

Χωρικές και Πολυμεσικές Βάσεις Δεδομένων (ΠΜΣ)

Ενδεικτικές ερωτήσεις-θέματα για την εξέταση της θεωρίας

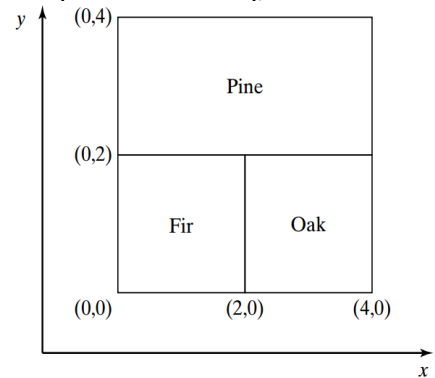
1. Ποια είναι τα βασικά πλεονεκτήματα ενός παραδοσιακού σχεσιακού συστήματος βάσεων δεδομένων και ποια ενός συστήματος επεξεργασίας χωρικών βάσεων δεδομένων; Διατυπώστε με κείμενο δύο παραδείγματα χωρικών ερωτημάτων και δύο μη χωρικών. Αναφέρετε τρεις τύπους μη χωρικών δεδομένων και τρεις χωρικών.

2. Ποιες είναι οι ομοιότητες και ποιες οι διαφορές ενός συστήματος επεξεργασίας χωρικών βάσεων δεδομένων (SDBMS) και ενός GIS; Ποιες ερμηνείες δόθηκαν για τα αρχικά GIS στις δεκαετίες που πέρασαν και ποιος ήταν ο κυριότερος στόχος των συστημάτων αυτών σε κάθε περίπτωση;

3. Ποιες γλώσσες / πρότυπα καθιερώθηκαν για την διατύπωση των ερωτημάτων στα συστήματα επεξεργασίας χωρικών βάσεων δεδομένων (SDBMS); Γιατί απαιτούσε η χωρική επεξεργασία ερωτημάτων νέες τεχνικές και αλγόριθμους; Αναφέρετε σχηματικά ένα παράδειγμα της τεχνικής filter and refine.

4. Γιατί τα συστήματα επεξεργασίας χωρικών βάσεων δεδομένων (SDBMS) απαιτούσαν νέες δομές δεικτοδότησης και αποθήκευσης των δεδομένων; Αναφέρετε τέσσερα παραδείγματα τέτοιων δομών: δύο χωροπληρωτικών και δύο ιεραρχικών δομών. Ποια τα πλεονεκτήματα της κάθε κατηγορίας; Ποια δομή από τις τέσσερις που αναφέρατε είναι τελικά η πιο αποδοτική;

5. Με ποιους βασικούς δύο τρόπους μοντελοποιείται η χωρική πληροφορία; Ποια τα βασικά χαρακτηριστικά και οι ιδιότητες του κάθε μοντέλου; Καταγράψτε τα συγκεκριμένα στοιχεία και δεδομένα των δύο μοντέλων για το παράδειγμα των δασικών συστάδων που εμφανίζεται στην διπλανή εικόνα.



6. Αναφέρετε τουλάχιστον πέντε διαφορετικούς βασικούς τύπους χωρικών αντικειμένων του μοντέλου OGIS. Με ποιο τρόπο παριστάνονται τα σύνθετα γεωμετρικά αντικείμενα στο μοντέλο αυτό (οι τύποι τους); Αναφέρετε τουλάχιστον πέντε διαφορετικούς χωρικούς τελεστές (τοπολογικούς ή κατεύθυνσης ή μέτρησης) και διατυπώστε με κείμενο αντίστοιχα παραδείγματα ερωτημάτων που απαιτείται η χρήση τους.

7. Περιγράψτε το μοντέλο των 9 τομών για τις τοπολογικές σχέσεις μεταξύ δύο αντικειμένων A και B (μπορείτε να χρησιμοποιήσετε και δικούς σας συμβολισμούς για το εσωτερικό, το εξωτερικό και το περίγραμμα ενός αντικειμένου). Σε κάθε περίπτωση σχεδιάστε την θέση των αντικειμένων καθώς και τις τιμές του πίνακα του μοντέλου.

8. Δίνεται το ER-διάγραμμα της χωρικής βάσης του διπλανού σχήματος με pictograms. Να αναγνωριστούν και να καταγραφούν όλες οι οντότητες που υπάρχουν, τα χαρακτηριστικά της κάθε οντότητας, καθώς και όλες οι συσχετίσεις που υπάρχουν στο διάγραμμα με τις αντίστοιχες πληθικότητές τους. Επίσης να καταγραφεί ο τύπος και οι ιδιότητες των χωρικών αντικειμένων.

