

Τα βασικά της XML

Δευτέρα, 20 Οκτωβρίου 2008
Εισηγητής: Παπαδάκης Γιάννης

Η XML είναι μια μετα-γλώσσα

- Δεν είναι σαν την HTML
- Δημιουργούμε τα συστατικά - elements (ετικέτες - tags αρχής - τέλους) όταν τα χρειαζόμαστε
- Οι προδιαγραφές των συστατικών καταγράφονται σε αρχεία Document Type Definition (DTD) ή XML schemas
- Είναι μια μεθοδολογία σύνταξης για συγκεκριμένες γλώσσες σημείωσης όπως MusicML, MathML, EAD, TEI

Τί είναι η XML?

- **Extensible Markup Language**
- Μια σύνταξη για έγγραφα
- Μια **Meta** γλώσσα σημείωσης
- Μια δομική και θεματική γλώσσα, όχι μια γλώσσα μορφοποίησης
- Μια γλώσσα με πολλές εφαρμογές

Η XML περιγράφει δομή και περιεχόμενο, όχι μορφοποίηση

- Τα XML έγγραφα σχηματίζουν μια δενδροειδή μορφή
- Τα συστατικά και τα χαρακτηριστικά τους προσδιορίζουν το είδος των συστατικών
- Η μορφοποίηση προστίθεται π.χ. Με ένα style sheet

Για ποιό λόγο χρησιμοποιείται η XML?

- Για την κατασκευή συγκεκριμένων γλωσσών σημείωσης
- Για αυτο-περιγραφή δεδομένων
- Ανταλλαγή δεδομένων μεταξύ εφαρμογών
- Δομημένα και ολοκληρωμένα δεδομένα

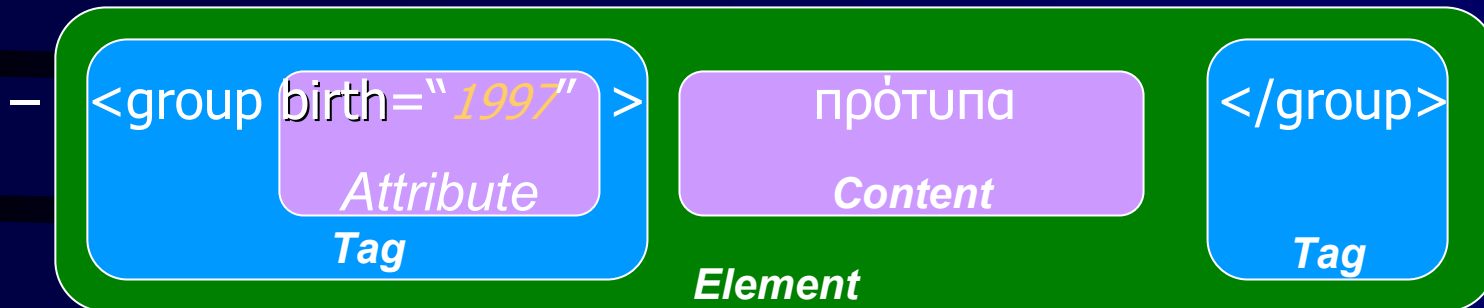
Η δήλωση XML

```
<?xml version="1.0" standalone="yes" encoding = "UTF-8"?>
```

- version χαρακτηριστικό
 - απαιτούμενο
 - Έχει πάντα την τιμή 1.0
- standalone χαρακτηριστικό
 - yes
 - no
- encoding χαρακτηριστικό
 - UTF-8
 - 8859_1
 - κλπ.

