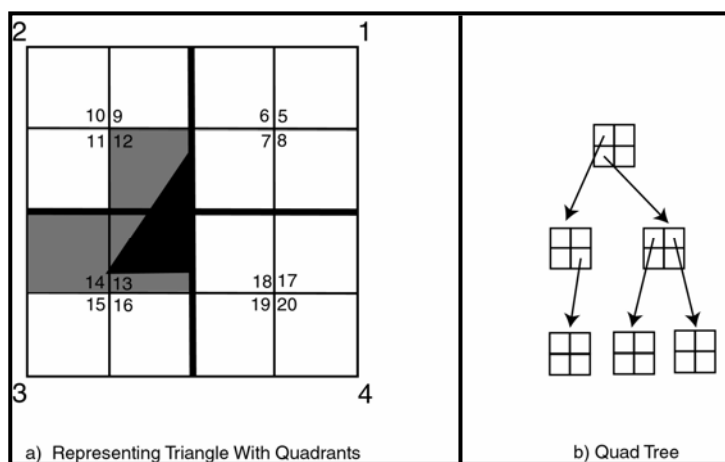


Χωρικά ευρετήρια



- Η δομή ενός index εξαρτάται από τον τύπο δεδομένων του δεικτοδοτούμενου χαρακτηριστικού (attribute)
- Χαρακτηριστικά με ολική διάταξη
 - αριθμοί, συμβολοσειρές, σημεία ταξινομημένα με καμπύλες διάσχισης χώρου
 - Το B-tree είναι μία δημοφιλής δομή οργάνωσης τέτοιων τύπων
- Χωρικά αντικείμενα (π.χ. πολύγωνα)
 - Η χωρική οργάνωση είναι πιο αποτελεσματική
 - Εκατοντάδες προτάσεις στη βιβλιογραφία
 - Δύο βασικές οικογένειες:
 - Τετραδικά δέντρα (Quadtrees) και
 - R-δέντρα (R-trees)

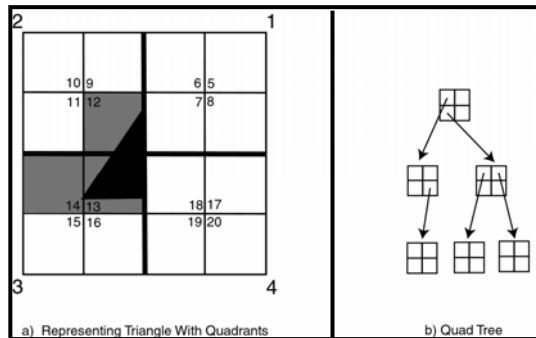
Παράδειγμα Quadtree



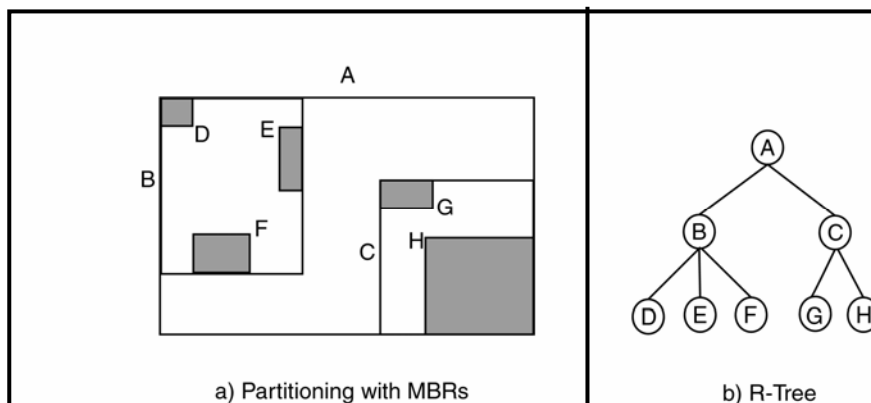
Ιδιότητες Quadtree



- Ιεραρχική αποσύνθεση του χώρου σε τεταρτημόρια (quadrants)
- Κάθε επίπεδο στο Quadtree αναπαριστά το αντικείμενο ως σύνολο MBRs με βάση τα τεταρτημόρια ανάλυσης του χώρου.
 - Κάθε επίπεδο είναι μια πιο ακριβής αναπαράσταση του αντικειμένου.
- Ο αριθμός των απαιτούμενων επιπέδων εξαρτάται από την επιθυμητή ακρίβεια.

