

Εργασία 4

Οι ασκήσεις προέρχονται από το βιβλίο Θεμελιώδεις Αρχές Συστημάτων Βάσεων Δεδομένων έκδοση 5^η των Elmasri και Navathe, που είναι ένα από τα προτεινόμενα συγγράμματα στην πλατφόρμα Εύδοξος

Άσκηση 1

13.27. Ένα αρχείο ΠΡΟΪΟΝ με το Π# ως κλειδί κατακερματισμού περιλαμβάνει εγγραφές με τις ακόλουθες τιμές Π#: 2369, 3760, 4692, 4871, 5659, 1821, 1074, 7115, 1620, 2428, 3943, 4750, 6975, 4981, 9208. Τα αρχείο χρησιμοποιεί οκτώ κάδους, αριθμημένους από 0 ως 7. Κάθε κάδος καταλαμβάνει ένα μπλοκ δίσκου και χωρά δύο εγγραφές. Φορτώστε αυτές τις εγγραφές στο αρχείο με τη σειρά που δίνονται, χρησιμοποιώντας τη συνάρτηση κατακερματισμού $h(K) = K \bmod 8$. Υπολογίστε το μέσο αριθμό προσπελάσεων για τυχαία ανάκτηση όταν δίνεται το Π#.

Άσκηση 2

13.28. Φορτώστε τις εγγραφές της Άσκησης 13.27 σε επεκτάσιμα αρχεία κατακερματισμού με χρήση i) δυναμικού κατακερματισμού και ii) επεκτατού κατακερματισμού. Δείξτε τη δομή του καταλόγου σε κάθε βήμα. Για τον επεκτατό κατακερματισμό, δείξτε το ολικό και τα τοπικά βάθη σε κάθε στάδιο. Χρησιμοποιήστε ως συνάρτηση κατακερματισμού την $h(K) = K \bmod 128$.

Άσκηση 3

13.29. Φορτώστε τις εγγραφές της Άσκησης 13.27 σε ένα επεκτάσιμο αρχείο κατακερματισμού, χρησιμοποιώντας γραμμικό κατακερματισμό. Αρχίστε με ένα μπλοκ δίσκου, χρησιμοποιώντας τη συνάρτηση κατακερματισμού $h_0 = K \bmod 2^0$, και δείξτε πώς μεγαλώνει το αρχείο και πώς αλλάζουν οι συναρτήσεις κατακερματισμού καθώς εισάγονται εγγραφές. Υποθέστε ότι τα μπλοκ διασπώνται με κάθε υπερχείλιση και δείξτε την τιμή του n σε κάθε φάση.

Άσκηση 4

- 13.26. Υποθέστε ότι μια μονάδα δίσκου έχει τις ακόλουθες παραμέτρους: χρόνος εντοπισμού $s = 20$ msec, καθυστέρηση περιστροφής $rd = 10$ msec, ταχύτητα μεταφοράς μπλοκ $bt = 1$ msec, μέγεθος μπλοκ $B = 2.400$ μπάιτ, διάκενο μεταξύ μπλοκ $G = 600$ μπάιτ. Ένα αρχείο ΕΡΓΑΖΟΜΕΝΟΣ έχει τα κατωτέρω πεδία: AP_TAYT (9 μπάιτ), ΕΠΙΘΕΤΟ (20 μπάιτ), ΟΝΟΜΑ (20 μπάιτ), APX_ΠΑΤΕΡΑ (1 μπάιτ), ΗΜΕΡ_ΓΕΝ (10 μπάιτ), ΔΙΕΥΘΥΝΣΗ (35 μπάιτ), ΤΗΛΕΦΩΝΟ (12 μπάιτ), AP_TAYT_ΠΡΟΪΣΤΑΜΕΝΟΥ (9 μπάιτ), ΤΜΗΜΑ (4 μπάιτ), ΚΩΔ_ΕΡΓΑΣΙΑΣ (4 μπάιτ) και *σημάδι διαγραφής* (1 μπάιτ). Το αρχείο ΕΡΓΑΖΟΜΕΝΟΣ έχει $r = 30.000$ εγγραφές, μορφοποίηση σταθερού μήκους και μη εκτεινόμενη οργάνωση. Διατυπώστε κατάλληλους τύπους και υπολογίστε τις κατωτέρω τιμές για το αρχείο ΕΡΓΑΖΟΜΕΝΟΣ:
- Το μέγεθος εγγραφής R (συμπεριλαμβανομένου του σημαδιού διαγραφής), τον παράγοντα ομαδοποίησης bfr , και το πλήθος b των μπλοκ δίσκου.
 - Υπολογίστε τον χρησιμοποιήσιμο χώρο σε κάθε μπλοκ δίσκου λόγω της μη εκτεινόμενης οργάνωσης.
 - Υπολογίστε το ρυθμό μεταφοράς tr και τον ολικό ρυθμό μεταφοράς btr για αυτή τη μονάδα δίσκου (βλ. παράρτημα Γ για τον ορισμό των tr και btr).
 - Υπολογίστε το μέσο αριθμό προσπελάσεων μπλοκ που απαιτούνται για την εύρεση μιας τυχαίας εγγραφής του αρχείου με γραμμική αναζήτηση.
 - Υπολογίστε το μέσο χρόνο σε msec που απαιτείται για την εύρεση μιας τυχαίας εγγραφής του αρχείου με γραμμική αναζήτηση, αν τα μπλοκ του αρχείου αποθηκεύονται σε διαδοχικά μπλοκ δίσκου και χρησιμοποιείται διπλή ενδιάμεση μνήμη.
 - Υπολογίστε το μέσο χρόνο σε msec που απαιτείται για την εύρεση μιας τυχαίας εγγραφής του αρχείου με γραμμική αναζήτηση, αν τα μπλοκ του αρχείου δεν αποθηκεύονται σε διαδοχικά μπλοκ δίσκου.
 - Υποθέστε ότι οι εγγραφές είναι διατεταγμένες ως προς κάποιο πεδίο-κλειδί. Υπολογίστε το μέσο αριθμό προσπελάσεων μπλοκ και το μέσο χρόνο που απαιτείται για την εύρεση μιας τυχαίας εγγραφής του αρχείου με δυαδική αναζήτηση.