



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI



Instrumente Structurale
2007-2013



Platformă de e-learning și curriculum e-content pentru învățământul superior tehnic

Proiect nr. 154/323 cod SMIS – 4428 cofinanțat de prin Fondul European de Dezvoltare Regională “Investiții pentru viitorul dumneavoastră”.

Programul Operațional Sectorial Creșterea Competitivității Economice - POS CCE



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI



Instrumente Structurale
2007-2013



Platformă de e-learning și curriculum e-content
pentru învățământul superior tehnic

Interacțiunea om-calculator

6. XML

Perspective asupra XML

- este un limbaj universal de adnotare a documentelor (derivat din SGML);
- permite declararea unei gramatici pentru un limbaj de adnotări:
 - explicit, printr-un DTD sau o schemă,
 - implicit, pe baza structurii de adnotări, chiar dacă nu există un DTD
- este o modalitate foarte comodă de transfer al informațiilor pe Internet între aplicații de categorii foarte diferite;

Perspective asupra XML (cont.)

- este un limbaj adecvat aplicațiilor de baze de date federate;
- este o modalitate universală de a reprezenta liniar orice structură, oricât de complexă;
- este un limbaj cu tipuri manifeste, spre deosebire de formatul fix al fișierelor sau bazelor de date;
- este un limbaj de declarare formală a unui vocabular de elemente și attribute;

Perspective asupra XML (cont.)

- permite restricționarea și validarea documentelor care pot fi create;
- este o “ontologie de nivel 0”;
- este un prim pas către facilitarea achiziției cunoștințelor din document;

Structura documentelor XML

Fiecare document XML are o structură:

- fizică
- logică

Structura fizică a documentelor XML

Fizic, documentele XML sunt compuse din unități numite **entități**.

O entitate are un **nume** și un **conținut**.

O entitate poate face referire la alte entități pentru a le include în document.

Un document începe printr-o **rădăcină** (“root”) sau **entitate document** (“document entity”).

Entități

Din punct de vedere al plasării în fișierul documentului:

- **interne** (în același fișier)
- **externe** (aflate în alt fișier).

Entități (cont.)

- **prelucrate** (analizate, “parsed”) de către procesoarele XML; acestea pot fi atât interne cât și externe; - conțin **text**, adică o secvență de caractere, ce pot reprezenta **adnotări** sau **date de tip caracter** (“character data”). Caracterele constituie atomii unui text, conform specificațiilor standardului ISO/IEC 10646. Caracterele legale sunt tab, “enter” (CR), sfârșit de linie LF și caracterele legale ale Unicode și ISO/IEC 10646.
- **neprelucrate** (neanalizate, “unparsed”), aceste entități sunt întotdeauna externe.

Entități (cont.)

- **generale**, utilizabile fără restricții;
- **parametru**, utilizabile numai în cadrul unui DTD

Structura logică a documentelor XML

elemente, delimitate în text de adnotări (“markup”) de început și de sfârșit.

Adnotările sunt incluse între paranteze unghiulare (delimitate de caracterele “<” și “>”, care nu pot fi folosite în alt scop în documentele XML.

Dacă este necesară utilizarea “<” și “>”, se folosesc construcțiile: < și >;, adică referințe la entități special constituite. Adnotările apar în perechi, pentru a indica începutul respectiv sfârșitul elementului. Adnotarea de sfârșit se deosebește de cea de început printr-o bară oblică.

Adnotari XML

```
<pagtitlu>
```

```
<traducator>Andrei Cornea</traducator> a tradus  
<titlu>Republica</titlu> de <autor>Platon</autor>  
<ingrijit>pentru volumul ingrijit de Constantin  
Noica</ingrijit>.
```

```
</pagtitlu>
```

Există o infinitate de posibilități de adnotare a unui text:

```
<pagtitlu>
```

```
<traducator><substantivPropriu>Andrei</substantivPropriu>  
<substantivPropriu>Cornea</substantivPropriu></traducator>  
<verb>a tradus</verb>
```

```
<titlu>Republica</titlu> de  
<autor><substantivPropriu>Platon</substantivPropriu></autor>  
<ingrijit>pentru <substantiv>volumul</substantiv>  
<verb>ingrijit</verb> de  
<substantivPropriu>Constantin</substantivPropriu>  
<substantivPropriu>Noica</substantivPropriu>.</ingrijit>
```

```
</pagtitlu>
```

Elemente vide, fără conținut

<adnotare/>

Adnotările pot fi atributate

```
<pagtitlu>
```

```
<traducator>Andrei Cornea</traducator> a tradus  
  <titlu>Republica<titlu> de <autor>Platon</autor>
```

```
<ingrijit secret="da">
```

```
pentru volumul ingrijit de Constantin Noica.
```

```
</ingrijit>
```

```
</pagtitlu>
```

Structura logică a documentelor XML

- Fiecare document XML are o **structură de arbore**, cu un element rădăcină.
- Pentru fiecare element C care nu este radacina documentului, există un alt element P, **părintele lui C** (“parent”) în document, astfel încât C să fie conținut direct de acesta. Se spune că C este **copilul** lui P (“child”).

Structura logică a documentelor XML

```
<a><b> text1 <c> text2 </c> <d> text3 </d> </b> text4  
<b>text5 </b></a>
```

```
<a>  
  <b>  
    text1  
    <c> text2 </c>  
    <d> text3 </d>  
  </b>  
  text4  
  <b> text5 </b>  
</a>
```

Componente ale structurii logice a documentelor XML

- elemente, conform precizărilor anterioare;
- comentarii;
- secțiuni CDATA;
- indicații (instrucțiuni) de procesare;
- referințe la caractere sau la entități;
- declarații de prolog, care cuprind declarația XML și, eventual definirea unui DTD.