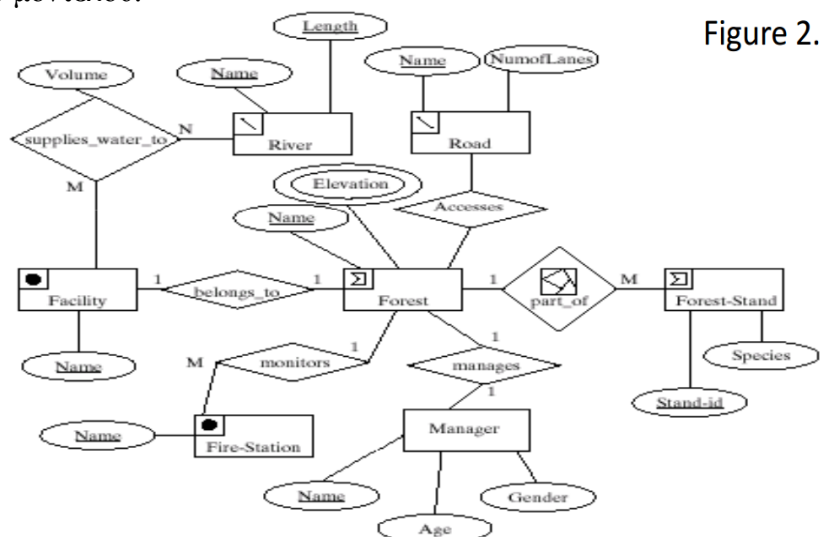
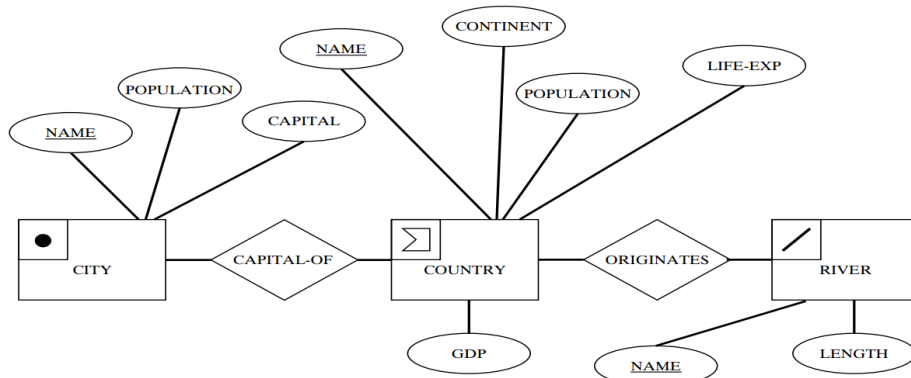


Figure 2.7

9. Δίνεται ένα διάγραμμα μίας χωρικής βάσης δεδομένων στο παρακάτω σχήμα:

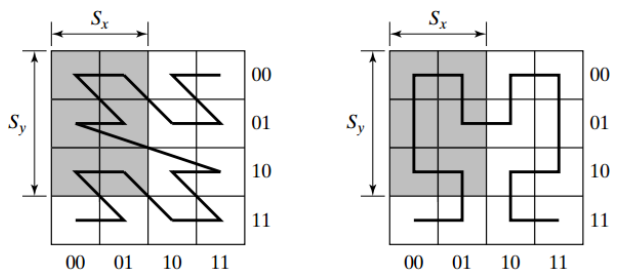


Για τα παρακάτω ερωτήματα SQL να απαντήσετε τι ακριβώς κάνουν περιγράφοντάς τα και τι επιστρέφουν τελικά στον χρήστη:

<pre>SELECT Ci.Name,Co.Pop FROM City Ci,Country Co WHERE Ci.Country=Co.Name AND Co.GDP >1000.0 AND Ci.Capital='Y'</pre>
<pre>SELECT Co.Name FROM Country Co WHERE Co.GDP >ANY(SELECT Co1.GDP FROM Country Co1 WHERE Co1.Name ='Canada')</pre>
<pre>SELECT C.Name, C.Pop, Area(C.Shape) AS "Area" FROM Country C</pre>
<pre>SELECT R.Name, C.Name FROM River R, Country C WHERE Cross(R.Shape,C.Shape)=1</pre>
<pre>SELECT Ci.Name FROM City Ci, River R WHERE Overlap(Ci.Shape, Buffer(R.Shape,300))=1 AND R.Name ='St.Lawrence '</pre>
<pre>SELECT Co.Name, Count(Co1.Name) FROM Country Co, Country Co1 WHERE Touch(Co.Shape,Co1.Shape) GROUP BY Co.Name ORDER BY Count(Co1.Name)</pre>
<pre>SELECT C1.Name, R1.Name FROM City C1, River R1 WHERE Distance (C1.Shape,R1.Shape) <= ALL (SELECT Distance(C2.Shape,R1.shape) FROM City C2 WHERE C1.Name <> C2.Name)</pre>

10. Τι είναι το φυσικό μοντέλο σε μία χωρική βάση δεδομένων και τι ιδιότητες και χαρακτηριστικά πρέπει να διαθέτει; Ποιες είναι οι υποθέσεις που γίνονται συχνά σε ένα τέτοιο μοντέλο για τα δεδομένα; Ποια είναι τα πιο κοινά είδη ερωτημάτων σε μία χωρική βάση δεδομένων; Ποιες είναι οι πιο κοινές δομές δεικτοδότησης που χρησιμοποιούνται στα μοντέλα αυτά;

