

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΡΧΕΙΩΝ
2^η άσκηση

Ημερομηνία παράδοσης: 20 Μαΐου 2016

Δυναμικός Κατακερματισμός (dynamic hashing) - Κεντρική Μνήμη (5 μονάδες)

Υλοποιείτε την μέθοδο «Δυναμικού Κατακερματισμού» (dynamic hashing) στην κεντρική μνήμη, με πράξεις αναζήτησης, εισαγωγής και διαγραφής στοιχείων. Ο στατικός πίνακας έχει μέγεθος 100 θέσεων. Οι σελίδες αποθήκευσης έχουν χωρητικότητα 2 στοιχεία (για να αποφεύγονται οι συχνές διασπάσεις).

- **Φτιάξτε την μέθοδο εισαγωγής.** Μετρήστε το μέσο αριθμό συγκρίσεων κατά την εισαγωγή 20 νέων κλειδιών.
- **Φτιάξτε την μέθοδο αναζήτησης.** Αφού έχει γίνει η εισαγωγή κλειδιών, μετρήστε τον μέσο αριθμό συγκρίσεων πάνω σε 20 αναζητήσεις κλειδιών.
- **Φτιάξτε την μέθοδο διαγραφής.** Αφού έχει γίνει η εισαγωγή κλειδιών, μετρήστε τον μέσο αριθμό συγκρίσεων πάνω σε 20 διαγραφές κλειδιών.

Επαναλάβετε Μετρήστε για $N=100, 1000, 10000$. Για τον υπολογισμό του αριθμού συγκρίσεων, αυξάνεται η τιμή ενός μετρητή κάθε φορά που γίνεται χρήση ενός κλειδιού (πχ σε μία σύγκριση).

Σχολιάστε (και συγκρίνετε) τα αποτελέσματα και προσπαθήστε να ερμηνεύσετε λογικά την απόδοση της μεθόδου σε κάθε περίπτωση. Εξηγήστε γιατί και πως (από τι) εξαρτάται η απόδοση της μεθόδου σε εισαγωγές ή διαγραφές γιατί διαφέρει η απόδοση σε αναζητήσεις, εισαγωγές και σε διαγραφές. Εξετάστε αν εξαρτάται η απόδοση από το N (δηλαδή από το μέγεθος της δομής δεδομένων).

Δυναμικός Κατακερματισμός (dynamic hashing) - Δίσκος (5 μονάδες)

Υλοποιείτε την μέθοδο στον δίσκο. Ο στατικός πίνακας και το δένδρο δεικτοδότησης παραμένουν στην κεντρική μνήμη. Τα δεδομένα αποθηκεύονται στο δίσκο. Το μέγεθος της σελίδας δίσκου είναι $b=128$ bytes (το b είναι παράμετρος που ορίζεται στην αρχή ώστε η μέθοδος να μπορεί να λειτουργεί για κάθε τιμή). Επομένως, υπάρχουν δείκτες από την κεντρική μνήμη προς σελίδες δεδομένων.

Κάντε τις ίδιες μετρήσεις όπως στο πρώτο μέρος με την διαφορά ότι τώρα πρέπει να μετρήσετε πως τον αριθμός των προσβάσεων σε σελίδες δίσκου σε πράξεις εισαγωγής, αναζήτησης και διαγραφής για τις διαφορετικές τιμές του N . Δηλαδή ο μετρητής

αυξάνεται μόνο σε περιπτώσεις ανάγνωσης ή γραφής σελίδας στο δίσκο. Σχολιάστε την απόδοση της μεθόδου όπως κάνατε στο πρώτο ερώτημα.

Στην περίπτωση διαγραφής, αγνοήστε την συγχώνευση (δηλαδή επιτρέπεται δύο «γειτονικές» σελίδες να έχουν λιγότερα κλειδιά ή ακόμα και να είναι άδειες).

Υποδείξεις: Κατά την μέτρηση της απόδοσης, κάθε πράξη στην οποία συμμετέχει το κλειδί, μετράται ως μια σύγκριση. Δηλαδή η $h(key) = key \% 100$ μετράει ως μια πράξη σύγκρισης, το ίδιο κάθε δυαδική πράξη με το κλειδί (στο δυαδικό δένδρο) και κάθε σύγκριση με κλειδιά που βρίσκονται στις σελίδες αποθήκευσης.

Για τον δίσκο, για την εκτίμηση της απόδοσης μετράτε μόνο τις προσβάσεις σε σελίδες δίσκου (δηλαδή μόνο πράξεις read, write σε σελίδες δίσκου).

Επίσης, για την κατασκευή του δυαδικού δένδρου δεικτών στην κεντρική μνήμη, θα πρέπει να διαχειριστείτε την δυαδική παράσταση του κλειδιού και να διαβάζετε πάντα το λιγότερο σημαντικό bit του κλειδιού (least significant bit). Για κάθε ενδεχόμενο είναι καλύτερα να δηλώσετε τα κλειδιά ως αριθμούς χωρίς πρόσημο (unsigned integers) αν και το πιο πιθανό είναι ότι δεν επηρεάζει σε κάτι την υλοποίηση. Για δυαδικές πράξεις στην java όπως (AND, OR, XOR κ.λ.π) δείτε τα παρακάτω

<http://docs.oracle.com/javase/tutorial/java/nutsandbolts/op3.html>

http://www.tutorialspoint.com/java/java_basic_operators.htm

Παραδοτέα: Ένα συμπίεσμένο zip αρχείο που περιέχει ότι ζητείται παρακάτω:

A) ο κώδικας περιέχει σχόλια που εξηγούν την υλοποίηση

B) μία έκθεση που περιγράφει σε 1-2 σελίδες πως φτιάχτηκε ο κώδικας (δηλ. για κάθε ερώτημα ποια είναι η γενική ιδέα της λύσης σε 3-4 προτάσεις), υπάρχουν σαφείς οδηγίες μετάφρασης από compiler και εκτέλεσης, τι λάθη έχει (αν έχει, περιπτώσεις που δεν δουλεύει το πρόγραμμα, ή περιπτώσεις που κάνει περισσότερα από όσα σας ζητεί η άσκηση, τι χρησιμοποιήσατε από έτοιμα προγράμματα ή πηγές πληροφόρησης. Υποδείξτε ακόμα και πηγές στο WWW όπως Wikipedia ή ακόμα και συναδέλφους που σας βοήθησαν στην άσκηση.

Οι ασκήσεις βαθμολογούνται με άριστα εφόσον

A) το zip είναι πλήρες

B) Οι κώδικες περνούν από compiler και εκτελούνται κανονικά και σωστά σε windows ή Linux περιβάλλον