

Validarea Documentelor XML

Mihai Gabroveanu

Cuprins

- 1 Introducere
 - Specificarea modelului de date
 - Documente XML Valide
- 2 Definirea Tipului de Document
 - Referirea DTD-urilor
 - Definirea unui DTD intern
 - Definirea unui DTD extern
 - Sintaxa DTD-urilor
- 3 Bibliografie
- 4 Intrebari si Raspunsuri

De ce sa specificam un model de date?

Sa consideram urmatorul document XML:

```
<?xml version="1.0"?>
<BIBLIOTECA>
  <CARTE>
    <TITLU>XML Bible</TITLU>
    <AUTOR>Elliotte Rusty Harold</AUTOR>
    <EDITURA>IDG Books Worldwide</EDITURA>
    <AN_APARITIE>2002</AN_APARITIE>
  </CARTE>
</BIBLIOTECA>
```

De ce sa specificam un model de date?

In cadrul acestui document avem:

Elementul radacina **BIBLIOTECA**, care consta din unul sau mai multe elemente de tip **CARTE**. Fiecare element **CARTE** are patru sub-elemente **TITLU**, **AUTOR**, **EDITURA**, **AN_APARITIE**.

Sa consideram, in continuare urmatorul element:

```
<CARTE>
  <TITLU></TITLU>
  <AUTOR>Elliotte Rusty Harold</AUTOR>
  <EDITURA>IDG Books Worldwide</EDITURA>
  <AN_APARITIE>2002</AN_APARITIE>
</CARTE>
```

De ce sa specificam un model de date?

In acest caz

- Documentul este bine format
- Dar, elementul obligatoriu **TITLU** este vid
- Singurul mod prin care o aplicatie sa poata sa elimine o astfel de situatie este sa se defineasca un model de date.

Specificarea modelului de date

Pentru specificarea modelului putem utiliza:

- Document Type Definition (DTD)
- XML Schema

Documente XML Valide

Un document XML este **valid** daca:

- este un document "bine format" (well-formed)
- refera o gramatica (DTD, XML Schema, etc.), si
- respecta acea gramatica

Ce este un DTD?

DTD - Document Type Definition

DTD-ul (Definitia Tipului de Document) este un set de reguli care definesc modul de structurare a documentelor XML.

Un DTD descrie:

- Elementele
- Atributele
- Entitatile

Standardele DTD-ului sunt definite de [World Wide Web Consortium \(W3C\)](#).

Declararea tipului de document

Declararea tipului de document:

- Specifica de ce tip este documentul XML
- Aceasta declaratie "spune" analizorului unde gaseste DTD-ul

Un DTD poate fi definit:

- **intern** - definit in interiorul fisierului XML
- **extern** - definit intr-un fisier separat

Definirea unui DTD intern

Sintaxa declararii unui DTD intern este urmatoarea:

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

```
<!DOCTYPE element_radacina [  
    declaratii de elemente, attribute, entitati,  
    instructiuni de procesare, notatii  
>]
```

```
<element_radacina>
```

```
. . .
```

```
</element_radacina>
```

Definirea unui DTD intern

Exemplu de DTD intern:

```
<?xml version="1.0"?>
<!DOCTYPE CARTE [
  <!ELEMENT CARTE (TITLU, AUTOR+, EDITURA, AN_APARITIE)>
  <!ELEMENT TITLU (#PCDATA)>
  <!ELEMENT AUTOR (#PCDATA)>
  <!ELEMENT EDITURA (#PCDATA)>
  <!ELEMENT AN_APARITIE (#PCDATA)>
]>
<CARTE>
  <TITLU>XML Bible</TITLU>
  <AUTOR>Elliotte Rusty Harold</AUTOR>
  <EDITURA>IDG Books Worldwide</EDITURA>
  <AN_APARITIE>2002</AN_APARITIE>
</CARTE>
```

Definirea unui DTD extern

Un DTD extern:

- este definit intr-un alt fisier text care trebuie sa se afle la o adresa specificata
- va fi referit printr-un identificator public si/sau unul sistem
- are avantajul ca poate fi partajat de mai multe aplicatii

Sintaxa declararii unui DTD extern este urmatoarea:

```
<!DOCTYPE element_radacina SYSTEM "SYSTEM-URI">
```

sau

```
<!DOCTYPE element_radacina PUBLIC "PUBLIC-URI" "SYSTEM-URI">
```

unde, element_radacina este numele elementului radacina

Declararea unui DTD extern

Exemplu de declarare a unui DTD extern:

Presupunand ca avem un DTD pentru un document de tip CARTE care este disponibil la adresa URL <http://inf.ucv.ro/CARTE.dtd> atunci un declararea tipului unui document XML de acest tip este de forma:

```
<?xml version="1.0">
<!DOCTYPE CARTE PUBLIC "http://inf.ucv.ro/CARTE.dtd" >
<CARTE>
  <TITLU>XML Bible</TITLU>
  <AUTOR>Elliotte Rusty Harold</AUTOR>
  <EDITURA>IDG Books Worldwide</EDITURA>
  <AN_APARITIE>2002</AN_APARITIE>
</CARTE>
```

