

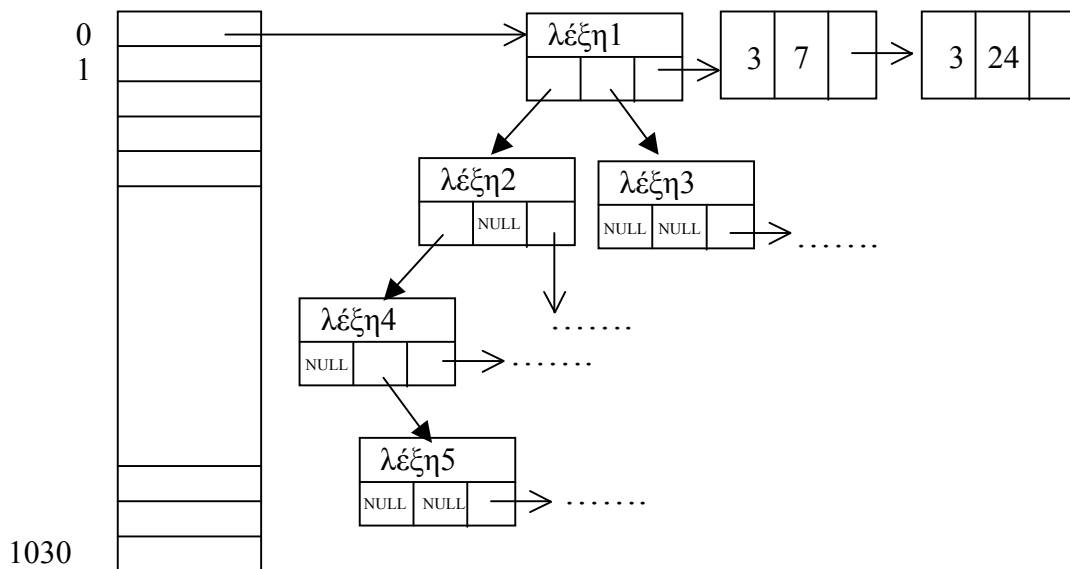


## ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΕΣ ΤΕΧΝΙΚΕΣ 2<sup>η</sup> ΑΣΚΗΣΗ 2012-2013

Σημείωση: Οι ασκήσεις αποτελούν μέρος της εργαστηριακής εξάσκησης στο μάθημα "Προγραμματιστικές Τεχνικές". Η υποβολή του σχετικού προγράμματος έμμεσα δηλώνει ότι είστε οι μοναδικοί συγγραφείς της λύσης της άσκησης. Εάν το πρόγραμμα ή μέρος του έχει παρθεί από άλλη πηγή θα πρέπει να αναφέρεται η πηγή και ο συγγραφέας του.

Στο πλαίσιο της Άσκησης 2 καλείστε να αναπτύξετε ένα ευρετήριο λέξεων για κείμενα. Πιο συγκεκριμένα, το πρόγραμμα θα δέχεται στην είσοδό του ένα κείμενο και θα προσδιορίζει τις θέσεις εμφάνισης της κάθε λέξης, δηλαδή τις δυάδες γραμμών-στηλών εμφάνισης του πρώτου χαρακτήρα της λέξης. Ως κείμενο θεωρείται μία ακολουθία χαρακτήρων, ενώ ως λέξη θεωρείται μία ακολουθία χαρακτήρων ενός κειμένου που δεν περιέχει κενό και βρίσκεται ανάμεσα σε κενά ή αλλαγές γραμμής ή αρχή ή τέλος αρχείου. Το πρόγραμμά σας θα αποθηκεύει και θα αναζητά τις θέσεις αυτές με χρήση Πινάκων Κατακερματισμού, ενώ για την επίλυση των συγκρούσεων θα χρησιμοποιούνται Δυαδικά Δέντρα Αναζήτησης. Για την αποθήκευση των θέσεων μίας λέξης χρησιμοποιήστε μία απλά συνδεδεμένη λίστα, στην οποία οι θέσεις θα είναι ταξινομημένες, δηλαδή για παράδειγμα αν μία λέξη εμφανίζεται στις θέσεις (2,3) (γραμμή, στήλη) και (5,2) του κειμένου, το (2,3) θα προηγείται του (5,2) στη λίστα των θέσεων της λέξης αυτής.

Η δομή θα φαίνεται ως εξής



Όπου η λέξη1, λέξη2 ... λέξη 5, έχουν όλες τιμή hash 0 και η λέξη1 υπάρχει στο κείμενο στη γραμμή 3, στήλη 7 και, στη γραμμή 3, στήλη 24.

Για τον έλεγχο, το πρόγραμμά σας θα διαβάζει αρχικά το κείμενο από ένα αρχείο το οποίο θα ξεκινά με μία λέξη (λέξη αναζήτησης), θα αλλάζει γραμμή και στη συνέχεια θα παραθέτει το κείμενο (οι θέσεις των λέξεων θα προσδιορίζονται με βάση το κείμενο και όχι με βάση το αρχείο, θεωρώντας ότι η λέξη αναζήτησης δεν αποτελεί μέρος του κειμένου αλλά της εισόδου). Στη συνέχεια, θα εκτυπώνει τη λέξη αναζήτησης και όλες τις θέσεις εμφάνισής της στο κείμενο, αλλάζοντας γραμμές. Αν δεν εμφανίζεται η λέξη αναζήτησης στο κείμενο θα εμφανίζει τη συμβολοσειρά NOT FOUND.

### Παρατηρήσεις

1. Για τον υπολογισμό της hash τιμής για λέξη μπορείτε να χρησιμοποιήσετε τη συνάρτηση

$$f(s) = \left( \sum_{i=0}^{n-1} B^{n-i-1} s_i \right) \bmod W$$

όπου

s: είναι η λέξη που αποτελείται από n χαρακτήρες

$s_i$ : είναι η ASCII τιμή, του  $i^{\text{th}}$  χαρακτήρα της λέξης

W: είναι το μέγεθος του πίνακα κατακερματισμού ( $W = 1031$ )

B: είναι ίσο με  $256 = 2^8$  (θεωρώντας ότι κάθε χαρακτήρας είναι 8 bits)

2. Θεωρήστε ότι η είσοδος που δίνεται είναι πάντα σύμφωνη με τις προδιαγραφές.
3. Θεωρήστε ότι κάθε λέξη δεν μπορεί να είναι μεγαλύτερη από 20 χαρακτήρες και κάθε γραμμή μεγαλύτερη από 120 χαρακτήρες.
4. Για την υποβολή στο αυτόματο σύστημα υποβολής και ελέγχου (grader), θα χρειαστεί να τροποποιήσετε λίγο το πρόγραμμά σας ώστε να διαβάζει το αρχικό κείμενο από το πληκτρολόγιο και να τυπώνει το ευρετήριο στην οθόνη.

### Παράδειγμα

#### Είσοδος

of

Words are flowing out like endless rain into a paper cup

They slither wildly as they slip away across the universe

Pools of sorrow, waves of joy are drifting through my open mind

Possessing and caressing me

#### Έξοδος

of

(3,7)

(3,24)