

Δομές Δεδομένων και Αρχείων

Επεξεργασία Αρχείων

Φλέγγας Γιώργος
2014030161

1)Κατασκευή της δομής δεδομένων στην Κεντρική Μνήμη :

Σε αυτό το κομμάτι λαμβάνουμε ως είσοδο τα 3 αρχεία και από αυτά διαβάζουμε όλες τις λέξεις και τις μεταφέρουμε στο λεξικό και ταυτόχρονα δημιουργούμε το ευρετήριο . Για την υλοποίηση βασιστήκαμε σε 2 classes τις Word(String name, **int** page) και Reference(String fileName, **long** position, **int** page). Το λεξικό(Lexicon) αποτελεί μια Object Array με αρχικό μέγεθος 9 το οποίο ανάλογα με τις ανάγκες που προκύπτουν αλλάζει κατά την διάρκεια της εκτέλεσης, ενώ το ευρετήριο (Index) είναι μια arraylist. Για την ανάγνωση από ένα αρχείο χρησιμοποιούμε την μέθοδο AddFile η οποία έχει ως όρισμα το όνομα του αρχείου που θα διαβάσουμε . Κατά την εκτέλεση διαβάζουμε μια γραμμή την φορά, την οποία μεταφέρουμε στην κεντρική μνήμη , όπου στην συνέχεια διαβάζουμε μια λέξη κάθε φορά μέχρι να τελειώσει η γραμμή και να συνεχίσουμε στην επόμενη. Αν η λέξη που διαβάζουμε υπάρχει ήδη στο λεξικό τότε απλά επεκτείνουμε τις πληροφορίες της υπάρχουσας έγγραφης στο ευρετήριο . Σε αντίθετη περίπτωση δημιουργούμε νέες έγγραφες στο λεξικό αλλά και στο ευρετήριο (οι έγγραφες στο ευρετήριο γίνονται σε Object Array μέσα στην arraylist και αντιστοιχεί από μια Object Array για κάθε μια λέξη του λεξικού). Εν τέλει αφού διαβάσουμε και τα 3 αρχεία χρησιμοποιούμε την μέθοδο sorting για να ταξινομήσουμε αλφαβητικά το λεξικό.

2)Κατασκευή Δομής Αρχείου :

Για να μεταφέρουμε το λεξικό στο αρχείο χρησιμοποιήσουμε την μέθοδο GenerateLex, η οποία αντιγράφει επαναληπτικά ένα Object Word κάθε φορά στο αρχείο μέχρι να έχει μεταφέρει όλο το λεξικό έχοντας υπόψιν ότι η σελίδα αρχείου έχει μέγεθος 128 bytes. Για να βεβαιωθούμε ότι το περιεχόμενο κάθε σελίδας είναι ≤ 128 bytes έχουμε μια μεταβλητή (size) τύπου long , στην οποία πριν μεταφέρουμε τη λέξη στο αρχείο προσθέτουμε το μέγεθος του Word και αν δεν ξεπερνάει τα 128 bytes τότε την καταχωρούμε στην σελίδα αυτή αλλιώς, αλλάζουμε σελίδα μηδενίζουμε το size δίνοντας του για αρχική τιμή το μέγεθος της λέξης και καταχωρούμε την λέξη στην νέα αυτή σελίδα.

Αντίστοιχα για το ευρετήριο χρησιμοποιήσουμε την μέθοδο GenerateInd, η οποία μεταφέρει ένα ένα τα αντικείμενα των Object Array που βρίσκονται μέσα στην arraylist.

3)Αναζήτηση στη δομή δεδομένων :

Αρχικά ο χρήστης δίνει μια λέξη μέσω του πληκτρολογίου η οποία περνά ως όρισμα στην μέθοδο readFromLexicon που διαβάζει από το αρχείο Lexicon objects τύπου Word και αναζητάει αυτό το οποίο αντιστοιχεί η λέξη του χρήστη. Όταν την βρει δίνει στο p(που έχει αρχική τιμή -1) την τιμή της σελίδας που θα βρούμε στο ευρετήριο την καταχωρισμένη για την λέξη αυτή και τερματίζει την αναζήτηση. Ακόμα υπολογίζει τον αριθμό προσβάσεων στο δίσκο μέχρι να βρεθεί η λέξη και τον επιστρέφει. Στην συνέχεια εκτυπώνετε ο αριθμός αυτός . Μετά καλούμε την μέθοδο locateOnIndex με την οποία, αν το p δεν είναι -1, βρίσκουμε όλες τις έγγραφες για την λέξη που ζητήθηκε και τις εκτυπώνουμε με το κατάλληλο μήνυμα, ενώ παράλληλα κρατάμε τον αριθμό προσβάσεων στον δίσκο, που εκτυπώνετε στην συνέχεια.

Εν τέλη εκτυπώνουμε τον συνολικό αριθμό προσβάσεων στο δίσκο που πραγματοποιήθηκαν από το πρόγραμμα.

Το ακόλουθο είναι ένα παράδειγμα από την εκτέλεση των 2 προγραμμάτων :

```
Importing from :Kennedy
Importing from :MartinLutherKing
Importing from :Obama
Sorting Lexicon...
Generating Lexicon File...
Lexicon size: 15682 bytes
Generating Index File...
Index size: 102183 bytes
Done!
```

```
Which word do you want to search?
end
Number of times accessed the drive during readFromLexicon:388
Text:Kennedy contains the word:END at the place:509
Text:MartinLutherKing contains the word:END at the place:714
Text:MartinLutherKing contains the word:END at the place:3355
Text:Obama contains the word:END at the place:2140
Text:Obama contains the word:END at the place:6280
Text:Obama contains the word:END at the place:9112
Text:Obama contains the word:END at the place:13600
Number of times accessed the drive during locateOnIndex:546
Total Number of times accessed the drive:934
```