Declararea unui DTD extern

Fisierul CARTE.dtd ce contine definirea tipului de fisier:

```
<?xml version="1.0"?>  
<!ELEMENT CARTE (TITLU, AUTOR, EDITURA, AN_APARITIE)>  
<!ELEMENT TITLU (#PCDATA)>  
<!ELEMENT AUTOR (#PCDATA)>  
<!ELEMENT EDITURA (#PCDATA)>  
<!ELEMENT AN_APARITIE (#PCDATA)>
```

Sintaxa DTD-urilor

Un fisier DTD contine urmatoarele tipuri de definitii:

- ELEMENT
- ATTLIST
- ENTITY
- NOTATION

Standardele DTD-ului sunt definite de [World Wide Web Consortium \(W3C\)](#).

Declararea elementelor - ELEMENT

Sintaxa declararii unui tip de element nevid:

```
<!ELEMENT nume-element tip-informatie>
```

sau

```
<!ELEMENT nume-element (continut-element)>
```

Sintaxa declararii unui tip de element vid:

```
<!ELEMENT nume-element EMPTY>
```

Declararea elementelor - ELEMENT

Sintaxa declararii elementelor numai cu date caracter:

```
<!ELEMENT nume-element (#PCDATA)>
```

sau

```
<!ELEMENT nume-element (#CDATA)>
```

Elementul AUTOR din exemplul anterior se declara astfel:

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

Declararea elementelor - ELEMENT

Sintaxa declararii elementelor cu orice tip de continut:

```
<!ELEMENT nume-element ANY>
```

Sintaxa declararii elementelor ce au copii:

```
<!ELEMENT nume-element (nume-copil-1,nume-copil-2, ...)>
```

Elementele copil trebuie declarate si ele la randul lor.

Elementul CARTE din exemplul anterior se declara astfel:

```
<!ELEMENT CARTE (TITLU, AUTOR, EDITURA, AN_APARITIE)>
```

Declararea elementelor - ELEMENT

Simbol	Numar de aparitii
nimic	elementul apare o singura data
*	0 sau mai multe ori
+	1 sau mai multe ori
?	0 sau 1 ori

Table: Specificarea cardinalitatii

Declararea elementelor - ELEMENT

Elementul CARTE din exemplul anterior il putem declara astfel:

```
<!ELEMENT CARTE (TITLU, (AUTOR+ | EDITOR),  
                  EDITURA, AN_APARITIE, REZUMAT?)>
```

O carte are un titlu, cel putin un autor sau un editor, o editura, un an al aparitiiei, si eventual un rezumat.

Declararea atributelor - ATTLIST

O declarare a unui atribut specifica elementul caruia ii este asociat, numele lui, tipul, si posibilele valori implicite.

Sintaxa de declarare este urmatoarea:

```
<!ATTLIST elm-name attr-name attr-type #DEFAULT def-value>
```

sau

```
<!ATTLIST elm-name attr-name attr-type #FIXED fixed_value>
```

sau

```
<!ATTLIST elm-name attr-name  
                attr-type (Val1|Val2|..) def-value>
```

sau

```
<!ATTLIST elm-name attr-name attr-type #IMPLIED>
```

sau

```
<!ATTLIST elm-name attr-name attr-type #REQUIRED>
```

Declararea atributelor - ATTLIST

Exemple:

```
<!ATTLIST persoana sex CDATA #DEFAULT "masculin">  
<!ATTLIST persoana sex CDATA #FIXED "masculin">  
<!ATTLIST persoana sex CDATA #REQUIRED>  
<!ATTLIST persoana sex CDATA #IMPLIED>  
<!ATTLIST persoana sex (masculin|feminin) "masculin">
```

Declararea entitatilor - ENTITY

O entitate poate fi declarata *local* sau *extern*.

Sintaxa declararii entitatilor locale (in interiorul DTD-ului) este:

```
<!ENTITY entity-name entity-value>
```

Exemplu:

```
<!ENTITY website "http://inf.ucv.ro">  
<!ENTITY copyright "Copyright (c) 2011 M. Gabroveanu">
```

Utilizare intr-un fisier XML:

```
<author>&copyright;&website;</author>
```

Declararea entitatilor - ENTITY

Sintaxa declararii entitatilor externe DTD-ului este:

```
<!ENTITY entity-name SYSTEM "entity-URL">
```

Exemplu:

```
<!ENTITY website SYSTEM "http://inf.ucv.ro/entity.dtd">
```

Bibliografie

XML

- Extensible Markup Language (XML) 1.0 (Third Edition) W3C Recommendation 4th February 2004, Francois Yergeau, Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler
- Elliotte Rusty Harold, XML Bible. IDG Books Worldwide, Inc, 919 E. Hillsdale Blvd., Suite 400, Foster City, CA 94404
- <http://www.w3schools.com/dtd/>

Q & A

- Intrebari?
- Comentarii?