



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ II

ΚΑΘΗΓΗΤΗΣ : Β. ΜΕΓΑΛΟΟΙΚΟΝΟΜΟΥ

ΟΝΟΜΑΤΕΠΩΝΥΜΟ:.....

ΑΜ:.....ΕΤΟΣ:.....

Θέμα 1 (15%)

Εξηγήστε γιατί οι ημιενώσεις (semijoins) θεωρούνται χρήσιμες στη βελτιστοποίηση επερωτήσεων σε κατανεμημένες βάσεις δεδομένων.

Θέμα 2 (15%)

Δώστε τον ορισμό του lower bounding lemma και αναφέρετε πως βοηθάει στην αναζήτηση ομοιότητας (βάση της προσέγγισης GEMINI- Generic Multimedia Indexing) πολυμεσικών δεδομένων.

Θέμα 3 (15%)

Αναφέρετε επιγραμματικά τις διαφορές SQL και NoSQL database systems.

Θέμα 4: Βελτιστοποίηση Επερωτήσεων (30%)

Έστω η πράξη επιλογής:

$\sigma_{\text{publisher}='PublishEnterprise'}(\text{books}),$

όπου *books*(ISBN, title, price, publisher, writer, description). Για τον υπολογισμό της επιλεκτικότητας (selection cardinality) θεωρείστε ότι έχουμε ομοιόμορφη κατανομή των τιμών.

Θεωρείστε:

$N_{\text{books}} = 20.480$ εγγραφές,

$F_{\text{books}} = 40$ εγγραφές ανά μπλοκ,

$V(\text{price}, \text{books}) = 60$ διακριτές εγγραφές,

$V(\text{writer}, \text{books}) = 80$ διακριτές εγγραφές,

$V(\text{publisher}, \text{books}) = 128$ διακριτές εγγραφές.

Βρείτε τον αριθμό προσπελάσεων σε μπλοκς:

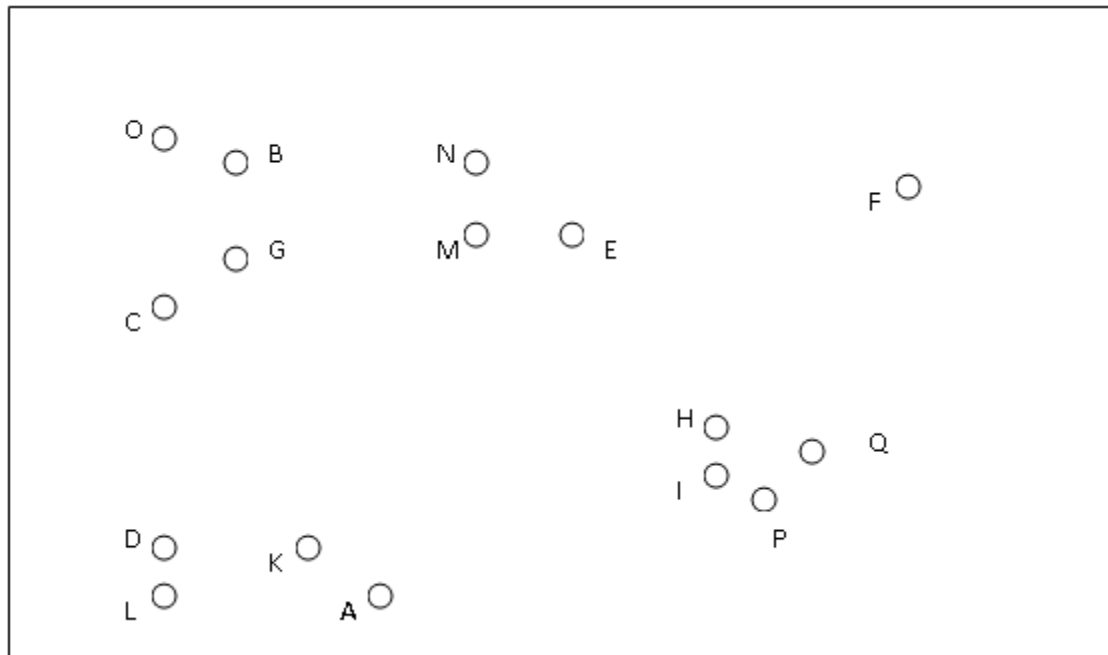
1. μιας ακολουθιακής σάρωσης στην μέση περίπτωση.

2. όταν πρόκειται να χρησιμοποιήσουμε δυαδική αναζήτηση, αν θεωρήσουμε ότι οι εγγραφές είναι ταξινομημένες ως προς το πεδίο publisher και ότι τα κλειδιά είναι ταξινομημένα και βρίσκονται σε διαδοχικά μπλοκς.

Θέμα 5: Χωρικές Δομές Δεδομένων – SAMs (25%)

A) Βάσει του παρακάτω σχήματος σχεδιάστε (επάνω στην κόλλα των θεμάτων) τα βέλτιστα ελάχιστα περιοριστικά ορθογώνια (*optimal minimum bounding rectangles*) που αντιστοιχούν σε R-tree με **fan-out** = 4 προκειμένου να δεικτοδοτηθούν όλα τα δεδομένα-σημεία του σχήματος. Να δικαιολογήσετε την απάντησή σας.

B) Σχεδιάστε το R-tree που προκύπτει.



Καλή επιτυχία!