Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Αλγοριθμική Επιστήμη Δεδομένων 2019-20

Διδάσκοντες: Α. Παγουρτζής, Θ. Σούλιου, Δ. Φωτάκης

1η Σειρά Ασκήσεων

Άσκηση 1. Να λύσετε τις ασκήσεις 6.3.1 και 6.4.2 του βιβλίου LRU.

Άσκηση 2. Έστω ένα σύνολο U. Μια οικογένεια συναρτήσεων κατακερματισμού $\mathcal{H} = \{h : U \to [m]\}$ λέγεται *καθολική* αν

$$\forall x, y \in U,\ x \neq y : \quad \Pr_{h \in \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{m}$$

(*σημείωση: χρησιμοποιούμε τον συμβολισμό* $[m] = \{0, \dots m-1\}$)

Ισοδύναμα, για κάθε δύο διαφορετικές τιμές $x, y \in U$, υπάρχουν το πολύ $|\mathcal{H}|/m$ συναρτήσεις $h \in \mathcal{H}$ για τις οποίες $h(x) = h(y)$.

(α) Αποδείξτε ότι για $a \in [m] \setminus \{0\}, b \in [m]$ η οικογένεια συναρτήσεων $h_{a,b}(x) = (ax + b) \bmod m$ δεν έχει την ιδιότητα της καθολικότητας για $U = [m^2]$.

Υπόδειξη: μελετήστε τη συμπεριφορά των συναρτήσεων της οικογένειας για $x, y : x \neq y, x \equiv y \pmod m$ *(δηλ.* $x \bmod m = y \bmod m$*).*

(β) Αποδείξτε ότι για πρώτο αριθμό $p \geq m^2$ και για $a \in [p] \setminus \{0\}, b \in [p]$ η οικογένεια συναρτήσεων $h_{a,b}(x) = ((ax + b) \bmod p) \bmod m$ έχει την ιδιότητα της καθολικότητας για $U = [m^2]$.

Άσκηση 3. Εξετάστε την μέθοδο κατακερματισμού ανοιχτής διευθυνσιοδότησης (open addressing) και:

(α) Εξηγήστε γιατί ο μέσος χρόνος επιτυχούς αναζήτησης, μετά από εισαγωγή n στοιχείων, είναι ίδιος με τον μέσο χρόνο εισαγωγής των στοιχείων στον πίνακα.

(β) Αποδείξτε ότι ο χρόνος αυτός φράσσεται άνω από την ποσότητα $\min\{\frac{1}{\alpha} \ln \frac{1}{1-\alpha} + \frac{1}{\alpha}, \frac{1}{1-\alpha}\}$, όπου $\alpha = n/m$ ο παράγοντας φόρτου.

Υπόδειξη: ξεκινήστε αποδεικνύοντας ότι κατά την εισαγωγή του i*-οστού στοιχείου ο αναμενόμενος αριθμός δοκιμών μέχρι να βρεθεί κενή θέση είναι* $m/(m-i)$*, υποθέτοντας uniform hashing.*

Άσκηση 4. Απαντήστε στα παρακάτω ερωτήματα σχετικά με τον αλγόριθμο DGIM (που μετράει το πλήθος των '1' στα τελευταία k bits ενός stream ($k \leq N$, όπου N ένα δεδομένο μέγεθος 'παραθύρου').

(α) Ποιος είναι ο λόγος προσέγγισης του αλγορίθμου (στη χειρότερη περίπτωση); Θεωρήστε ως λόγο προσέγγισης την τιμή $\max(\frac{A}{C}, \frac{C}{A})$, όπου C η πραγματική τιμή του μετρητή και A η τιμή που επιστρέφει ο αλγόριθμος.

(β) Μπορείτε να βελτιώσετε τον λόγο αυτό (στη χειρότερη περίπτωση πάντα) τροποποιώντας απλώς και μόνο την εκτίμηση για το πλήθος '1' στον παλαιότερο κάδο (η εκτίμηση στις διαφάνειες είναι 2^{i-1}); Ποιος είναι ο μικρότερος λόγος που μπορείτε να πετύχετε με αυτόν τον τρόπο;

(γ) Εξηγήστε πώς μπορούμε να πετύχουμε λόγο προσέγγισης ε για οποιαδήποτε σταθερά $\varepsilon > 0$. Αποδείξτε τον ισχυρισμό σας.

(δ) Δώστε ένα άνω φράγμα για το μέγεθος της μνήμης που θα χρειαστεί ο αλγόριθμος με την τροποποίηση του ερωτήματος (γ), σαν συνάρτηση των N και ε, και εξηγήστε την απάντησή σας.

Προθεσμία υποβολής και οδηγίες. Οι απαντήσεις θα πρέπει να υποβληθούν έως τις 14/4/2020, σε ηλεκτρονική μορφή. Για απορίες / διευκρινίσεις: στείλτε μήνυμα στη διεύθυνση `ads@corelab.ntua.gr`.