

3ο Σύνολο Ασκήσεων

Ημερομηνία Παράδοσης: 16/1/2008, πριν το μάθημα.

Θεματική Ενότητα: Αποθήκευση. Ευρετήρια. Εισαγωγή στην Επεξεργασία Ερωτήσεων.

Για γενικές οδηγίες σχετικά με τις ασκήσεις, συμβουλευτείτε την ιστοσελίδα του μαθήματος.

Λύστε οποιοσδήποτε 6 από τις παρακάτω ασκήσεις. Είναι όλες βαθμολογικά ισοδύναμες

Άσκηση 1. Θεωρείστε επεκτατό κατακερματισμό όπου κάθε κάδος χωρά μέχρι 3 εγγραφές. Χρησιμοποιείστε τα πρώτα bits.

- (α) Δώστε το ευρετήριο μετά την εισαγωγή των παρακάτω κλειδιών: 1, 3, 6, 14, 15, 17, 21, 23, 24, 25, 29, 31. Προσδιορίστε το ολικό και το τοπικό βάθος.
- (β) Δώστε το ευρετήριο μετά την εισαγωγή των παρακάτω κλειδιών: 17, 15, 14, 6, 3, 1, 31, 29, 25, 24, 23, 21. Προσδιορίστε το ολικό και το τοπικό βάθος.

Άσκηση 2. Θεωρείστε γραμμικό κατακερματισμό όπου κάθε κάδος χωρά μέχρι 2 εγγραφές και ως αρχική συνάρτηση κατακερματισμού τη συνάρτηση $h(x) = x \bmod 2$.

- (α) Δώστε το ευρετήριο μετά την εισαγωγή των παρακάτω κλειδιών: 32, 9, 31, 25, 7, 35.
- (β) Περιγράψτε έναν αλγόριθμο για τη μείωση του μεγέθους του ευρετηρίου στην περίπτωση γραμμικού κατακερματισμού. Εξηγήστε πότε αυτό είναι δυνατόν να γίνει και με ποιον τρόπο. Δώστε ένα παράδειγμα χρησιμοποιώντας τα δεδομένα του προηγούμενου ερωτήματος (ερώτημα (α)).

Άσκηση 3. Θεωρείστε ένα B-δέντρο τάξεως 3 αρχικά άδειου.

- (α) Θεωρείστε ότι εισάγουμε τιμές αναζήτησης από το 1 έως το 18 με τη σειρά. Δώστε το δέντρο που προκύπτει. Ποιο είναι το μέγεθός του σε blocks;
- (β) Δώστε το δέντρο που προκύπτει αν διαγράψουμε από το δέντρο του ερωτήματος (α) όλους τους πρώτους αριθμούς (από τον μικρότερο στο μεγαλύτερο).

Άσκηση 4. Επαναλάβετε τα ερωτήματα της Άσκησης 3 για ένα B+-δέντρο τάξεως 3 για τα φύλλα και τους εσωτερικούς κόμβους. Θεωρείστε ότι τιμή ίση με την τιμή αναζήτησης σε ένα κόμβο αποθηκεύεται στο υποδέντρο στα δεξιά του.

Άσκηση 5. Ποιο είναι το μεγαλύτερο μέγεθος αρχείου δεδομένων (σε αριθμό blocks) που μπορεί να δεικτοδοτήσει ένα B+-δέντρο τάξης $p = 28$ και $p_{leaf} = 45$, στις παρακάτω περιπτώσεις. Θεωρείστε ότι ο παράγοντας ομαδοποίησης (σελιδοποίησης) του αρχείου είναι 35. Κάντε όποιες άλλες υποθέσεις θεωρείστε αναγκαίες.

- (α) το αρχείο δεδομένων είναι ταξινομημένο ως προς το πεδίο ευρετηριοποίησης και το πεδίο ευρετηριοποίησης είναι κλειδί

- (β) το αρχείο δεν είναι ταξινομημένο ως προς το πεδίο ευρετηριοποίησης και το πεδίο ευρετηριοποίησης είναι κλειδί.
- (γ) το αρχείο είναι ταξινομημένο ως προς το πεδίο ευρετηριοποίησης αλλά το πεδίο ευρετηριοποίησης δεν είναι κλειδί και υπάρχουν 100 εγγραφές ανά τιμή του πεδίου.
- (δ) το αρχείο δεν είναι ταξινομημένο ως προς το πεδίο ευρετηριοποίησης, το πεδίο ευρετηριοποίησης δεν είναι κλειδί και υπάρχουν 100 εγγραφές ανά τιμή του πεδίου.

Άσκηση 6 Επαναλάβετε το ερώτημα (α) της Άσκησης 5 για B-δέντρο τάξης 35.

Άσκηση 7. Άσκηση 14.20 του βιβλίου (σελ 602).

Άσκηση 8. Θεωρείστε ότι θέλουμε να υπολογίσουμε μια συνθήκη διαστήματος (δηλαδή, να βρούμε όλες τις εγγραφές με τιμή για το πεδίο ευρετηριοποίησης x , τέτοιες ώστε $a \leq x \leq b$, για κάποιες τιμές a και b). Υποθέστε ότι τη συνθήκη την ικανοποιούν το 10% των εγγραφών του αρχείου δεδομένων. Υπολογίστε το κόστος (σε αριθμό blocks) για το αρχείο δεδομένων και το ευρετήριο του ερωτήματος (β) της Άσκησης 5.

Άσκηση 9. Θεωρείστε την πράξη $R \text{ Join}_{R.A=S.B} S$. Υποθέστε ότι: το μέγεθος του block είναι 1024 bytes, η σχέση R έχει 7.000 πλειάδες και κάθε πλειάδα έχει μέγεθος 90 bytes, η σχέση S έχει 350.000 πλειάδες και κάθε πλειάδα έχει μέγεθος 100 bytes. Το πεδίο (γνώρισμα) A είναι κλειδί για την R και το πεδίο (γνώρισμα) B είναι κλειδί για την S . Θεωρείστε μη εκτεινόμενη αποθήκευση. Τα γνωρίσματα A και B έχουν το καθένα μέγεθος 8 bytes. Υποθέστε ότι η σχέση S είναι αποθηκευμένη σε αρχείο διατεταγμένο ως προς το B . Δεν υπάρχουν ευρετήρια ούτε για το A , ούτε για το B . Θεωρείστε ότι υπάρχουν διαθέσιμοι 5 καταχωρητές στη μνήμη μεγέθους ενός block ο καθένας.

- (α) Υπολογίστε το μεγαλύτερο δυνατό μέγεθος του αποτελέσματος της συνένωσης (σε blocks).
- (β) Περιγράψτε έναν αλγόριθμο υπολογισμού της συνένωσης με την σχέση S στον εσωτερικό βρόγχο. Ο αλγόριθμος πρέπει να χρησιμοποιεί αποδοτικά τους καταχωρητές και το γεγονός ότι η S είναι αποθηκευμένη σε ταξινομημένο αρχείο. Υπολογίστε το κόστος του (σε blocks).