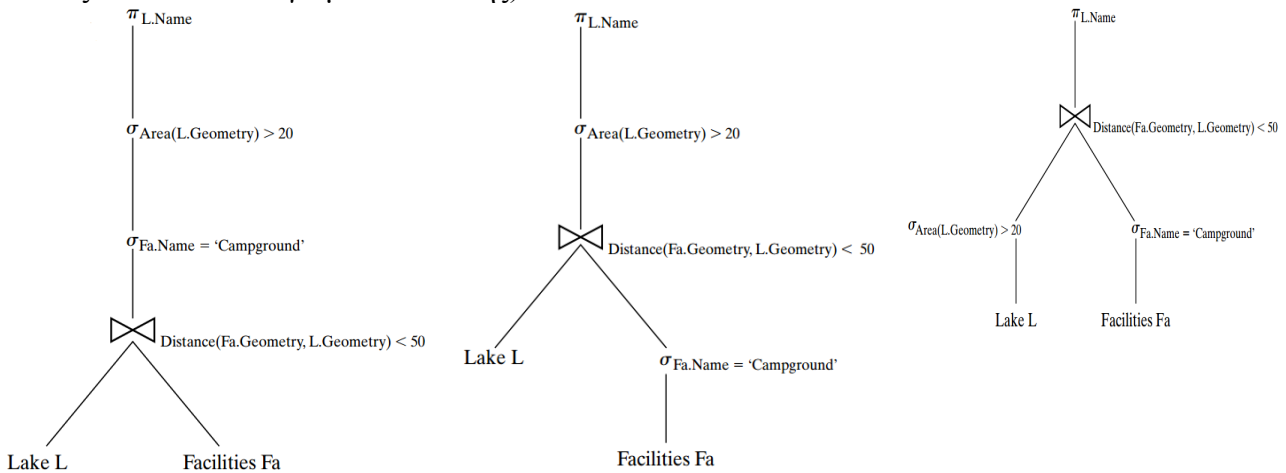
11. Τι είναι οι χωροπληρωτικές καμπύλες Z και Hilbert; Ποιους χώρους μπορούν να δεικτοδοτήσουν; Ποιες είναι οι ομοιότητες και ποιες οι διαφορές τους; Στο παράδειγμα δεξιά ποια είναι η καμπύλη Z και ποια η Hilbert; Ποιες είναι οι τιμές Z και Hilbert για το γκρι χωρικό αντικείμενο αντίστοιχα; Ποια από τις δύο καμπύλες το δεικτοδοτεί πιο αποδοτικά και γιατί;



12. Τι είναι τα R-Trees και ποιες είναι οι βασικές τους ιδιότητες; Σχεδιάστε ένα δικό σας απλό μικρό παράδειγμα. Ποιες βασικές διαφορές έχουν τα R+Trees από τα R-Trees; Πως γίνεται η αναζήτηση ενός αντικειμένου σε ένα R-Tree και πως σε ένα R+Tree;

13. Τι είναι και γιατί καθιερώθηκε το MBR (minimum bounding rectangle) ως μέθοδος προσέγγισης και επεξεργασίας χωρικών αντικειμένων; Δώστε ένα παράδειγμα MBR για ένα αντικείμενο τεθλασμένης γραμμής και για ένα πολύγωνο. Η επεξεργασία ερωτημάτων που αφορά χωρικά αντικείμενα στα οποία υπολογίστηκαν τα MBR τους γίνεται συνήθως με την μέθοδο filtering and refinement. Περιγράψτε αναλυτικά τα βήματα της μεθόδου αυτής κατά τη φάση του filtering και κατά τη φάση του refinement (είτε γενικά με διάγραμμα είτε με δικό σας παράδειγμα).

14. Ποια είναι η διαδικασία της βελτιστοποίησης και επεξεργασίας ερωτημάτων; Ποια βασικά στάδια εφαρμόζει ο βελτιστοποιητής; Το πλάνο εκτέλεσης ενός ερωτήματος δίνεται συνήθως με τη μορφή ενός δέντρου. Ποιο από τα παρακάτω πλάνα είναι πιο αποδοτικό και γιατί; (να εξεταστούν και οι δύο περιπτώσεις: όταν ο υπολογισμός του εμβαδού έχει περισσότερο και όταν έχει λιγότερο κόστος από τον υπολογισμό απόστασης)



