



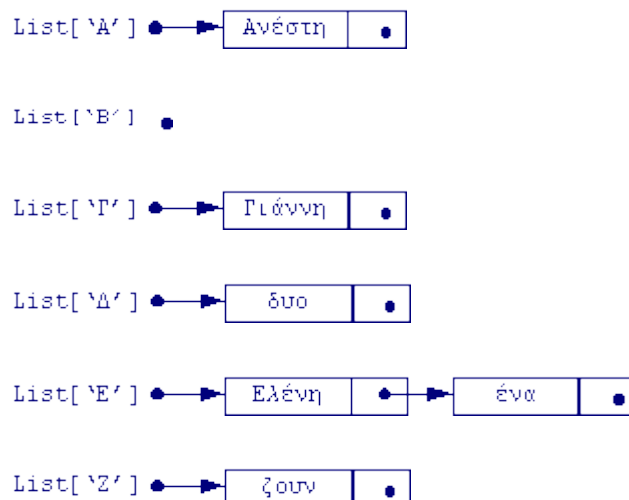
## 4.7 Εφαρμογή Συνδεδεμένων Λιστών: Αλφαβητικό ευρετήριο κειμένου

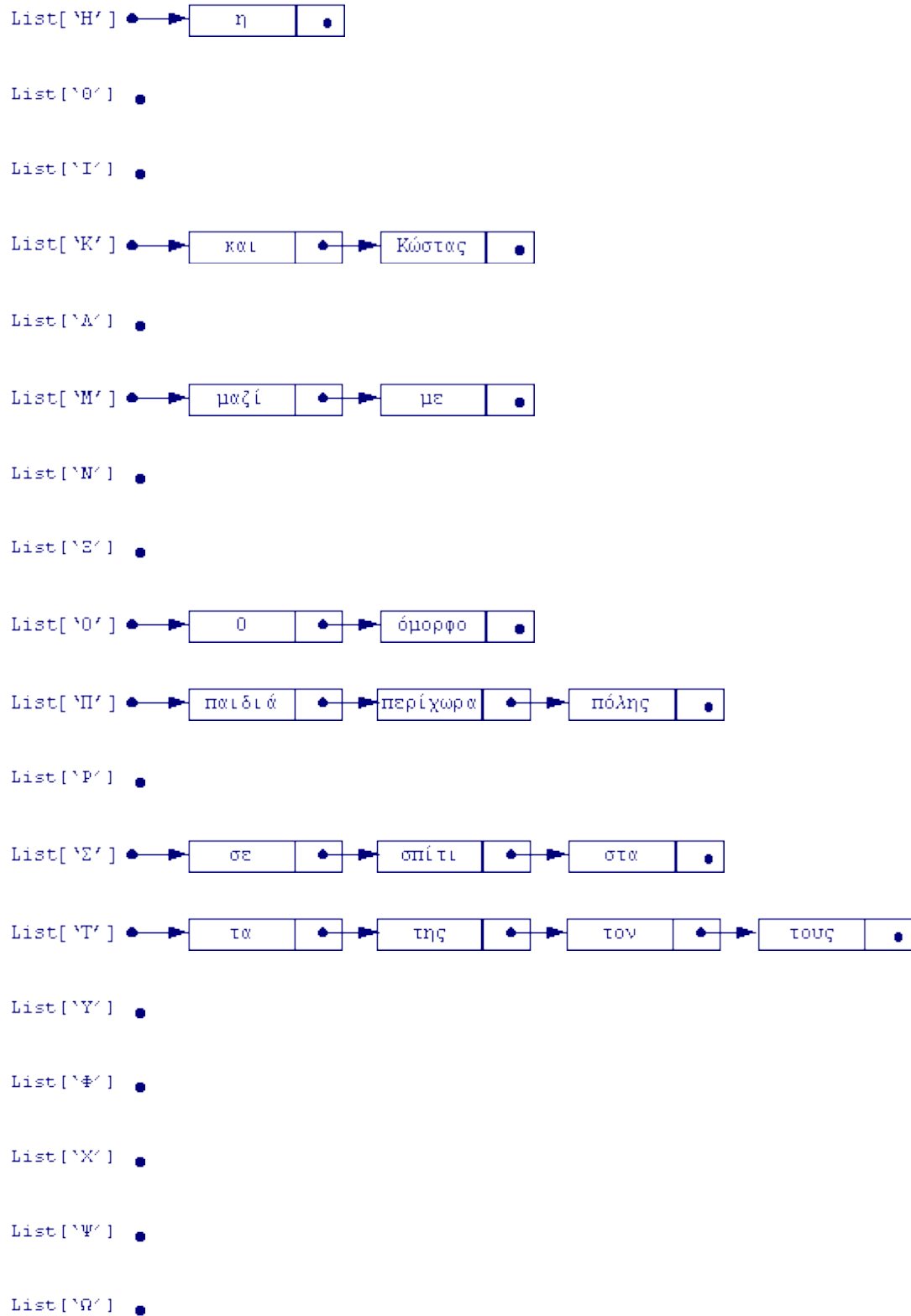
Ένα *αλφαβητικό ευρετήριο κειμένου* (*text concordance*) είναι μια αλφαβητική λίστα όλων των διαφορετικών λέξεων από ένα κομμάτι κειμένου. Το γεγονός ότι οι λέξεις σε ένα αλφαβητικό ευρετήριο είναι ταξινομημένες με αλφαβητική σειρά σημαίνει ότι ένα αλφαβητικό ευρετήριο είναι μια ταξινομημένη λίστα. Για να κατασκευάσουμε ένα τέτοιο ευρετήριο, ξεκινάμε με μια άδεια λίστα. Κάθε λέξη διαβάζεται και τοποθετείται στη σωστή θέση μέσα στη λίστα, εφόσον δεν υπάρχει ήδη στη λίστα. Προφανώς, μπορεί να χρειαστεί η εισαγωγή λέξεων να γίνει σε οποιοδήποτε σημείο της λίστας, οπότε χρειαζόμαστε μια συνδεδεμένη λίστα.

