

**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@dblab.ece.ntua.gr

**ΠΕΡΙΛΗΨΗ:** Η εργασία αποσκοπεί σε σχεδιασμό και υλοποίηση μεθοδολογίας διαγραφής πλεονάζοντος όγκου online δεδομένων σχετικών με την πρόσφατη κίνηση μεγάλου αριθμού 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 δευτερόλεπτα). Ο όγκος της πληροφορίας που συσσωρεύεται σ' ένα τέτοιο ρεύμα δεδομένων (*data stream*) γίνεται τότε υπερβολικά μεγάλος. Εφόσον το σύστημα οφείλει να διαχειριστεί online τα δεδομένα, ενδεχομένως να καθυστερεί η επεξεργασία ποικίλων ερωτημάτων διαρκείας (*continuous queries*), τα οποία πρέπει να παρέχουν συνεχώς επίκαιρες απαντήσεις (π.χ., «ποιοί 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 των εισερχόμενων στοιχείων. Θα πρέπει επίσης να συνεξετάζεται η πρόσφατη τροχιά καθενός αντικειμένου (λ.χ. κατά την τελευταία μία ώρα), ώστε να διατηρείται η πληροφορία κίνησης εκείνων που πιθανόν παίζουν ρόλο σε πολλά εκκρεμή ερωτήματα (λ.χ. ποιοι διέσχισαν πρόσφατα τα όρια του δακτυλίου). Προφανώς, οι απαντήσεις θα είναι πλέον αναγκαστικά προσεγγιστικές.

Τα στάδια εκπόνησης της διπλωματικής εργασίας προβλέπονται ως εξής:

- Βιβλιογραφική επισκόπηση 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, κατανεμημένη, ιεραρχική).
- Υλοποίηση δομών τήρησης των δεδομένων στην κύρια μνήμη, π.χ. με χρήση τετραδικού δένδρου (*quad-tree*), σε κάρναβο (*grid*) με κατακερματισμό ή άλλων παραπλήσιων καταταμίσεων του χώρου σε περιοχές ενδιαφέροντος (*θα αξιοποιηθεί υπάρχων κώδικας*).
- Υλοποίηση και συγκριτική αποτίμηση 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. Εκ των πραγμάτων, ο αλγόριθμος θα υιοθετεί *πιθανοτικό χειρισμό* των τροχιών, ενώ μπορεί επίσης να ενσωματώνει και προτιμήσεις των χρηστών όσον αφορά το επιθυμητό *περιθώριο εμπιστοσύνης* στις παρεχόμενες απαντήσεις.
- Αξιολόγηση πειραματικών αποτελεσμάτων σε δοκιμαστικά σύνολα δεδομένων με μέτρηση επιδόσεων για χρόνους απόκρισης, απαιτήσεις σε μνήμη, ακρίβεια απαντήσεων κ.ά.

Για περισσότερες πληροφορίες σχετικά με την διαχείριση ρευμάτων κινούμενων αντικειμένων, επισκεφθείτε την ιστοσελίδα: <http://www.dblab.ece.ntua.gr/~kpatro/TSV/>