

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΡΧΕΙΩΝ

3^η άσκηση

Ημερομηνία παράδοσης: 31 Μάϊου 2018

Η άσκηση είναι ατομική

Γραμμικός Κατακερματισμός

Υλοποιείτε την μέθοδο «γραμμικού κατακερματισμού» (linear hashing) στον **δίσκο** με πράξεις αναζήτησης και εισαγωγής στοιχείων. Υπάρχουν κώδικες (κυρίως για κεντρική μνήμη) στο διαδίκτυο και στον ιστοχώρο μαθήματος. Η άσκηση είναι διατυπωμένη με τέτοιο τρόπο ώστε θα χρειαστεί να καταλάβετε και να αλλάξετε τους κώδικες προκειμένου να πάρετε τα αποτελέσματα που ζητούνται. **Διαφορετικά υπάρχει κίνδυνος η άσκηση να μοιάζει με κώδικες που υπάρχουν και να μηδενιστεί ως αντιγραφή.**

Αρχικά το αρχείο περιέχει 10 σελίδες. Οι κύριες σελίδες όπως και οι σελίδες υπερχειλίσης έχουν μέγεθος **b=512 bytes**. Αυτό σημαίνει ότι σε κάθε σελίδα χωράνε 127 κλειδιά (ακέραιοι αριθμοί) και ένα ακέραιο που χρησιμοποιείται για να δείξει σε μία θέση δίσκου. Αυτή είναι θέση στην οποία αποθηκεύεται μία σελίδα υπερχειλίσης (αν υπάρχει, αλλιώς ο δείκτης είναι 0) ή η επόμενη μίας σελίδας υπερχειλίσης (αν έχει δημιουργηθεί λίστα από σελίδες υπερχειλίσης).

Όλες οι κύριες σελίδες αποθηκεύονται σε ένα αρχείο. Όλες οι σελίδες υπερχειλίσης (στις οποίες δείχνουν κύριες σελίδες) αποθηκεύονται σε ένα δεύτερο αρχείο. Για την διαχείριση των αρχείων μπορείτε να χρησιμοποιήσετε τις μεθόδους της 1^{ης} άσκησης.

Κατασκευή Αρχείου Γραμμικού Κατακερματισμού (50 μονάδες)

Υλοποιείτε τις πράξεις αναζήτησης τυχαίου κλειδιού και εισαγωγής νέου κλειδιού. Το κλειδί εισάγεται μόνο αν δεν υπάρχει στο αρχείο. Ένα νέο κλειδί μπορεί να εισαχθεί είτε σε μια κύρια σελίδα ή σε μια σελίδα υπερχειλίσης. Επομένως, μία πράξη εισαγωγής προϋποθέτει μια αναζήτηση για την τιμή του κλειδιού, που με την σειρά της απαιτεί μία πρόσβαση στην κύρια σελίδα (που είναι η σελίδα που δείχνει η συνάρτηση hash), και ενδεχομένως επιπλέον (μία ή περισσότερες) προσβάσεις στις σελίδες υπερχειλίσης που συνδέονται με τις κύριες σελίδες. Για κάθε πρόσβαση σε κύρια σελίδα ή σε σελίδα υπερχειλίσης αυξάνεται κατά ένα ο αριθμός προσβάσεων στο δίσκο.

Στο αρχείο εισάγονται κλειδιά (ακέραιοι) που παράγονται από μια γεννήτρια τυχαίων αριθμών. Διάσπαση σελίδων γίνεται όταν ο συντελεστής πληρότητας **u** ξεπεράσει ένα ποσοστό (**u > 80%**).

Ένας μετρητής μετράει τον αριθμό προσβάσεων στο δίσκο σε πράξεις εισαγωγής νέων κλειδιών. Ο αριθμός αυτός είναι 1 (αν το κλειδί εισάγεται σε κύρια σελίδα) ή μεγαλύτερος από 1 (αν το κλειδί εισάγεται σε σελίδα υπερχείλισης). Μετά από κάθε εισαγωγή υπολογίστε το u . Κάθε φορά που συμβαίνει $u > 80\%$ το αρχείο επεκτείνεται με την προσθήκη μίας νέας κύριας σελίδας. Όταν συμβεί αυτό, μετρήστε τον μέσο αριθμό προσβάσεων ανά εισαγωγή για τα στοιχεία που εισήχθησαν ανάμεσα σε δύο επεκτάσεις αρχείου. Δηλαδή, για τις εισαγωγές που μεσολάβησαν διαιρέστε τον συνολικό αριθμό προσβάσεων με τον αριθμό των στοιχείων που εισήχθησαν. Μετά μηδενίστε τον μετρητή ώστε να ξεκινήσει να μετράει από το μηδέν και μέχρι να γίνει η επόμενη επέκταση αρχείου.

Μετρήστε πως μεταβάλλεται ο αριθμός των προσβάσεων σε σελίδες του δίσκου σε πράξεις εισαγωγής μέχρι να δημιουργηθούν 40 κύριες σελίδες (στο αριθμό αυτό δεν συμπεριλαμβάνονται σελίδες υπερχείλισης). Όταν δημιουργηθούν 40 κύριες σελίδες το αρχείο έχει ήδη διπλασιαστεί δύο φορές (από 10 σελίδες που ήταν αρχικά): Την πρώτη φορά δημιουργήθηκαν 20 κύριες σελίδες και την δεύτερη φορά 40 κύριες σελίδες.

Φτιάξτε το διάγραμμα που δείχνει πως μεταβάλλεται ο αριθμός προσβάσεων στο δίσκο (Y άξονας) ως συνάρτηση του μεγέθους του αρχείου (X άξονας με τον αριθμό σελίδων του αρχείου).

Κριτήριο $u > 50\%$ (20 μονάδες)

Επαναλάβετε το παραπάνω πείραμα με κριτήριο $u > 50\%$. Θα φτιάξετε ένα διάγραμμα που θα αντιστοιχεί σε $u > 50\%$.

Τεκμηρίωση Αποτελεσμάτων (30 μονάδες)

Σχολιάστε τα αποτελέσματα προσπαθήστε να ερμηνεύσετε λογικά την απόδοση της μεθόδου. Εξηγήστε τις τιμές τον τρόπο που μεταβάλλεται η απόδοση του γραμμικού κατακερματισμού με το u και την αύξηση του μεγέθους του αρχείου.

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

Γ) **Ασκήσεις που μοιάζουν ή δεν τεκμηριώνονται σωστά ελέγχονται για αντιγραφή ή copy & paste από το διαδίκτυο. Αν διαπιστωθεί αντιγραφή μηδενίζονται οι ασκήσεις του έτους.**