

## Μεταπτυχιακό Πρόγραμμα Σπουδών στην Πληροφορική Θέματα Διπλωματικών Εργασιών

### ΘΕΜΑ 1

#### «Αλγόριθμοι ερωτημάτων skyline σε ροές πολυδιάστατων δεδομένων»

Σε μία ροή πολυδιάστατων δεδομένων αναζητούνται διάφορα ερωτήματα που αφορούν τα πολυδιάστατα αντικείμενα ή ιδιότητές τους. Ένα από τα ερωτήματα αυτά είναι και το skyline. Ένα πολυδιάστατο αντικείμενο θα ανήκει στο skyline όταν δεν κυριαρχείται από κανένα άλλο αντικείμενο. Ένα αντικείμενο κυριαρχεί πάνω σε ένα άλλο όταν οι τιμές σε όλες τις διαστάσεις του είναι καλύτερες (ή ίδιες) σε σχέση με τις αντίστοιχες τιμές του άλλου, ενώ σε μία τουλάχιστον διάσταση είναι σαφώς καλύτερο. Για να απαντήσουμε σε ένα τέτοιο ερώτημα απαιτούνται ειδικές μέθοδοι διαχείρισης της ροής και ειδικά σχήματα δεικτοδότησης σε περιορισμένο χώρο μνήμης. Στόχος της εργασίας αυτής είναι να επιλύσουμε προσεγγιστικά το πρόβλημα αυτό υλοποιώντας συγκεκριμένους αλγορίθμους και κάνοντας πειραματική μελέτη τους σε πολύ μεγάλα σύνολα δεδομένων.

**Απαιτήσεις:** πολύ καλή γνώση C++ και αλγορίθμων ροών δεδομένων.

### ΘΕΜΑ 2

#### «Αλγόριθμοι ερωτημάτων ομοιότητας σε μεγάλες βάσεις δεδομένων εικόνων»

Σε εφαρμογές αναζήτησης σε μεγάλες βάσεις δεδομένων εικόνων η αναζήτηση δεν γίνεται πάνω στα πρωτότυπα αντικείμενα των εικόνων αλλά τα αντικείμενα μετασχηματίζονται σε διανύσματα πραγματικών (ή ακεραίων) αριθμών μεγάλου πλήθους διαστάσεων (descriptor vectors), με κατάλληλους μετασχηματισμούς (π.χ. SIFT, CEDD, GIST, SURF κλπ). Τα διανύσματα αυτά συνήθως δεικτοδοτούνται από ειδικές μεθόδους κατακερματισμού καθώς οι ιεραρχικές δομές δεν είναι αποδοτικές σε τόσο μεγάλο πλήθος διαστάσεων. Στόχος της εργασίας αυτής είναι η υλοποίηση βασικών σχημάτων κατακερματισμού και αλγορίθμων για την αναζήτηση των  $k$  πιο όμοιων εικόνων της βάσης σε σχέση με μία δεδομένη εικόνα, καθώς και η σύγκριση της απόδοσής τους με πειραματική μελέτη.

