



XML Οντότητες (Entities)

Παπαδάκης Γιάννης

26.10.08

XML οντότητες - entities

- Οι οντότητες στην XML χρησιμεύουν για να αποθηκεύσουν μικρά (ή και μεγάλα) τμήματα κώδικα
- Οι οντότητες απλουστεύουν τη δομή στην XML
- Οι οντότητες μπορούν να περιέχουν από μια φράση, μια εικόνα, ένα ολόκληρο dtd, ένα video κλπ

XML οντότητες - entities

- Υπάρχουν 2 ειδών οντότητες:
 - General entities
Αναφέρονται στο XML αρχείο.
Αναφερόμαστε σ' αυτές ως εξής:
&entity;
 - Parameter entities
Αναφέρονται στο dtd αρχείο.
Αναφερόμαστε σ' αυτές ως εξής:
%entity;

XML οντότητες - General entities(1/7)

- Υπάρχουν 3 ειδών General entities:
 - General entities/ Character entities
Αναφέρονται σε ένα και μόνο χαρακτήρα.
Έχουν 3 υποδιαιρέσεις:
 - Predefined character entities
 - **& = &**
 - **< = <**
 - **> = >**
 - **” = "**
 - **` = '**
 - **Τις μαθαίνουμε απ’ έξω, δε χρειάζεται να τις δηλώσουμε στο dtd**

XML οντότητες - General entities(2/7)

- General entities/Character entities:
 - Numbered character entities
 - Εξαιτίας της υποστήριξης από την XML του προτύπου UNICODE, έχουμε τη δυνατότητα να αναφερθούμε μέσα στο xml έγγραφο σε οποιοδήποτε UNICODE χαρακτήρα ως εξής:
 - **&#number;** Π.χ, ᇉ
δε χρειάζεται να τις δηλώσουμε στο dtd.

XML οντότητες - General entities(3/7)

- General entities /Character entities:

- Name character entities

- Το ίδιο με τις Numbered character entities, μόνο που δε χρειάζεται να θυμόμαστε κάποιον αριθμό, μόνο ένα όνομα. Αναφερόμαστε σε αυτές στο XML ως εξής:

&name; Όπου *name*, η οντότητα που έχουμε ορίσει στο dtd. Για παράδειγμα, η οντότητα

```
<!ENTITY Lambda "&#923;">
```

Αναφέρεται στο xml ως εξής: Λ

XML οντότητες - General entities(4/7)

- General entities/Mixed Content Entities
Μπορούν να περιέχουν άπειρο κείμενο ή/και σημείωση (markup). Εχουμε 2 υποδιαιρέσεις:
 - Internal Entities
Περιέχουν κείμενο ή/και σημείωση που το χρησιμοποιούμε πολλές φορές σε ένα xml και βαριόμαστε να το γράφουμε ξανά και ξανά:

XML οντότητες - General entities(5/7)

```
<?xml version="1.0">
<!DOCTYPE TEI.2 SYSTEM "teixlite.dtd" [
<!ENTITY etext "<p>This text is copyright 2001 of The Electronic Text
Center<lb /> University of Virginia</p>">
]>
<TEI.2>
<text>
<div1 type="copyright_statement"><head>Copyright Statement</head>
&etext;
</div1>
</text>
</TEI.2>
```

When processed by the XML parser, the text would then expand to the following:

```
<?xml version="1.0">
<!DOCTYPE TEI.2 SYSTEM "teixlite.dtd" [
<!ENTITY etext "<p>This text is copyrighted 2001 of The Electronic
Text Center
<lb />University of Virginia Library</p>">
]>
<TEI.2>
<text>
<div1 type="copyright_statement"><head>Copyright Statement</head>
○<p>This text is copyright 2001 of The Electronic Text Center<lb /> University
of Virginia Library</p> </div1></text></TEI.2>
```


XML οντότητες - General entities(6/7)