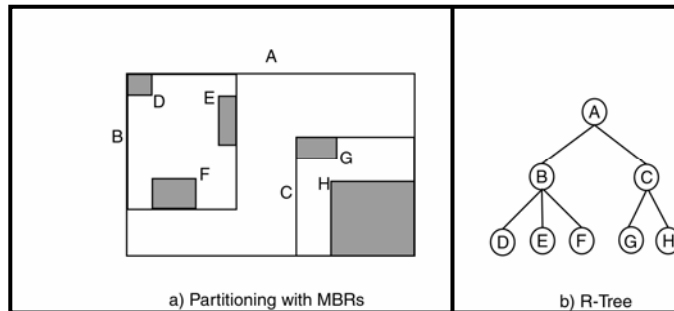


Παράδειγμα R-tree



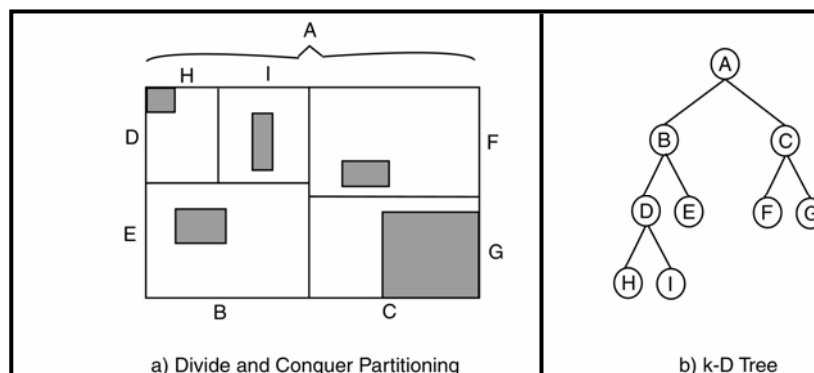
Ιδιότητες R-tree

- Βασική ιδέα
 - Γενίκευση B-tree σε χωρικά δεδομένα
- Ιδιότητες των R-trees
 - Ισοζυγισμένα / Οι κόμβοι είναι ορθογώνια / Το ορθογώνιο-παιδί περιέχεται εντός του ορθογωνίου-πατέρα
 - Η αναζήτηση γίνεται με κατάβαση στο δένδρο
 - Μπορεί να προκύψει χωρική επικάλυψη μεταξύ των κόμβων!



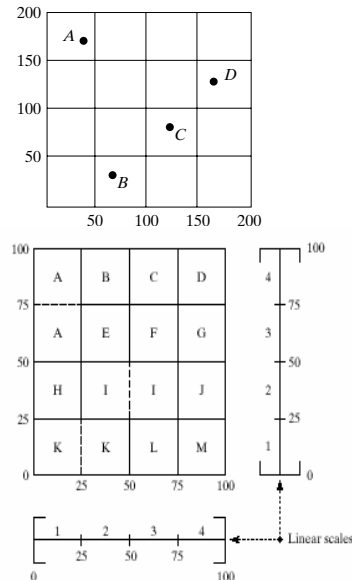
K-D tree

- Παραλλαγή του δυαδικού δένδρου αναζήτησης (BST).
- Κάθε επίπεδο χρησιμοποιείται για να δεικτοδοτήσει μια από τις διαστάσεις του χωρικού αντικειμένου.
- Κάθε κελί του κατώτατου επιπέδου έχει ένα μόνο αντικείμενο εντός του.
- Οι διαιρέσεις δεν γίνονται με χρήση MBR αλλά με διαδοχικές διαιρέσεις του χώρου της διάστασης.