Για κάθε λέξη που διαβάζεται πρέπει να γίνει σειριακή αναζήτηση της συνδεδεμένης λίστας ξεκινώντας από τον πρώτο κόμβο. Ωστόσο, αν πρόκειται για μεγάλο κείμενο, το ευρετήριο μπορεί να γίνει τόσο μεγάλο που η αναζήτηση μιας και μόνο λίστας δεν είναι αποδοτική. Για να μειωθεί ο χρόνος αναζήτησης, μπορούμε να χρησιμοποιήσουμε πολλές μικρές συνδεδεμένες λίστες. Για το συγκεκριμένο πρόβλημα μπορούμε να κατασκευάσουμε μια συνδεδεμένη λίστα για κάθε γράμμα της αλφαβήτου, δηλαδή μια λίστα με λέξεις που ξεκινάν από "Α", μια λίστα με λέξεις που ξεκινάν από "Β", κ.ο.κ. Επομένως, συνολικά θα χρειαστούν 24 συνδεδεμένες λίστες καθώς και 24 δείκτες, ένας για κάθε λίστα, οπότε μπορούμε να χρησιμοποιήσουμε έναν πίνακα δεικτών με δείκτες από "Α" ως "Ω". Έστω, για παράδειγμα ότι έχουμε το παρακάτω κείμενο:

*Ο Κώστας και η Ελένη ζουν σε ένα όμορφο σπίτι στα περίχωρα της πόλης μαζί με τα δυο τους παιδιά, τον Γιάννη και τον Ανέστη.*

Οι παρακάτω συνδεδεμένες λίστες δείχνουν πώς μπορεί να αποθηκευτεί το αλφαβητικό ευρετήριο αυτού του κειμένου.





Για να εμφανιστεί το ευρετήριο αρκεί να διασχίσουμε τις 24 αυτές συνδεδεμένες λίστες.

Παρακάτω φαίνεται ο αλγόριθμος για την κατασκευή ενός τέτοιου ευρετηρίου:

**Αλγόριθμος κατασκευής ενός Αλφαβητικού Ευρετηρίου Κειμένου**

**/\*Είσοδος:** Ένα αρχείο κειμένου.

**Λειτουργία:** Κατασκευάζει ένα αλφαβητικό ευρετήριο κειμένου από ένα έγγραφο που είναι αποθηκευμένο σε ένα αρχείο. Το ευρετήριο αποθηκεύεται σε μια δομή δεδομένων που αποτελείται από έναν πίνακα συνδεδεμένων λιστών, *List*, όπου *List[Ch]* είναι μια ταξινομημένη συνδεδεμένη λίστα λέξεων που ξεκινούν από το γράμμα *Ch*.

**Έξοδος:** Μια λίστα των διαφορετικών λέξεων του αρχείου.\*/\*

**1. Για Ch από 'Α' μέχρι 'Ω'**

Δημιούργησε μια κενή συνδεδεμένη λίστα, *List[Ch]*

**Τέλος\_επανάληψης**

**2. Άνοιξε το αρχείο και πάρε την πρώτη λέξη Word**

**3. Όσο υπάρχουν λέξεις προς επεξεργασία επανάλαβε**

α. Ψάξε στη λίστα των λέξεων που ξεκινούν από το ίδιο γράμμα που ξεκινά και η *Word* για να δεις αν ήδη υπάρχει εκεί η λέξη αυτή

β. Αν δεν υπάρχει η λέξη *Word* στην αντίστοιχη λίστα των λέξεων τότε εισήγαγε την *Word* στη λίστα αυτή

γ. Πάρε την επόμενη λέξη *Word*

**Τέλος\_επανάληψης**

**4. Για Ch από 'Α' μέχρι 'Ω'**

**Αν** η *List[Ch]* δεν είναι κενή **τότε**

Διάσχισε τη λίστα εμφανίζοντας κάθε λέξη της

**Τέλος\_αν**

**Τέλος\_επανάληψης**

Η κατασκευή ενός τέτοιου αλφαβητικού ευρετηρίου υλοποιείται από το πρόγραμμα-πελάτη TextConcordance.c, εφόσον υπάρχει και ένα αρχείο κειμένου atext.txt, με βάση το οποίο να κατασκευαστεί το αλφαβητικό ευρετήριο. Για την συνδεδεμένη λίστα χρησιμοποιείται η υλοποίηση με δείκτες και το κατάλληλο αρχείο διασύνδεσης και υλοποίησης είναι το LPWordADT.h και το LPWordADT.c αντίστοιχα, όπου τα στοιχεία της λίστας είναι τύπου πίνακα χαρακτήρων ή αλφαριθμητικό:

```
typedef char ListElementType[MaxWord];
```

όπου *MaxWord* είναι σταθερά και δηλώνει το μέγιστο μέγεθος της λέξης.

