

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ
Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

ΕΞΑΤΟΜΙΚΕΥΜΕΝΑ ΣΥΣΤΗΜΑΤΑ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ
ΕΙΣΑΓΩΓΙΚΟ ΣΗΜΕΙΩΜΑ

ΠΛΗΡΟΦΟΡΙΕΣ: Τάσος Αρβανίτης, 210 772 1402, anarv-at-dblab-dot-ntua-dot-gr

ΠΕΡΙΛΗΨΗ: Στόχος των συγκεκριμένων διπλωματικών εργασιών είναι η κατασκευή ενός πλήρως λειτουργικού εξατομικευμένου Συστήματος Διαχείρισης Βάσεων Δεδομένων.

ΑΡΙΘΜΟΣ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ: 3 (συζητήσιμος αριθμός ατόμων σε κάθε εργασία)

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Διάφορες (σε συνεννόηση με τους υποψηφίους), SQL/pgSQL

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Η αλματώδης ανάπτυξη του Παγκοσμίου Ιστού τον έχει καταστήσει μία τεράστια δεξαμενή πληροφοριών για την ανθρωπότητα. Βασικά χαρακτηριστικά της ανάπτυξης αυτής είναι: (α) η υπερπροσφορά πληροφοριών και υπηρεσιών και (β) η ανομοιογένεια μεταξύ των χρηστών. Σε αυτό το περιβάλλον, οι χρήστες δυσκολεύονται να ξεχωρίσουν τα προϊόντα ή υπηρεσίες που τους ενδιαφέρουν, ενώ οι επιχειρήσεις αναζητούν τρόπους πιο αποτελεσματικής προσέλευσης πελατών με διαφορετικά χαρακτηριστικά. Παράλληλα, η εξέλιξη των κοινωνικών δικτύων και του Web 2.0 και η αυξανόμενος όγκος συναλλαγών μέσω διαδικτύου (π.χ. αγορές προϊόντων) παρέχουν άφθονα δεδομένα για τα χαρακτηριστικά, τα ενδιαφέροντα και τις καταναλωτικές συνήθειες των χρηστών. Για παράδειγμα, σε ένα κοινωνικό δίκτυο οι χρήστες συνηθίζεται να μοιράζονται πληροφορίες για τις αγαπημένες τους ταινίες, μουσική, βιβλία, ενώ οι χρήστες ενός διαδικτυακού καταστήματος αφήνουν ίχνη από τα προϊόντα που αγόρασαν ή αναζήτησαν στο παρελθόν. Τα δεδομένα αυτά είναι χρήσιμο να συλλεχθούν και να αξιοποιηθούν για διάφορες εφαρμογές όπως μηχανές αναζήτησης, διαφημιστές κλπ. Για όλους αυτούς τους παραπάνω λόγους, η ανάγκη για συστήματα και τεχνικές που να παρέχουν στοχευμένες, εξατομικευμένες υπηρεσίες για χρήστες και επιχειρήσεις γίνεται επιτακτική.

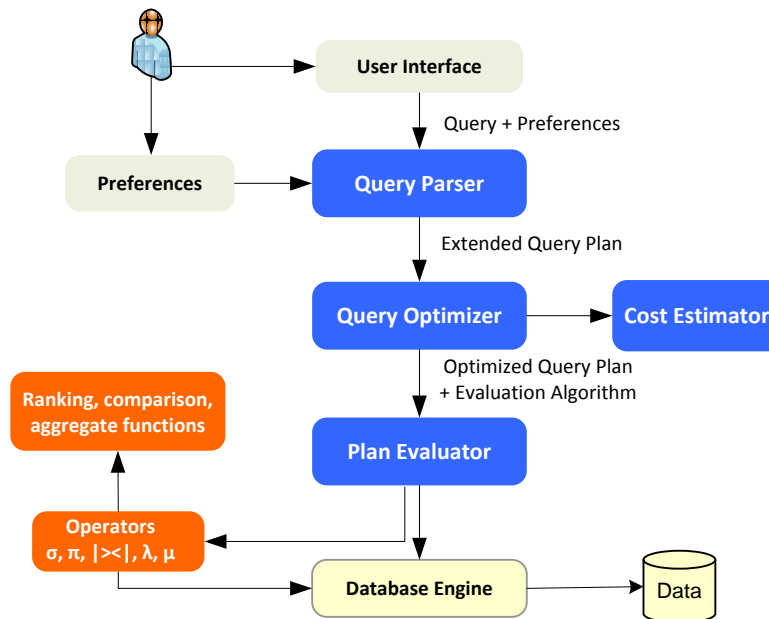
Τα *συστήματα εξατομίκευσης* προσαρμόζουν τη λειτουργία τους στα ενδιαφέροντα και τις προτιμήσεις των χρηστών. Χαρακτηριστικά παραδείγματα εμπορικών συστημάτων εξατομίκευσης είναι το [Amazon](#) και το [last.fm](#). Το πρώτο είναι μία υπηρεσία πώλησης προϊόντων (βιβλία, DVD κλπ.) ενώ το δεύτερο μία διαδικτυακή υπηρεσία ακρόασης μουσικής. Η ιστοσελίδα του Amazon προσαρμόζει τις προτάσεις για την αγορά βιβλίων που εμφανίζει σε κάθε χρήστη, με βάση τα προϊόντα που ο χρήστης αγόρασε στο παρελθόν ή τις αγορές άλλων χρηστών με παρόμοια ενδιαφέροντα. Αντίστοιχα, η εφαρμογή last.fm επιτρέπει στον χρήστη να ορίσει τα αγαπημένα του συγκροτήματα, είδη μουσικής κλπ. και στη συνέχεια «μαθαίνει» τις προτιμήσεις του και του προτείνει μουσική που είναι πιθανό να του αρέσει επίσης.