Grid File

- Βασική ιδέα – Διαίρεση του χώρου σε κελιά μέσω ενός πλέγματος
 - Παράδειγμα: συντεταγμένες lat/lon
 - Αποθήκευση δεδομένων κάθε κελιού σε χωριστό τομέα δίσκου
 - Πρόβλημα: Ανομοιόμορφη κατανομή δεδομένων στο χώρο
- Βελτίωση αρχικής ιδέας
 - Χρήση ανομοιόμορφων πλεγμάτων
 - Επιτρέπουμε διαμοιρασμό των τομέων δίσκου μεταξύ κελιών του πλέγματος



Επεξεργασία & βελτιστοποίηση ερωτήσεων

Query Processing and Optimization (QPO)

- Βασική ιδέα QPO
 - Στην SQL, οι ερωτήσεις εκφράζονται σε μία υψηλού επιπέδου δηλωτική μορφή
 - Η διαδικασία QPO μεταφράζει μία SQL ερώτηση σε ένα «σχέδιο εκτέλεσης» (execution plan)
 - Πάνω στο φυσικό μοντέλο δεδομένων
 - Χρησιμοποιώντας λειτουργίες σε δομές αρχείων, ευρετήρια κτλ.
 - Ένα ιδανικό σχέδιο εκτέλεσης απαντά στην ερώτηση Q στο συντομότερο δυνατό χρόνο
 - Περιορισμοί: Το κόστος της διαδικασίας QPO πρέπει να είναι μικρό
 - Υπολογιστικός χρόνος του QPO << χρόνος εκτέλεσης της Q

Τρεις έννοιες-κλειδιά για QPO



- Δομικά συστατικά (building blocks)
 - Τα σχεσιακά ΣΔΒΔ έχουν λίγα δομικά συστατικά:
 - select (match query, interval query), join, sorting, ...
 - Μία ερώτηση SQL αναλύεται σε δομικά συστατικά
- Στρατηγικές επεξεργασίας ερωτήσεων για δομικά συστατικά
 - Τα ΣΔΒΔ διατηρούν ένα μικρό αριθμό στρατηγικών για κάθε δομικό συστατικό
 - Π.χ. ένα match query (δηλ. μια συνθήκη ισότητας στο Where clause) μπορεί να απαντηθεί μέσω ενός index (αν υπάρχει) ή διατρέχοντας όλο το αρχείο δεδομένων (αν δεν υπάρχει index)
- Βελτιστοποίηση ερωτήσεων
 - Για κάθε δομικό συστατικό ενός ερωτήματος, το ΣΔΒΔ προσπαθεί να επιλέξει
 - Την πιο “αποτελεσματική” στρατηγική, δεδομένων κάποιων παραμέτρων
 - Παραδείγματα παραμέτρων: μέγεθος πίνακα, διαθέσιμα ευρετήρια, ...

Προκλήσεις της διαδικασίας QPO



- Επιλογή δομικών συστατικών
 - Οι ερωτήσεις SQL βασίζονται στη σχεσιακή άλγεβρα
 - Τα δομικά συστατικά της ΣΑ είναι τα select / project / join
- Επιλογή στρατηγικών επεξεργασίας δομικών συστατικών
 - Περιορισμοί: Πολλές στρατηγικές → μεγαλύτερη πολυπλοκότητα
 - Τα εμπορικά ΣΔΒΔ διαθέτουν 10-30 στρατηγικές
 - 2-4 στρατηγικές για κάθε δομικό συστατικό
- Πώς επιλέγεται η καλύτερη στρατηγική
 - Χρήση προκαθορισμένου πλάνου προτεραιότητας
 - Χρήση απλού μοντέλου κόστους (cost model) βασισμένου στις παραμέτρους της ΒΔ

