



Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Αλγοριθμική Επιστήμη Δεδομένων 2020-21

Διδάσκοντες: Α. Παγουρτζής, Θ. Σούλιου, Β. Νάκος

1η Σειρά Ασκήσεων

**Άσκηση 1.** Να λύσετε τις ασκήσεις 6.3.1, 6.3.2 και 6.4.2 του βιβλίου LRU.

**Άσκηση 2.** Μελετήστε τις υλοποιήσεις αλγορίθμων εξόρυξης συχνών συνόλων στοιχείων που υπάρχουν στο αποθετήριο <http://fimi.uantwerpen.be/> και προσπαθήστε να επαναλάβετε ορισμένα από τα πειράματα που περιγράφονται εκεί (τουλάχιστον 2 αλγορίθμους με τουλάχιστον 3 dataset, της επιλογής σας). Περιγράψτε συνοπτικά τι κάνατε και παρουσιάστε με λεπτομέρειες τα αποτελέσματα που πήρατε.

**Άσκηση 3.** Έστω ένα σύνολο  $U$ . Μια οικογένεια συναρτήσεων κατακερματισμού  $\mathcal{H} = \{h : U \rightarrow [m]\}$  λέγεται *καθολική* αν

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

(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  $U = [m^k], k \geq 2$ .

(β) Αποδείξτε ότι για πρώτο αριθμό  $p > m^k, k \geq 2$  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  $U = [m^k]$ .

(γ) Εξακολουθεί να ισχύει η ιδιότητα της καθολικότητας αν στο ερώτημα (β) θέσουμε  $U = [p]$ ;

**Άσκηση 4.** 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  $\frac{1}{\alpha} \ln \frac{1}{1-\alpha}$ , όπου  $\alpha = n/m$  ο παράγοντας φόρτου.

Υπόδειξη: ξεκινήστε με μια εκτίμηση για το αναμενόμενο πλήθος δοκιμών κατά την εισαγωγή του  $i$ -οστού στοιχείου υποθέτοντας *uniform hashing*.

**Προθεσμία υποβολής και οδηγίες.** Οι απαντήσεις θα πρέπει να υποβληθούν έως τις 23/4/2020, σε ηλεκτρονική μορφή. Για απορίες / διευκρινίσεις: στείλτε μήνυμα στη διεύθυνση [ads@corelab.ntua.gr](mailto:ads@corelab.ntua.gr).