XML έγγραφα

- Σύνταξη σημείωσης - markup



- Τα πεζά είναι διαφορετικά από τα κεφαλαία
- Υποστήριξη Unicode
- **Ονόματα** ::= (γράμμα ή `'_'` ή `':'`) (Nmtoken)+, όπου *Nmtoken* είναι μια συμβολοσειρά που αποτελείται από:
 $(\text{γράμμα ή ψηφίο ή } ':' \text{ ή } '-' \text{ ή } '_' \text{ ή } ':')^*$
- Τα ονόματα των χαρακτηριστικών - attributes (**AttValue**), εσωτερικών οντοτήτων - internal entities (**EntityValue**), και εξωτερικών προσδιοριστών external identifiers (**SystemLiteral**) δέχονται μόνο γράμματα εντός εισαγωγικών
- Τα έγγραφα πρέπει να είναι well-formed
- Τα μέρη ενός εγγράφου:
 - $document ::= prolog^* element epilog^*$
 - Υπάρχει μόνο ένα συστατικό: "root" ή "document" συστατικό

Document

Prolog

```
<?xml version="1.0"?>  
<!-- σχόλια και processing instructions -->  
<!DOCTYPE sdsc_play_groups SYSTEM "http://localhost/spg.dtd">  
<!-- σχόλια και processing instructions -->
```

Body

```
<sdsc_play_groups>  
  <play_group ID="Data-issues">  
    <member_groups>  
      <group>Scientific Computing</group>  
      <group>Data Intensive Computing</group>  
      <group>Security Technologies</group>  
    </member_groups>  
    <charter source="XPG"/>  
    <url>http://www.sdsc.edu/~marciano/XML/xpg.html</url>  
    <title>XML Play Group</title>  
  </play_group>  
</sdsc_play_groups>
```

Epilog

```
<!-- σχόλια και processing instructions -->
```


Η XML βασίζεται στη σημείωση

```
<bibliography>
```

```
<paper ID= "object-fusion">
```

```
<authors>
```

```
<author>Y.Papakonstantinou</author>
```

```
<author>S. Abiteboul</author>
```

```
<author>H. Garcia-Molina</author>
```

```
</authors>
```

```
<fullPaper source="fusion"/>
```

```
<title>Object Fusion in Mediator Systems</title>
```

```
<booktitle>VLDB 96</booktitle>
```

```
</paper>
```

```
</bibliography>
```

*Η σημείωση υποδεικνύει
δομή και περιεχόμενο*

*Η μορφοποίηση
παρέχεται αλλού*

Elements και περιεχόμενο

<bibliography>

τύπος element

element

<paper ID="object-fusion">

περιεχόμενο
element

<authors>

<author>Y.Papakonstantinou</author>

<author>S. Abiteboul</author>

<author>H. Garcia-Molina</author>

</authors>

<fullPaper source="fusion"/>

<title>Object Fusion in Mediator Systems</title>

<booktitle>VLDB 96</booktitle>

άδειο
element

</paper>

</bibliography>

κείμενο

Element Attributes

όνομα Attribute

τιμή Attribute

<bibliography>

<paper pid="object-fusion">

<authors>

<author>Y.Papakonstantinou</author>

<author>S. Abiteboul</author>

<author>H. Garcia-Molina</author>

</authors>

<fullPaper source="fusion"/>

<title>Object Fusion in Mediator Systems</title>

<booktitle>VLDB 96</booktitle>

</paper>

</bibliography>

Άλλα συντακτικά στοιχεία της XML

- Τα *σχόλια* (comments) επιτρέπονται οπουδήποτε εκτός από το εσωτερικό των ετικετών. Ένα σχόλιο ξεκινά με το <!-- και τελειώνει με το -->
 - <!-- Αυτό είναι ένα σχόλιο -->
- Οι *οδηγίες επεξεργασίας* (Processing Instructions) PI επιτρέπουν σε ένα XML τεκμήριο να περιέχει οδηγίες που απευθύνονται σε προγράμματα εφαρμογών. Μια οδηγία επεξεργασίας περιλαμβάνει το όνομα μιας εφαρμογής στην οποία απευθύνεται, ακολουθούμενο από πληροφορίες (οδηγίες επεξεργασίες, παραμέτρους κ.λ.π.) οι οποίες θέλουμε να περάσουν στην εφαρμογή
 - <?xml-stylesheet href="book.css" type="text/css"?>

