



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ- Η.Μ.Μ.Υ ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΡΧΕΙΩΝ – ΠΛΗ 201

ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ 3

ΑΤΟΜΙΚΗ ΑΝΑΦΟΡΑ ΥΛΟΠΟΙΗΣΗΣ

Αντώνης Χρυσοφάκης 2015030116

**Αναζήτηση σε Αρχεία κειμένου με χρήση Δεικτοδότησης
(indexing)**

Το πρόγραμμα χρησιμοποιεί τις ακόλουθες δομές δεδομένων:

- B+ Tree (Λεξικό): Αποθηκεύει τις λέξεις ως κλειδιά και τις αντίστοιχες θέσεις που έχουν στο ευρετήριο, ως τιμές, διευκολύνοντας την αποτελεσματική αναζήτηση.
- Συνδεδεμένη λίστα (Ευρετήριο): Αποθηκεύει πρόσθετες πληροφορίες σχετικά με τις λέξεις, όπως το αρχείο και τους αριθμούς γραμμής όπου βρίσκονται.

Ροή εργασιών προγράμματος:

Το πρόγραμμα ξεκινά με την εμφάνιση ενός μενού:

A) Import file

B) Search a word from file

C) Search 100 words.

Ο χρήστης μπορεί να επιλέξει να εισαγάγει ένα αρχείο, να αναζητήσει μια λέξη ή να πραγματοποιήσει μια αναζήτηση σε ένα σύνολο 100 λέξεων.

A) Κατά την εισαγωγή ενός αρχείου, το πρόγραμμα διαβάζει το αρχείο, ελέγχει εάν κάθε λέξη υπάρχει στο δέντρο B+ και ενημερώνει ανάλογα το δέντρο B+ και τη συνδεδεμένη λίστα.

B) Η αναζήτηση μιας λέξης περιλαμβάνει την αναζήτηση της λέξης στο δέντρο B+ και την εμφάνιση σχετικών πληροφοριών από το ευρετήριο εάν βρεθεί.

Γ) Η εκτέλεση 100 αναζητήσεων περιλαμβάνει τη φόρτωση 50 λέξεων από κάθε αρχείο, την ανακάτεμα και τον συνδυασμό τους και την αναζήτηση για κάθε λέξη στο συνδυασμένο σύνολο χρησιμοποιώντας το δέντρο B+. Όταν καλείται αυτή η διαδικασία δημιουργούνται 2 καινούρια δέντρα (τάξης 10 και 20) και για κάθε δέντρο η συνδεδεμένη λίστα που του αντιστοιχεί.

A) Εισαγωγή αρχείων (functionA)

- Ζητείται από τον χρήστη να καθορίσει ένα αρχείο για εισαγωγή.
- Το πρόγραμμα διαβάζει το αρχείο, λέξη προς λέξη, και ελέγχει εάν κάθε λέξη υπάρχει στο δέντρο B+.
- Εάν η λέξη δεν βρεθεί, προστίθεται στο δέντρο B+ μαζί με την αντίστοιχη θέση της στο ευρετήριο. Επιπλέον, η λέξη και οι πληροφορίες θέσης προσαρτώνται στη συνδεδεμένη λίστα.
- Εάν η λέξη υπάρχει ήδη στο δέντρο B+, οι πληροφορίες θέσης ενημερώνονται στη συνδεδεμένη λίστα.

B) Αναζήτηση λέξης (functionB)

- Ο χρήστης καλείται να εισαγάγει μια λέξη για αναζήτηση.
- Το πρόγραμμα αναζητά τη λέξη στο δέντρο B+.
- Εάν βρεθεί η λέξη, το πρόγραμμα ανακτά την αντίστοιχη θέση στο ευρετήριο από το δέντρο B+ και εμφανίζει τις σχετικές πληροφορίες από τη συνδεδεμένη λίστα.
- Εάν η λέξη δεν βρεθεί, εμφανίζεται ένα κατάλληλο μήνυμα.

C) Αναζήτηση 100 λέξεων (functionC)

- Οι λέξεις από κάθε αρχείο φορτώνονται στο δέντρο B+ και στη συνδεδεμένη λίστα και ένα υποσύνολο 50 λέξεων επιλέγεται τυχαία από κάθε αρχείο.
- Οι επιλεγμένες λέξεις συνδυάζονται σε μια ενιαία λίστα.
- Για κάθε λέξη στη συνδυασμένη λίστα, το πρόγραμμα αναζητά το δέντρο B+.
- Το πρόγραμμα υπολογίζει στατιστικά στοιχεία για το δέντρο B+, συμπεριλαμβανομένου του αριθμού των κόμβων και των κλειδιών που έχουν πρόσβαση κατά τη διάρκεια της αναζήτησης για κάθε λέξη.

Τέλος, το πρόγραμμα εμφανίζει τον μέσο αριθμό κόμβων στους οποίους έχετε πρόσβαση και τον μέσο αριθμό συγκρίσεων μεταξύ κλειδιών για ολόκληρο το σύνολο των 100 λέξεων, χωριστά για δύο διαφορετικές σειρές δέντρων B+ (τάξης 10 και 20).

Αποτελέσματα 100 αναζητήσεων σε πίνακα:

A1. Μέσος αριθμός προσβάσεων κόμβων στο Ευρετήριο (B+-tree) για αναζήτηση Τάξη M=10	A2. Μέσος αριθμός συγκρίσεων με κλειδιά για αναζήτηση Τάξη M=10	B1. Μέσος αριθμός προσβάσεων κόμβων στο Ευρετήριο (B+-tree) για αναζήτηση Τάξη M=20	B2. Μέσος αριθμός συγκρίσεων με κλειδιά για αναζήτηση Τάξη M=20
2.96	1.96	14.27	18.64

Όταν ο χρήστης επιλέγει τις 100 αναζητήσεις τα A1 και B1 είναι πάντα τα ίδια.

Συνολικά, οι διαφορές υποδεικνύουν ότι το δέντρο B+ με τάξη 20 απέδωσε καλύτερα από αυτό με τάξη 10 όσον αφορά την αποτελεσματικότητα. Είχε πρόσβαση σε λιγότερους κόμβους και, παρόλο που είχε πρόσβαση σε περισσότερα κλειδιά κατά μέσο όρο, η συνολική βελτίωση της απόδοσης προτείνει μια πιο βελτιστοποιημένη δομή για την αποθήκευση και την αναζήτηση των λέξεων.

Αν το Λεξικό και το Ευρετήριο ήταν αποθηκευμένα στον δίσκο, τότε θα υπάρχουν ζητήματα απόδοσης λόγω των σχετικά πιο αργών λειτουργιών του δίσκου. Ωστόσο, επιτρέπει επίσης τη δυνατότητα χειρισμού μεγαλύτερων συνόλων δεδομένων που μπορεί να υπερβαίνουν τη διαθέσιμη χωρητικότητα μνήμης.