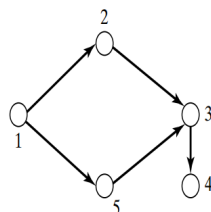
15. Ποια είναι τα βασικά είδη συστημάτων παράλληλων χωρικών βάσεων; Τι διαμοιράζεται το κάθε ένα; Πως γίνεται συνήθως η κατανομή των δεδομένων και της δεικτοδότησής τους στην περίπτωση χρήσης πολλών δίσκων; Ποια είναι τα πλεονεκτήματα ενός τέτοιου συστήματος; Έστω π.χ. μία περιοχή 64 θέσεων η οποία δεικτοδοτείται με την καμπύλη Z και έχουμε σύστημα με οκτώ δίσκους (ids 0 ως 7). Αν η κατανομή γίνεται με βάση την τιμή $Z \bmod 8$ να συμπληρώσετε τον παρακάτω πίνακα δεικτοδότησης στους δίσκους:

42	43	46	47	58	59	62	63
40	41	44	45	56	57	60	61
34	35	38	39	50	51	54	55
32	33	36	37	48	49	52	53
10	11	14	15	26	27	30	31
8	9	12	13	24	25	28	29
2	3	6	7	18	19	22	23
0	1	4	5	16	17	20	21

Z-Curve Method -> disk-id = $Z(x, y) \bmod 8$

16. Τι είναι ένα χωρικό δίκτυο και πως μοντελοποιείται; Αναφέρετε τα βασικά low-level operations στον γράφο ενός δικτύου καθώς και τα δομικά στοιχεία για την εκτέλεση ερωτημάτων. Αναφέρετε ένα παράδειγμα χωρικού δικτύου και διατυπώστε με κείμενο τρία χωρικά ερωτήματα πάνω σε αυτό.

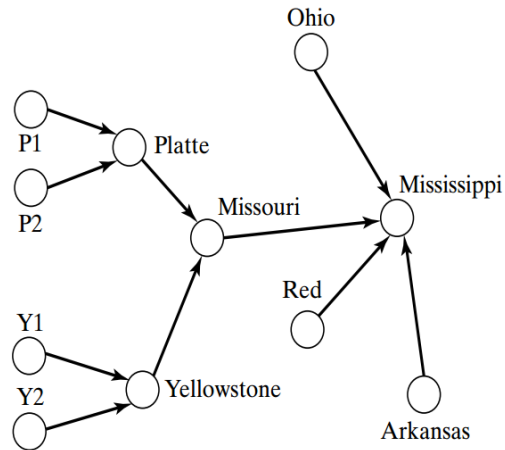
17. Τι είναι η δομή transitive closure T ενός γράφου G και πως υπολογίζεται; Δώστε έναν σαφή ορισμό της. Έστω το χωρικό δίκτυο G του παρακάτω παραδείγματος και ο αντίστοιχος πίνακας των ακμών του. Να κατασκευάσετε το transitive closure T του G και τον πίνακα με τις ακμές του T:



R	
SOURCE	DEST
1	2
1	5
2	3
3	4
5	3

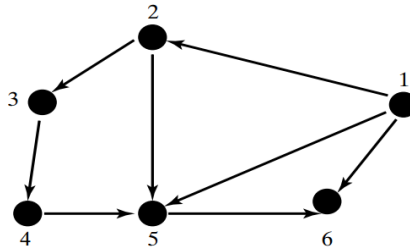
18. Δίνεται ένα SQL ερώτημα connect για το δίκτυο των ποταμών της διπλανής εικόνας. Να περιγράψετε την εκτέλεση του ερωτήματος αυτού και να καταγράψετε τον πίνακα των αποτελεσμάτων του σε κάθε επίπεδο αναδρομής. (1=Mississippi)

SELECT source **FROM** FallsInto
CONNECT BY PRIOR source = dest
START WITH dest =1



19. Ποιες 4 βασικές μέθοδοι/στρατηγικές υπάρχουν για τον υπολογισμό του ελαχίστου μονοπατιού μεταξύ δύο κόμβων σε χωρικά δίκτυα (single pair shortest path) όταν τα δεδομένα του δικτύου βρίσκονται στην μνήμη και ποια όταν τα δεδομένα βρίσκονται εξωτερικά σε δίσκο; Περιγράψτε τα βασικά αλγοριθμικά βήματα της κάθε μεθόδου.

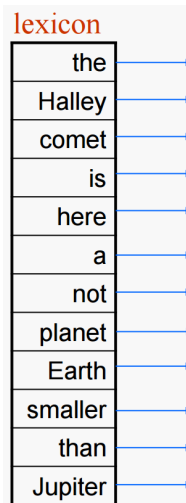
20. Στην περίπτωση που ένα χωρικό δίκτυο δεν χωράει στην μνήμη πως μπορεί αποδοτικά να αποθηκευτεί εξωτερικά σε δίσκο; Γιατί ένα R-Tree δεν είναι αποδοτικό στην περίπτωση αυτή; Τι είναι το CRR, πως υπολογίζεται και γιατί οι μέθοδοι εξωτερικής αποθήκευσης βασίζονται σε αυτό; Για το παρακάτω τμήμα χωρικού δικτύου δίνεται ότι σε έναν τομέα δίσκου είναι αποθηκευμένοι οι κόμβοι 1,2,3 και σε έναν άλλο οι κόμβοι 4,5,6. Να υπολογιστεί η τιμή του CRR για το τμήμα αυτό:



