

PROJECT: “DATA STRUCTURES 2025”

PART I: “Sorting and Searching Algorithms”

Στο αρχείο `tempm.txt` περιέχονται μετρήσεις για τη θερμοκρασία (σε °C) στην πόλη του Aarhus την περίοδο από 13/2/2014 έως και 8/6/2014. Όμοια, στο αρχείο `hum.txt` υπάρχουν οι καταγραφές των τιμών της υγρασίας (σε ποσοστό %) για την ίδια χρονική περίοδο. Κάθε γραμμή των αρχείων αυτών αντιστοιχεί και σε μία ημέρα μετρήσεων, ενώ οι γραμμές έχουν την παρακάτω μορφή:

```
{“timestamp1”: “value1”, “timestamp2”: “value2”, ... , “timestampN”: “valueN”}
```

Σας ζητείται να υλοποιήσετε τρία διαφορετικά προγράμματα σε γλώσσα επιλογής σας (C/Java/C++/Python), που να χρησιμοποιούν ως είσοδο τα παραπάνω αρχεία και το καθένα να υλοποιεί τις παρακάτω λειτουργίες:

- (1) Ταξινόμηση κατά αύξουσα σειρά των timestamps με βάση τις τιμές των θερμοκρασιών 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) Εύρεση θερμοκρασίας ή/και υγρασίας για συγκεκριμένη χρονική στιγμή (timestamp) που θα δίνεται από το χρήστη, σύμφωνα με τους αλγορίθμους **Δυικής Αναζήτησης Παρεμβολής (BIS)** και της παραλλαγής του (**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. 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. Συγκρίνατε πειραματικά τους παραπάνω δύο αλγορίθμους. Τί παρατηρείτε ως προς τους χρόνους μέσης και χειρότερης περίπτωσης?
- (3) ¹Υλοποιήστε τον BIS σε python και επιπλέον την Learning Augmented έκδοση του (LABIS), στην οποία το interpolation step αντικαθίσταται από ένα oracle που κάνει πρόβλεψη της θέσης με κάποιο μοντέλο μηχανικής μάθησης (Linear Regression, KNeighbors Regressor, κ.τ.λ...). Κάντε χρήση των βιβλιοθηκών `numpy` και `sklearn`.

PART II: “BSTs & HASHING”

Με τον κατάλληλο ορισμό δομών (structs) και συναρτήσεων (functions), να υλοποιήσετε μια εφαρμογή σε γλώσσα επιλογής σας (C/Java/C++/Python), που θα επεξεργάζεται τα δεδομένα του αρχείου `temp.txt`. Θυμίζουμε ξανά ότι κάθε γραμμή του αρχείου αυτού αντιστοιχεί σε μία ημέρα μετρήσεων θερμοκρασιών, ενώ οι γραμμές έχουν την παρακάτω μορφή:

```
{“timestamp1”: “value1”, “timestamp2”: “value2”, ... , “timestampN”: “valueN”}
```

- (A) Η εφαρμογή διαβάζει αρχικά το αρχείο και δημιουργεί ένα **Δυαδικό Δένδρο Αναζήτησης (ΔΔΑ)**² στο οποίο κάθε κόμβος του διατηρεί την εγγραφή (ΗΜΕΡΑ, ΜΕΣΗ ΘΕΡΜΟΚΡΑΣΙΑ της ημέρας αυτής). Το ΔΔΑ διατάσσεται ως προς την ΗΜΕΡΑ και υλοποιείται με δυναμική διαχείριση μνήμης. Μετά την δημιουργία του ΔΔΑ η εφαρμογή εμφανίζει ένα μενού με τις ακόλουθες επιλογές:

¹ Το ερώτημα είναι προαιρετικό και συνεισφέρει μισή μονάδα επιπλέον bonus στον βαθμό του project. Σκεφτείτε και εφαρμογές της τεχνικής σε διάταξη (learning augmented sorting).

² Σκεφτείτε το δέντρο αναζήτησης να είναι ζυγισμένο (AVL, ή red-black ή (α,β) δέντρα). Αν το υλοποιήσετε έχετε μισή μονάδα επιπλέον bonus στον βαθμό του project.

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 ΜΕΣΗΣ ΘΕΡΜΟΚΡΑΣΙΑΣ βάσει ΗΜΕΡΑΣ που θα δίνεται από το χρήστη.
3. Τροποποίηση του πεδίου της ΜΕΣΗΣ ΘΕΡΜΟΚΡΑΣΙΑΣ που αντιστοιχεί σε συγκεκριμένη ΗΜΕΡΑ.
4. Διαγραφή μιας εγγραφής που αντιστοιχεί σε συγκεκριμένη ΗΜΕΡΑ.
5. Έξοδος από την εφαρμογή.

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

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

(Γ) Υλοποιήστε το (Α) 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:

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

Το πρόγραμμα θα εμφανίζει ένα μενού με τις ακόλουθες επιλογές:

1. Αναζήτηση ΜΕΣΗΣ ΘΕΡΜΟΚΡΑΣΙΑΣ βάσει της ΗΜΕΡΑΣ που θα δίνεται από το χρήστη.
2. Τροποποίηση των στοιχείων εγγραφής βάσει ΗΜΕΡΑΣ που θα δίνεται από το χρήστη. Η τροποποίηση προφανώς αφορά ΜΟΝΟ το πεδίο της ΜΕΣΗΣ ΘΕΡΜΟΚΡΑΣΙΑΣ.
3. Διαγραφή μιας εγγραφής από τον πίνακα κατακερματισμού βάσει ΗΜΕΡΑΣ που θα δίνεται από το χρήστη.
4. Έξοδος από την εφαρμογή.

Ενοποιήστε τα (Α), (Β) και (Γ) σε ένα πρόγραμμα στο οποίο ο χρήστης θα ερωτάται αν θέλει τη φόρτωση του αρχείου 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 ΗΜΕΡΑ ή τη ΜΕΣΗ ΘΕΡΜΟΚΡΑΣΙΑ ανά ημέρα.

DEADLINE: ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ ΤΟΝ ΙΟΥΝΙΟ

Η παράδοση της άσκησης θα πραγματοποιείται με αποστολή μηνύματος ηλεκτρονικού ταχυδρομείου ΚΑΙ ΣΤΙΣ ΠΕΝΤΕ ακόλουθες διευθύνσεις με ένα μήνυμα (με τέσσερις παραλήπτες και όχι τέσσερα διακριτά μηνύματα):

sioutas@ceid.upatras.gr, makri@ceid.upatras.gr, mvonitsanos@ceid.upatras.gr, st1057466@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.

ΑΡΙΘΜΟΣ ΦΟΙΤΗΤΩΝ ΑΝΑ ΟΜΑΔΑ <=4

ΣΥΝΟΛΙΚΟ (ΜΑΖΙ ΜΕ ΤΗΝ ΠΡΩΤΗ ΕΡΓΑΣΙΑ) ΠΟΣΟΣΤΟ ΕΠΙ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΒΑΘΜΟΥ: 30%

ΚΑΛΗ ΕΠΙΤΥΧΙΑ!!!