

### 3<sup>ο</sup> Σύνολο Ασκήσεων

Ημερομηνία Παράδοσης: Πέμπτη 14 Ιανουαρίου, μέχρι τις 14μμ (στο B15)  
Θεματική Ενότητα: Οργάνωση Αρχείων. Ευρετήρια.

Για γενικές οδηγίες σχετικά με τις ασκήσεις, συμβουλευτείτε την ιστοσελίδα του μαθήματος.  
Εξηγήστε τις απαντήσεις σας με σαφήνεια. Αν κάνετε οποιαδήποτε υπόθεση, καταγράψτε την.

#### Άσκηση 1

(α) Θεωρείστε ένα ευρετήριο επεκτατού κατακερματισμού.

(i) Έστω ότι κάθε κάδος (bucket/block) μπορεί να χωρέσει έως 3 τιμές. Θεωρείστε ότι κάποια στιγμή ο κατάλογος του ευρετηρίου έχει 1024 εγγραφές (θέσεις). Ποιος είναι ο μεγαλύτερος αριθμός εγγραφών (θέσεων) του καταλόγου που μπορεί να δείχνουν στον ίδιο κάδο.

(ii) Θεωρείστε ότι κάποια στιγμή το ευρετήριο έχει 200 κάδους. Ποιο είναι το μέγεθος του μικρότερου δυνατού καταλόγου;

(β) Θεωρείστε γραμμικό κατακερματισμό, αρχικά άδειο. Εισάγετε με αυτή τη σειρά τις τιμές: 0, 1, 2, 3, 5, 4, 24, 16, 9. Θεωρείστε τη συνάρτηση  $h_0(k) = k \bmod 2$ , και ότι κάθε κάδος χωρά 2 τιμές. Δείξτε το ευρετήριο μετά την εισαγωγή κάθε τιμής.

#### Άσκηση 2

Ποιο είναι το μικρότερο και ποιο το μεγαλύτερο (σε αριθμό κόμβων) B-δέντρο τάξης 4 που μπορεί να δεικτοδοτήσει 100 τιμές δεδομένων.

#### Άσκηση 3

Θεωρείστε ένα B+-δέντρο τάξης  $p = 25$ ,  $p_{leaf} = 40$  και ύψους 3.

(α) Υπολογίστε το μεγαλύτερο αριθμό τιμών δεδομένων που μπορεί να ευρετηριοποιησει. Δώστε το μέγεθος του δέντρου (σε αριθμό blocks) σε αυτήν την περίπτωση.

(β) Εξηγήστε ποιο είναι το μεγαλύτερο μέγεθος αρχείου δεδομένων (σε αριθμό blocks) που μπορεί ένα τέτοιο ευρετήριο να δεικτοδοτήσει για κάθε μία από τις παρακάτω περιπτώσεις. Υποθέστε ότι ο παράγοντας ομαδοποίησης (σελιδοποίησης) για το αρχείο δεδομένων είναι 30. Κάντε όποιες άλλες υποθέσεις θεωρείστε αναγκαίες. (Υπόδειξη: σκεφτείτε ποιες τιμές του πεδίου ευρετηριοποίησης θα πρέπει να εισάγετε στο ευρετήριο σε κάθε περίπτωση).

(i) Το αρχείο δεδομένων είναι αποθηκευμένο ως αρχείο σωρού και το πεδίο ευρετηριοποίησης είναι κλειδί.

(ii) Το αρχείο δεδομένων είναι αποθηκευμένο ως αρχείο σωρού και το πεδίο ευρετηριοποίησης δεν είναι κλειδί. Έστω ότι στο αρχείο δεδομένων, υπάρχουν 150 εγγραφές για κάθε διαφορετική τιμή του πεδίου.

(iii) Το αρχείο δεδομένων είναι διατεταγμένο ως προς το πεδίο ευρετηριοποίησης και το πεδίο ευρετηριοποίησης είναι κλειδί.

(iv) Το αρχείο δεδομένων είναι διατεταγμένο ως προς το πεδίο ευρετηριοποίησης και το πεδίο ευρετηριοποίησης δεν είναι κλειδί. Έστω ότι στο αρχείο δεδομένων, υπάρχουν 150 εγγραφές για κάθε διαφορετική τιμή του πεδίου.

(γ) Θεωρείστε την περίπτωση (i) και (iv) του ερωτήματος (β). Για κάθε μία από τις δύο αυτές περιπτώσεις, δώστε το I/O κόστος της αναζήτησης (σε αριθμό blocks) αν (i) δε χρησιμοποιήσουμε το B+ ευρετήριο και (ii) αν το χρησιμοποιήσουμε.

(δ) Εξηγήστε πως μπορείτε να χρησιμοποιήσετε το B+ ευρετήριο στην περίπτωση (i) και (iii) του ερωτήματος (β) για να απαντήσετε σε μια ερώτηση αναζήτησης των τιμών που ικανοποιούν μια συνθήκη σύγκρισης:  $A > a$ , όπου  $A$  το πεδίο ευρετηριοποίησης και  $a$  μια τιμή του πεδίου τιμών του  $A$ .