

Αναφορά 1^η Άσκησης για το μάθημα Δομές Δεδομένων και Αρχείων

Περιγραφή του Κώδικα

Δημιουργήθηκε η Κλάση Tuple, η οποία κάνει χρήση του interface Comparable, για να υλοποιηθούν οι εγγραφές του πίνακα δεικτοδότησης, με μορφή:

Λέξη Αριθμός της γραμμή στο κείμενο, που βρίσκεται η λέξη

Υλοποιήθηκε και η abstract μέθοδος compareTo για την συγκεκριμένη κλάση, ώστε να υπάρχει η δυνατότητα η αλφαβητική ταξινόμηση των εγγραφών, με βάση το πεδίο Λέξη.

Ως βάση την Κλάση Tuple, δημιουργήθηκε η Κλάση IndexingTable, η οποία χρησιμοποιώντας δύο ArrayList ίσου μήκους, το ένα με τις λέξεις του αρχείου txt και το άλλο με τους αριθμούς των γραμμών, δημιουργεί ένα Vector, που περιέχει τις εγγραφές Tuple και λειτουργεί ως ο πίνακας δεικτοδότησης του αρχείου txt, με δυνατότητα αλφαβητικής ταξινόμησης. Η ταξινόμηση απαιτείται για την εγγραφή του πίνακα σε αρχείο, ενώ απουσιάζει κατά την εμφάνιση του πίνακα στην οθόνη του χρήστη.

Για την παραμετροποίηση του προγράμματος πριν την μεταγλώττιση, δημιουργήθηκε η final Κλάση SizeConstants, η οποία περιέχει τις σταθερές για τα μεγέθη που χρειάζεται το πρόγραμμα. Κατά την εκκίνηση του προγράμματος, δημιουργείται και αρχικοποιείται το μοναδικό στιγμιότυπο, αυτής της κλάσης, καθολικά προσβάσιμο από τις άλλες κλάσεις του προγράμματος.

Με σκοπό την εγγραφή του αρχείου δεικτοδότησης, δεδομένου του αρχείου txt που υπάρχει στην διπλά διασυνδεδεμένη λίστα του προγράμματος, δημιουργήθηκαν η Κλάση IndexFileBufferFileWriter. Η συγκεκριμένη κλάση, χρησιμοποιεί το στιγμιότυπο του IndexingTable και μετατρέπει τις εγγραφές σε byte μορφή, τις ομαδοποιεί και τις αποθηκεύει ως byte array, ένα array ανά σειρά, στο αρχείο δεικτοδότησης.

Προκειμένου, το αρχείο txt εισόδου, να αποθηκεύεται στην μνήμη του προγράμματος ως Linked List, δημιουργήθηκε η Κλάση FileScanner. Κάνοντας χρήση της Κλάσης Scanner, διαβάζεται γραμμή προς γραμμή το αρχείο εισόδου και η κάθε γραμμή τοποθετείται ως εγγραφή στην διπλά διασυνδεδεμένη λίστα, διατηρώντας την αρίθμηση γραμμών του αρχείου εισόδου.

Για να υλοποιηθούν οι ζητούμενες δυνατότητες της γραμμής εντολών, δημιουργήθηκε η Κλάση CMD. Αφού διαβαστεί η επιθυμητή δυνατότητα από τον χρήστη, εκτελείται το αντίστοιχο τμήμα κώδικα και εμφανίζεται κατάλληλο μήνυμα στην οθόνη του χρήστη.

Καταλήγοντας, δημιουργήθηκε η Κλάση και η συνάρτηση Main, η οποία, αφού δεχτεί το αρχείο εισόδου από τον χρήστη, δημιουργεί το CMD στιγμιότυπο και ενεργοποιεί την διαδικασία επιλογής εντολών και αλληλεπίδρασης με το πρόγραμμα, μέχρις ότου ο χρήστης δώσει την εντολή q ή x, για να τερματίσει επιτυχώς το πρόγραμμα.

Οδηγίες Μετάφρασης:

Το πρόγραμμα χρησιμοποιεί βασικές βιβλιοθήκες της Java, οπότε δεν απαιτείται τίποτα περισσότερο από την τυπική διαδικασία μεταγλώττισης των αρχείων με επέκταση . java. Το αρχείο εισόδου πρέπει να βρίσκεται στο φάκελο CommandLine, που περιέχεται ως μοναδικός φάκελος στο zip.

Πληροφορίες Προγράμματος

Έχουν υλοποιηθεί όλες οι ζητούμενες δυνατότητες του προγράμματος και όχι κάτι επιπλέον. Δεν υπάρχουν τμήματα του κώδικα που να εμφανίζουν προβλήματα.

Για την υλοποίηση του κώδικα, χρησιμοποιήθηκαν πληροφορίες άλλα όχι έτοιμα τμήματα κώδικα από τις παρακάτω ιστοσελίδες:

Oracle docs: <https://docs.oracle.com/javase/8/docs/api/allclasses-noframe.html>

Stack Overflow: <https://stackoverflow.com/>

Tutorials Point: <https://www.tutorialspoint.com/java/index.htm>

Μετρήσεις και Πορίσματα Πειράματος

Παρατήρηση: Οι τιμές των δεκαδικών αποτελεσμάτων είναι στρογγυλοποιημένες στο πρώτο δεκαδικό ψηφίο.