21. Τι είναι η ανάκτηση πληροφορίας (information retrieval) και ποιος ο σκοπός της; Αναφέρετε τα στάδιά της είτε με τη μορφή κειμένου είτε με τη μορφή διαγράμματος. Ποιες είναι οι βασικές διαφορές ενός συστήματος ανάκτησης πληροφορίας και ενός παραδοσιακού συστήματος βάσης δεδομένων;

22. Πως γίνεται η αναπαράσταση ενός κειμένου στην ανάκτηση πληροφορίας; Περιγράψτε τα βήματα της αναπαράστασης είτε με τη μορφή κειμένου είτε με τη μορφή διαγράμματος. Τι είναι το inverted index; Για την παρακάτω συλλογή κειμένων D1,D2,D3 και το λεξικό που έχει δημιουργηθεί δίπλα να συμπληρώσετε την δομή του inverted index:

D1: the Halley comet is here
 D2: a comet is not a planet
 D3: planet Earth is smaller than planet Jupiter



23. Ποια είναι τα βασικά χαρακτηριστικά του Boolean retrieval model; Αναφέρετε τρία παραδείγματα ερωτημάτων για το μοντέλο αυτό. Ποια είναι τα βασικά χαρακτηριστικά του Vector-space model; Πως υπολογίζονται τα βάρη των όρων στα κείμενα; Πως παριστάνονται τα κείμενα στον χώρο και πως υπολογίζεται η ομοιότητά τους; Αναφέρετε ένα παράδειγμα. Πως ορίζονται τα μέτρα αποτελεσματικότητας recall και precision σε ένα σύστημα ανάκτησης πληροφορίας;

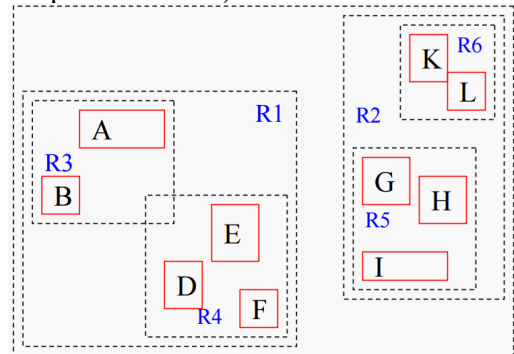
24. Ποια είναι τα βασικά προβλήματα που έχουμε όταν σε ένα σύστημα ανάκτησης πληροφορίας έχουμε πολυμεσικά αντικείμενα; Ποιες είναι οι βασικές ιδιότητες που πρέπει να έχει ένα τέτοιο σύστημα; Γιατί η λεκτική περιγραφή των αντικειμένων και οι τεχνικές κειμένου πάνω στις περιγραφές δεν είναι καλή λύση; Γιατί η υλοποίηση ερωτημάτων περιοχής και κοντινότερου γείτονα στα συστήματα αυτά δεν πρέπει να γίνεται με brute-force μεθόδους; Ποια λύση τελικά προτείνεται για την αποδοτική επεξεργασία ερωτημάτων στα συστήματα αυτά;

25. Ποια είναι τα πιο σημαντικά χαρακτηριστικά σε εικόνες για ένα σύστημα ανάκτησης πληροφορίας εικόνων; Ποια είναι τα δομικά στοιχεία ενός τέτοιου συστήματος; Για τη σύγκριση δύο εικόνων με βάση το χρώμα γίνεται συνήθως η σύγκριση των ιστογραμμάτων τους. Πως κατασκευάζεται αυτό το ιστογράμμα; Αναφέρατε τρεις τουλάχιστον μετρικές απόστασης για δύο ιστογράμματα εικόνων.

26. Τι είναι η υφή (texture) μίας εικόνας; Ποια τα βασικά της χαρακτηριστικά; Πως κατηγοριοποιούνται τα textures; Ποιες είναι οι 3 σημαντικότερες κατηγορίες τους; Με ποιες μορφές εφαρμόζονται στα συστήματα ανάκτησης πληροφορίας εικόνων; Τι είναι η μορφή (shape) μίας εικόνας; Ποια τα βασικά της χαρακτηριστικά; Πως κατηγοριοποιούνται τα shapes; Με ποιες μορφές εφαρμόζονται στα συστήματα ανάκτησης πληροφορίας εικόνων;

27. Ποιες είναι οι ομοιότητες ενός συστήματος ανάκτησης πληροφορίας ήχων και ενός εικόνων; Ποια είναι τα βασικά χαρακτηριστικά ενός ήχου; (περιγράψτε σύντομα το καθένα). Ποια είναι τα σημασιολογικά χαρακτηριστικά ενός ήχου; (περιγράψτε σύντομα το καθένα).

