

# 2<sup>η</sup> Ανάθεση για το Σπίτι

## (σελ. 1 από 2)

(1) Να υλοποιήσετε (σε **Python 2.7**) αλγόριθμο που εκτιμά την  $K$ -στή στιγμή (για κάποιον ακέραιο  $K \geq 2$ ) ως προς τις συχνότητες εμφάνισης στοιχείων σε μια ροή της οποίας τα στοιχεία προέρχονται από το από το λατινικό αλφάβητο (θεωρούμε μόνο πεζά γράμματα). Συγκεκριμένα, το πρόγραμμά σας:

(1α) Θα διαβάζει στην είσοδο:

- \* την τιμή  $K$  (για την  $K$ -στιγμή που θα υπολογίσουμε προσεγγιστικά),
- \* το πλήθος  $\Lambda$  των μεταβλητών που θα χρησιμοποιήσουμε,
- \* το όνομα-αρχείου ροής (διαδοχικοί χαρακτήρες λατινικού αλφάβητου, που χωρίζονται με κόμμα, πχ, “a,a,z,b,d,d,d,e,f,a,a,a,g,l,k,j”), ή εναλλακτικά,
- \* τη ροή, στοιχείο-προς-στοιχείο, από το πληκτρολόγιο.

(1β) Θα υλοποιεί τον αλγόριθμο **AMS** (βλ. ενότητα 4.5 του MMDS βιβλίου) για υπολογισμό της  $K$ -στιγμής της ροής. Προς τούτο:

- \* Θα χρησιμοποιεί μητρώο  $S$  με  $\Lambda$  θέσεις (για τις μεταβλητές που χρησιμοποιούμε).
- \* Καθώς η ροή διαβάζεται (από το αρχείο ροής, ή από το πληκτρολόγιο) στοιχείο-προς-στοιχείο, θα αποφασίζεται αν θα αρχικοποιείται μια καινούργια μεταβλητή από εκείνη τη χρονική στιγμή και μετά. Εφόσον το  $S$  είναι ήδη πλήρες (με ήδη χρησιμοποιούμενες μεταβλητές) και αποφασιστεί η εκκίνηση καινούργιας μεταβλητής, επιλέγεται ισοπίθανα τυχαία θέση στο  $S$  που αρχικοποιείται η καινούργια μεταβλητή.

(1γ) Ανά πάσα χρονική στιγμή (δηλαδή, μετά την άφιξη και επεξεργασία του καινούργιου στοιχείου), θα δίνεται η τρέχουσα εκτίμηση της  $K$ -στής στιγμής, όπως έχει διαμορφωθεί μέχρι στιγμής με βάση τις τιμές των μεταβλητών που συντηρούμε στο  $S$ .

# 2<sup>η</sup> Ανάθεση για το Σπίτι

## (σελ. 2 από 2)

(2) Έστω ότι μας δίνεται η ροή «3,1,4,1,5,9,2,6,5». Θεωρούμε ότι οι συναρτήσεις κατακερματισμού που χρησιμοποιούμε για τον αλγόριθμο FM (βλ. Ενότητα 4.4 του MMDS βιβλίου) που μετρά τα διαφορετικά στοιχεία στη ροή) είναι της μορφής  $h(x) = (ax + b) \bmod 32$ , για κάποιους φυσικούς θετικούς ακέραιους αριθμούς  $a$  και  $b$ . Η έξοδος κάθε συνάρτησης κατακερματισμού θεωρούμε ότι δίνεται σε 5-ψήφια δυαδική αναπαράσταση.

(2α) Να υπολογίσετε την τιμή του  $r(a)$  για κάθε στοιχείο της ροής, καθώς επίσης και την **εκτίμηση του πλήθους διαφορετικών στοιχείων** που επιστρέφεται στο τέλος από τον FM αλγόριθμο, αν η συνάρτηση κατακερματισμού είναι η εξής:

$$(2\alpha 1) \quad h(x) = (2x + 1) \bmod 32.$$

$$(2\alpha 2) \quad h(x) = (3x + 7) \bmod 32.$$

$$(2\alpha 3) \quad h(x) = (4x) \bmod 32.$$

(2β) Παρατηρείτε κάποιο πρόβλημα ως προς την επιλογή συνάρτησης κατακερματισμού στις παραπάνω περιπτώσεις? Τι θα συστήνατε σε κάποιον που θέλει να χρησιμοποιήσει τον αλγόριθμο FM με μια συνάρτηση κατακερματισμού της μορφής  $h(x) = (ax + b) \bmod 2^K$ ?