Στοιχεία Αρχείων Εισόδου

Όνομα Αρχείου	Μέγεθος σε bytes	Αριθμός Σελίδων
testfile1.txt	68000	131
testfile_x2.txt	135.000	262
testfile_x5.txt	337.000	655
testfile_x10.txt	675.000	1.310
testfile_x1000.txt	68.476.000	131.000

Μετρήσεις για το αρχείο testfile1.txt

Αποτελέσματα Επιτυχών Αναζητήσεων

Λέξη	Εξαντλητική Αναζήτηση (Προσβάσεις)	Δυναμική Αναζήτηση (Προσβάσεις)
Rectors	41	7
laboratories	95	5
Technical	49	11
Venetian	56	7

Αποτελέσματα Ανεπιτυχών Αναζητήσεων

Εξαντλητική Αναζήτηση (Μέσος όρος προσβάσεων)	Δυναμική Αναζήτηση (Μέσος όρος προσβάσεων)
133,7	6,2

Μετρήσεις για το αρχείο testfile_x2.txt

Αποτελέσματα Επιτυχών Αναζητήσεων

Λέξη	Εξαντλητική Αναζήτηση (Προσβάσεις)	Δυναμική Αναζήτηση (Προσβάσεις)
Rectors	80	8
laboratories	188	7
Technical	96	14
Venetian	111	12

Αποτελέσματα Ανεπιτυχών Αναζητήσεων

Εξαντλητική Αναζήτηση (Μέσος όρος προσβάσεων)	Δυναμική Αναζήτηση (Μέσος όρος προσβάσεων)
227,2	7,1

Μετρήσεις για το αρχείο testfile_x5.txt

Αποτελέσματα Επιτυχών Αναζητήσεων

Λέξη	Εξαντλητική Αναζήτηση (Προσβάσεις)	Δυναμική Αναζήτηση (Προσβάσεις)
Rectors	198	12
laboratories	468	12
Technical	237	23
Venetian	274	13

Αποτελέσματα Ανεπιτυχών Αναζητήσεων

Εξαντλητική Αναζήτηση (Μέσος όρος προσβάσεων)	Δυναμική Αναζήτηση (Μέσος όρος προσβάσεων)
567,8	9,3

Μετρήσεις για το αρχείο testfile_x10.txt

Αποτελέσματα Επιτυχών Αναζητήσεων

Λέξη	Εξαντλητική Αναζήτηση (Προσβάσεις)	Δυναμική Αναζήτηση (Προσβάσεις)
Rectors	395	12
laboratories	935	13
Technical	473	39
Venetian	547	17

Αποτελέσματα Ανεπιτυχών Αναζητήσεων

Εξαντλητική Αναζήτηση (Μέσος όρος προσβάσεων)	Δυναμική Αναζήτηση (Μέσος όρος προσβάσεων)
1.135,5	9,7

Μετρήσεις για το αρχείο testfile_x1000.txt

Αποτελέσματα Επιτυχών Αναζητήσεων

Λέξη	Εξαντλητική Αναζήτηση (Προσβάσεις)	Δυναμική Αναζήτηση (Προσβάσεις)
Rectors	39.401	212
laboratories	93.401	211
Technical	47.201	3.207
Venetian	54.601	809

Αποτελέσματα Ανεπιτυχών Αναζητήσεων

Εξαντλητική Αναζήτηση (Μέσος όρος προσβάσεων)	Δυναμική Αναζήτηση (Μέσος όρος προσβάσεων)
113.533,5	16,8

Με βάση τα αποτελέσματα του πειράματος, διαφαίνεται ότι η δυναμική αναζήτηση είναι πιο αποδοτική έναντι της εξαντλητικής αναζήτησης. Η παρατήρηση αυτή συνάγει με την θεωρία, καθώς η εξαντλητική αναζήτηση έχει γραμμική πολυπλοκότητα $O(n)$, ενώ η δυναμική αναζήτηση έχει λογαριθμική πολυπλοκότητα $O(\log n)$.

Επίσης, με βάση τα αποτελέσματα, καταλήγω στο συμπέρασμα ότι η αύξηση του μεγέθους του αρχείου εισόδου, προκαλεί γραμμική αύξηση του του μέσου αριθμού προσβάσεων στην περίπτωση της εξαντλητικής αναζήτησης, ενώ στην περίπτωση της δυναμικής αναζήτησης, ο μέσος αριθμός προσβάσεων παραμένει πρακτικά αμετάβλητος. Παρακάτω σας παραθέτω την γραφική παράσταση για τα τέσσερα αρχεία εισόδου, παρουσιάζοντας την σχέση μεγέθους και μέσου αριθμού προσβάσεων, ανά είδος αναζήτησης.

Τέλος, συμπεραίνω ότι η θέση της λέξης μέσα στο αρχείο δεικτοδότησης, επηρεάζει αισθητά την εξαντλητική αναζήτηση, ενώ ελάχιστα την δυναμική αναζήτηση.

Μέγεθος Αρχείου - Μέσος Αριθμός Προσβάσεων

