

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΑΣΚΗΣΗ ΔΕΥΤΕΡΗ

ΕΠΙΣΤΡΕΦΕΤΑΙ

10.1.2013

Η άσκηση αυτή είναι **ατομική**, δε θα γίνουν δεκτές ασκήσεις από ομάδες φοιτητών.

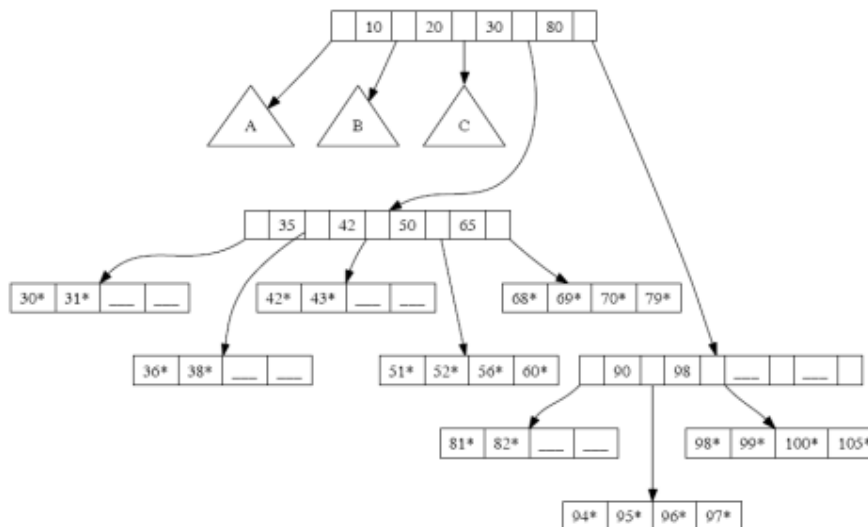
ΜΑΘΗΜΑ
ΑΚΑΔ. ΕΤΟΣ
ΔΙΔΑΣΚΟΝΤΕΣ

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ
 2012-13
 Ιωάννης Βασιλείου *Καθηγητής*, Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
 Τιμολέων Σελλής *Καθηγητής*, Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Ερώτημα 1.

Θεωρείστε το B+ δέντρο της Εικόνας 1.

- (i) Σε ποιο υπο-δέντρο θα αναζητήσετε την τιμή 9 και σε ποιο την τιμή 25;
 - (ii) Εισάγετε την τιμή 101 και δώστε το αποτέλεσμα της εισαγωγής.
 - (iii) Διαγράψτε την τιμή 43 και δώστε το αποτέλεσμα της διαγραφής.
 - (iv) Εισάγετε την τιμή 80 και δώστε το αποτέλεσμα της εισαγωγής.
 - (v) Διαγράψτε την τιμή 30 και δώστε το αποτέλεσμα της διαγραφής.
- Οι παραπάνω εισαγωγές και διαγραφές να γίνουν όλες στο αρχικό ευρετήριο της Εικόνας 1.



Εικόνα 1. B+ δέντρο για το Ερώτημα 1. Ο μέγιστος αριθμός δεικτών για τους εσωτερικούς κόμβους είναι 5 και για τα φύλλα 4. Τα φύλλα είναι διπλά συνδεδεμένα μεταξύ τους (δε φαίνεται στην εικόνα). Τα A, B, C είναι υπο-δέντρα στα οποία δείχνουν οι αντίστοιχοι δείκτες.

Ερώτημα 2.

Θεωρείστε στατικό πίνακα κατακερματισμού (οργάνωση αρχείου) όπου κάθε κάδος χωρά μέχρι 2 εγγραφές. Η συνάρτηση κατακερματισμού είναι $h(k) = k \bmod 3$. Ως συνήθως ένας κάδος έχει την χωρητικότητα ενός μπλοκ στον δίσκο.

- (i) Δείξτε τα περιεχόμενα του πίνακα κατακερματισμού σε κάθε βήμα κατά τη εισαγωγή των ακόλουθων κλειδιών: 27, 5, 18, 30, 10, 32, 38. Για τη διαχείριση υπερχειλίσεων χρησιμοποιείστε αλυσιδωτή σύνδεση (chaining).

