

Θέματα Πτυχιακών Εργασιών για το Ακαδημαϊκό Έτος 2013-2014

Χειμερινό Εξάμηνο

Ευαγγελία Πιτουρά

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ 1: Ποικιλομορφία

Πρόσφατα, η ποικιλομορφία (diversity) έχει προσελκύσει το ερευνητικό ενδιαφέρον ως ένα μέσο διαβάθμισης των δεδομένων συμπληρωματικού αυτού της σχετικότητας (relevance). Η ποικιλομορφία αναφέρεται στην επιλογή αποτελεσμάτων τέτοιων ώστε πέρα από τη σχετικότητά τους ως προς την ερώτηση να καλύπτουν διαφορετικά θέματα, να είναι καινούργια (novel) ή διαφορετικά μεταξύ τους (dissimilar). Έχουμε αναπτύξει ένα σύστημα [1] που υλοποιεί και οπτικοποιεί μερικούς κλασικούς ορισμούς ποικιλομορφίας βασισμένους σε απόσταση (δηλαδή, στην επιλογή των πιο διαφορετικών/μακρινών σημείων).

Πιθανές επεκτάσεις τους συστήματος αφορούν:

- Υποστήριξη νέου είδους αποστάσεων (για παράδειγμα, cosine για έγγραφα, shortest path σε γραφήματα, κλπ) και οπτικοποίηση των αποτελεσμάτων
- Βελτίωση της διεπαφής του συστήματος
- Υποστήριξη πολλαπλών ακτινών

[1] Marina Drosou, Evaggelia Pitoura: POIKILO: A Tool for Evaluating the Results of Diversification Models and Algorithms. 1246-1249 <http://www.vldb.org/pvldb/vol6/p1246-drosou.pdf>

Τα κοινωνικά δίκτυα (π.χ., Twitter, Facebook) συνήθως αναπαρίστανται ως γράφοι όπου οι κόμβοι αντιστοιχούν στους χρήστες και οι ακμές στις μεταξύ τους σχέσεις. Οι παρακάτω πτυχιακές 2 έως 5 αφορούν στην αποθήκευση και επεξεργασία τέτοιων γραφημάτων.

- Θα εξεταστεί η δυνατότητα η υλοποίηση να χρησιμοποιήσει το Graphlab και ειδικά το <http://graphlab.org/graphchi/>.
- Ένα ενδιαφέρον σύστημα οπτικοποίησης γράφων είναι το Gephi <https://gephi.org/>
- Μια μεγάλη συλλογή δεδομένων γράφου μπορείτε να βρείτε στο SNAP - Stanford Large Network Dataset Collection <http://memetracker.org/data/index.html>

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ 2 : Εύρεση αντιπροσωπευτικών κόμβων (χρηστών) σε γράφους κοινωνικών δικτύων

Στην εργασία αυτή θα εξετάσουμε τον εντοπισμό αντιπροσωπευτικών κόμβων με βάση την ποικιλομορφία [1, 2] και τους διαφορετικούς ορισμούς της.

Τα αποτελέσματα θα αξιολογηθούν με χρήση πραγματικών δεδομένων.

[1] Marina Drosou, Evaggelia Pitoura: DisC diversity: result diversification based on dissimilarity and coverage. PVLDB 6(1): 13-24 (2012) <http://www.vldb.org/pvldb/vol6/p13-drosou.pdf>

[2] Marina Drosou, Evaggelia Pitoura: Search result diversification. SIGMOD Record 39(1): 41-47 (2010) <http://dl.acm.org/citation.cfm?doid=1860702.1860709>

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ 3: : Σύγκριση Διαφορετικών Μεθόδων Αποθήκευσης Γραφημάτων

Η εργασία θα εξετάσει διαφορετικούς μεθόδους για την αποθήκευση μεγάλων γραφημάτων. Η επιθυμητή υλοποίηση θα πρέπει να είναι κατανομημένη σε cluster. Θα εξεταστεί η χρήση μη σχεσιακών βάσεων δεδομένων, αρχικά της Neo4j

Ως αρχή, κοιτάξτε το:

[1] B Renzo Angles; Arnau Prat; David Dominguez; Josep Lluís Larriba Pey. *Benchmarking database systems for social network applications*, GRADES Workshop, 2013
<http://event.cwi.nl/grades2013/15-Angles.pdf>

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ 4: Ευρετήρια για Μεγάλα Χρονικά Μεταβαλλόμενα Γραφήματα

Ένα χαρακτηριστικό των κοινωνικών γράφων είναι ότι το περιεχόμενο και η δομή τους αλλάζει συνεχώς, π.χ. με την προσθήκη νέων χρηστών, τη προσθήκη και τη διαγραφή συσχετίσεων μεταξύ τους κλπ.

Στη βιβλιογραφία έχουν προταθεί ευρετήρια έτσι ώστε να απαντώνται γρήγορα ερωτήσεις τόσο προσπέλασης (reachability) όσο και κοντινότερης διαδρομής (shortest path). Θα εξετάσουμε πως οι αλγόριθμοι αυτοί συμπεριφέρονται στις περιπτώσεις που αποθηκεύουμε και την ιστορία των γράφων.

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ 5: Ευρετήρια για Γραφήματα με Ετικέτες

Θα εξετάσουμε την περίπτωση που οι κόμβοι ή οι ακμές του δικτύου περιέχουν δεδομένα ή ετικέτες. Θα δούμε γνωστά ευρετήρια και πως αυτά τροποποιούνται σε αυτήν την περίπτωση.

ΓΙΑ ΕΡΩΤΗΣΕΙΣ ΠΕΡΑΣΤΕ ΑΠΟ ΤΟ ΓΡΑΦΕΙΟ ΜΟΥ