

Τεχνολογίες Εφαρμογών Διαδικτύου

XML

Java Architecture for XML Binding (JAXB)

Δρ. Ι. Χαμόδρακας

Μέλος του Εργαστηριακού Διδακτικού Προσωπικού

Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

* Μέρος του υλικού προέρχεται από τη Wikipedia και το site w3schools.org

XML: eXtensible Markup Language

- Η XML (αγγλ. αρκτ. από το Extensible Markup Language) είναι μία γλώσσα σήμανσης, που περιέχει ένα σύνολο κανόνων για την ηλεκτρονική κωδικοποίηση κειμένων.
- Οι χαρακτήρες που απαρτίζουν ένα κείμενο XML, αποτελούν είτε τη **σήμανση** είτε το **περιεχόμενό** του.
- Όλα τα αλφαριθμητικά που συνιστούν τη σήμανση, είτε ξεκινούν με το χαρακτήρα "<" και καταλήγουν στο χαρακτήρα ">", είτε ξεκινούν με το χαρακτήρα "&" και καταλήγουν στο χαρακτήρα ";".
- Ακολουθίες χαρακτήρων που δε συνιστούν τη σήμανση, αποτελούν το περιεχόμενο ενός κειμένου XML.

XML: eXtensible Markup Language

- **Ετικέτα:** ένα στοιχείο σήμανσης που ξεκινά με το χαρακτήρα "<" και καταλήγει στο χαρακτήρα ">".
- Υπάρχουν τρία είδη ετικέτας: *ετικέτες-αρχής*, για παράδειγμα <section>, *ετικέτες-τέλους*, για παράδειγμα </section>, και *ετικέτες-χωρίς-περιεχόμενο*, για παράδειγμα <line-break/>.
- **Στοιχείο:** Ένα λογικό απόσπασμα ενός κειμένου, που είτε ξεκινά με μία ετικέτα-αρχής και καταλήγει σε μία ετικέτα-τέλους, είτε αποτελείται μόνο από μία ετικέτα-χωρίς-περιεχόμενο.
- Οι χαρακτήρες που υπάρχουν μεταξύ μιας ετικέτας-αρχής και μιας ετικέτας-τέλους, συνιστούν το **περιεχόμενο** του στοιχείου
- Το περιεχόμενο μπορεί να περιέχει σήμανση, συμπεριλαμβανομένων και άλλων στοιχείων, που ονομάζονται **στοιχεία-παιδιά**.

XML: eXtensible Markup Language

- **Χαρακτηριστικό:** ένα στοιχείο σήμανσης που αποτελείται από ένα ζευγάρι όνομα/τιμή, το οποίο υπάρχει μέσα σε μία ετικέτα-αρχής ή σε μία ετικέτα-χωρίς-περιεχόμενο.
- Παράδειγμα: το στοιχείο *img* έχει δύο χαρακτηριστικά, τα *src* και *alt*: ``.
- Παράδειγμα πλήρους εγγράφου XML:

```
<?xml version="1.0" encoding='UTF-8'?>
<painting>
    
    <caption>This is Raphael's "Foligno" Madonna,
    painted in <date>1511</date> <date>1512</date>.
    </caption>
</painting>
```

DTD: Document Type Definition

- Ένα αρχείο **DTD** (ορισμός τύπου εγγράφου) ορίζει τα νόμιμα δομικά στοιχεία ενός εγγράφου XML, HTML και γενικότερα **SGML** (Standard Generalized Markup Language)
- Ορίζει τη δομή του εγγράφου και τη λίστα των νόμιμων στοιχείων και των χαρακτηριστικών τους.
- Ένα έγγραφο DTD συσχετίζεται με ένα έγγραφο XML (ή SGML) μέσω μιας δήλωσης τύπου εγγράφου (DOCTYPE)

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0  
Transitional//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-  
transitional.dtd">
```

Παράδειγμα δήλωσης DOCTYPE

```
<?xml version="1.0" encoding="utf-8"?>
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0  
Transitional//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-  
transitional.dtd">
```

```
<!-- the XHTML document body starts here-->
```

```
<html xmlns="http://www.w3.org/1999/xhtml"> ...  
</html>
```


Element Types

- Τα έγγραφα DTD περιέχουν τουλάχιστον τους τύπους των στοιχείων που περιέχονται στα έγγραφα και τη λίστα των χαρακτηριστικών τους.

