

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ
Εργαστήριο Συστημάτων Βάσεων Γνώσεων & Δεδομένων

**p-XQuery: ΕΠΕΚΤΑΣΗ ΤΗΣ ΓΛΩΣΣΑΣ ΕΡΩΤΗΣΕΩΝ XQuery ΓΙΑ ΥΠΟΣΤΗΡΙΞΗ
ΕΡΩΤΗΣΕΩΝ ΜΕ ΜΕΡΙΚΟ ΚΑΘΟΡΙΣΜΟ ΔΟΜΙΚΩΝ ΠΕΡΙΟΡΙΣΜΩΝ**

ΠΛΗΡΟΦΟΡΙΕΣ: Θ. Δαλαμάγκας, 210 7721402, dalamag@dblab.ntua.gr, Σ. Σουλδάτος, 210 7721402
stef@dblab.ece.ntua.gr.

ΠΕΡΙΛΗΨΗ: Η διπλωματική εργασία στοχεύει στο σχεδιασμό και την υλοποίηση επέκτασης της γλώσσας ερωτήσεων XQuery. Με τη νέα μορφή της γλώσσας και τις δυνατότητες για νέες εντολές και δηλώσεις που θα προσφέρονται, θα είναι δυνατή (α) η εύκολη διατύπωση ερωτήσεων στις οποίες οι δομικοί περιορισμοί ορίζονται μερικώς και όχι πλήρως, και (β) η αποδοτική αποτίμηση των ερωτήσεων αυτών σε συλλογές από αρχεία XML.

ΑΤΟΜΑ: 2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java/C++

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Η γλώσσα XML είναι πλέον πρότυπο ανταλλαγής δεδομένων σε εφαρμογές του Παγκόσμιου Ιστού. Η χρήση ετικετών (tags/elements) για το μαρκάρισμα και τον χαρακτηρισμό των δεδομένων σε ένα XML αρχείο διευκολύνει την επεξεργασία του από προγράμματα. Δείτε ένα πολύ απλό παράδειγμα XML αρχείου που αποθηκεύει πληροφορίες για φορητούς υπολογιστές:

```
<?xml version="1.0">
<Doc>
  <Type> Laptop
    <Brand> Sony
      <Item PRICE=1200> VAIO VGN-B3VP </Item>
      <Item PRICE=1300> VAIO VGN-S1XP </Item>
    </Brand>
    <Brand> Dell
      <Item PRICE=1500> X1 </Item>
    </Brand>
  </Type>
</Doc>
```

Για την διαχείριση XML αρχείων έχουν προταθεί αρκετές γλώσσες ερωτήσεων, με την XQuery να είναι η πιο δημοφιλής. Οι ερωτήσεις σε XQuery βασίζονται σε *εκφράσεις μονοπατιών*. Για παράδειγμα, για να βρει κανείς το κομμάτι του αρχείου που περιέχει τους φορητούς υπολογιστές Dell πρέπει να καθορίσει ένα μονοπάτι ορίζοντας (α) *δομικούς περιορισμούς*, π.χ. /Doc/Type/Brand (δηλ. το Brand να είναι παιδί του Type, το οποίο να είναι παιδί του Doc) και (β) *περιορισμούς τιμών*, π.χ. Type[text()=Laptop] και Brand[text()=Dell].

Πολλά όμως αρχεία XML, αν και αποθηκεύουν ίδιου τύπου αντικείμενα, μπορεί να έχουν διαφορετική οργάνωση. Για παράδειγμα, σε ένα XML αρχείο κάποιος μπορεί να έχει πρώτα την μάρκα Brand και μετά τον τύπο Type, όπως φαίνεται στο διπλανό XML αρχείο. Το γεγονός αυτό δυσκολεύει την διατύπωση ερωτήσεων, με ενιαίο τρόπο, που να αφορούν και στα δύο XML αρχεία.

```
...
<Brand> Dell
  <Type> Laptop
    <Item PRICE=1500> X1 </Item>
  </Type>
</Brand>
...
```

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ

Εργαστήριο Συστημάτων Βάσεων Γνώσεων & Δεδομένων

Η διπλωματική εργασία στοχεύει στον *σχεδιασμό και στην υλοποίηση επέκτασης της γλώσσας ερωτήσεων XQuery*. Η νέα μορφή της γλώσσας, με τις δυνατότητες για νέες εντολές και δηλώσεις που θα προσφέρονται, θα δίνει τις εξής δυνατότητες:

- (α) εύκολη διατύπωση ερωτήσεων στις οποίες οι δομικοί περιορισμοί ορίζονται μερικώς και όχι πλήρως. Έτσι ο χρήστης θα μπορεί να διατυπώνει για παράδειγμα την ερώτηση ‘Βρες φορητούς υπολογιστές Dell, ανεξάρτητα από το αν θα πρέπει να χρησιμοποιηθεί η έκφραση μονοπατιού /Doc/Type/Brand ή /Doc/Brand/Type, ανάλογα με το αρχείο XML στο οποίο απευθύνεται η ερώτηση, και
- (β) αποδοτική αποτίμηση των ερωτήσεων αυτών σε συλλογές από αρχεία XML με διαφορετική οργάνωση.

Η διπλωματική αυτή επομένως, θα ασχοληθεί με τα παρακάτω θέματα:

1. Σχεδιασμός p-XQuery, μιας επέκτασης της γλώσσας ερωτήσεων XQuery για υποστήριξη ερωτήσεων με μερικό καθορισμό δομικών περιορισμών.
2. Προσθήκη των κατάλληλων νέων εντολών/δηλώσεων σε υλοποιήσεις XQuery ανοικτού κώδικα.
3. Πλήρης υλοποίηση της γλώσσας.
4. Υλοποίηση μηχανισμού αποτίμησης ερωτήσεων σε συλλογές από αρχεία XML
5. Αξιολόγηση, μέτρηση απόδοσης.

Για περισσότερες πληροφορίες, δείτε τις αναφορές:

α) <http://www.dbnet.ece.ntua.gr/~dalamag/pub/dtk-t-jods04.pdf>

β) <http://www.dbnet.ece.ntua.gr/~dalamag/pub/wise05.pdf>

γ) <http://www.dbnet.ece.ntua.gr/~dalamag/pub/cikm05.pdf>