

ΠΡΩΤΗ ΕΡΓΑΣΙΑ ΒΑΣΕΙΣ-II

Να δημιουργήσετε ένα πρόγραμμα που υλοποιεί την ευρετηρίαση αρχείων με τη χρήση δέντρων B+. Θα διαβάζει ένα αρχείο κειμένου που περιέχει δεδομένα και θα δημιουργήσει το ευρετήριο, αντιμετωπίζοντας τους πρώτους 14 χαρακτήρες κάθε γραμμής στο αρχείο ως το κλειδί ευρετηρίασης. Επιτρέπονται οι εξής γλώσσες προγραμματισμού: C, C++, Python, C#, Java (αρκεί να δημιουργήσετε όλες τις δομές δεδομένων που θα χρησιμοποιήσετε από μόνοι σας και όχι μέσω βιβλιοθήκης)

Η δομή του ευρετηρίου έχει ως εξής:

Τα πρώτα 256 bytes: Όνομα του αρχείου κειμένου που έχετε κάνει ευρετηρίαση. Αυτό πρέπει να είναι κενό στα δεξιά. Το υπόλοιπο του αρχείου αποτελείται από δεδομένα ευρετηρίου που έχουν μέγεθος 1K μπλοκ, ανάλογα με τον τρόπο δομής ενός δέντρου B+. Αυτό σημαίνει ότι το πρόγραμμά σας θα πρέπει να διαβάζει ένα μπλοκ δεδομένων, να το χειρίζεται, και ενδεχομένως (εάν εισάγετε ζεύγη κλειδιών / δεικτών) **να το γράφει ξανά πίσω ως μπλοκ**. Χρησιμοποιήστε μια μακρά διεύθυνση εγγραφής (8-byte) για τους δείκτες σας. Αυτοί οι "δείκτες" είναι η μετατόπιση του byte στο αρχείο κειμένου (byte offset) του αρχείου δεδομένων. Σημειώστε ότι η δομή αναμειγνύει τα δεδομένα κειμένου (το κλειδί) και τα δυαδικά δεδομένα (ο δείκτης.) Η μετατροπή του δείκτη σε ψηφία κειμένου δεν επιτρέπεται.

Κάθε μία από τις λειτουργίες και οι παράμετροι της γραμμής εντολών δίνονται παρακάτω, υποθέτοντας ότι το όνομα του προγράμματος σας είναι BTREE:

1. Δημιουργία ευρετηρίου. Αυτό παίρνει μια παράμετρο (το όνομα του αρχείου txt και το όνομα του αρχείου ευρετηρίου που δημιουργείται). Εάν το αρχείο ευρετηρίου υπάρχει, το αντικαθιστά. Ενδιάμεσο output μπορεί να δημιουργηθεί κατά περίπτωση. Υπάρχουν δύο τρόποι δημιουργίας ενός ευρετηρίου. Ο πρώτος είναι να ξεκινήσετε με μια κενή δομή και να εισάγετε νέα κλειδιά. Ο δεύτερος είναι να χρησιμοποιήσετε ένα ταξινομημένο αρχείο εισόδου των κλειδιών και να δημιουργήσετε το δέντρο από αριστερά προς τα δεξιά. Αυτή η δεύτερη μέθοδος, ενώ είναι πιο δύσκολη, αλλά είναι πολύ πιο γρήγορη. Ωστόσο, είναι πολύ πιο δύσκολο να εφαρμοστεί και δεν χρειάζεται να το κάνετε αυτό. Ενδέχεται να υπάρχουν διπλά κλειδιά στο αρχείο εισόδου. Το πρόγραμμά σας θα πρέπει να τα απαριθμήσει και να εμφανίζει επίσης και τον αριθμό γραμμής του αρχείου εισόδου στο οποίο εμφανίζονται. Δεν πρέπει να εισαχθούν διπλότυπα κλειδιών στο ευρετήριο. Η γραμμή εντολών είναι:

BTREE - create <αρχείο εισόδου> <αρχείο ευρετηρίου>

2. Αναζήτηση εγγραφής με το κλειδί. Αυτή η λειτουργία πρέπει να εμφανίζει ολόκληρη την εγγραφή, συμπεριλαμβανομένου του κλειδιού, και να δίνει τη θέση της, σε bytes, μέσα στο αρχείο. Εάν το κλειδί δεν βρίσκεται στο ευρετήριό σας, το πρόγραμμα πρέπει να δώσει ένα μήνυμα που να ενημερώνει τον χρήστη για αυτό το γεγονός/αποτέλεσμα. Ο λόγος που έχουμε το όνομα του αρχείου_ευρετηρίου παρακάτω στην γραμμή εντολών είναι για να μπορείτε να χρησιμοποιείτε διαφορετικά ευρετήρια (πρωταρχικά και δευτερεύοντα). Η γραμμή εντολών είναι:

BTREE – find <όνομα αρχείου ευρετηρίου> <κλειδί>

για παράδειγμα, εάν εκτελέσατε αυτήν τη γραμμή εντολών:

BTREE – find indexfile.indx 100000

το πρόγραμμά σας πρέπει να επιστρέψει ένα μήνυμα ότι το κλειδί δεν βρέθηκε. Η γραμμή εντολών

BTREE – find indexfile.indx 645416

θα πρέπει να επιστρέψει:

Εγγραφή 2127, ΠΕΡΙΕΧΟΜΕΝΑ ΕΓΓΡΑΦΗΣ: 645416 "RECORD CONTENTS"

(Αυτό προϋποθέτει ότι το offset σε byte αυτής της εγγραφής από την αρχή του αρχείου είναι 2127).

