

# ΡΕΥΜΑΤΑ ΔΕΔΟΜΕΝΩΝ

## DATA STREAMS

*Rien ne va plus:* ΑΠΟΡΡΙΨΗ ΦΟΡΤΟΥ ΑΠΟ ΡΕΥΜΑΤΑ ΚΙΝΟΥΜΕΝΩΝ  
ΑΝΤΙΚΕΙΜΕΝΩΝ ..... 2

**ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ**  
Εργαστήριο Συστημάτων Βάσεων Γνώσεων & Δεδομένων

**Rien ne va plus: ΑΠΟΡΡΙΨΗ ΦΟΡΤΟΥ ΑΠΟ ΡΕΥΜΑΤΑ ΚΙΝΟΥΜΕΝΩΝ ΑΝΤΙΚΕΙΜΕΝΩΝ**

ΠΛΗΡΟΦΟΡΙΕΣ: Κώστας Πατρούμπας, 210 772 1446, [kpatro@dbl-lab.ece.ntua.gr](mailto:kpatro@dbl-lab.ece.ntua.gr)

**ΠΕΡΙΛΗΨΗ:** Η εργασία αποσκοπεί σε σχεδιασμό και υλοποίηση μεθοδολογίας διαγραφής πλεονάζοντος όγκου δεδομένων τα οποία συλλέγονται online σχετικά με την τρέχουσα θέση μεγάλου αριθμού κινούμενων αντικειμένων.

**ΑΤΟΜΑ:** 1-2

**ΠΕΡΙΒΑΛΛΟΝ ΕΡΓΑΣΙΑΣ:** C++/ JAVA

**ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ:** Στις σύγχρονες εφαρμογές παρακολούθησης κινούμενων αντικειμένων (λ.χ. πολυάριθμα οχήματα στο οδικό δίκτυο), υπάρχει ενδεχόμενο το στίγμα των αντικειμένων να καταγράφεται με πολύ μεγάλη συχνότητα (λ.χ. κάθε 10 sec). Ο όγκος της πληροφορίας ενός τέτοιου *ρεύματος δεδομένων (data stream)* που το σύστημα οφείλει να διαχειριστεί online γίνεται τότε υπερβολικά μεγάλος, καθυστερώντας σημαντικά την επεξεργασία *ερωτημάτων διαρκείας (continuous queries)*, τα οποία πρέπει να παρέχουν συνεχώς επίκαιρες απαντήσεις (π.χ., «*ποιο είναι το πιο κοντινό φαρμακείο στην τωρινή μου θέση;*»).

Εφόσον συνήθως όλη η επεξεργασία διεξάγεται στην κύρια μνήμη, πολλές τεχνικές προτείνουν την *επιλεκτική (à la carte) διαγραφή* μέρους των στοιχείων (*απόρριψη φόρτου – load shedding*). Βάσει εκτιμήσεων της ποιότητας των απαντήσεων, ορισμένο τμήμα των πρωτογενών καταγραφών απορρίπτεται εντελώς πριν καν δοθεί προς επεξεργασία, ώστε το σύστημα να αντεπεξέλθει σε απροσδόκητες αυξομειώσεις του ρυθμού άφιξης των στοιχείων. Εξυπακούεται ότι θα είναι αναγκαστικά προσεγγιστικές οι απαντήσεις που πλέον θα δίνονται στα ερωτήματα.

Τα στάδια εκπόνησης της διπλωματικής εργασίας προβλέπονται ως εξής:

- Βιβλιογραφική ανασκόπηση των κυριότερων μεθόδων διαχείρισης και επεξεργασίας τυπικών ερωτημάτων διαρκείας σε χωροχρονικά ρεύματα δεδομένων.
- Σχεδιασμός του μοντέλου διακίνησης και καταμερισμού των δεδομένων, καθώς και της κατάλληλης αρχιτεκτονικής επεξεργασίας τους (λ.χ. συγκεντρωτική, κατανενημένη, ιεραρχική).
- Υλοποίηση δομών τήρησης των δεδομένων στην κύρια μνήμη, π.χ. με χρήση τετραδικού δένδρου (quad-tree), σε πλέγματα (grid) με κατακερματισμό ή άλλων παραπλήσιων καταταμίσεων του χώρου σε περιοχές ενδιαφέροντος.
- Υλοποίηση αλγορίθμου απόρριψης φόρτου, συνυπολογίζοντας παραμέτρους όπως: ο όγκος των δεδομένων, το πλήθος των ενεργών ερωτημάτων, η ακρίβεια των απαντήσεων που δίνονται μέχρι στιγμής, ο βαθμός κατάτμησης του χώρου κ.ά.
- Συγκριτική αποτίμηση του αλγορίθμου με άλλες τεχνικές από την βιβλιογραφία, καθώς και με ενδεχόμενες παραλλαγές του, λ.χ. τυχαία, ομοιόμορφη ή πιθανοτική διαγραφή.
- Αξιολόγηση πειραματικών αποτελεσμάτων σε δοκιμαστικά σύνολα δεδομένων με μέτρηση επιδόσεων για χρόνους απόκρισης, απαιτήσεις σε μνήμη, ακρίβεια απαντήσεων κ.ά.