Δεδομένου ότι μια τυπική βάση δεδομένων δεν παρέχει λειτουργίες εξατομίκευσης, στις περισσότερες περιπτώσεις η εξατομίκευση επιτυγχάνεται σε επίπεδο εφαρμογής, δηλαδή αναπτύσσονται ειδικοί αλγόριθμοι οι οποίοι βασισμένοι στις προτιμήσεις των χρηστών επιλέγουν τα δεδομένα (προϊόντα, πληροφορίες κλπ.) που θα παρουσιαστούν στον χρήστη ή τα κατατάσσουν με κριτήριο το βαθμό με τον οποίο συμφωνούν με το προφίλ του. Αυτή η προσέγγιση έχει σημαντικά μειονεκτήματα καθώς χρειάζεται να αναπτυχθούν εξ αρχής νέες μέθοδοι ανά περίπτωση, ενώ συνήθως οι μέθοδοι αυτές έχουν εφαρμογή μόνο για δεδομένα μιας σχέσης (π.χ. ταινίες). Έτσι, είναι πρακτικά δύσκολο να αναπτυχθούν εφαρμογές που να συνδυάζουν δεδομένα και προτιμήσεις από πολλές σχέσεις (π.χ. ταινίες, σκηνοθέτες και ηθοποιούς) ώστε να απαντούν πιο σύνθετα ερωτήματα

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ
Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

εξατομίκευσης που περιέχουν συνενώσεις (joins), συναθροίσεις (aggregations) κλπ. Επιπλέον, τίθενται περιορισμοί όσον αφορά την αποδοτικότητά τους καθώς δεν αλληλεπιδρούν άμεσα με τον πυρήνα μιας βάσης δεδομένων που είναι ο βελτιστοποιητής ερωτήσεων (query optimizer).

Στο εργαστήριο ΣΒΓΔ έχει αναπτυχθεί ένα σύστημα διαχείρισης βάσης δεδομένων που επιτρέπει την εξατομίκευση εφαρμογών σε επίπεδο συστήματος [1]. Συγκεκριμένα, έχουμε επεκτείνει το σχεσιακό μοντέλο ώστε να επιτρέπει τον χειρισμό προτιμήσεων και σχέσεων με προτιμήσεις, ενώ έχουν οριστεί νέοι τελεστές προτίμησης και αλγόριθμοι αποτίμησης ερωτημάτων με προτιμήσεις. Το μοντέλο που προτείνουμε στο [1] είναι η πρώτη ολοκληρωμένη πρόταση ενός σχεσιακού μοντέλου δεδομένων που ενσωματώνει τις προτιμήσεις χρηστών στο επίπεδο της διαχείρισης των δεδομένων. Το πρωτότυπο σύστημα έχει βασιστεί στην open-source βάση δεδομένων PostgreSQL.



Σχήμα 1. Σχηματική αναπαράσταση συστήματος

Μια υψηλού επιπέδου σχηματική αναπαράσταση του συστήματος φαίνεται στο Σχήμα 1. Ένας χρήστης ορίζει τις προτιμήσεις του στο γραφικό περιβάλλον διεπαφής και θέτει ένα ερώτημα με προτιμήσεις. Το ερώτημα είναι εκφρασμένο σε μια εκτεταμένη σύνταξη της SQL, την Preference-SQL, η οποία ενσωματώνει τους νέους τελεστές προτίμησης. Στη συνέχεια, το ερώτημα αναλύεται συντακτικά από τον query parser και δημιουργείται ένα πλάνο εκτέλεσης (preference query plan). Όπως και σε ένα τυπικό σύστημα διαχείρισης βάσεων δεδομένων, ο query planner, μετασχηματίζει το πλάνο εκτέλεσης στο (κατά το δυνατό) βέλτιστο πλάνο και στη συνέχεια το ερώτημα εκτελείται και τα αποτελέσματα επιστρέφονται πίσω στον χρήστη.