```
<!ELEMENT html (head, body)>
```

```
<!ELEMENT p (#PCDATA | p | ul | ol | table |  
h1|h2|h3)*>
```

- Το στοιχείο html περιέχει τα στοιχεία head και body.
- Το στοιχείο p έχει περιεχόμενο #PCDATA: parsed character data, δηλαδή κείμενο που στο εσωτερικό του μπορεί να περιέχει άλλα στοιχεία (τα στοιχεία p, ul, ol, table, h1, h2 και h3).
- CDATA: unparsed character data, δηλαδή κείμενο του οποίου το περιεχόμενο δεν αναλύεται για σήμανση άλλων στοιχείων.
- * : μηδέν ή περισσότερα αντικείμενα, +: τουλάχιστον ένα, ?: μηδέν ή ένα. Αν δεν υπάρχει τίποτε από τα παραπάνω: ακριβώς ένα αντικείμενο στο περιεχόμενο του στοιχείου.
- <!ELEMENT square EMPTY>: στοιχείο χωρίς περιεχόμενο <square />

Attribute list

- Οι λίστες χαρακτηριστικών ορίζουν τα χαρακτηριστικά των στοιχείων. Περιέχουν: το όνομα του στοιχείου, τα ονόματα των χαρακτηριστικών, τον τύπο δεδομένων των χαρακτηριστικών (ή απαρίθμηση των επιτρεπτών τιμών), πιθανές default τιμές.

```
<!ATTLIST element-name attribute-name attribute-type  
attribute-value>
```

```
<!ATTLIST img  
    src CDATA #REQUIRED  
    id ID #IMPLIED  
    sort CDATA #FIXED "true"  
    print (yes | no) "yes">
```

ID: μοναδική τιμή

attribute-value: REQUIRED: απαραίτητο, FIXED: συγκεκριμένη τιμή που ακολουθεί, IMPLIED: προαιρετικό

XSD: XML Schema Definition

- XSD: Εναλλακτικός **πιο ισχυρός** τρόπος ορισμού της δομής εγγράφων XML μέσω της ίδιας της γλώσσας XML.

```
<?xml version="1.0"?>
```

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
targetNamespace="http://www.w3schools.com">
```

```
...
```

```
...
```

```
</xs:schema>
```

- Πάντοτε ένα σχήμα XML έχει ως στοιχείο-ρίζα το schema
- Το πρώτο χαρακτηριστικό `xmlns:xs` δηλώνει ότι τα στοιχεία και οι τύποι δεδομένων στο σχήμα προέρχονται από τον χώρο ονομάτων (namespace) `http://www.w3.org/2001/XMLSchema` και ότι ξεκινούν πάντοτε με `xs:`

XSD: XML Schema Definition

- Το επόμενο χαρακτηριστικό δηλώνει ότι τα στοιχεία που θα οριστούν στο σχήμα ανήκουν στο namespace
targetNamespace=http://www.w3schools.com
- Σε ένα άλλο έγγραφο XML δηλώνεται συμμόρφωση προς ένα σχήμα XSD ως εξής:

```
<?xml version="1.0"?>
```

```
<note xmlns="http://www.w3schools.com"  
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
xsi:schemaLocation="http://www.w3schools.com note.xsd">
```

```
<to>Tove</to>
```

```
<from>Jani</from>
```

```
<heading>Reminder</heading>
```

```
<body>Don't forget me this weekend!</body>
```

```
</note>
```

XSD: XML Schema Definition

- Το χαρακτηριστικό xmlns εντός ενός στοιχείου δηλώνει το namespace των στοιχείων που θα χρησιμοποιηθούν.
- Το χαρακτηριστικό xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" δηλώνει τη χρήση των στοιχείων από το αναγραφόμενο namespace που ξεκινούν από xsi:
- Το χαρακτηριστικό xsi:schemaLocation="http://www.w3schools.com note.xsd" δηλώνει το σχήμα που θα χρησιμοποιηθεί για το namespace (πρώτη τιμή είναι το namespace και δεύτερη η θέση του σχήματος)

Στοιχεία XSD

- Απλά στοιχεία: `<xs:element name="xxx" type="yyy"/>`
`<xs:element name="lastname" type="xs:string"/>`
`<xs:element name="age" type="xs:integer"/>`
`<xs:element name="dateborn" type="xs:date"/>`
`xs:string`, `xs:decimal`, `xs:integer`, `xs:boolean`,
`xs:date`, `xs:time` (συνήθεις τύποι)

Τα στοιχεία μπορούν να έχουν default ή fixed τιμές

```
<xs:element name="color" type="xs:string"
default="red"/>
<xs:element name="color" type="xs:string"
fixed="red"/>
```

(δεν μπορεί να οριστεί άλλη τιμή εκτός από την προκαθορισμένη)

Σύνθετα στοιχεία XSD

- **Σύνθετα στοιχεία:** στοιχεία που περιέχουν άλλα στοιχεία, στοιχεία που περιέχουν κείμενο και άλλα στοιχεία, στοιχεία που περιέχουν μόνο κείμενο, κενά στοιχεία.

```
<xs:element name="employee">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="firstname" type="xs:string"/>
      <xs:element name="lastname" type="xs:string"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

```
<employee>
  <firstname>John</firstname>
  <lastname>Smith</lastname>
</employee>
```

XML Schema Definition: περισσότερα

<http://www.w3schools.com/schema/default.asp>