

3ο Σύνολο Ασκήσεων

Ημερομηνία Παράδοσης: Οι ασκήσεις θα λυθούν στο μάθημα της Δευτέρας 25/1/2004 ή/και της Τετάρτης 27/1/2004

Θεματική Ενότητα: Αποθήκευση. Ευρετήρια. Επεξεργασία Ερωτήσεων.

Θέμα 1. Έστω ένα B+-δέντρο τάξης 400 (και για τους εσωτερικούς κόμβους και για τα φύλλα).

(i) Ποιος είναι ο μικρότερος αριθμός κλειδιών που μπορεί ένα τέτοιο B+-δέντρο να κρατήσει (ευρετηριοποιήσει); Ποιος είναι ο μεγαλύτερο αριθμός κλειδιών που μπορεί ένα τέτοιο B+-δέντρο να κρατήσει (ευρετηριοποιήσει);

(ii) Θεωρείστε ότι κάθε κόμβος του B+-δέντρου είναι γεμάτος κατά 70%. Πόσα επίπεδα χρειάζονται για $6 * 10^9$ κλειδιά; Ποιο ποσοστό του συνολικού χώρου (blocks) είναι φύλλα σε αυτήν την περίπτωση;

Θέμα 2. Θεωρείστε επεκτατό δυναμικό κατακερματισμό.

(i) Υποθέστε ότι κάθε κάδος (bucket) χωρά το πολύ 3 κλειδιά. Δώστε τη δομή που προκύπτει με την εισαγωγή των 1, 2, ..., 14 με αυτή τη σειρά. Δώστε το τοπικό και το ολικό βάθος για τον κατάλογο. Χρησιμοποιήστε τα τελευταία bits της δυαδικής αναπαράστασης.

(ii) Ο διπλασιασμός του καταλόγου απαιτεί την εξέταση όλων των κάδων με τοπικό β βάθος ίσο με το ολικό βάθος; Εξηγήστε.

Θέμα 3. Θεωρείστε τη πράξη $R \bowtie_{R.a=S.b} S$. Το κόστος που μας ενδιαφέρει είναι ο αριθμός των blocks που μεταφέρονται μεταξύ δίσκου και μνήμης. Αγνοήστε το κόστος της εγγραφής του αποτελέσματος στο δίσκο (εκτός αν ζητείτε ρητά). Υποθέστε ότι: το μέγεθος του block είναι 1024 bytes, η σχέση R έχει 10.000 πλειάδες και κάθε πλειάδα έχει μέγεθος 50 bytes, η σχέση S έχει 2.000 πλειάδες και κάθε πλειάδα έχει μέγεθος 48 bytes. Το γνώρισμα a είναι το πρωτεύον κλειδί για τη σχέση R . Το γνώρισμα a και το γνώρισμα b έχουν μέγεθος 8 bytes. Ένας δείκτης block έχει μήκος $P = 6$ bytes και ένας δείκτης εγγραφής έχει μήκος $P_R = 7$ bytes.

(α) Ποιός είναι ο μέγιστος αριθμός πλειάδων που παράγεται από τη συνένωση και ποιό το κόστος εγγραφής αυτού του αποτελέσματος στο δίσκο;

(β) Υποθέστε ότι η σχέση S είναι αποθηκευμένη σε μη ταξινομημένο αρχείο, ενώ η σχέση R είναι αποθηκευμένη σε ταξινομημένο αρχείο ως προς το γνώρισμα a . Υπάρχει ένα ευρετήριο ως προς a για τη σχέση R και ένα ευρετήριο ως προς b για τη σχέση S . Υποθέστε ότι κάθε πλειάδα του R συνενώνεται με ακριβώς 5 πλειάδες της σχέσης S . Θεωρείστε ότι υπάρχουν 25 καταχωρητές μεγέθους ενός block.

(β1) Υπολογίστε το μέγεθος αυτών των ευρετηρίων.

(β2) Δώστε τον αλγόριθμο της συνένωσης σε ψευτογλώσσα χρησιμοποιώντας το ευρετήριο στο a .

