

Document Type Definition – DTD μέρος 1

Εισηγητής: Παπαδάκης Γιάννης

Δευτέρα, 27.10.2008

DTDs και Εγκυρότητα (Validity)

- Ένα **Document Type Definition** περιγράφει τα συστατικά – **elements** και τα χαρακτηριστικά – **attributes** που έχουν το δικαίωμα να εμφανιστούν σε ένα XML έγγραφο
- Η διαδικασία της εγκυρότητας – **Validation** συγκρίνει ένα συγκεκριμένο XML έγγραφο με ένα DTD
- Η **Well-formedness** είναι προαπαιτούμενη για την εγκυρότητα

Τί είναι ένα DTD;

- Μια λίστα από elements, tags, attributes, and entities που (μπορούν να) περιέχονται σε ένα έγγραφο όπως επίσης και η σχέση μεταξύ τους
- εσωτερικά – εξωτερικά DTDs

Η σημασία της εγκυρότητας – validation

- Προεξοφλεί ότι τα δεδομένα είναι ορθά πριν την επεξεργασία τους από κάποιο πρόγραμμα
- Διασφαλίζει την ύπαρξη συγκεκριμένης δομής
- Καθορίζει τι πρέπει να υποστηρίζεται
- Δε χρειάζεται όλα τα έγγραφα να είναι έγκυρα: μερικές φορές αρκεί να είναι well-formed

Ενα DTD για το greeting.xml

■ greeting.xml:

```
<?xml version="1.0"?>  
<GREETING>  
Hello XML!  
</GREETING>
```

■ greeting.dtd:

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

Document Type Declarations

```
<?xml version="1.0"?>  
<!DOCTYPE GREETING SYSTEM  
  "greeting.dtd">  
<GREETING>  
Hello XML!  
</GREETING>
```

- προσδιορίζει το συστατικό ρίζα – root element
- Προσδιορίζει μια διεύθυνση – URL για το DTD

Invalid έγγραφα

■ Έγκυρα:

```
<GREETING>
```

```
various random text but no markup
```

```
</GREETING>
```

■ Ακυρα:

```
<GREETING>
```

```
  <someTag>μπλα - μπλα</someTag>
```

```
  <someEmptyTag/>
```

```
</GREETING>
```

– ή

```
<GREETING>
```

```
  <GREETING>μπλα - μπλα</GREETING>
```

```
</GREETING>
```

Εργαλεία για την εγκυρότητα

- Ειδικά προγράμματα XJParse
- Online εργαλεία
 - <http://www.stg.brown.edu/service/xmlvalid/>
 - <http://www.cogsci.ed.ac.uk/%7Eerichard/xml-check.html>
- Web Browsers (IE 7.0)

Ορισμοί συστατικών – elements

- Κάθε συστατικό πρέπει να οριστεί σε μια δήλωση `<!ELEMENT>`
- Μια δήλωση `<!ELEMENT>` ορίζει το όνομα και το περιεχόμενο του συστατικού
- Η μοντελοποίηση του περιεχομένου χρησιμοποιεί μια απλή γραμματική για να περιγράψει με ακρίβεια τι επιτρέπεται και τι δεν επιτρέπεται σε ένα συστατικό

Προδιαγραφές Περιεχομένου

- ANY
- #PCDATA
- Sequences
- Choices
- Mixed Content
- Modifiers
- Empty

ANY

<!ELEMENT SEASON ANY>

- Το συστατικό SEASON μπορεί να περιέχει οποιοδήποτε (any) συστατικό – παιδί ή/και απλό κείμενο (parsed character data)

#PCDATA

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

- **Parsed Character Data:** απλό κείμενο, χωρίς σημείωση - markup

#PCDATA

■ Εγκυρο:

```
<YEAR>1999</YEAR>
```

```
<YEAR>99</YEAR>
```

```
<YEAR>1999
```

```
  C.E.</YEAR>
```

```
<YEAR>
```

```
The year of our  
Lord one  
thousand, nine  
hundred, and  
ninety-nine
```

```
</YEAR>
```

