

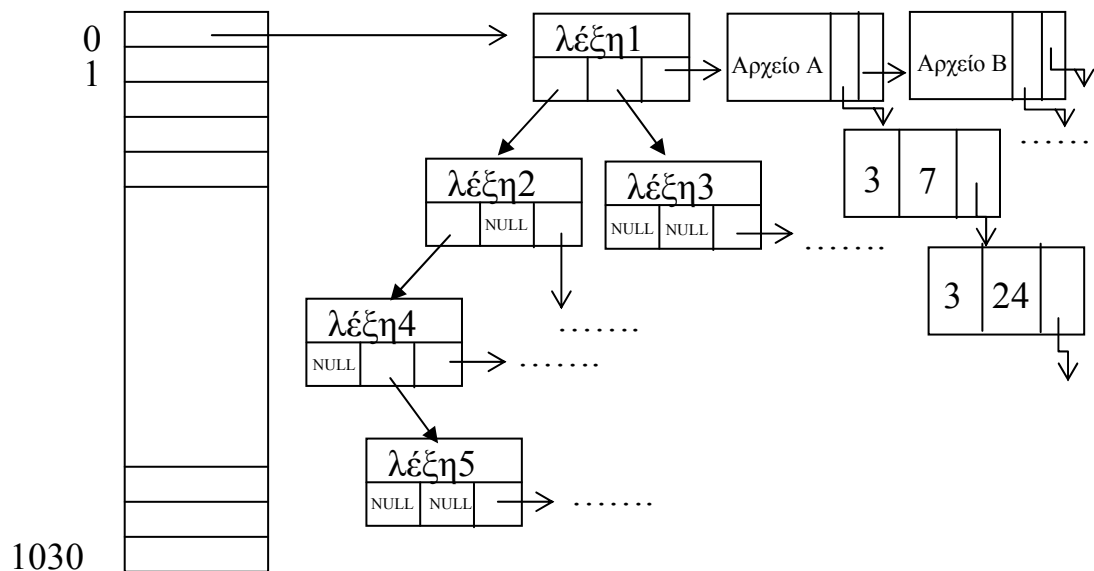


## ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΕΣ ΤΕΧΝΙΚΕΣ 2<sup>η</sup> ΑΣΚΗΣΗ 2013-2014

**Σημείωση:** Οι ασκήσεις αποτελούν μέρος της εργαστηριακής εξάσκησης στο μάθημα “Προγραμματιστικές Τεχνικές”. Η υποβολή του σχετικού προγράμματος έμμεσα δηλώνει ότι είσθε 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 μια λίστα από αρχεία κειμένων και θα προσδιορίζει τις θέσεις εμφάνισης κάθε λέξης, δηλαδή τις δυάδες γραμμής-στήλης εμφάνισης του πρώτου χαρακτήρα της λέξης για κάθε αρχείο. (Θεωρούμε ότι οι λέξεις μπορεί να περιέχονται σε περισσότερα του ενός αρχεία.) Ως κείμενο του κάθε αρχείου θεωρείται μία ακολουθία χαρακτήρων, ενώ ως λέξη θεωρείται μία ακολουθία χαρακτήρων ενός κειμένου που δεν περιέχει κενό και βρίσκεται ανάμεσα σε δύο από τα ακόλουθα: κενό, αλλαγή γραμμής, αρχή ή τέλος αρχείου. Το πρόγραμμά σας θα αποθηκεύει και θα αναζητά τις θέσεις αυτές με χρήση Πινάκων Κατακερματισμού. 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 του αρχείου A, στη γραμμή 3, στήλη 7 και, στη γραμμή 3, στήλη 24.

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

Το πρόγραμμά σας θα διαβάζει αρχικά το κείμενο από κάθε αρχείο και θα κατασκευάζει τη δομή που φαίνεται στο παραπάνω σχήμα. Για τον έλεγχο, το πρόγραμμά σας θα διαβάζει μία λέξη αναζήτησης και στη συνέχεια θα εκτυπώνει τη λέξη αναζήτησης και όλες τις θέσεις εμφάνισής της σε κάθε αρχείο, 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<sub>i</sub>: είναι η ASCII τιμή, του i<sup>th</sup> χαρακτήρα της λέξης

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

B: είναι ίσο με 256 = 2<sup>8</sup> (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. Στα αρχεία εισόδου θεωρούμε ότι έχουν αφαιρεθεί τα σημεία στίξης (. , ; “! κ.λπ.) και ότι μεταξύ των λέξεων μπορεί να υπάρχουν περισσότερα του ενός κενά, ή αλλαγή γραμμής.

### Παράδειγμα

#### Διάβασμα Αρχείων

Αρχείο A

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

Αρχείο B

Pools of sorrow, waves of joy are drifting through my open mind  
Possessing and caressing me

#### Έλεγχος Προγράμματος

##### Είσοδος

of

##### Έξοδος

of

Αρχείο A (3,7)

Αρχείο A (3,24)

.....

Αρχείο B (1, 7)

.....

Αρχείο C (xx, yy)

.....