- General entities (συνέχεια):
 - Mixed Content Entities.
Μπορούν να περιέχουν άπειρο κείμενο ή/και σημείωση (markup). Εχουμε 2 υποδιαιρέσεις:
 - External Entities. Το ίδιο με τις εσωτερικές οντότητες, μόνο που σε αυτήν την περίπτωση αναφερόμαστε σε εξωτερικά αρχεία του συστήματος:

XML οντότητες - General entities(7/7)

```
<?xml version="1.0">
<!DOCTYPE TEI.2 SYSTEM "teixlite.dtd" [
<!ENTITY letter1 SYSTEM "letter1.txt">
]>
<TEI.2>
<text>
<body><head>Edgar Allan Poe's Early Life</head>
  <div1 type="section"><head>Letter</head>
  <p>In the following letter, Edgar Allan Poe pleads with his father for
money.</p>
  </div1>
  <div1 type="letter"><head>Letter from Poe</head>
&letter1;
  </div1>
</body>
</text>
</TEI.2>
```

Το αρχείο letter1.txt θα πάρει τη θέση του μέσα στο αρχείο xml

XML οντότητες - Parameter entities(1/6)

- Όταν στο dtd επαναλαμβάνονται δηλώσεις attributes (ATTLIST) ή δηλώσεις elements (content models), μπορούμε να γλυτώσουμε πληκτρολόγηση χρησιμοποιώντας parameter entities. Προσοχή, οι οντότητες αυτές δηλώνονται **και** αναφέρονται στο dtd
- Οι parameter entities δηλώνονται ως εξής στο dtd:
 - %entity;

XML οντότητες - Parameter entities(2/6)

- Κατηγοριοποιούνται σε 2 κατηγορίες:
 - Internal Parameter Entities
Δήλωση:
<!ENTITY %entity_name "entity_content">
Το μόνο που αλλάζει με τις **general entities** είναι η χρήση του % αντί του &

XML οντότητες - Parameter entities(3/6)

○ Παράδειγμα:

```
<!ELEMENT apartment (address, footage, rooms, baths, rent)>
<!ELEMENT sublet (address, footage, rooms, baths, rent)>
<!ELEMENT house (address, footage, rooms, baths, price)>
<!ELEMENT condo (address, footage, rooms, baths, price)>

<!ENTITY %phys_desc "address, footage, rooms, baths">

<!ELEMENT apartment (%phys_desc;, rent)>
<!ELEMENT sublet (%phys_desc;, rent)>
<!ELEMENT house (%phys_desc;, price)>
<!ELEMENT condo (%phys_desc;, price)>
```

XML οντότητες - Parameter entities(4/6)

- External Parameter Entities
 - Το ίδιο με τις αντίστοιχες εσωτερικές, μόνο που αντλούν το περιεχόμενό τους από εξωτερικά αρχεία
 - Δηλώνονται ως εξής (στο dtd):
 - `<!ENTITY %EntityName SYSTEM URL>`
 - Αναφέρονται ως εξής (στο dtd):
 - `%EntityName;`

XML οντότητες - Parameter entities(5/6)

- Παράδειγμα

Εχω το εξής αρχείο cust.ent (η κατάλληλη δεν παίζει ρόλο)

```
<!ELEMENT Name (FirstName , LastName)>
```

```
<!ELEMENT FirstName (#PCDATA)>
```

```
<!ELEMENT LastName (#PCDATA)>
```

```
<!ELEMENT Phone (#PCDATA)>
```

```
<!ELEMENT Customer (Name , Phone)>
```

XML οντότητες - Parameter entities(6/6)

○ Παράδειγμα (συνέχεια...)

Στο αρχείο order.dtd, έχουμε:

```
<!ENTITY %CustomerInfo SYSTEM "cust.ent">
```

```
%CustomerInfo;
```

```
<!ELEMENT Order (Customer ,OrderDate , PaymentMethod)>
```

```
<!ELEMENT OrderDate (#PCDATA)>
```

```
<!ELEMENT PaymentMethod(#PCDATA)>
```