■ Ακυρο:

```
<YEAR>
```

```
<MONTH>January</MONTH>
```

```
<MONTH>February</MONTH>
```

```
<MONTH>March</MONTH>
```

```
<MONTH>April</MONTH>
```

```
<MONTH>May</MONTH>
```

```
<MONTH>June</MONTH>
```

```
<MONTH>July</MONTH>
```

```
<MONTH>August</MONTH>
```

```
<MONTH>September</MONTH>
```

```
<MONTH>October</MONTH>
```

```
<MONTH>November</MONTH>
```

```
<MONTH>December</MONTH>
```

```
</YEAR>
```

Child Elements

- Για να ορίσουμε ότι το συστατικό LEAGUE πρέπει να έχει το LEAGUE_NAME ως συστατικό – παιδί:

```
<!ELEMENT LEAGUE (LEAGUE_NAME)>
```

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

Sequences

- Διαχωρίζουμε πολλαπλά, υποχρεωτικά συστατικά – παιδιά με κόμμα:

```
<!ELEMENT SEASON (YEAR, LEAGUE,  
LEAGUE) >
```

```
<!ELEMENT LEAGUE (LEAGUE_NAME,  
DIVISION, DIVISION, DIVISION) >
```

Ένα ή περισσότερα παιδιά +

```
<!ELEMENT DIVISION_NAME (#PCDATA)>  
<!ELEMENT DIVISION (DIVISION_NAME,  
TEAM+)>
```


Μηδέν ή περισσότερα παιδιά *

```
<!ELEMENT TEAM (TEAM_CITY,  
TEAM_NAME, PLAYER*)>
```

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

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

Μηδέν ή ένα παιδί ?

```
<!ELEMENT PLAYER (GIVEN_NAME,  
SURNAME, POSITION, GAMES,  
GAMES_STARTED, AT_BATS?, RUNS?,  
HITS?, DOUBLES?, TRIPLES?,  
HOME_RUNS?, RBI?, STEALS?,  
CAUGHT_STEALING?, SACRIFICE_HITS?,  
SACRIFICE_FLIES?, ERRORS?, WALKS?,  
STRUCK_OUT?, HIT_BY_PITCH?, WINS?,  
LOSSES?, SAVES?, COMPLETE_GAMES?,  
SHUT_OUTS?, ERA?, INNINGS?,  
EARNED_RUNS?, HIT_BATTER?,  
WILD_PITCHES?, BALK?, WALKED_BATTER?,  
STRUCK_OUT_BATTER?)
```

Choices

```
<!ELEMENT PAYMENT (CASH |  
CREDIT_CARD) >
```

```
<!ELEMENT PAYMENT (CASH |  
CREDIT_CARD | CHECK) >
```

Ομαδοποίηση με παρενθέσεις

- Οι παρενθέσεις συνδυάζουν πολλά συστατικά σε ένα συστατικό
- Συστατικά που αναπαρίστανται με παρενθέσεις μπορούν να ενθυλακωθούν μέσα σε άλλες παρενθέσεις
- Το συστατικό που αναπαριστάται με μια παρένθεση μπορεί να συνοδεύεται από τα σύμβολα: + , ? , *

```
<!ELEMENT dl (dt, dd)*>  
<!ELEMENT ARTICLE (TITLE, (P | PHOTO | GRAPH  
| SIDEBAR | PULLQUOTE | SUBHEAD)*,  
BYLINE?)>
```

Μεικτό περιεχόμενο

- Οχι μόνο #PCDATA αλλά και συστατικά – παιδιά σε μια choice

```
<!ELEMENT TEAM (#PCDATA | TEAM_CITY  
| TEAM_NAME | PLAYER)*>
```

- Το #PCDATA πρέπει να προηγείται
- Το #PCDATA δεν μπορεί να χρησιμοποιηθεί σε μια sequence

Empty συστατικά

```
<!ELEMENT BR EMPTY>
```

```
<!ELEMENT IMG EMPTY>
```

```
<!ELEMENT HR EMPTY>
```