

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ
Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

ΕΞΕΡΕΥΝΗΣΗ, ΟΠΤΙΚΟΠΟΙΗΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΜΕΓΑΛΩΝ ΔΕΔΟΜΕΝΩΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Σταύρος Μαρούλης, 210 6875403, stavmars[at]imis.athena-innovation.gr
Γιώργος Παπαστεφανάτος, 210 6875403, gparas[at]imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Σκοπός της εργασίας είναι η ανάπτυξη καινοτόμων εφαρμογών και τεχνικών για την εξερεύνηση (*exploration*), οπτικοποίηση (*visualization*) και ανάλυση (π.χ., ανάλυση συσχετίσεων, στατιστική ανάλυση, κτλ.) μεγάλων όγκων δεδομένων (*Big Data*).

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Θα καθοριστεί σε συνεννόηση με τον ενδιαφερόμενο. Java κατά προτίμηση

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Καθώς ο όγκος των δεδομένων που παράγονται καθημερινά μεγαλώνει, το κόστος φόρτωσης (*loading & storing*) σε ένα παραδοσιακό σύστημα βάσεων δεδομένων (π.χ. σχεσιακών βάσεων) γίνεται απαγορευτικό. Το πρόβλημα καθίσταται ιδιαίτερα έντονο στο πλαίσιο εφαρμογών εξερεύνησης και ανάλυσης δεδομένων (*data exploration and analysis*), όπου προσπαθούμε να εξάγουμε γνώση και αξία από δεδομένα (π.χ. ανακάλυψη ενδιαφέροντων μοτίβων, στατιστικών ιδιοτήτων, κτλ.) χωρίς να είναι πάντα γνωστό το τί αναζητούμε σε αυτά. Τέτοια σενάρια είναι αρκετά συνηθισμένα στην περίπτωση των επιστημονικών δεδομένων, όπως για παράδειγμα η ανάλυση των μεγάλων όγκων αστρονομικών ή βιολογικών δεδομένων που παράγονται καθημερινά από τηλεσκόπια και εργαστηριακά πειράματα. Στα σενάρια αυτά, οι τεχνικές των παραδοσιακών συστημάτων βάσεων δεδομένων (π.χ. δεικτοδότηση) δεν είναι αποδοτικές. Στόχος της διπλωματικής είναι η ανάπτυξη καινοτόμων τεχνικών για την εξερεύνηση, οπτικοποίηση και ανάλυση (π.χ., ανάλυση συσχετίσεων, εντοπισμός τάσεων, στατιστική ανάλυση, κτλ.) μεγάλων όγκων δεδομένων. Ιδιαίτερη έμφαση θα δοθεί στην ανάπτυξη μιας *web-based* εφαρμογής που θα επιτρέπει τη χρήση οπτικών μεθόδων για την πλοήγηση και ανάλυση δεδομένων.

Πιο συγκεκριμένα, σε συνεννόηση με τους ενδιαφερόμενους θα μελετηθούν θέματα σχετιζόμενα με: (α) την σταδιακή δημιουργία οπτικοποιήσεων, βασιζόμενες στην πλοήγηση του χρήστη (*incremental visualization*), (β) εξερεύνηση η οποία προσαρμόζεται στις προτιμήσεις και στις ανάγκες των χρηστών (*personalized exploration*), (γ) εξερεύνηση και πλοήγηση μέσω όψεων (*facets browsing*) (δ) δομές δεδομένων και ευρετηρίων (*data structures & indexes*) για την αποδοτική εξερεύνηση και ανάλυση δεδομένων, (ε) δημιουργία περιλήψεων (*data summarization/abstraction*), κτλ.

Η συγκεκριμένη εργασία θα πραγματοποιηθεί σε συνεργασία με το Ινστιτούτο Πληροφοριακών Συστημάτων του Ερευνητικού Κέντρου "Αθηνά".

ΣΧΕΤΙΚΟ ΥΛΙΚΟ

- Big Data: <http://www.ibm.com/big-data>
- S. Idreos. [Big Data Exploration](#). Taylor and Francis, 2013.
- I. Alagiannis, R. Borovica, M. Branco, S. Idreos, and A. Ailamaki, [NoDB: efficient query execution on raw data files](#).