

Δομές Δεδομένων και Τεχνικές Προγραμματισμού (Τμήμα Αρτίων ΑΜ)

Εργαστήριο 5: Κατακερματισμός (hashing)

Στο εργαστήριο αυτό θα μελετήσουμε την τεχνική του κατακερματισμού και θα χρησιμοποιήσουμε τον σχετικό κώδικα που παρουσιάστηκε στην δωδέκατη ενότητα διαλέξεων και βρίσκεται στην ιστοσελίδα του μαθήματος: <http://cgi.di.uoa.gr/~k08/code.html>.

Ο κώδικας που σας έχει δοθεί υλοποιεί την τεχνική της ανοικτής διευθυνσιοδότησης με διπλό κατακερματισμό (open addressing with double hashing). Τα κλειδιά των στοιχείων που αποθηκεύουμε στον πίνακα κατακερματισμού είναι ακέραιοι.

Ο κώδικας σας αποτελείται από τα αρχεία `HashTableTypes.h`, `HashTableInterface.h`, `menu.h`, `HashTableImplementation.c`, `menu.c` και `main.c`.

Μελετήστε τον κώδικα και κατανοήστε τη λειτουργία του. Δείτε προσεκτικά τις συναρτήσεις κατακερματισμού που χρησιμοποιούνται. Καταλαβαίνετε τη λειτουργία τους; Μεταγλωττίστε τον κώδικα και εκτελέστε τον. Πειραματιστείτε με όλες τις δυνατότητες του μενού λειτουργιών. Για παράδειγμα, να αποθηκεύσετε μερικούς ακεραίους στον πίνακα κατακερματισμού και να ψάξετε να βρείτε κάποιους από αυτούς.

Παρατηρείστε ότι τα αρχεία επικεφαλίδας (π.χ., `HashTableTypes.h`) που σας δίνονται περιέχουν την οδηγία `#ifndef`. Πότε χρησιμοποιείται αυτή η οδηγία και γιατί;

Αφού κατανοήσετε πλήρως τον κώδικα που σας δόθηκε, να κάνετε τις ακόλουθες αλλαγές και επεκτάσεις.

1. Να αλλάξετε τον κώδικα σας ώστε τα στοιχεία που αποθηκεύονται στον πίνακα κατακερματισμού να έχουν κλειδί μια συμβολοσειρά (και όχι ένα ακέραιο όπως είναι στον κώδικα που σας δόθηκε). Για τον πειραματισμό σας, δίνεται το αρχείο `EnglishWords.txt` που περιέχει λέξεις της Αγγλικής γλώσσας.
2. Να επεκτείνετε τον κώδικα σας ώστε να μετράει πόσες συγκρούσεις (collisions) γίνονται συνολικά κατά την εισαγωγή στοιχείων στον πίνακα κατακερματισμού. Επίσης, ποιος είναι ο μέγιστος και ποιος ο ελάχιστος αριθμός συγκρούσεων που συμβαίνουν, και για ποιο κουβά του πίνακα κατακερματισμού.
3. Να επεκτείνετε τον κώδικα σας ώστε να μετράει το συνολικό χρόνο που πήρε η αποθήκευση ενός αρχείου με λέξεις καθώς και τον μέσο χρόνο αποθήκευσης για μια λέξη (επιλογή 1 από το μενού λειτουργιών).

4. Να πειραματιστείτε με διαφορετικές συναρτήσεις κατακερματισμού από αυτές που συζητήθηκαν στην τάξη.
5. Στον κώδικα, υπάρχει ένα λάθος, σχετικό με μία λειτουργία του πίνακα κατακερματισμού. Ποιο είναι αυτό; Τεκμηριώστε την απάντησή σας με ένα παράδειγμα.
6. Φτιάξτε μια συνάρτηση κατακερματισμού για συμβολοσειρές η οποία δουλεύει ως εξής:
 - a. Προσθέτει τις τιμές των χαρακτήρων της συμβολοσειράς
 - b. Στη συνέχεια επιστρέφει το άθροισμα modulo TABLESIZE

Θέστε το TABLESIZE ως 31393.

Μετρήστε το χρόνο που χρειάζεται η αποθήκευση των λέξεων του αρχείου `Hard.txt` με αυτή τη συνάρτηση κατακερματισμού. Συγκρίνετε με άλλες συναρτήσεις κατακερματισμού που συζητήσαμε στην τάξη.