

ΘΕΜΑΤΑ ΕΞΕΤΑΣΗΣ 2021-2022

Το διαγώνισμα χωριζόταν σε δύο μέρη: Μέρος Α' με πολλαπλής επιλογής όπου κάθε σωστή επιλογή έπαινε 3 μονάδες και κάθε λάθος επιλογή έπαινε -1. Με μια εξαίρεση όπου δόθηκε διευκρίνιση, σε κάθε ερώτημα υπήρχε ακριβώς μία σωστή απάντηση. Τα ερωτήματα ήταν 12 και μερικά ήταν κατευθείαν από τις διαφάνειες copy paste, ενώ άλλα όχι. Περιλάμβαναν κατά κύριο λόγο δέντρα απόφασης, outlier detection και ύλη σχετική με MapReduce, μιας και ήταν αυτά στα οποία δεν υπήρχαν ασκήσεις.

Μέρος Β': 9 Ασκήσεις σύντομης ανάπτυξης.

[1] Δινόταν ένα διάγραμμα με 4 εμφανή clusters και στη μέση ένα μοναδικό outlier. Τα σημεία είχαν προκύψει από προβολή πολυδιάστατων δεδομένων σε έναν 2D χώρο και το ερώτημα ήταν τι μέθοδο ανίχνευσης outlier θα χρησιμοποιούσαμε για το detection του μοναδικού outlier.

[2] Κατασκευή 3D-Cube βάσει ενός πίνακα με 4 στήλες και 6-7 εγγραφές. Ακριβώς ίδια δομή με το παράδειγμα στις διαφάνειες του Τσουμάκου.

[3] Υπολογισμός της (α) Ευκλείδειας και (β) Manhattan απόστασης μεταξύ δύο σημείων της μορφής $x = (a,b,c,d)$ και $y = (e,f,g,h)$.

[4] Δινόταν ένας πίνακας με ID, Τίτλο βιβλίου, Όνομα Συγγραφέα, Κατηγορία Βιβλίου, Πλήθος Σελιδών και 4 εγγραφές, στα πλαίσια της ύλης του Στάμου. Ρωτούσε αρχικά ποιο πεδίο είναι καλό για να στήσουμε οντολογία και μετά έλεγε να προταθεί για αυτήν την οντολογία 1 άτομο, 2 ρόλοι και 7 έννοιες.

[5] Άσκηση με Bloom Filter ίδια με θέματα 2020-2021 αλλά με fixed νούμερα (όχι ανάλογα τον AM).

[6] Μια άσκηση με HMM, πάλι ίδια με θέματα 2020-2021, αλλά με fixed νούμερα. Και ακριβώς το ίδιο ζητούμενο (πιθανότητα ακολουθίας bca).

[7] Εφαρμογή στον a-priori, έδινε ένα δέντρο και ρωτούσε (α) Εάν το "ade" είναι συχνό, τότε τι άλλο είναι συχνό; (β) Εάν το "ace" δεν είναι συχνό, τότε τι άλλο δεν είναι συχνό;

[8] και [9] μου διαφεύγουν, όποι@ θυμάται ας συμπληρώσει.