Προκλήσεις QPO σε ΧΣΔΒΔ

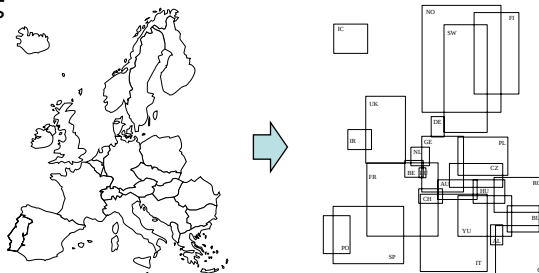


- Επιλογή δομικών συστατικών
 - Πλούσιο σύνολο χωρικών τύπων δεδομένων, λειτουργιών
 - Έλλειψη κοινά αποδεκτών δομικών συστατικών
 - Ποικίλουν ανάλογα με τον κατασκευαστή και το προϊόν
 - Ενδεικτικά: spatial selection (point / range / nearest neighbor query), spatial join (overlap / distance join)
- Επιλογή στρατηγικών
 - Δεν υπάρχουν πολλές επιλογές!
- Πώς επιλέγεται η καλύτερη στρατηγική
 - Τα μοντέλα κόστους είναι πιο πολύπλοκα καθώς
 - Οι χωρικές ερωτήσεις έχουν υψηλές απαιτήσεις, τόσο σε CPU όσο και σε I/O
 - Σε αντίθεση με την παραδοσιακή QPO που έχει μόνο I/O απαιτήσεις
 - Στρατηγικές βάσει μοντέλων κόστους είναι ακόμα σε ερευνητικό επίπεδο

Επιλογή δομικών συστατικών



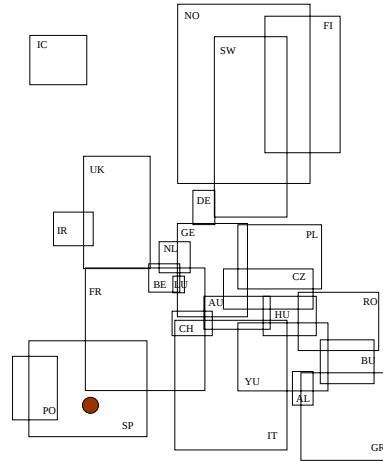
- Αντιπροσωπευτικά δομικά συστατικά (με παραδείγματα από χάρτη Ευρώπης)
 - **Point Query** – π.χ. να βρεθεί σε ποιο κράτος ανήκει ένα τονισμένο σημείο στο χάρτη
 - **Range Query** – π.χ. να βρεθούν όλα τα κράτη που εμφανίζονται στο παράθυρο
 - **Nearest Neighbor Query** – π.χ. να βρεθεί το κράτος που βρίσκεται πλησιέστερα σε ένα τονισμένο σημείο
 - **Spatial Join Query** – π.χ. να βρεθούν τα ζεύγη των κρατών που συνορεύουν μεταξύ τους



Στρατηγικές για Point Queries



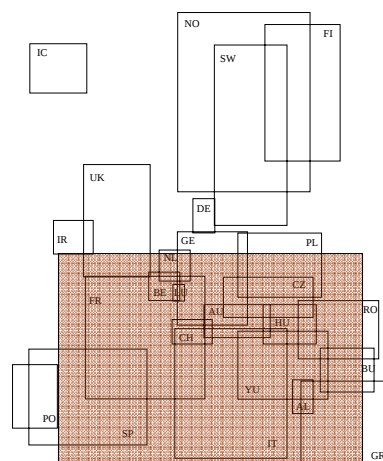
- **Point Query** – π.χ. να βρεθεί σε ποιο κράτος ανήκει ένα τονισμένο σημείο στο χάρτη
 - Είσοδος: ένα σύνολο χωρικών αντικειμένων D και ένα σημείο Q
 - Έξοδος: ένα ή περισσότερα αντικείμενα στα οποία ανήκει το σημείο
- Στρατηγικές:
 - Αν οι εγγραφές του D είναι διατεταγμένες (π.χ. σε Z-order)
 - Χρησιμοποίησε δυαδική αναζήτηση στη Z-order
 - Κόστος = $\log_2 B$ τομών δίσκου
 - Αν υπάρχει χωρικό ευρετήριο πάνω στο D
 - Χρησιμοποίησε τη λειτουργία point query του ευρετηρίου
 - Κόστος = ύψος ευρετηρίου (τυπικές τιμές, 4 με 5)