3. Εισαγωγή μιας νέας εγγραφής. Όπως και με τη δημιουργία ενός νέου ευρετηρίου, τα πρώτα 14 byte πρέπει να περιέχουν ένα μοναδικό κλειδί. Αν το κλειδί δεν βρίσκεται στο ευρετήριο, γράψτε την εγγραφή στο τέλος του αρχείου δεδομένων (θυμηθείτε τη θέση, αφού θα το χρειαστείτε) και στη συνέχεια εισαγάγετε το κλειδί στη δομή ευρετηρίου με το δείκτη στην αρχή της νέας εγγραφής. Επιστρέψτε ένα μήνυμα που υποδεικνύει την επιτυχία και τη θέση της νέας εγγραφής στο αρχείο κειμένου. Αν το κλειδί βρίσκεται ήδη στο ευρετήριο, επιστρέψτε ένα μήνυμα για αυτό το αποτέλεσμα και μην εισάγετε την εγγραφή. Το πρόγραμμα πρέπει να εισάγει το κλειδί στο ευρετήριο και την εγγραφή στο αρχείο κειμένου. Τα περιεχόμενα της εγγραφής πρέπει να είναι σε εισαγωγικά, επειδή μπορεί να περιέχουν κενά και άλλα σημεία στίξης. Αυτό έχει την ακόλουθη γραμμή εντολών:

BTREE – insert <όνομα αρχείου ευρετηρίου> "νέα γραμμή κειμένου που θα εισαχθεί."

Για παράδειγμα, η ακόλουθη γραμμή εντολών θα προσπαθήσει να εισαγάγει μια εγγραφή στο αρχείο. Από το προηγούμενο παράδειγμα, γνωρίζετε ότι το κλειδί του είναι ήδη στο ευρετήριο σας, οπότε θα πρέπει να λάβετε ένα μήνυμα σφάλματος:

BTREE – insert indexfile.indx "645416 some record data"

Ενώ η παρακάτω γραμμή εντολών θα επέστρεφε με επιτυχία:

BTREE – insert indexfile.indx "111111 some record data"

4. Καταγραφή διαδοχικών εγγραφών. Το πρόγραμμα πρέπει να εμφανίζει την εγγραφή που περιέχει το δεδομένο κλειδί και τις επόμενες COUNT εγγραφές που ακολουθούν, αν υπάρχουν. Εάν δεν υπάρχει εγγραφή με το δεδομένο κλειδί, τότε το πρόγραμμα πρέπει να εμφανίζει τις εγγραφές που περιέχουν το επόμενο (μεγαλύτερο) κλειδί και δώσει ένα μήνυμα ότι το πραγματικό κλειδί δεν βρέθηκε. Όπως και με τη λειτουργία -find, εμφανίστε τις εγγραφές σε σειρά με το κλειδί και τη θέση τους μέσα στο αρχείο κειμένου. Προφανώς, αυτή πρέπει να διασχίσει το δέντρο σε σειρά με βάση τον αριθμό του κλειδιού. Η γραμμή εντολών θα πρέπει να λειτουργεί ως εξής:

BTREE – list <όνομα αρχείου ευρετηρίου> <αρχικό κλειδί> <COUNT>

5. Διαγραφή μίας εγγραφής και ανανέωση του ευρετηρίου. Η λειτουργία αυτή από την γραμμή εντολών θα πρέπει να διαγράφει την ΠΡΩΤΗ εγγραφή που έχει κλειδί αυτό που δίνεται. Αυτή η διαγραφή θα γίνεται στο αρχείο δεδομένων (δηλαδή το αρχείο κειμένου) και κατόπιν πρέπει να γίνεται επίσης και το ευρετήριο (και η δομή αλλά και το αρχείο του ευρετηρίου). Σε περιπτώσεις που το κλειδί δεν βρίσκεται το πρόγραμμα θα βγάζει το κατάλληλο μήνυμα. Σε περίπτωση που η διαγραφή της εγγραφής σηματοδοτεί το rearrangement των κόμβων στο B+Tree αυτό πρέπει να γίνεται αντιληπτό (με κατάλληλο μήνυμα). Η εντολή θα είναι ως εξής:

BTREE – delete <όνομα αρχείου ευρετηρίου> <κλειδί>

6 EXTRA CREDIT: Να γίνει προσπάθεια εμφάνισης του δέντρου κατά την εισαγωγή των κλειδιών όπως γίνεται στη σελίδα <https://www.cs.usfca.edu/~galles/visualization/BPlusTree.html>

Δίνεται το dataset αρχείο

ΠΑΡΑΔΟΤΕΑ

1. Το πρόγραμμα με ένα αρχείο οδηγιών εγκατάστασης
2. Περιγραφή των λειτουργιών και αποτελέσματα με βάση το αρχείο test που θα σας δοθεί