

Άσκηση 1 ΠΛΗ202

Ιωαννίδης Χρήστος 2018030006

Μέρος 1ο:

Για το πρώτο μέρος χρειάστηκε η υλοποίηση μιας διπλά διασυνδεδεμένης λίστας η οποία περιεχί συναρτήσεις για την επεξεργασία της(διαγραφή,προσθήκη κ.λ.π). Η λίστα περιεχί κόμβους με δυνατότητα ενός αποθήκευσης String.

Η φόρτωση στην λίστα γίνεται μεσώ ξεχωριστής κλάσης που χρησιμοποιείται για input από το αρχείο. Με την χρήση αντίστοιχης συνάρτησης το αρχείο αποθηκεύεται γραμμή-γραμμή στην λίστα αφού πρώτα κάθε γραμμή περάσει έλεγχο και επεξεργασία έτσι ώστε το μέγιστο μέγεθος της να είναι σύμφωνο με αυτό ορισμένο από τον χρήστη.

Η επικοινωνία του χρήστη και του προγράμματος γίνεται μεσώ ενός βρόχου που διαβάσει έναν χαρακτήρα από τον χρήστη και εμπεριέχει ένα switch statement μέσα από το οποίο προσδιορίζεται και η συνάρτηση που θα κληθεί.

Ο βρόχος τερματίζεται όταν δοθεί η αντίστοιχη εντολή από τον χρήστη και επίσης υπάρχει δυνατότητα αποθήκευσης των αλλαγών που έχουν γίνει στο αρχείο.

Μέρος 2ο:

Για την δημιουργία του αρχείου δεικτοδότησης έχει υλοποιηθεί η αντίστοιχη συνάρτηση που παίρνει γραμμή γραμμή το κείμενο, το χωρίζει σε λέξεις και της αποθηκεύει μαζί με τον αριθμό γραμμής του σε ένα array list τύπου Word. Το Word είναι μια κλάση που αποθηκεύει ένα string και έναν ακέραιο και υλοποιεί το interface comparable. Η εντολή sort χρησιμοποιείται για την αλφαβητική ταξινόμηση του αρχείου. Μέτα την ταξινόμηση δεσμεύεται χώρος στην μνήμη για την δημιουργία ενός array από bytes στο οποίο αποθηκεύονται οι λέξεις με τον αριθμό γραμμής τους. Για λογούς καλύτερης απόδοσης το array αποθηκεύεται σε ένα αρχείο κομμάτι-κομμάτι με την χρήση ενός buffer.

Η ανάγνωση του αρχείου από τον δίσκο γίνεται με την χρήση buffer. Το αρχείο διαβάζεται σελίδα σελίδα και διαχωρίζεται σε λέξεις και αριθμούς γραμμής που μετά αποθηκεύονται σε μια λίστα από words.

Για την σειριακή αναζήτηση το αρχείο διαβάζεται σελίδα-σελίδα και γίνεται σύγκριση των περιεχόμενων της κάθε σελίδας με την λέξη προς αναζήτηση μέχρι να ανευρεθεί η λέξη η να διαβαστεί όλο το αρχείο. Αν βρεθεί η λέξη η αναζήτηση συνεχίζεται για τη περίπτωση που υπάρχει παραπάνω φορές. Επειδή η το αρχείο είναι ταξινομημένο οι ίδιες λέξεις είναι αποθηκευμένες διπλά η μια στην άλλη και

έτσι όταν βρει την πρώτη φορά την λέξη ενεργοποιείται ένας έλεγχος που σταματάει την αναζήτηση όταν σταματήσει να εμφανίζεται η επιθυμητή λέξη.

Για την δυαδική αναζήτησης η ανάγνωση του αρχείου ξεκινάει από την μεσαία σελίδα και η αναζήτηση συνεχίζεται φορτώνοντας σελίδες με βάση τον αλγόριθμο της δυαδικής αναζήτησης όπου συγκρίνονται τα περιεχόμενα στις άκρες της υπάρχουσας σελίδας για τον προσδιορισμό την επόμενης. Στην περίπτωση που βρεθεί η λέξη σε ακριανή θέση στη σελίδα ελέγχεται και η διπλανή της.

Λάθη στο πρόγραμμα

Στην δυαδική αναζήτηση υπάρχει μικρή περίπτωση να μην βρεθεί μια λέξη που υπάρχει στο αρχείο. Επίσης αν βρεθεί η ζητούμενη λέξη και στις δυο άκρες της σελίδας ελέγχονται οι σελίδες μόνο από την πλευρά της μιας άκρης.

Κατα την δημιουργία του index αρχείου οι τελευταίες λέξεις της λίστας δεν αποθηκεύονται.

Για την διπλά διασυνδεδεμένη λίστα χρησιμοποιήθηκε υλοποίηση από αυτήν την ιστοσελίδα <https://dzone.com/articles/doubly-linked-list-in-java>

Πείραμα:

testfile1.txt
μέγεθος του αρχείου:13656
αριθμός σελιδών:106

προσβάσεις στον δίσκο

testfile1.txt	Technical	Rector	laboratories	Venetian	Ανεπιτυχής Αναζήτηση(M.O)
εξαντλητική	42	35	84	50	55
δυαδική	3	5	5	5	5

testfile_x2.txt
μέγεθος του αρχείου:27384
αριθμός σελιδών:213

προσβάσεις στον δίσκο

testfile_x2.txt	Technical	Rectors	laboratories	Venetian	Ανεπιτυχής Αναζήτηση(M.O)
εξαντλητική	85	70	169	100	111
δυαδική	3	7	5	5	6

testfile_x5.txt
μέγεθος του αρχείου:68568
αριθμός σελιδών:535

προσβάσεις στον δίσκο

testfile_x5.txt	Technical	Rectors	laboratories	Venetian	Ανεπιτυχής Αναζήτηση(M.O)
εξαντλητική	215	177	423	252	278
δυναμική	3	8	7	6	7

testfile_x10.txt
μέγεθος του αρχείου:137208
αριθμός σελιδών:1071

προσβάσεις στον δίσκο

testfile_x10.txt	Technical	Rectors	laboratories	Venetian	Ανεπιτυχής Αναζήτηση(M.O)
εξαντλητική	431	355	847	505	557
δυναμική	4	9	9	8	8

testfile_x1000.txt
μέγεθος του αρχείου:
αριθμός σελιδών:

προσβάσεις στον δίσκο

testfile_x1000.txt	Technical	Rector	laboratories	Venetian	Ανεπιτυχής Αναζήτηση(M.O)
εξαντλητική	-	-	-	-	-
δυναμική	-	-	-	-	-

Από το πείραμα παρατηρείται ότι η δυναμική αναζήτηση είναι πολύ πιο αποδοτική από την σειριακή. Παρατηρείται επίσης ότι όσο μεγαλώνει το αρχείο δεικτοδότησης αυξάνονται και οι προσβάσεις στον δίσκο. Για το testfile_x1000.txt το πρόγραμμα δεν ξεκίνησε.