Με τις συγκεκριμένες διπλωματικές εργασίες, στόχος μας είναι να επεκτείνουμε το πρωτότυπο αυτό σύστημα εξατομίκευσης. Στο Σχήμα 1 φαίνονται με μπλε χρώμα τα επιμέρους υποσυστήματα που θα αποτελέσουν και αντίστοιχες διπλωματικές εργασίες. Κάθε διπλωματική εργασία είναι ανεξάρτητη από τις υπόλοιπες και εστιάζει στην υλοποίηση του αντίστοιχου υποσυστήματος.

ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ/ΑΝΑΦΟΡΕΣ:

[1] Koutrika, G. and Ioannidis, Y. Personalization of Queries in Database Systems, ICDE 2004

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ
Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

ΕΞΑΤΟΜΙΚΕΥΜΕΝΑ ΣΥΣΤΗΜΑΤΑ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ
ΓΡΑΦΙΚΟ ΠΕΡΙΒΑΛΛΟΝ ΧΡΗΣΤΗ

ΠΛΗΡΟΦΟΡΙΕΣ: Τάσος Αρβανίτης, 210 772 1402, anarv-at-dblab-dot-ntua-dot-gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η υλοποίηση του γραφικού περιβάλλοντος διασύνδεσης με τους χρήστες και του υποσυστήματος αξιολόγησης των αποτελεσμάτων για ένα εξατομικευμένο Σύστημα Διαχείρισης Βάσεων Δεδομένων.

ΑΡΙΘΜΟΣ ΑΤΟΜΩΝ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Ανοικτή για το UI (θα καθοριστεί από τους υποψηφίους), Java, SQL/pgSQL

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Παραπέμπουμε στο εισαγωγικό σημείωμα για λεπτομέρειες σχετικά με τη χρησιμότητα και αρχιτεκτονική ενός εξατομικευμένου συστήματος διαχείρισης βάσεων δεδομένων. Η συγκεκριμένη διπλωματική εστιάζει στην υλοποίηση ενός γραφικού περιβάλλοντος για τη διασύνδεση των χρηστών με το σύστημα. Συγκεκριμένα, μέσω του γραφικού περιβάλλοντος ένας χρήστης θα μπορεί να δηλώνει τις προτιμήσεις του, να χρησιμοποιεί τις προτιμήσεις αυτές για να συνθέτει ερωτήματα, να τα εκτελεί και να λαμβάνει τα αποτελέσματα της εκτέλεσης. Αντίστοιχα, ο διαχειριστής του συστήματος θα μπορεί να ρυθμίζει διάφορες παραμέτρους του συστήματος (π.χ. τον αλγόριθμο αποτίμησης που χρησιμοποιείται) και να προσθέτει επιπλέον συναρτήσεις κατάταξης και συνάθροισης (library ranking/aggregation functions). Επιπλέον, θα υλοποιηθεί ένα υποσύστημα για την αξιολόγηση της χρηστικότητας του συστήματος σε επίπεδο χρήστη (user evaluation module). Συγκεκριμένα, το υποσύστημα αυτό θα καταγράφει τα ερωτήματα των χρηστών στη βάση δεδομένων του συστήματος και θα επικοινωνεί με το γραφικό περιβάλλον, ώστε να δίνει τη δυνατότητα στους ίδιους τους χρήστες να βαθμολογούν τις απαντήσεις που τους έδωσε το σύστημα με κριτήριο το βαθμό συνάφειας των αποτελεσμάτων με τις προτιμήσεις τους. Στόχος του υποσυστήματος αυτού είναι να διευκολύνει τη χρήση και αξιολόγηση του συστήματος εξατομικεύσης από τους χρήστες και να εντοπίσει τυχόν μη αναμενόμενη συμπεριφορά με στόχο τη βελτίωση της λειτουργικότητάς του. Η συγκεκριμένη διπλωματική είναι ανεξάρτητη από τις υπόλοιπες και θα αποτελέσει μέρος ενός πλήρως λειτουργικού εξατομικευμένου συστήματος διαχείρισης δεδομένων το οποίο, με την ολοκλήρωσή του, θα επιδιωχθεί να παρουσιαστεί σε διεθνές συνέδριο βάσεων δεδομένων.

ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ/ΑΝΑΦΟΡΕΣ:

[1] Koutrika, G. and Ioannidis, Y. Personalization of Queries in Database Systems, ICDE 2004

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ
Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

ΕΞΑΤΟΜΙΚΕΥΜΕΝΑ ΣΥΣΤΗΜΑΤΑ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ
ΥΠΟΣΥΣΤΗΜΑ ΑΝΑΛΥΣΗΣ ΕΡΩΤΗΜΑΤΩΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Τάσος Αρβανίτης, 210 772 1402, anarv-at-dblab-dot-ntua-dot-gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η υλοποίηση ενός υποσυστήματος ανάλυσης ερωτημάτων (query parser) με προτιμήσεις για ένα εξατομικευμένο Σύστημα Διαχείρισης Βάσεων Δεδομένων.

ΑΡΙΘΜΟΣ ΑΤΟΜΩΝ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Ανοικτή (θα καθοριστεί από τους υποψηφίους), SQL/pgSQL

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Παραπέμπουμε στο εισαγωγικό σημείωμα για λεπτομέρειες σχετικά με τη χρησιμότητα και αρχιτεκτονική ενός εξατομικευμένου συστήματος διαχείρισης βάσεων δεδομένων. Η συγκεκριμένη διπλωματική εστιάζει στην υλοποίηση ενός υποσυστήματος συντακτικής ανάλυσης ερωτημάτων (query parser). Συγκεκριμένα, θα οριστεί η Preference-SQL η οποία θα ενσωματώνει τους νέους τελεστές προτίμησης που έχουν οριστεί στο [1] στη standard SQL. Επιπλέον, θα υλοποιηθεί ένας πλήρως λειτουργικός *query parser* ο οποίος θα λαμβάνει ως είσοδο ένα ερώτημα σε μορφή Preference-SQL, θα το αναλύει συντακτικά και θα κατασκευάζει ένα πλάνο εκτέλεσης της ερώτησης. Η συγκεκριμένη διπλωματική είναι ανεξάρτητη από τις υπόλοιπες και θα αποτελέσει μέρος ενός πλήρως λειτουργικού εξατομικευμένου συστήματος διαχείρισης δεδομένων το οποίο, με την ολοκλήρωσή του, θα επιδιωχθεί να παρουσιαστεί σε διεθνές συνέδριο βάσεων δεδομένων.

ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ/ΑΝΑΦΟΡΕΣ:

[1] Koutrika, G. and Ioannidis, Y. Personalization of Queries in Database Systems, ICDE 2004

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ
Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

ΕΞΑΤΟΜΙΚΕΥΜΕΝΑ ΣΥΣΤΗΜΑΤΑ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ
ΥΠΟΣΥΣΤΗΜΑ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ ΕΡΩΤΗΜΑΤΩΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Τάσος Αρβανίτης, 210 772 1402, anarv-at-dblab-dot-ntua-dot-gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η υλοποίηση ενός υποσυστήματος σχεδιασμού και βελτιστοποίησης ερωτημάτων (query planner/optimizer) για ένα εξατομικευμένο Σύστημα Διαχείρισης Βάσεων Δεδομένων.

ΑΡΙΘΜΟΣ ΑΤΟΜΩΝ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, SQL/pgSQL

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Παραπέμπουμε στο εισαγωγικό σημείωμα για λεπτομέρειες σχετικά με τη χρησιμότητα και αρχιτεκτονική ενός εξατομικευμένου συστήματος διαχείρισης βάσεων δεδομένων. Η συγκεκριμένη διπλωματική εστιάζει στην επέκταση του υπάρχοντος βελτιστοποιητή ερωτημάτων του συστήματος (query planner/optimizer). Συγκεκριμένα, ο query planner θα λαμβάνει ως είσοδο το πλάνο εκτέλεσης που παράγει ο query parser και θα εκτελεί αναζήτηση του βέλτιστου πλάνου εκτέλεσης της ερώτησης στον χώρο των πιθανών πλάνων χρησιμοποιώντας τους μετασχηματισμούς ισοδυναμίας που έχουν οριστεί στο [1]. Θα διερευνηθούν διάφορες επιλογές υλοποίησης που θα περιλαμβάνουν (α) βελτιστοποιήσεις που βασίζονται σε κανόνες (rule-based), (β) βελτιστοποιήσεις βασισμένες στο κόστος (cost-based) καθώς και (γ) υβριδικές προσεγγίσεις. Η συγκεκριμένη διπλωματική είναι ανεξάρτητη από τις υπόλοιπες και θα αποτελέσει μέρος ενός πλήρως λειτουργικού εξατομικευμένου συστήματος διαχείρισης δεδομένων το οποίο, με την ολοκλήρωσή του, θα επιδιωχθεί να παρουσιαστεί σε διεθνές συνέδριο βάσεων δεδομένων.

ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ/ΑΝΑΦΟΡΕΣ:

[1] Koutrika, G. and Ioannidis, Y. Personalization of Queries in Database Systems, ICDE 2004