Στρατηγικές για Range Queries

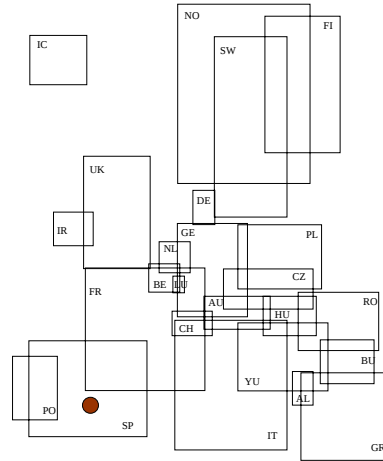


- **Range Query** - π.χ. να βρεθούν όλα τα κράτη που εμφανίζονται στο παράθυρο
 - Είσοδος: ένα σύνολο χωρικών αντικειμένων D και ένα παράθυρο Q
 - Έξοδος: τα αντικείμενα που έχουν επικάλυψη με το παράθυρο
- Στρατηγικές:
 - Αν οι εγγραφές του D είναι διατεταγμένες (π.χ. σε Z-order)
 - Προσδιόρισε το εύρος των τιμών του Z-order που ικανοποιούν την ερώτηση
 - Χρησιμοποίησε δυαδική αναζήτηση για την εύρεση της μικρότερης αποδεκτής τιμής Z
 - Συνέχισε με γραμμική αναζήτηση μέχρι τη μεγαλύτερη αποδεκτή τιμή Z
 - Αν υπάρχει χωρικό ευρετήριο πάνω στο D
 - Χρησιμοποίησε τη λειτουργία range query του ευρετηρίου



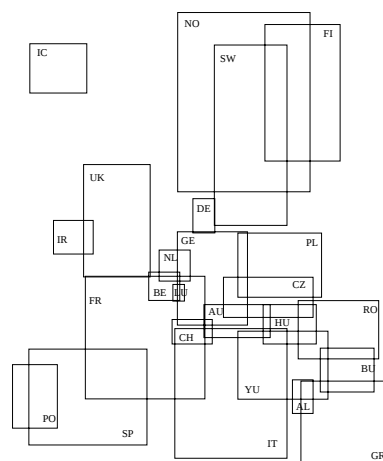
Στρατηγικές για NN Queries

- **NN Query** - π.χ. να βρεθεί το κράτος που βρίσκεται πλησιέστερα σε ένα τονισμένο σημείο
 - Είσοδος: ένα σύνολο χωρικών αντικειμένων D και ένα σημείο Q
 - Έξοδος: ένα αντικείμενο, αυτό που βρίσκεται πλησιέστερα στο σημείο
 - Παραλλαγή: k -NN Query
- Στρατηγικές:
 - Προσέγγιση σε δύο φάσεις
 - Φόρτωσε τις εγγραφές της σελίδας δίσκου του D που περιέχει την τοποθεσία του σημείου Q (Point Query)
 - M = minimum distance(Q , αντικείμενα που φορτώθηκαν)
 - Έλεγχε όλα τα αντικείμενα σε απόσταση M από το Q (Range Query)
 - Αν υπάρχει χωρικό ευρετήριο πάνω στο D
 - Χρησιμοποίησε τη λειτουργία NN query του ευρετηρίου



Στρατηγικές για Spatial Joins

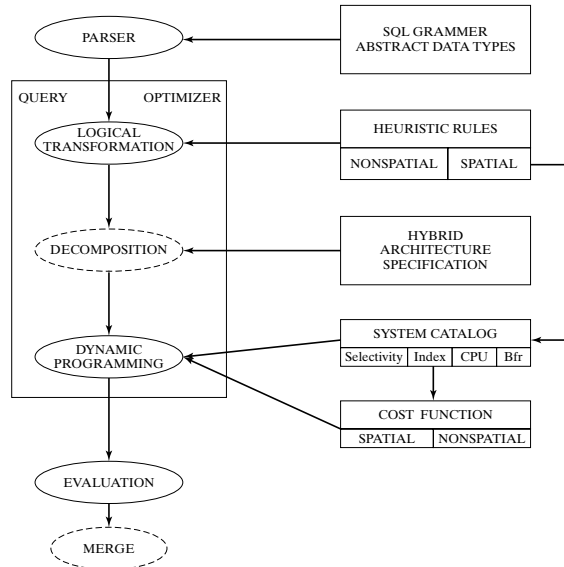
- **Spatial Join Query** – π.χ. να βρεθούν τα ζεύγη των κρατών που συνορεύουν μεταξύ τους
 - Είσοδος: δύο σύνολα χωρικών αντικειμένων D_1 και D_2 και ένας χωρικός τελεστής σύνδεσης τ
 - Έξοδος: ένα σύνολο από ζεύγη αντικειμένων (r, s) , $r \in D_1$ και $s \in D_2$, για τα οποία ικανοποιείται ο τελεστής τ
- Στρατηγικές:
 - φωλιασμένοι βρόγχοι (nested loops):
 - Έλεγχε όλα τα πιθανά ζευγάρια βάσει του τελεστή τ
 - Διαμερισμός χώρου (space partitioning):
 - Έλεγχε ζευγάρια αντικειμένων μόνο από κοινές περιοχές
 - Αν υπάρχει χωρικό ευρετήριο πάνω στο D_1 ή/και το D_2
 - Διάφορες τεχνικές, με βάση το αν υπάρχουν και τα δύο ευρετήρια ή μόνο το ένα.



