

Δομές Δεδομένων και Αρχείων

3η Άσκηση

Σκοπός της άσκησης είναι να υλοποιήσουμε ένα B+ Tree στο δίσκο. Συγκεκριμένα καλούμαστε να κατασκευάσουμε ένα δένδρο(m-ary tree) το οποίο χρησιμοποιείται ευρέως στις βάσεις δεδομένων και στα συστήματα αρχείων. Στην άσκηση ζητήθηκε να υλοποιηθούν οι εξής λειτουργίες:

- Εισαγωγή τυχαίου κλειδιού
- Αναζήτηση τυχαίου κλειδιού
- Αναζήτηση εύρους τιμών κλειδιών
- Διαγραφή τυχαίου κλειδιού

Για την υλοποίηση του B+Tree στο δίσκο πρέπει να μετρηθεί :

- Μέσος αριθμός προσβάσεων δίσκου ανά εισαγωγή για 20 εισαγωγές
- Μέσος αριθμός προσβάσεων δίσκου ανά αναζήτηση και διαγραφή για 20 αναζητήσεις και 20 διαγραφές αντίστοιχα.
- Μέσος αριθμός προσβάσεων δίσκου για 20 αναζητήσεις εύρους τιμών κλειδιών.

• Απόδοση και τεκμηρίωση αποτελεσμάτων

| Μέθοδος | Μέσος αριθμός προσβάσεων ν/εισαγωγή | Μέσος αριθμός προσβάσεων /τυχαία αναζήτηση | Μέσος αριθμός προσβάσεων ανά διαγραφή | Μέσος αριθμός συγκρίσεων /αναζήτηση εύρους K=10 | Μέσος αριθμός συγκρίσεων ν/αναζήτηση εύρους K=1000 |
|---------|-------------------------------------|--|---------------------------------------|---|--|
| | 5 | 2 | 5 | - | - |

• Εισαγωγή τυχαίου κλειδιού

Γνωρίζουμε ότι σε μια μέση περίπτωση η χρονική πολυπλοκότητα της εισαγωγής τυχαίου κλειδιού και για το B+ Tree είναι $O(\log(N))$, όπου N ο αριθμός των κλειδιών που εισάγονται, δηλαδή 100000 στην περίπτωση μας. Άρα $\log(100000) = 5$. Το οποίο είναι πολύ λογικό καθώς η δομή του B+ Tree είναι φτιαγμένη έτσι ώστε να εξοικονομούμε χρόνο στην αναζήτηση έχοντας όλους τους εξωτερικούς κόμβους (leaf Nodes) στο ίδιο επίπεδο. Όπως φαίνεται οι μετρήσεις που πάρθηκαν είναι πολύ κοντά στη θεωρητική τιμή.

• Αναζήτηση τυχαίου κλειδιού

Από την θεωρία προκύπτει ότι σε μια μέση περίπτωση η χρονική πολυπλοκότητα της αναζήτησης τυχαίου κλειδιού είναι $O(\log(N))$ όπου N ο αριθμός των κλειδιών που εισάγονται. Άρα όπως προηγουμένως $\log(100000) = 5$. Σε αντίθεση με την εισαγωγή τυχαίου κλειδιού παρατηρούμε μια απόκλιση από τις μετρήσεις που πήραμε.

• Διαγραφή τυχαίου κλειδιού

Σύμφωνα με την θεωρία για μια μέση περίπτωση η πολυπλοκότητα για την διαγραφή ενός τυχαίου κλειδιού

είναι $O(\log(N))$ όπου N ο αριθμός των κλειδιών που εισάγονται. Άρα όπως προηγουμένως $\log(100000) = 5$. Εδώ βλέπουμε πως οι μετρήσεις με την θεωρία συμβαδίζουν σχεδόν τέλεια. Εδώ να σημειώσουμε ότι επειδή διαλέγουμε τυχαίους αριθμούς που έχουν παραχθεί στο εύρος 1-1000000 υπάρχει το ενδεχόμενο να μην βρεθούν και τα 20 κλειδιά που αναζητούμε να διαγράψουμε .

- **Αναζήτηση εύρους τιμών κλειδιού**

Οι μετρήσεις για την αναζήτηση εύρους τιμών κλειδιού λείπουν καθώς, λόγω τεχνικών προβλημάτων δεν έφτασε ο χρόνος να την υλοποιήσουμε, παρόλα αυτά επειδή ο αλγόριθμος υλοποίησης της μεθόδου είναι πιο πολύπλοκος περιμένουμε περισσότερες αναζητήσεις ειδικά στην περίπτωση που $K=1000$.