Πανεπιστημιο Πελοποννησου
Σχολή Θετικών Επιστημών και Τεχνολογίας
Τμήμα Επιστήμης και Τεχνολογίας Υπολογιστών
Τέρμα Καραϊσκάκη, 22100, Τρίπολη
Τηλ. 2710-372227
Φαξ. 2710-372225

Θεματα Πτυχιακων Εργασιων

GRAPHIT-db: Πρότυπο σύστημα διαχείρισης γράφων

(Πληροφορίες: Σπύρος Σκιαδόπουλος και Θοδωρής Δαλαμάγκας, {spiros,dalamag}@uop.gr)

Περίληψη:

Η πτυχιακή εργασία στοχεύει στην υλοποίηση ενός συστήματος το οποίο θα μπορεί να αποθηκεύει, επεξεργάζεται και αναζητά γράφους.

Άτομα: 1-2

Πλατφόρμα εργασίας: JAVA ή C++

Σύντομη περιγραφή:

Οι γράφοι, ως δομή, χρησιμοποιούνται εδώ και πόλλες δεκαετίες για την μοντελοποίηση περίπλοκων δεδομένων. Στην βιβλιογραφία υπάρχουν πάρα πολλοί αλγόριθμοι που απαντούν ερωτήσεις για δεδομένα οργανωμένα σε γράφους, όπως: «Ποιος είναι ο πιο σύντομος δρόμος μεταξύ δύο κόμβων;», «Ποια είναι τα συνεκτικά κομμάτια του γράφου;» κ.λ.π.

Τον τελευταίο καιρό, λόγω της εξάπλωσης των εφαρμογών Σημασιολογικού Ιστού (Semantic Web), όπως τα Κοινωνικά Δίκτυα (Social Networks), δίκτυα από blogs (Blog Networks), Δίκτυα Εννοιών (Concept Networks), έχουν προκύψει νέες κατηγορίες ερωτήσεων για δεδομένα οργανωμένα σε γράφους. Για παράδειγμα, είναι πολύ χρήσιμο σε ένα Κοινωνικό Δίκτυο που έχει τη μορφή ενός γράφου να μπορεί κάποιος να βρει, δοθέντων τριών κόμβων n_1, n_2 και n_3 που συμβολίζουν τρία διαφορετικά πρόσωπα, τον αριθμό των συντομότερων μονοπατιών που ξεκινούν από τον n_1 και καταλήγουν στον n_3, διαμέσου του n_2. Αν ο αριθμός αυτός είναι μεγάλος, τότε υπάρχει ένδειξη ισχυρής φιλίας του n_2 με τους n_1 και n_3!

Επιπλέον, οι γράφοι των εφαρμογών αυτών είναι συνήθως πολύ μεγάλοι. Η πλειοψηφία των γνωστών αλγόριθμων γράφων μπορεί να εφαρμοστεί αποδοτικά είτε σε γράφους με μικρό μέγεθος (δηλαδή σε γράφους που είναι δυνατό να αποθηκευθούν στην κύρια μνήμη) είτε σε μη-μεταβαλλόμενους γράφους (δηλαδή σε γράφους στους οποίους το σύνολο των κόμβων, των ακμών αλλά και των βαρών των ακμών δεν μεταβάλλονται).

Ο σκοπός επομένως της παρούσας πτυχιακής είναι να μελετήσει διάφορες κατηγορίες ερωτημάτων πάνω σε δεδομένα γράφων, να επεκτείνει τις υπάρχουσες μεθόδους και αλγορίθμους αποτίμησης ερωτήσεων ώστε να μπορούν να εφαρμοστούν αποδοτικά σε μεγάλους και μεταβαλλόμενους γράφους ή/και να προτείνει νέες μεθόδους αποτίμησης ερωτήσεων. Τα παραπάνω θα αποτελέσουν μέρος του GRAPHIT-db, ενός πρότυπου σύστημα διαχείρισης βάσης δεδομένων οργανωμένων σε γράφους.

Πανεπιστημιο Πελοποννησου
Σχολή Θετικών Επιστημών και Τεχνολογίας
Τμήμα Επιστήμης και Τεχνολογίας Υπολογιστών
Τέρμα Καραϊσκάκη, 22100, Τρίπολη
Τηλ. 2710-372227
Φαξ. 2710-372225

Θεματα Πτυχιακων Εργασιων

Σύστημα διαχείρισης ροών δεδομένων (data streams)

(Πληροφορίες: Σπύρος Σκιαδόπουλος, spiros@uop.gr)

Περίληψη:

Σχεδιασμός και υλοποίηση συστήματος για την σύνδεση ροών δεδομένων με σχέσεις αποθηκευμένες σε ένα σύστημα διαχείρισης βάσεων δεδομένων.

Άτομα: 1-2

Πλατφόρμα εργασίας: C ή C++

Σύντομη περιγραφή:

Οι ροές δεδομένων (data streams) απασχολούν την ερευνητική κοινότητα λόγο της πληθώρας των εφαρμογών στις οποίες δεδομένα καταφθάνουν για επεξεργασία με αδιάλειπτη ροή. Στην παρούσα πτυχιακή εργασία θα ασχοληθούμε με ένα πολύ συνηθισμένο πρόβλημα σχετικό με την επεξεργασία ροών δεδομένων: την σύνδεση ροών δεδομένων με μεγάλες σχέσεις αποθηκευμένες σε ένα σύστημα διαχείρισης βάσεων δεδομένων.

