

**ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ**  
Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

**XML Wrap&Roll: ΣΥΣΤΗΜΑ ΣΥΝΤΑΞΗΣ, ΔΙΑΧΕΙΡΙΣΗΣ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ  
ΕΡΩΤΗΣΕΩΝ XML ΑΡΧΕΙΩΝ**

ΠΛΗΡΟΦΟΡΙΕΣ: Θ. Δαλαμάγκας, 210 7721402, [dalamag@dblabb.ntua.gr](mailto:dalamag@dblabb.ntua.gr), Σ. Σουλδάτος, 210 7721402  
[stef@dblabb.ece.ntua.gr](mailto:stef@dblabb.ece.ntua.gr)

**ΠΕΡΙΛΗΨΗ:** Η διπλωματική εργασία στοχεύει στο σχεδιασμό ενός πρότυπου συστήματος κατασκευής και διαχείρισης δεδομένων από βάσεις δεδομένων κωδικοποιημένων σε XML αρχεία.

**ΑΤΟΜΑ:** 1-2

**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

```
<?xml version="1.0">
<Doc>
<Type> Laptop
<Brand> Sony
<Item PRICE=1200> VAI0 VGN-B3VP </Item>
<Item PRICE=1300> VAI0 VGN-S1XP </Item>
</Brand>
<Brand> Dell
<Item PRICE=1500> X1 </Item>
</Brand>
</Type>
</Doc>
```

πολύ απλό παράδειγμα XML αρχείου που αποθηκεύει πληροφορίες για φορητούς υπολογιστές. Θεωρούμε ότι μια σειρά από ετικέτες, όπως π.χ. Type, Brand, δημιουργούν μια ιεραρχία στο XML αρχείο, ενώ οι ετικέτες item περικλείουν τα δεδομένα που ανακτώνται από μια βάση δεδομένων.

Στο σύστημα **XML Wrap&Roll**, πριν την δημιουργία των XML αρχείων, ο χρήστης θα μπορεί να ορίσει **διαστάσεις (dimensions)** που περιγράφουν το πεδίο γνώσης που χαρακτηρίζει τα δεδομένα, και επιτρεπτές **τιμές διαστάσεων**. Για παράδειγμα, στο πεδίο γνώσης των υπολογιστών, παραδείγματα διαστάσεων είναι οι παρακάτω:

Type={desktop, laptop}, Brand={Sony, IBM, HP, Dell}, Monitor\_type={17'', 19''}

Επίσης, ο χρήστης θα μπορεί να ορίσει όπου θέλει **περιορισμούς δομής (structural constraints)**. Για παράδειγμα, θα μπορεί να ορίσει ότι Type→Brand (δηλαδή ότι η διάσταση Brand θα πρέπει να είναι φωλιασμένη ως παιδί στην διάσταση Type), ή ότι Type ⇒ Brand (δηλαδή ότι η διάσταση Brand θα πρέπει να είναι φωλιασμένη ως απόγονος στην διάσταση Type). Το σύστημα θα επιτρέπει την κατασκευή XML αρχείων που θα τηρούν τους παραπάνω περιορισμούς όσον αφορά τις ετικέτες που αντιστοιχούν στις διαστάσεις. Αυτή είναι η λειτουργία **Wrap** στο **Wrap&Roll!**

**ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ**  
Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

Τα XML αρχεία που θα κατασκευάζονται με το σύστημα **XML Wrap&Roll**, αν και

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

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. Η διπλωματική θα υλοποιήσει γλώσσα ερωτήσεων που θα βασίζεται στις διαστάσεις και τις τιμές τους και 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.

(β) Αποδοτική αποτίμηση των ερωτήσεων αυτών σε συλλογές από αρχεία XML με διαφορετική οργάνωση.

Αυτή είναι η λειτουργία **Roll** στο **Wrap&Roll**!

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

1. Ανάλυση απαιτήσεων συστήματος.
2. Σχεδίαση συστήματος.
3. Υλοποίηση διαπροσωπείας, αλγορίθμων επεξεργασίας και αποτίμησης ερωτήσεων, και ανάπτυξη πρότυπου συστήματος σύνταξης, διαχείρισης και επεξεργασίας ερωτήσεων σε XML αρχεία.

Για περισσότερες πληροφορίες, δείτε την αναφορά

<http://www.dbnet.ece.ntua.gr/~dalamag/pub/dtkt-jods04.pdf>