

PROJECT: “DATA STRUCTURES 2021”

PART I: “Sorting and Searching Algorithms”

Τα αρχεία `agn.us.txt`, `ainv.us.txt`, `ale.us.txt` περιέχουν πλήρη ιστορικά στοιχεία με τις ημερήσιες τιμές για τρεις μετοχές που διαπραγματεύονται στα χρηματιστήρια αξιών NYSE και NASDAQ από το 2005 έως το 2017. Κάθε γραμμή των αρχείων αυτών αντιστοιχεί και σε μία ημέρα καταγραφής της κίνησης της κάθε μετοχής, και έχει την παρακάτω μορφή:

Date, Open, High, Low, Close, Volume, OpenInt

Σας ζητείται να υλοποιήσετε τέσσερα διαφορετικά προγράμματα σε γλώσσα C, που να χρησιμοποιούν ως είσοδο τα παραπάνω αρχεία και το καθένα να υλοποιεί τις παρακάτω λειτουργίες:

- (1) **Ταξινόμηση** κατά αύξουσα σειρά των ημερομηνιών (πεδίο **Date**) με βάση τις τιμές ανοίγματος (πεδίο **Open**) των μετοχών κάνοντας χρήση των αλγορίθμων **Merge Sort** και **Quick Sort**, σύμφωνα με τον ψευδοκώδικα που σας επεξηγήθηκε στη θεωρία (για λεπτομέρειες δείτε τις σχετικές διαφάνειες στο e-class). Συγκρίνατε πειραματικά τους δύο (2) αλγορίθμους. Τι παρατηρείτε?
- (2) **Ταξινόμηση** κατά **αύξουσα** σειρά των ημερομηνιών (πεδίο **Date**) με βάση τις τιμές **κλεισίματος** (πεδίο **Close**) των μετοχών κάνοντας χρήση των αλγορίθμων **Heap Sort** και **Counting Sort**, σύμφωνα με τον ψευδοκώδικα που σας επεξηγήθηκε στη θεωρία (για λεπτομέρειες δείτε τις σχετικές διαφάνειες στο e-class). Για να κάνετε χρήση του **Counting Sort** θα θεωρήσετε ότι γίνεται στρογγυλοποίηση στον πλησιέστερο ακέραιο. Συγκρίνατε πειραματικά τους δύο (2) αλγορίθμους. Τι παρατηρείτε?
- (3) **Εύρεση** συνολικού όγκου συναλλαγών (πεδίο **Volume**) στις μετοχές του αρχείου **agn.us.txt** για συγκεκριμένη ημερομηνία (**Date**) που θα δίνεται από το χρήστη, σύμφωνα με τους αλγορίθμους **Διαδικής Αναζήτησης** και **Αναζήτησης με Παρεμβολή**. Τί παρατηρείτε ως προς τους χρόνους μέσης περίπτωσης? Πόσο η **KATANOMH** του Data Set επηρεάζει την απόδοση του κάθε αλγορίθμου?
- (4) Υλοποιήστε το ζητούμενο του ερωτήματος (3) κάνοντας χρήση του αλγορίθμου **Διαδικής Αναζήτησης Παρεμβολής (BIS)**. Συμβουλευτείτε τον ψευδοκώδικα της σελίδας 82 του βιβλίου «Δομές Δεδομένων», Α.Κ. Τσακαλίδης, Πανεπιστήμιο Πατρών, Τμήμα Μηχανικών Η/Υ και Πληροφορικής καθώς και τις διαφάνειες *9.searching.pdf* που είναι διαθέσιμα στο e-class. Επαληθεύστε πειραματικά τη χρονική πολυπλοκότητα που ισχύει για την μέση (expected) και χειρότερη περίπτωση (worst-case). Η βελτίωση της χειρότερης περίπτωσης επιτυγχάνεται με μία παραλλαγή του BIS. Συμβουλευτείτε τη σελίδα 85 του βιβλίου «Δομές Δεδομένων», Α.Κ. Τσακαλίδης, Πανεπιστήμιο Πατρών, Τμήμα Μηχανικών Η/Υ και Πληροφορικής καθώς και τις διαφάνειες *9.searching.pdf* που είναι διαθέσιμα στο e-class και υλοποιήστε τον αλγόριθμο της συγκεκριμένης παραλλαγής του **BIS**. Συγκρίνατε πειραματικά τους παραπάνω δύο αλγορίθμους. Τί παρατηρείτε ως προς τους χρόνους χειρότερης περίπτωσης?

PART II: “BSTs & HASHING”

Με τον κατάλληλο ορισμό δομών (structs) και συναρτήσεων (functions), να υλοποιήσετε μια εφαρμογή (να γράψετε ένα πρόγραμμα σε γλώσσα C) που θα επεξεργάζεται τα δεδομένα του αρχείου **agn.us.txt**.

- (A) Η εφαρμογή διαβάσει αρχικά το αρχείο και δημιουργεί ένα **Διαδικό Δένδρο Αναζήτησης (ΔΔΑ)** στο οποίο κάθε κόμβος του διατηρεί την εγγραφή (**Date, Volume**). Το ΔΔΑ διατάσσεται ως προς το πεδίο Date και υλοποιείται με δυναμική διαχείριση μνήμης. Το ΔΔΑ θα είναι **ΕΙΤΕ AVL δέντρο ΕΙΤΕ red-black δέντρο**. Μετά την δημιουργία του ΔΔΑ η εφαρμογή εμφανίζει ένα μενού με τις ακόλουθες επιλογές:
 1. Απεικόνιση του ΔΔΑ με ενδο-διατεταγμένη διάσχιση. Κάθε απεικόνιση θα πρέπει να περιέχει μια επικεφαλίδα με τους τίτλους των στοιχείων των εγγραφών που απεικονίζονται.