Διεργασία QPO



- Η διαδικασία
 - Αφετηρία : Μία SQL ερώτηση
 - Προορισμός: Ένα «καλό» σχέδιο εκτέλεσης
 - Ενδιάμεσες στάσεις
 - Δένδρο ερώτησης
 - Λογικός μετασχηματισμός δένδρου
 - Επιλογή στρατηγικής



Δένδρα ερωτήσεων

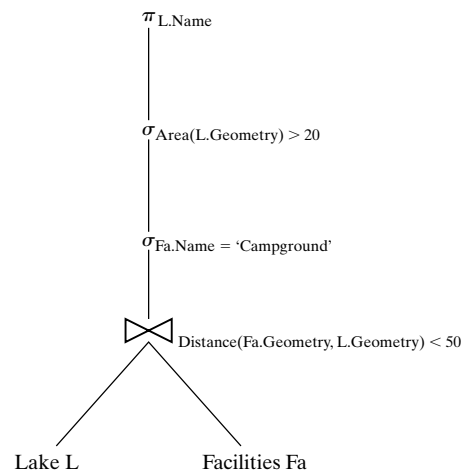


- Κόμβοι = δομικά συστατικά χωρικών ερωτήσεων
- Παιδιά = είσοδος στα δομικά συστατικά
- Φύλλα = Πίνακες
- Παράδειγμα SQL ερώτησης και του δένδρου αυτής:
 - Βρες όλες τις λίμνες με εμβαδόν μεγαλύτερο από 20 km² σε ακτίνα 50 km από το χώρο κατασκήνωσης

```

SELECT L.Name
FROM   Lake L, Facilities Fa
WHERE  Area(L.Geometry) > 20
      AND Fa.Name = 'Campground'
      AND Distance(Fa.Geometry, L.Geometry) < 50

```

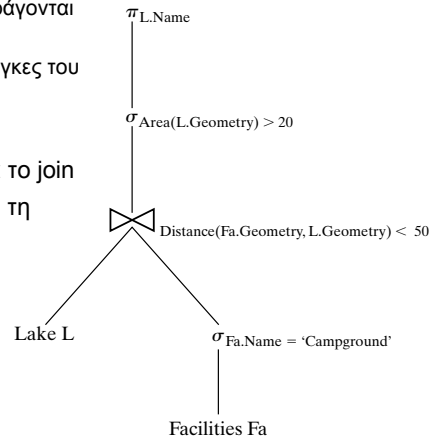


Λογικοί μετασχηματισμοί δένδρων ερώτησης

- Κίνητρο
 - Ο μετασχηματισμός δεν αλλάζει την απάντηση στην ερώτηση
 - Αλλά αλλάζει το υπολογιστικό κόστος με
 - Μειώνοντας τα δεδομένα που παράγονται από τις υπο-ερωτήσεις
 - Μειώνοντας τις υπολογιστικές ανάγκες του πατρικού κόμβου

- Παράδειγμα μετασχηματισμού
 - Εκτέλεσε τη λειτουργία select μετά το join
 - Μειώνει το μέγεθος του πίνακα για τη λειτουργία join

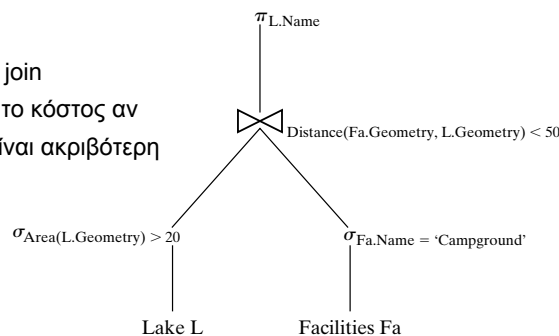
- Άλλοι κοινοί μετασχηματισμοί
 - Μετάθεση της λειτουργίας project
 - Επαναδιάταξη λειτουργιών join
 - ...