Για παράδειγμα, σκεφτείτε το κεντρικό εξυπηρέτη ενός μεγάλου τηλεπικοινωνιακού οργανισμού, ο οποίος λαμβάνει τις τηλεφωνικές κλήσεις υπό τη μορφή συνεχόμενης ροής δεδομένων. Κάθε κλήση περιέχει πληροφορίες αναφορικά με τη συνδιάλεξη, όπως τους αριθμούς τηλεφώνων, τη χρονική διάρκεια κ.λ.π. Ο παραπάνω εξυπηρέτης συνδυάζει τις πληροφορίες της συνδιάλεξης με πληροφορίες αποθηκευμένες στη βάση δεδομένων του τηλεπικοινωνιακού οργανισμού. Για παράδειγμα, ο εξυπηρέτης ελέγχει αν η κλήση επιτρέπεται με βάση το πρόγραμμα του χρήστη και αποφασίζει με ποιο τρόπο θα την δρομολογήσει.

Στα πλαίσια της παρούσας πτυχιακής εργασίας θα σχεδιαστεί και θα υλοποιηθεί ένα σύστημα για την σύνδεση γρήγορα μεταβαλλόμενων ροών δεδομένων με μεγάλες σχέσεις αποθηκευμένες σε ένα σύστημα διαχείρισης βάσεων δεδομένων. Πιο συγκεκριμένα, θα μελετηθούν υπάρχοντες στατικοί αλγόριθμοι για τη σύνδεση ροών δεδομένων με σχέσεις. Με βάση τους παραπάνω αλγόριθμους μπορούμε να υπολογίσουμε τους απαραίτητους πόρους του συστήματος (π.χ., την ελάχιστη μνήμη) που χρειάζονται για να επεξεργαστούμε ροές δεδομένων με συγκεκριμένο ρυθμό έλευσης. Το πρόβλημα με τους παραπάνω αλγόριθμους παρουσιάζεται όταν η ροή δεδομένων δεν έχει σταθερό ρυθμό. Στόχος της παρούσας πτυχιακής εργασίας είναι η σχεδίαση και η υλοποίηση δυναμικών αλγορίθμων σύνδεσης. Σε αντίθεση με τους στατικούς αλγόριθμους, οι δυναμικοί αλγόριθμοι θα δεσμεύουν πόρους όταν ο ρυθμός έλευσης της ροής αυξάνεται ενώ θα ελευθερώνουν πόρους όταν ο ρυθμός μειώνεται.

Πανεπιστημιο Πελοποννησου
Σχολή Θετικών Επιστημών και Τεχνολογίας
Τμήμα Επιστήμης και Τεχνολογίας Υπολογιστών
Τέρμα Καραϊσκάκη, 22100, Τρίπολη
Τηλ. 2710-372227
Φαξ. 2710-372225

Θεματα Πτυχιακων Εργασιων

Διαχείριση αναφορών από το Google Scholar

(Πληροφορίες: Σπύρος Σκιαδόπουλος, spiros@uop.gr)

Περίληψη:

Σχεδιασμός και υλοποίηση συστήματος για τον διαχωρισμό των αναφορών που παρουσιάζει το Google Scholar

Άτομα: 1-2

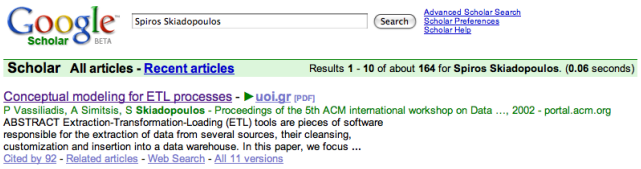
Πλατφόρμα εργασίας: C ή C++

Σύντομη περιγραφή:

Το Google Scholar (http://scholar.google.com) αποτελεί μηχανή αναζήτησης για δημοσιεύσεις. Αντίστοιχα με το Google, το Google Scholar σταχυολογεί τις δημοσιεύσεις από τις βάσεις δεδομένων μεγάλων οργανισμών (όπως η ACM, η IEEE, το JSTOR και το SCOPUS) αλλά και από το διαδίκτυο. Χρησιμοποιώντας το Google Scholar μπορούμε να βρούμε χρήσιμες πληροφορίες

- για μια δημοσίευση (όπως τους συγγραφείς της, το περιοδικό ή το συνέδριο που δημοσιεύθηκε, τη χρονολογία δημοσίευσης, τον αριθμό των αναφορών της κλπ) και
- για ένα συγγραφέα (όπως τον αριθμό των δημοσιεύσεών του).

Δοκιμάστε για παράδειγμα να αναζητήσετε τις δημοσιεύσεις του Spiros Skiadopoulos. Ως απάντηση θα έχετε μια λίστα με δημοσιεύσεις όπως στο επόμενο σχήμα.



Το Google Scholar για κάθε δημοσίευση συλλέγει τις εργασίες που την αναφέρουν. Για παράδειγμα στο παραπάνω σχήμα αν επιλέξετε το σύνδεσμο Cited by 92, θα εμφανιστούν οι εργασίες οι οποίες αναφέρουν την εργασία με τίτλο 'Conceptual modeling for ETL processes'.

Το πρόβλημα με την λίστα που εμφανίζει το Google Scholar είναι ότι δεν γίνεται διαχωρισμός μεταξύ *αυτοαναφορών* (δηλαδή αναφορών που προέρχονται από τους συγγραφείς της δημοσίευσης) και *ετεροαναφορών* (δηλαδή αναφορών που **δεν** προέρχονται από τους συγγραφείς της δημοσίευσης).

Η παρούσα πτυχιακή εργασία σκοπεύει στο να καλύψει το παραπάνω κενό. Χρησιμοποιώντας τα αποτελέσματα που παρέχει το Google Scholar, επιθυμούμε να σχεδιάσουμε και να υλοποιήσουμε ένα σύστημα που θα διαχωρίσει τις αναφορές σε ετεροεαναφορές και αυτοαναφορές.