Άλλα συντακτικά στοιχεία της XML

- Ένα τμήμα CDATA ξεκινά με `<![CDATA[` και τελειώνει με `]]>`. Το περιβάλλον CDATA χρησιμοποιείται όταν θέλουμε να συμπεριλαμβάνουμε ετικέτες και άλλα ειδικά σύμβολα τα οποία θέλουμε να εκληφθούν απλά σαν ακολουθίες χαρακτήρων (συμβολοσειρές)
 - `<![CDATA[<αρχή> αυτό είναι ένα λανθασμένο στοιχείο </τέλος>]]>`

Αναφορές Χαρακτήρων

- Οι αναφορές χαρακτήρων χρησιμοποιούνται για την εισαγωγή οποιουδήποτε χαρακτήρα του ISO/IEC 10646 συνόλου χαρακτήρων σε ένα XML τεκμήριο. Αυτό γίνεται περικλείοντας το κωδικό του χαρακτήρα ανάμεσα σε & και ;
- Αν η αναφορά χαρακτήρα ξεκινά με &#x τότε τα ψηφία που ακολουθούν μέχρι το σύμβολο τερματισμού ; παρέχουν τη δεκαεξαδική αναπαράσταση του χαρακτήρα στο ISO/IEC 10646
 - ℞
- Αν όμως ξεκινά απλά με το &# τότε τα ψηφία που ακολουθούν μέχρι το σύμβολο τερματισμού ; παρέχουν τη δεκαδική αναπαράσταση του χαρακτήρα στο ISO/IEC 10646
 - το ℞

Βασικοί κανόνες για *well-formed* XML (υποχρεωτικοί)

- Κάθε αρχική ετικέτα πρέπει να συνοδεύεται από την αντίστοιχη τελική
- Τα άδεια tags – ετικέτες τελειώνουν με `</>`
- Οι ετικέτες πρέπει να ενθυλακώνονται
- Οι τιμές των χαρακτηριστικών πρέπει να μπαίνουν σε εισαγωγικά
- Τα ονόματα των ετικετών είναι ευαίσθητα στη χρήση πεζών-κεφαλαίων
- Κάθε έγγραφο επιτρέπεται να έχει ακριβώς ένα συστατικό-ρίζα
- `<` και `&` χρησιμοποιούνται μόνο ως start tags και οντότητες - entities
- Επιτρέπονται αναφορές σε πέντε μόνο προκαθορισμένες οντότητες - entities

Βασικοί κανόνες για *well-formed* XML

- Κάθε αρχική ετικέτα πρέπει να συνοδεύεται από την αντίστοιχη τελική
 - Για ετικέτες με περιεχόμενο: `<div>[...]</div>`
 - Για άδειες ετικέτες : `
` ή `
</br>` (αντί `
`)

Βασικοί κανόνες για *well-formed* XML

Οι ετικέτες πρέπει να ενθυλακώνονται:

Σωστό:

```
<work><author>Tolstoy</author></work>
```

Λάθος:

```
<work><author>Tolstoy</work></author>
```

Βασικοί κανόνες για *well-formed* XML

Οι τιμές των χαρακτηριστικών πρέπει να μπαίνουν σε εισαγωγικά:

`<page n="23"/>` ή `<page n='23'/>`

Όχι όμως

`<page n=23/>`

Βασικοί κανόνες για *well-formed* XML

Τα ονόματα των ετικετών είναι ευαίσθητα στη χρήση πεζών-κεφαλαίων:

`<title></title>` ή `<TITLE></TITLE>`

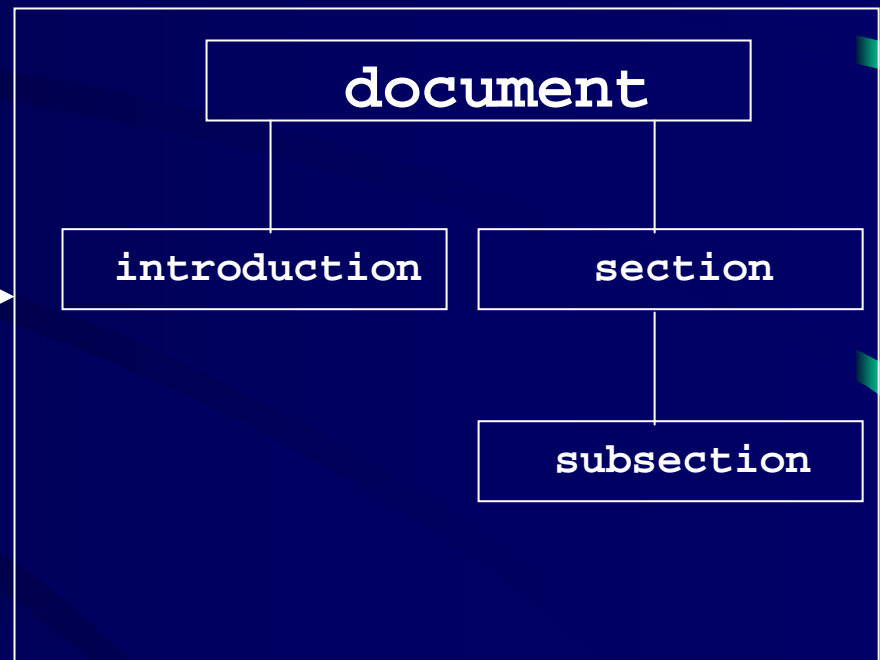
Όχι όμως

`<title></TITLE>`

Βασικοί κανόνες για *well-formed* XML

- Κάθε έγγραφο επιτρέπεται να έχει ακριβώς ένα συστατικό-ρίζα

```
<document>  
  <introduction>  
</introduction>  
  <section>  
    <subsection>  
  </subsection>  
  </section>  
</document>
```



Επιτρέπονται αναφορές σε πέντε μόνο προκαθορισμένες οντότητες

- Σωστό:

- `&`
- `<`
- `>`
- `"`
- `'`

- Λάθος:

- `©`
- `®`
- `&tm;`
- `α`
- `é`
- ` `
- etc.

< και & χρησιμοποιούνται μόνο ως start tags και οντότητες - entities

- **Σωστό:** <H1>O'Reilly & Associates</H1>
- **Λάθος:** <H1> O'Reilly & Associates</H1>
- **Σωστό:**
 - <CODE>for (int i = 0; i <= args.length; i++) {
</CODE>
- **Λάθος:** <CODE>for (int i = 0; i <= args.length;
i++) { </CODE>

Μια περιγραφή τραγουδιού σε HTML

```
<dt>Hot Cop
```

```
<dd> by Jacques Morali, Henri Belolo, and Victor  
Willis
```

```
<ul>
```

```
<li>Producer: Jacques Morali
```

```
<li>Publisher: PolyGram Records
```

```
<li>Length: 6:20
```

```
<li>Written: 1978
```

```
<li>Artist: Village People
```

```
</ul>
```



Μια περιγραφή τραγουδιού σε XML

```
<SONG>  
  <TITLE>Hot Cop</TITLE>  
  <COMPOSER>Jacques Morali</COMPOSER>  
  <COMPOSER>Henri Belolo</COMPOSER>  
  <COMPOSER>Victor Willis</COMPOSER>  
  <PRODUCER>Jacques Morali</PRODUCER>  
  <PUBLISHER>PolyGram Records</PUBLISHER>  
  <LENGTH>6:20</LENGTH>  
  <YEAR>1978</YEAR>  
  <ARTIST>Village People</ARTIST>  
</SONG>
```



Άλλο ένα παράδειγμα σε XML

```
<PERSON ID="p1100" SEX="M">  
  <NAME>  
    <GIVEN>Judson</GIVEN>  
    <SURNAME>McDaniel</SURNAME>  
  </NAME>  
  <BIRTH>  
    <DATE>21 Feb 1834</DATE>  
  </BIRTH>  
  <DEATH>  
    <DATE>9 Dec 1905</DATE>  
  </DEATH>  
</PERSON>
```