28. Περιγράψτε το R-Tree. Ποια είναι τα είδη των κόμβων που έχει και ποια είναι η δομή τους; Τι ισχύει για το ελάχιστο και μέγιστο πλήθος αντικειμένων μέσα σε έναν κόμβο; Ποιες είναι οι επιμέρους ιδιότητες της δομής του R-Tree και πως συνδέονται οι κόμβοι του; Για το παράδειγμα της εικόνας δίπλα δημιουργείστε την ιεραρχική μορφή του R-Tree και περιγράψτε τα βήματα της αναζήτησης του αντικειμένου E.

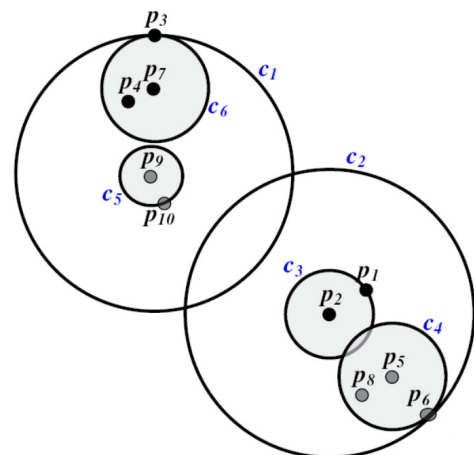


29. Τι είναι το R*-Tree; Ποια είναι τα κριτήρια βελτιστοποίησης που εφαρμόζει και σε τι διαφέρει από το R-Tree; Γιατί είναι πιο αποδοτικό από το R-Tree; Ποια είναι η πολιτική διαχωρισμού των κόμβων (split) στο R*-Tree; Σε τι βοηθάει η μέθοδος forced reinsert;

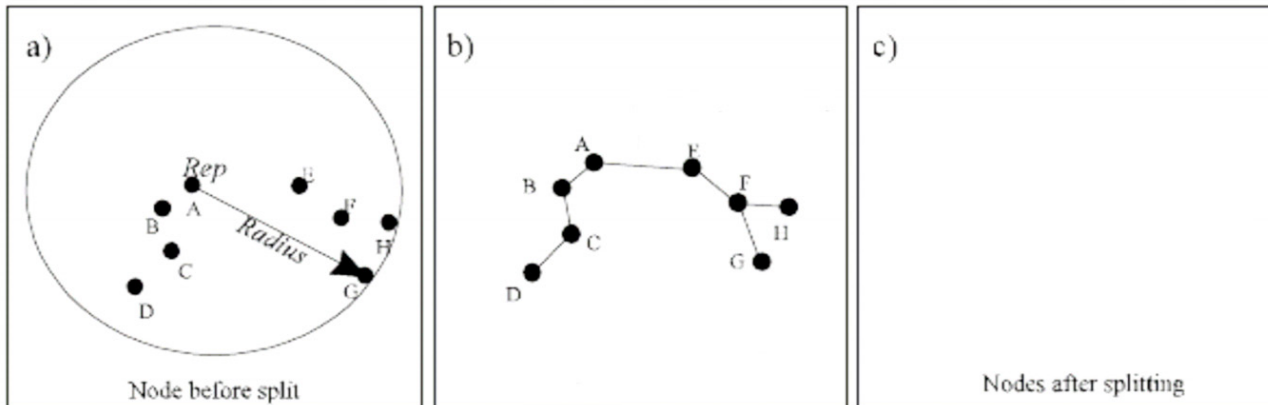
30. Τι είναι το X-Tree; Ποια προβλήματα των ιεραρχικών δομών αντιμετωπίζει, πως τα επιλύει και σε τι διαφέρει από τα R-Tree, R*-Tree; Πόσα είδη κόμβων υπάρχουν σε ένα X-Tree; Ποιες γενικές στρατηγικές εφαρμόζει για τον διαχωρισμό κόμβων (split);

31. Ποια είναι η βασική ιδέα της μεθόδου της πυραμίδας (Pyramid-Technique); Ποια δομή δεικτοδότησης χρησιμοποιεί; Ποια είναι τα βασικά στάδια της αναζήτησης ενός αντικειμένου q (point query) στην δομή της πυραμίδας; Πως αντιμετωπίζεται το πρόβλημα δύο ή περισσότερα αντικείμενα να έχουν το ίδιο pyramid value; Σε ποιες περιπτώσεις η μέθοδος της πυραμίδας είναι πιο αποδοτική;

32. Περιγράψτε το M-Tree και τις βασικές ιδιότητές του. Ποια είναι τα είδη των κόμβων που έχει και ποια είναι η δομή τους; Πως συνδέονται οι κόμβοι του; Το παράδειγμα της εικόνας δίπλα απεικονίζει ένα τμήμα ενός M-Tree σε Ευκλείδειο μετρικό χώρο. Δημιουργείστε την ιεραρχική μορφή του M-Tree και περιγράψτε τα βήματα της αναζήτησης του αντικειμένου p8.