(β3) Υπολογίστε το κόστος της συνένωσης αν χρησιμοποιηθεί ο αλγόριθμος (β2). Μη ξεχάστε να υπολογίστε το κόστος ανάγνωσης των απαραίτητων blocks του ευρετηρίου.

(γ) Υποθέστε ότι η σχέση S είναι αποθηκευμένη σε μη ταξινομημένο αρχείο, η σχέση R είναι αποθηκευμένη σε ταξινομημένο αρχείο ως προς το γνώρισμα a και ότι υπάρχει κατακερματισμός στο γνώρισμα a . Θεωρείστε ότι υπάρχουν 42 καταχωρητές μεγέθους ενός block.

(γ1) Δώστε σε ψευτογλώσσα έναν αλγόριθμο που να υπολογίζει τη συνένωση χρησιμοποιώντας τη συνάρτηση κατακερματισμού στο a .

(γ2) Υπολογίστε το κόστος της συνένωσης αν χρησιμοποιηθεί ο αλγόριθμος (γ1).

(δ) Εξηγήστε αν το αποτέλεσμα των αλγορίθμων (β2) και (γ1) είναι ταξινομημένο ή όχι.

Θέμα 4. Θεωρείστε την παρακάτω σχήμα σχέσης ΥΠΑΛΛΗΛΟΣ(Αριθμός_Ταυτότητας, Μισθός, Ηλικία, Αριθμός_Τμήματος). Υπάρχει ένα ευρετήριο στο Αριθμός_Ταυτότητας και ένα στην Ηλικία. Το αρχείο είναι διατεταγμένο στο γνώρισμα (πεδίο) Αριθμός_Ταυτότητας.

(i) Πως θα χρησιμοποιούσατε το ευρετήριο (ευρετήρια) για να εξασφαλίσετε τον περιορισμό του κλειδιού;

(ii) Δώστε μια πράξη τροποποίησης (update) της σχέσης που πάντα γίνεται πιο γρήγορα με την ύπαρξη αυτών των ευρετηρίων (περιγραφή της με λόγια αρχεί).

(iii) Δώστε μια πράξη τροποποίησης (update) της σχέσης που πάντα γίνεται πιο αργή με την ύπαρξη αυτών των ευρετηρίων (περιγραφή της με λόγια αρχεί).

(iv) Δώστε μια πράξη τροποποίησης (update) της σχέσης που δεν επηρεάζεται από την ύπαρξη αυτών των ευρετηρίων (περιγραφή της με λόγια αρχεί).

Θέμα 5. Θεωρείστε δυο σχέσεις R και S . Η R έχει 100.000 πλειάδες μεγέθους 100 bytes η καθεμία, και η S 20.000 πλειάδες μεγέθους 100 bytes. Κάθε καταχωρητής μνήμης έχει μέγεθος 1 block. 1 block έχει μέγεθος 512 bytes.

(i) Θέλουμε να υπολογίσουμε το $R \cap S$ χρησιμοποιώντας έναν αλγόριθμο που διαβάζει κάθε block της R και της S μόνο μια φορά. Περιγράψτε τον αλγόριθμο που χρησιμοποιεί τον ελάχιστο αριθμό απο καταχωρητές μνήμης (buffers). Ποιος είναι αυτός ο αριθμός; Οι R και S είναι αποθηκευμένες σε αρχεία σωρού.

(ii) Υποθέστε ότι η S είναι αποθηκευμένη σε διατεταγμένο αρχείο και το πεδίο διάτάξης είναι και πρωτεύον κλειδί για την S . Επίσης, στην S υπάρχει ευρετήριο (ενός επιπέδου) πάνω στο πρωτεύον κλειδί. Το μέγεθος κάθε εγγραφής του ευρετηρίου είναι 10 bytes. Περιγράψτε τον αλγόριθμο για τον υπολογισμό του $R \cap S$ που να χρησιμοποιεί το ευρετήριο στο S . Ποιο είναι το κόστος του; Θεωρείστε ότι έχετε μόνο δυο καταχωρητές μνήμης.