**Απαιτήσεις:** πολύ καλή γνώση C++ και μεθόδων κατακερματισμού.

### ΘΕΜΑ 3

#### «Αλγόριθμοι ερωτημάτων skyline σε μεγάλες βάσεις δεδομένων εικόνων»

Σε εφαρμογές αναζήτησης σε μεγάλες βάσεις δεδομένων εικόνων η αναζήτηση δεν γίνεται πάνω στα πρωτότυπα αντικείμενα των εικόνων αλλά τα αντικείμενα μετασχηματίζονται σε διανύσματα πραγματικών (ή ακεραίων) αριθμών μεγάλου πλήθους διαστάσεων (descriptor vectors), με κατάλληλους μετασχηματισμούς (π.χ. SIFT, CEDD, GIST, SURF κλπ). Τα διανύσματα αυτά συνήθως δεικτοδοτούνται από ειδικές μεθόδους κατακερματισμού καθώς οι ιεραρχικές δομές δεν είναι αποδοτικές σε τόσο μεγάλο πλήθος διαστάσεων. Μία εικόνα θα ανήκει στο skyline όταν δεν κυριαρχείται από καμία άλλη εικόνα. Μία εικόνα κυριαρχεί σε μία άλλη όταν οι τιμές σε όλες τις διαστάσεις του διανύσματός της είναι καλύτερες (ή ίδιες) σε σχέση με τις αντίστοιχες τιμές του διανύσματος της άλλης, ενώ σε μία τουλάχιστον διάσταση είναι σαφώς καλύτερη. Η εύρεση των εικόνων που βρίσκονται στο skyline βοηθάει σημαντικά στην ομαδοποίηση των εικόνων της βάσης ως προς την ομοιότητά τους σε σχέση με τις εικόνες που βρίσκονται στο skyline. Στόχος της εργασίας αυτής είναι η υλοποίηση βασικών σχημάτων κατακερματισμού και αλγορίθμων για την αναζήτηση των εικόνων της βάσης που βρίσκονται στο skyline, καθώς και η σύγκριση της απόδοσής τους με πειραματική μελέτη.

**Απαιτήσεις:** πολύ καλή γνώση C++ και μεθόδων κατακερματισμού.

#### **ΘΕΜΑ 4**

**«Αλγόριθμοι ερωτημάτων εγγύτητας τροχιών σε σημεία ενδιαφέροντος πάνω σε Ευκλείδειους χώρους».**  
Σε πολλές εφαρμογές αναζητούνται οι  $k$  εγγύτερες τροχιές σε προκαθορισμένα σημεία ενδιαφέροντος. Οι τροχιές και τα σημεία ενδιαφέροντος θεωρούνται ότι είναι πάνω σε Ευκλείδειους χώρους χωρίς περιορισμούς. Για να απαντηθούν τα ερωτήματα αυτά απαιτούνται ειδικές μέθοδοι δεικτοδότησης (όπως R-Trees, M-Trees κλπ), καθώς και I/O αποδοτικοί αλγόριθμοι. Στόχος της εργασίας αυτής είναι η υλοποίηση βασικών αλγορίθμων για το πρόβλημα αυτό καθώς και η σύγκριση της απόδοσής τους με πειραματική μελέτη.  
**Απαιτήσεις:** πολύ καλή γνώση C++ και I/O αποδοτικών αλγορίθμων.

#### **ΘΕΜΑ 5**

**«Αλγόριθμοι ερωτημάτων εγγύτητας τροχιών σε σημεία ενδιαφέροντος πάνω σε χωρικά δίκτυα».**  
Σε πολλές εφαρμογές χωρικών δικτύων αναζητούνται οι  $k$  εγγύτερες τροχιές σε προκαθορισμένα σημεία ενδιαφέροντος. Οι τροχιές και τα σημεία ενδιαφέροντος θεωρούνται ότι είναι πάνω σε ένα χωρικό δίκτυο και έτσι ακολουθούν τους περιορισμούς που αυτό θέτει (π.χ. ένα οδικό δίκτυο). Για να απαντηθούν τα ερωτήματα αυτά απαιτούνται ειδικές μέθοδοι δεικτοδότησης (όπως πχ M-Trees), καθώς και I/O αποδοτικοί αλγόριθμοι. Στόχος της εργασίας αυτής είναι η υλοποίηση βασικών αλγορίθμων για το πρόβλημα αυτό καθώς και η σύγκριση της απόδοσής τους με πειραματική μελέτη.  
**Απαιτήσεις:** πολύ καλή γνώση C++ και I/O αποδοτικών αλγορίθμων.

Παρακαλώ να δηλώσετε ενδιαφέρον μέχρι τις 14 Ιουνίου 2015 ([tiakas@csd.auth.gr](mailto:tiakas@csd.auth.gr)). Όλα τα θέματα απαιτούν καλές γνώσεις σε Αλγορίθμους, Δομές Δεδομένων, Βάσεις Δεδομένων και Προγραμματισμό.

**Ελευθέριος Τιάκας - Γιάννης Μανωλόπουλος**