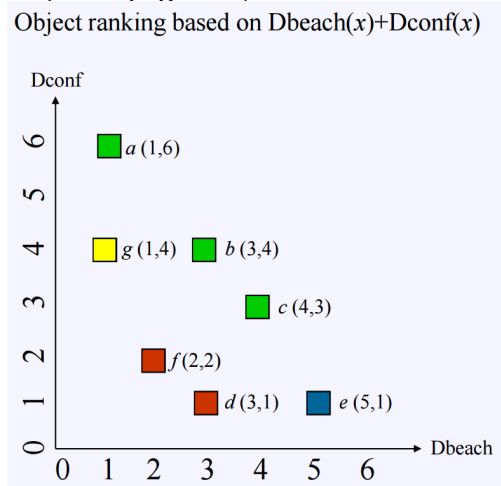
33. Τι είναι το Slim-Tree; Ποιες είναι οι βασικές διαφορές του από το M-Tree; Με ποια μέθοδο το Slim-Tree επιτυγχάνει πιο αποδοτική διάσπαση κόμβων (split); Εφαρμόστε την μέθοδο αυτή στο παράδειγμα της εικόνας και σχεδιάστε τους δύο κόμβους που θα δημιουργηθούν μετά το split:



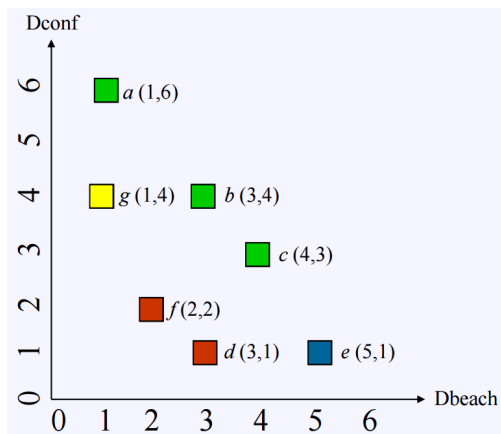
Τι είναι και τι μετράνε οι παράγοντες fat-factor και bloat-factor του Slim-Tree;

34. Τι είναι η ανάδραση σχετικότητας (relevance feedback); Ποιος είναι ο σκοπός και η στρατηγική της; Ποιες είναι οι τρεις βασικές μορφές της; Με ποιους τρόπους ο χρήστης μπορεί να ανατροφοδοτήσει ένα σύστημα που διαθέτει ανάδραση σχετικότητας; Κάθε ανατροφοδότηση που γίνεται για το ίδιο ερώτημα οδηγεί πάντα σε καλύτερα αποτελέσματα ή όχι και γιατί;

35. Τι είναι ένα ερώτημα top-k και ποια τα βασικά του χαρακτηριστικά; Ποια ιδιότητα πρέπει να ικανοποιεί η συνάρτηση βαθμολόγησής του ώστε να εξασφαλίζεται η ολική διάταξη των αντικειμένων της βάσης; Αναφέρετε τρία παραδείγματα τέτοιων συναρτήσεων. Ποια είναι τα μειονεκτήματα ενός ερωτήματος top-k; Στο παράδειγμα δίπλα δίνονται 7 αντικείμενα και αναφέρεται η συνάρτηση βαθμολόγησής τους. Να βρείτε τα σκορ για το κάθε αντικείμενο και να απαντήσετε σε ένα top-3 ερώτημα για τα αντικείμενα αυτά (το μικρότερο σκορ είναι και καλύτερο).



36. Τι είναι ένα ερώτημα skyline και ποια τα βασικά του χαρακτηριστικά; Πως ορίζεται η ιδιότητα της κυριαρχίας; Ποια είναι τα πλεονεκτήματα ενός ερωτήματος skyline σε σχέση με ένα ερώτημα top-k; Στο παράδειγμα δίπλα δίνονται 7 αντικείμενα. Να καταγράψετε όλες τις σχέσεις κυριαρχίας (ποιο αντικείμενο κυριαρχεί σε ποιο) και να απαντήσετε σε ένα skyline ερώτημα για τα αντικείμενα αυτά (και στις δύο διαστάσεις η μικρότερη τιμή είναι και καλύτερη).



37. Περιγράψτε τους naïve αλγόριθμους για την επεξεργασία των ερωτημάτων top-k και skyline. Ποια είναι η πολυπλοκότητά τους και ποια τα μειονεκτημά τους; Ποιες στρατηγικές εφαρμόζουν οι σύγχρονοι αλγόριθμοι για να μειώσουν την πολυπλοκότητα; Αναφέρετε ορισμένους σύγχρονους αλγόριθμους/μεθόδους που αντιμετωπίζουν αποδοτικά την επεξεργασία ερωτημάτων top-k και skyline.