Λογικοί μετασχηματισμοί και χωρικές ερωτήσεις

- Παραδοσιακοί κανόνες λογικών μετασχηματισμών
 - Για σχεσιακές ερωτήσεις με απλούς τύπους δεδομένων και λειτουργίες
 - Τα CPU κόστη είναι πολύ μικρότερα από τα I/O κόστη
 - Ανάγκη επαναθεώρησης για τις χωρικές ερωτήσεις
 - Πολύπλοκοι τύποι δεδομένων, λειτουργίες
 - Το κόστος CPU είναι υψηλότερο

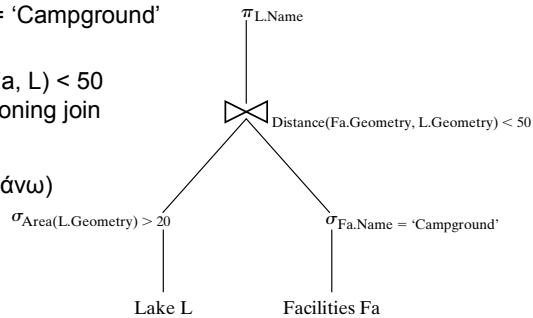
- Παράδειγμα:
 - Μετάθεσε χωρικό selection κάτω από το join
 - Μπορεί να μη μειώσει το κόστος αν
 - η λειτουργία area() είναι ακριβότερη
 - από την distance()



Σχέδια εκτέλεσης



- Ένα σχέδιο εκτέλεσης έχει τρία μέρη
 - Ένα δένδρο ερώτησης
 - Μία στρατηγική επιλεγμένη για κάθε μη-τερματικό κόμβο
 - Μία διάταξη αξιολόγησης των μη-τερματικών κόμβων
- Παράδειγμα
 - Ψάξε για $\text{Area}(\text{L.Geometry}) > 20$
 - Αναζήτηση $\text{Fa.Name} = \text{'Campground'}$ με χρήση index
 - Αναζήτηση $\text{Distance}(\text{Fa}, \text{L}) < 50$ με χρήση space-partitioning join
 - Προβολή on-the-fly
 - Διάταξη (όπως παραπάνω)



Επιλογή στρατηγικής



- Επιλογή βάσει κανόνων (rule-based προσέγγιση)
 - Το σύστημα έχει ένα σύνολο κανόνων που προβάλλουν περιπτώσεις σε επιλογές στρατηγικών
 - Παράδειγμα: Εξαντλητική αναζήτηση στην περίπτωση range query αν το μέγεθος του αποτελέσματος $> 10\%$ του αρχείου δεδομένων
- Επιλογή βάσει κόστους (cost-based προσέγγιση)
 - Για ένα δομικό συστατικό
 - Χρήση τύπων για τον υπολογισμό του κόστους κάθε στρατηγικής, δεδομένου του μεγέθους του πίνακα κτλ.
 - Επιλογή στρατηγικής με το μικρότερο κόστος
 - Για ένα δένδρο ερώτησης
 - Συνδυασμός στρατηγικών ελάχιστου κόστους για τους μη τερματικούς κόμβους
 - Χρήση αλγορίθμων δυναμικού προγραμματισμού
- Τι κάνουν τα εμπορικά συστήματα
 - Cost-based προσέγγιση για τα σχεσιακά (μη χωρικά) δομικά συστατικά
 - Rule-based προσέγγιση για τα χωρικά δομικά συστατικά
 - τα cost models για τα χωρικά δομικά συστατικά δεν είναι ακόμη ώριμα



-
- ΤΕΛΟΣ ΕΝΟΤΗΤΑΣ !