- (ii) Πόσο κοστίζει (σε I/O) η προσπέλαση της εγγραφής με κλειδί 30 και πόσο η προσπέλαση της εγγραφής με κλειδί 32;
- (iii) Ποιος ο μέσος αριθμός I/O μιας αποτυχημένης αναζήτησης;

Ερώτημα 3.

Θεωρείστε μια σχέση R που έχει 500.000 εγγραφές (πλειάδες) και είναι αποθηκευμένη σε ένα αρχείο σφρού. Κατασκευάζουμε ένα ευρετήριο για ένα γνώρισμα (πεδίο) A της σχέσης R, που δεν είναι κλειδί για τη σχέση. Έστω ότι υπάρχουν 1.000 διαφορετικές τιμές του A και οι εγγραφές είναι ομοιόμορφα κατανομημένες ως προς αυτές. Το μέγεθος του πεδίου A είναι 8 bytes και των δεικτών 16 bytes (όλων των ειδών δεικτών). Θεωρείστε μέγεθος block 2048 bytes.

- (α) Έστω ότι το ευρετήριο είναι ένα δευτερεύον ευρετήριο. Θεωρείστε ότι χρησιμοποιείται ένα επιπλέον επίπεδο έμμεσων δεικτών, εφόσον το κλειδί αναζήτησης (A) δεν είναι υποψήφιο κλειδί (βλέπε σελ. 483-485 από το βιβλίο των Silberschatz, Korth, Sudarshan).
 - (i) Δώστε το συνολικό μέγεθος του αρχείου ευρετηρίου (συμπεριλαμβανομένου και του επιπλέον επιπέδου) σε αριθμό blocks.
 - (ii) Υπολογίστε το κόστος σε αριθμό I/O (δηλαδή, σε αριθμό προσπελάσεων σελίδων) της αναζήτησης $A = a$, όπου a μια τιμή από το πεδίο ορισμού του A.
 - (iii) Έστω ότι εισάγουμε μια νέα πλειάδα στη σχέση R. Εξηγήστε πως αλλάζει το ευρετήριο, όταν η εγγραφή έχει μια τιμή στο A η οποία (1) δεν εμφανίζεται ήδη σε άλλη πλειάδα και (2) εμφανίζεται ήδη σε άλλη πλειάδα. Δώστε μια εκτίμηση του κόστους (πολυπλοκότητας) αυτών των αλλαγών (σε αριθμό I/O) για κάθε μία από τις δύο περιπτώσεις.
- (β) Υποθέστε τώρα ότι κατασκευάζουμε ένα B+-δέντρο. Θεωρείστε ότι κάθε τιμή κλειδιού αναζήτησης αποθηκεύεται μόνο μια φορά σε κάποιο από τα φύλλα του δέντρου και για κάθε τιμή διατηρείται ένας “κάδος” από δείκτες εγγραφών με αυτήν την τιμή κλειδιού αναζήτησης. Αυτοί οι κάδοι αποθηκεύονται σε ξεχωριστά block, δημιουργώντας ένα επιπλέον επίπεδο. Οι δείκτες των φύλλων δείχνουν στο επιπλέον επίπεδο (βλέπε σελ. 497-499 από το βιβλίο των Silberschatz, Korth, Sudarshan).
Επαναλάβετε τα υπο-ερωτήματα (i)-(iii) του ερωτήματος (α) θεωρώντας ότι το δέντρο είναι όσο το δυνατόν πιο γεμάτο.
- (γ) Υποθέστε τώρα ότι κατασκευάζουμε ένα δευτερεύον ευρετήριο όπου δεν υπάρχει το επιπλέον επίπεδο. Επαναλάβετε τα υπο-ερωτήματα (i) και (ii) του ερωτήματος (α). (βλέπε σελ. 483-485 από το βιβλίο των Silberschatz, Korth, Sudarshan και διαφάνειες διαλέξεων).
- (δ) Υποθέστε μια αναζήτηση $A > a$, όπου a μια τιμή από το πεδίο ορισμού του A, η οποία ικανοποιείται από το 10% των εγγραφών της σχέσης. Εξηγήστε πως μπορείτε να χρησιμοποιήσετε τα ευρετήρια των ερωτημάτων (α)-(γ) για αυτήν την αναζήτηση και δώστε μια εκτίμηση του κόστους της σε κάθε περίπτωση.