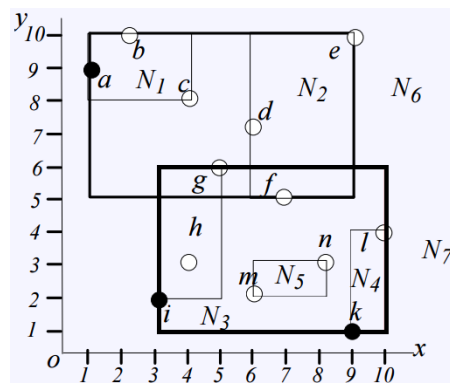
38. Περιγράψτε συνοπτικά τον αλγόριθμο του Fagin (FA) για ερωτήματα top-k. Εφαρμόστε τον αλγόριθμο αυτό για ένα top-2 ερώτημα στο παράδειγμα που δίνεται δίπλα, το οποίο απεικονίζει τα σκορ 5 πολυμεσικών αντικειμένων (O_0, O_1, O_2, O_3, O_4) σε κάθε ένα από τα 5 χαρακτηριστικά τους (a_1, a_2, a_3, a_4, a_5). Το συνολικό σκορ ενός αντικειμένου προκύπτει απλά από το άθροισμα των σκορ που έχει στα επιμέρους χαρακτηριστικά του.

a1	a2	a3	a4	a5
$O_3, 99$	$O_1, 91$	$O_1, 92$	$O_3, 74$	$O_3, 67$
$O_1, 66$	$O_3, 90$	$O_3, 75$	$O_1, 56$	$O_4, 67$
$O_0, 63$	$O_0, 61$	$O_4, 70$	$O_0, 56$	$O_1, 58$
$O_2, 48$	$O_4, 07$	$O_2, 16$	$O_2, 28$	$O_2, 54$
$O_4, 44$	$O_2, 01$	$O_0, 01$	$O_4, 19$	$O_0, 35$

39. Περιγράψτε συνοπτικά τον αλγόριθμο Threshold Algorithm (TA) για ερωτήματα top-k. Εφαρμόστε τον αλγόριθμο αυτό για ένα top-2 ερώτημα στο παράδειγμα που δίνεται δίπλα, το οποίο απεικονίζει τα σκορ 5 πολυμεσικών αντικειμένων (O_0, O_1, O_2, O_3, O_4) σε κάθε ένα από τα 5 χαρακτηριστικά τους (a_1, a_2, a_3, a_4, a_5). Το συνολικό σκορ ενός αντικειμένου προκύπτει απλά από το άθροισμα των σκορ που έχει στα επιμέρους χαρακτηριστικά του.

a1	a2	a3	a4	a5
$O_3, 99$	$O_1, 91$	$O_1, 92$	$O_3, 74$	$O_3, 67$
$O_1, 66$	$O_3, 90$	$O_3, 75$	$O_1, 56$	$O_4, 67$
$O_0, 63$	$O_0, 61$	$O_4, 70$	$O_0, 56$	$O_1, 58$
$O_2, 48$	$O_4, 07$	$O_2, 16$	$O_2, 28$	$O_2, 54$
$O_4, 44$	$O_2, 01$	$O_0, 01$	$O_4, 19$	$O_0, 35$

40. Περιγράψτε συνοπτικά τον αλγόριθμο BBS (Branch-and-Bound Skyline) για την επεξεργασία των ερωτημάτων skyline. Ποια δομή δεικτοδότησης, ποια μετρική απόστασης και ποιες βοηθητικές δομές χρησιμοποιεί; Γιατί είναι πιο αποδοτικός από τους παλιούς αλγορίθμους; Εφαρμόζοντας τον αλγόριθμο BBS για το παράδειγμα της διπλανής εικόνας καταγράψτε τα περιεχόμενα των βοηθητικών δομών σε κάθε στάδιο του αλγορίθμου.



41. Ποιες είναι οι γενικές κατηγορίες των μεθόδων επιλογής χαρακτηριστικών και μείωσης διαστάσεων; Στην κατηγορία των μεθόδων φιλτραρίσματος αναφέρετε παραδείγματα στατιστικών μέτρων που χρησιμοποιούνται. Ποιες γενικές στρατηγικές χρησιμοποιούνται για την επιλογή του υποχώρου των διαστάσεων και γιατί η εξαντλητική αναζήτηση δεν είναι εφικτή;

42. Τι είναι το πρόβλημα της «κατάρας των διαστάσεων»; Ποιο είναι το πρόβλημα του «κενού χώρου» και γιατί αυξάνεται καθώς αυξάνονται οι διαστάσεις; Αναφέρετε τις αναλυτικές ονομασίες 5 βασικών μεθόδων μείωσης των διαστάσεων. Ποια βασικά δομικά στοιχεία της Άλγεβρας εκμεταλλεύονται οι μέθοδοι αυτές; Σχεδιάστε ένα απλό παράδειγμα σημείων 2 διαστάσεων στο οποίο να χαράξετε (στο περίπου) τον πρώτο άξονα της μεθόδου PCA. Σχεδιάστε ένα απλό παράδειγμα με 2 κλάσεις σημείων 2 διαστάσεων στο οποίο να χαράξετε (στο περίπου) τον άξονα της μεθόδου LDA πάνω στον οποίο οι προβολές των σημείων των κλάσεων να διαχωρίζονται όσο το δυνατόν περισσότερο.