

PROJECT: “DATA STRUCTURES 2022”

PART I: “Sorting and Searching Algorithms”

Στο αρχείο ocean.csv περιέχονται μετρήσεις για τη θερμοκρασία (σε °C) στο νερό του ωκεανού καθώς και οι καταγραφές των τιμών Phosphate, Nitrite, Nitrate, Salinity, Oxygen για τη περίοδο 2000 έως 2015. Κάθε σειρά αντιπροσωπεύει ένα δείγμα νερού που συλλέχθηκε μεταξύ του 2000 και του 2015 από το California Cooperative Oceanic Fisheries Investigations (<https://calcofi.org/>). Μετρήθηκαν η θερμοκρασία, ο κορεσμός οξυγόνου, η συγκέντρωση διαφόρων χημικών ουσιών, ενώ οι γραμμές έχουν την παρακάτω μορφή:

Date, T_degC (Temp), POuM (Phosphate), SiO3uM (Silicate), NO2uM (Nitrite), NO3uM (Nitrate),
Salnty (Salinity), O2ml_L (Oxygen)

Σας ζητείται να υλοποιήσετε τέσσερα διαφορετικά προγράμματα σε γλώσσα C (ή σε γλώσσα επιλογής σας), που να χρησιμοποιούν ως είσοδο το παραπάνω αρχείο και το καθένα να υλοποιεί τις παρακάτω λειτουργίες:

- (1) Ταξινόμηση κατά αύξουσα σειρά των ημερομηνιών (Πεδίο **Date**) βάσει των τιμών των θερμοκρασιών (Πεδίο **Temp**) κάνοντας χρήση των αλγορίθμων **Insertion 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 βάσει των τιμών **Phosphate** 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). Συγκρίνατε πειραματικά τους δύο (2) αλγορίθμους. Τι παρατηρείτε?
- (3) Εύρεση θερμοκρασίας ή/και Phosphate για συγκεκριμένη ημερομηνία (**Date**) που θα δίνεται από το χρήστη, σύμφωνα με τους αλγορίθμους **Δυναμικής Αναζήτησης** και **Αναζήτησης με Παρεμβολή**. Τί παρατηρείτε ως προς τους χρόνους μέσης περίπτωσης? Πόσο η ΚΑΤΑΝΟΜΗ του Data Set επηρεάζει την απόδοση του κάθε αλγορίθμου?
- (4) Υλοποιήστε το ζητούμενο του ερωτήματος (3) κάνοντας χρήση του αλγορίθμου **Δυναμικής Αναζήτησης Παρεμβολής (BIS)**. Συμβουλευτείτε τον ψευδοκώδικα της σελίδας 80 του βιβλίου «Δομές Δεδομένων», Α.Κ. Τσακαλίδης, Πανεπιστήμιο Πατρών, Τμήμα Μηχανικών Η/Υ και Πληροφορικής καθώς και τις διαφάνειες *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 83 του βιβλίου «Δομές Δεδομένων», Α.Κ. Τσακαλίδης, Πανεπιστήμιο Πατρών, Τμήμα Μηχανικών Η/Υ και Πληροφορικής καθώς και τις διαφάνειες *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 ocean.csv. Θυμίζουμε ξανά ότι κάθε γραμμή του αρχείου αυτού αντιστοιχεί σε μία ημέρα μετρήσεων, ενώ οι γραμμές έχουν την παρακάτω μορφή:

Date, T_degC (Temp), PO4uM (Phosphate), SiO3uM (Silicate), NO2uM (Nitrite), NO3uM (Nitrate),
Salnty (Salinity), O2ml_L (Oxygen)

- (A) Η εφαρμογή διαβάζει αρχικά το αρχείο και δημιουργεί ένα **Δένδρο AVL** στο οποίο κάθε κόμβος του διατηρεί την εγγραφή (ΗΜΕΡΟΜΗΝΙΑ, ΘΕΡΜΟΚΡΑΣΙΑ της ημέρας αυτής). Το **AVL**

διατάσσεται ως προς την ΗΜΕΡΟΜΗΝΙΑ και υλοποιείται με δυναμική διαχείριση μνήμης. Μετά την δημιουργία του **AVL** η εφαρμογή εμφανίζει ένα μενού με τις ακόλουθες επιλογές:

1. **Απεικόνιση του AVL** με ενδο-διατεταγμένη διάσχιση. Κάθε απεικόνιση θα πρέπει να περιέχει μια επικεφαλίδα με τους τίτλους των στοιχείων των εγγραφών που απεικονίζονται.
2. **Αναζήτηση** της ΘΕΡΜΟΚΡΑΣΙΑΣ βάσει ΗΜΕΡΟΜΗΝΙΑΣ που θα δίνεται από το χρήστη.
3. **Τροποποίηση** του πεδίου της ΘΕΡΜΟΚΡΑΣΙΑΣ που αντιστοιχεί σε συγκεκριμένη ΗΜΕΡΟΜΗΝΙΑ.
4. **Διαγραφή** μιας εγγραφής που αντιστοιχεί σε συγκεκριμένη ΗΜΕΡΟΜΗΝΙΑ.
5. **Έξοδος** από την εφαρμογή.

(B) Τροποποιήστε κατάλληλα τον κώδικα του (A), ώστε το αρχείο να διαβάζεται στο **AVL** με βάση τη ΘΕΡΜΟΚΡΑΣΙΑ. Το **AVL** διατάσσεται ως προς τη ΘΕΡΜΟΚΡΑΣΙΑ ανά ΗΜΕΡΟΜΗΝΙΑ και υλοποιείται με δυναμική διαχείριση μνήμης. Μετά την δημιουργία του **AVL** η εφαρμογή εμφανίζει ένα μενού με τις ακόλουθες επιλογές:

1. Εύρεση ΗΜΕΡΑΣ/ΗΜΕΡΩΝ με την ΕΛΑΧΙΣΤΗ ΘΕΡΜΟΚΡΑΣΙΑ.
2. Εύρεση ΗΜΕΡΑΣ/ΗΜΕΡΩΝ με τη ΜΕΓΙΣΤΗ ΘΕΡΜΟΚΡΑΣΙΑ.

(Γ) Υλοποιήστε το (A) κάνοντας χρήση HASHING με αλυσίδες, αντί **AVL**. 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:

$\text{Hash}(\text{"2014-02-13"}) = [\text{ASCII}('2') + \text{ASCII}('0') + \text{ASCII}('1') + \text{ASCII}('4') + \text{ASCII}('-') + \text{ASCII}('0') + \text{ASCII}('2') + \text{ASCII}('-') + \text{ASCII}('1') + \text{ASCII}('3')] \bmod 11.$

Το πρόγραμμα θα εμφανίζει ένα μενού με τις ακόλουθες επιλογές:

1. **Αναζήτηση** ΘΕΡΜΟΚΡΑΣΙΑΣ βάσει της ΗΜΕΡΟΜΗΝΙΑΣ που θα δίνεται από το χρήστη.
2. **Τροποποίηση** των στοιχείων εγγραφής βάσει ΗΜΕΡΟΜΗΝΙΑΣ που θα δίνεται από το χρήστη. Η τροποποίηση προφανώς αφορά ΜΟΝΟ το πεδίο της ΘΕΡΜΟΚΡΑΣΙΑΣ.
3. **Διαγραφή** μιας εγγραφής από τον πίνακα κατακερματισμού βάσει ΗΜΕΡΟΜΗΝΙΑΣ 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 σε ένα AVL 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 AVL θα γίνει με βάση την ΗΜΕΡΟΜΗΝΙΑ ή τη ΘΕΡΜΟΚΡΑΣΙΑ ανά ημέρα.

DEADLINE: ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ ΕΑΡΙΝΟΥ ΕΞΑΜΗΝΟΥ

Η παράδοση της άσκησης θα πραγματοποιείται με ΑΝΑΡΤΗΣΗ ΣΤΟ E_CLASS και με αποστολή μηνύματος ηλεκτρονικού ταχυδρομείου ΚΑΙ ΣΤΙΣ ΤΡΕΙΣ ακόλουθες διευθύνσεις με ένα μήνυμα (με τρεις παραλήπτες και όχι τρία 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.

Προτείνουμε η υλοποίηση να γίνει σε γλώσσα C μπορείτε όμως και σε άλλη γλώσσα αρκεί να μην πάρετε έτοιμα τα κομμάτια που θέλουμε να υλοποιήσετε.

ΑΡΙΘΜΟΣ ΦΟΙΤΗΤΩΝ ΑΝΑ ΟΜΑΔΑ <=4

ΠΟΣΟΣΤΟΣΤΟ ΕΠΙ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΒΑΘΜΟΥ: 40%ΚΑΛΗ ΕΠΙΤΥΧΙΑ!!!