

Τεχνολογία Βάσεων Δεδομένων

Εργασία 2012-2013

Ποσοστό επί του βαθμού: 40%

Περιγραφή

Στην εργασία αυτή θα μελετήσετε βασικές μεθόδους επεξεργασίας σε πολύ μεγάλα σύνολα δεδομένων που βρίσκονται αποθηκευμένα στη δευτερεύουσα μνήμη. Καλείστε να υλοποιήσετε μερικές συναρτήσεις, οι οποίες υλοποιούν κάποιες βασικές λειτουργίες ενός DBMS. Πιο συγκεκριμένα, θα εστιάσουμε στις συναρτήσεις που δίνονται στο αρχείο dbtproj.h και αφορούν σε ταξινόμηση, απαλοιφή πολλαπλών εμφανίσεων, σύνδεση με συγχώνευση και σύνδεση με κατακερματισμό. Ο κώδικας που θα παραδώσετε μπορεί να αποτελείται από πολλά αρχεία .h και .c ή .cpp ανάλογα με τον τρόπο ανάπτυξης που θα ακολουθήσετε. Μοναδικός περιορισμός είναι ότι δεν επιτρέπεται να μεταβάλετε τα περιεχόμενα του αρχείου dbtproj.h

Βασικός στόχος είναι οι μέθοδοι που θα υλοποιηθούν να λειτουργούν σε περιορισμένο χώρο μνήμης και η αξιολόγηση θα πραγματοποιηθεί τόσο ως προς τη συνολική ταχύτητα εκτέλεσης όσο και ως προς την ποιότητα του κώδικα. Η αξιολόγηση των εργασιών θα βασιστεί σε μεγάλα σύνολα δεδομένων που θα καταλαμβάνουν αρκετά Gigabytes στο δίσκο.

Θα πραγματοποιηθεί επίσης και προφορική εξέταση της εργασίας. Οι εργασίες θα πρέπει να εκπονηθούν σε **ομάδες των δύο ατόμων**.

Παραδοτέα

Θα πρέπει να παραδώσετε:

- τον πηγαίο κώδικα σε C/C++
- έκθεση με τον τρόπο που υλοποιήσατε την εργασία και μετρήσεις που να φαίνεται η απόδοση των μεθόδων για διαφορετικά μεγέθη αρχείων και μέγεθος μνήμης.

Προθεσμία

Οι εργασίες πρέπει να ολοκληρωθούν μέχρι τα τέλη Ιουνίου. Η ημερομηνία της προφορικής εξέτασης θα οριστεί αργότερα.

Για όποιο πρόβλημα/απορία υπάρχει να επικοινωνήσετε στο papadopo@csd.auth.gr