2. Αναζήτηση της τιμής όγκου συναλλαγών μετοχής (**Volume**) βάσει της ημερομηνίας (**Date**) που θα δίνεται από το χρήστη.
 3. Τροποποίηση της τιμής όγκου συναλλαγών μετοχής (**Volume**) που αντιστοιχεί σε συγκεκριμένη ημερομηνία και η οποία θα δίνεται από το χρήστη.
 4. Διαγραφή μιας εγγραφής που αντιστοιχεί σε συγκεκριμένη ημερομηνία η οποία θα δίνεται από το χρήστη.
 5. Έξοδος από την εφαρμογή.
- (B) Τροποποιήστε κατάλληλα τον κώδικα του (A), ώστε το αρχείο να διαβάζεται στο ΔΔΑ με βάση την τιμή όγκου συναλλαγών μετοχής (πεδίο **Volume**). Το ΔΔΑ διατάσσεται ως προς την τιμή Volume ανά ημερομηνία και υλοποιείται με δυναμική διαχείριση μνήμης. Μετά την δημιουργία του ΔΔΑ η εφαρμογή εμφανίζει ένα μενού με τις ακόλουθες επιλογές:
1. Εύρεση ημέρας/ημερών με την ΕΛΑΧΙΣΤΗ τιμή όγκου συναλλαγών.
 2. Εύρεση ημέρας/ημερών με τη ΜΕΓΙΣΤΗ τιμή όγκου συναλλαγών.
- (Γ) Υλοποιήστε το (A) κάνοντας χρήση HASHING με αλυσίδες, αντί ΔΔΑ. Η συνάρτηση κατακερματισμού θα υπολογίζεται ως το υπόλοιπο (modulo) της διαίρεσης του αθροίσματος των κωδικών ASCII των επιμέρους χαρακτήρων που απαρτίζουν την ΗΜΕΡΟΜΗΝΙΑ με ένα περιττό αριθμό m που συμβολίζει το πλήθος των κάδων (buckets). Π.χ. για ΗΜΕΡΟΜΗΝΙΑ=" 2014-02-13" και m=11, ισχύει:
- $$\text{Hash}("2014-02-13") = [\text{ASCII}('2') + \text{ASCII}('0') + \text{ASCII}('1') + \text{ASCII}('4') + \text{ASCII}('-') + \text{ASCII}('0') + \text{ASCII}('2') + \text{ASCII}('-') + \text{ASCII}('1') + \text{ASCII}('3')] \bmod 11.$$
- Το πρόγραμμα θα εμφανίζει ένα μενού με τις ακόλουθες επιλογές:
1. Αναζήτηση τιμής όγκου συναλλαγών βάσει της ημερομηνίας που θα δίνεται από το χρήστη.
 2. Τροποποίηση των στοιχείων εγγραφής βάσει ημερομηνίας που θα δίνεται από το χρήστη. Η τροποποίηση προφανώς αφορά ΜΟΝΟ το πεδίο της τιμής όγκου συναλλαγών (Volume).
 3. Διαγραφή μιας εγγραφής από τον πίνακα κατακερματισμού βάσει ημερομηνίας που θα δίνεται από το χρήστη.
 4. Έξοδος από την εφαρμογή.

Ενοποιήστε τα (A), (B) και (Γ) σε ένα πρόγραμμα στο οποίο ο χρήστης θα ερωτάται αν θέλει τη φόρτωση του αρχείου **agn.us.txt** σε ένα ΔΔΑ ή σε μία δομή **Hashing** με αλυσίδες και στην περίπτωση που ο χρήστης επιλέξει το πρώτο να μπορεί εν συνεχεία να επιλέξει αν η φόρτωση στο ΔΔΑ θα γίνει με βάση την ημερομηνία (Date) ή την τιμή όγκου συναλλαγών (Volume) ανά ημερομηνία.

DEADLINE: ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ ΤΟΝ ΙΟΥΝΙΟ

Η παράδοση της άσκησης θα πραγματοποιείται με ΑΝΑΡΤΗΣΗ ΣΤΟ Ε_CLASS και με αποστολή μηνύματος ηλεκτρονικού ταχυδρομείου ΚΑΙ ΣΤΙΣ ΤΡΕΙΣ ακόλουθες διευθύνσεις με ένα μήνυμα (με πέντε παραλήπτες και όχι πέντε διακριτά μηνύματα):

sioutas@ceid.upatras.gr, makri@ceid.upatras.gr, mvonitsanos@ceid.upatras.gr,

Μπορείτε να συντάξετε την αναφορά σας σε όποια μορφή κειμένου επιθυμείτε (word, pdf, κ.λπ.). Στο ηλεκτρονικό μήνυμα που θα αποστείλετε θα έχετε συμπεριλάβει το αρχείο της αναφοράς σας καθώς και τα αρχεία των προγραμμάτων C.

ΑΡΙΘΜΟΣ ΦΟΙΤΗΤΩΝ ΑΝΑ ΟΜΑΔΑ <=4

ΠΟΣΟΣΤΟΣΤΟ ΕΠΙ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΒΑΘΜΟΥ: 40%

ΚΑΛΗ ΕΠΙΤΥΧΙΑ!!!