Comentarii

- Pot apare practic oriunde în documentele XML, în afara adnotărilor
- Sunt delimitate de “<!--“ și “-->”

```
<!-- Comentariu -->
```

Secțiuni CDATA

- Conțin date caracter (CDATA), care se iau ca atare (nu sunt analizate de procesoarele XML).
- Pot apare oriunde pot apare datele caracter. Sunt utilizate de obicei pentru a include blocuri de text ce conțin caractere ce ar fi altfel recunoscute drept adnotări.
- Încep cu șirul "<![CDATA[" si se termina cu șirul "]]>":

Secțiuni CDATA (exemplu)

```
<![CDATA[
```

```
Text
```

```
]]>
```

Indicații de procesare (PI)

- Modalitate de a include în documente indicații (instrucțiuni) pentru aplicațiile care prelucrează documentele XML. Ele nu sunt părți ale datelor caracter, dar sunt trimise aplicației.
- Indicațiile de procesare sunt delimitate de “<?” și “?>”, încep cu o țintă (“PITarget”) utilizată pentru a identifica aplicația căreia îi este destinată instrucțiunea, urmată de informații suplimentare de trimis aplicației.

Indicații de procesare (exemplu)

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

```
<?xml-stylesheet type="text/xsl"  
  href="b1.xsl" ?>
```

```
<adnotare> Text afisat </adnotare>
```

Referințe la caractere sau la entități interne

- Secvențe de caractere delimitate de “&#” sau “&#x” și “;”, pentru caractere,
- Secvențe de caractere delimitate de “&” și “;”, pentru entități.
- Tin locul, unui caracter, respectiv unui șir de caractere, cu care sunt înlocuite de procesoarele XML

Documente XML:

- Bine formate (pot sa nu aiba un DTD asociat)
- Valide (au un DTD asociat si il respecta)

Reguli pentru documentele XML bine formate

Un **document XML bine format** trebuie să respecte următoarele reguli:

- Documentul trebuie să aibă **adnotările în perechi** (nu pot lipsi adnotări de început sau sfârșit, ca în cazul SGML sau HTML), și contează dacă o literă este mare sau mică (<adn> este diferit de <ADN>, XML este “case sensitive”).
- Orice document XML trebuie să aibă un singur element superficial, numit **rădăcină** (“root”), sau element document, care nu apare în conținutul altui element.

Reguli pentru documentele XML bine formate (cont.)

- Dacă sunt mai multe perechi de adnotări, ele trebuie să fie imbricate (să respecte o structură de “paranteze”)
- Numele elementelor trebuie să satisfacă anumite reguli, de exemplu, să înceapă cu o literă sau cu “_”
- Valorile atributelor adnotărilor trebuie să fie puse între ghilimele (de exemplu, `<adn val="2">` și nu `<adn val=2>`).

Reguli pentru documentele XML bine formate (cont.)

- Nu trebuie ca un atribut să apară de mai multe ori în aceeași adnotare.
- Pot exista adnotări care nu includ o zonă de text. Aceste adnotări, denumite **adnotări vide**, au caracterul “/” la sfârșit, de exemplu: <adnotare/>.
- Entitățile amp, lt, gt, apos, quot (pentru caracterelor “&”, “<”, “>”, apostrof și ghilimea) pot fi utilizate fără a fi declarate.

```

<?xml version="1.0"?>

<bib>
<webdoc src="http://concept.cs.uah.edu/CG/cg-standard.html" abr="CG">
Conceptual Graphs </webdoc>
<webdoc abr="CYC" src="http://www.cyc.org"> CYC</webdoc>
<webdoc abr="DOM" src="http://www.w3.org/DOM"> DOM </webdoc>
<webdoc abr="KIF" src="http://logic.stanford.edu/kif/kif.html">
Knowledge Interchange Format </webdoc>
<webdoc abr="KQML" src="http://www.cs.umbc.edu/kqml"> KQML </webdoc>
<webdoc abr="RDF" src="http://www.w3.org/RDF"> RDF Specification
<an>1999</an></webdoc>
<webdoc abr="RDFS" src="http://www.w3.org/TR/WD-rdf-schema"> RDF Schema
Specification</webdoc>
<articol abr="Sho93">
  <autor> Y. Shoham, Agent-oriented programming </autor>
  <revista nr="60">
    <titlu>Artificial Intelligence</titlu>
    <an>1993</an>
    <pp> 51-92</pp>
  </revista>
</articol>
<webdoc abr="XML" src="http://www.w3.org/XML"> XML </webdoc>
<carte abr="XMLc" isbn="1-861003-4-12">
  <autor> Cagle, K. </autor><autor>Gibbons, D.</autor><autor> Hunter,
D.</autor><autor> Ozu, N.</autor><autor>Pinnock, J.
  </autor><autor>Spencer, P.</autor>
  <titlu> Beginning XML </titlu>
  <editura>Wrox Press,</editura>
  <an> 2000</an>
</carte>
</bib>

```

```

- <bib>
  <webdoc src="http://concept.cs.uah.edu/CG/cg-standard.html "
abr="CG">Conceptual Graphs</webdoc>
  <webdoc abr="CYC" src="http://www.cyc.org">CYC</webdoc>
  <webdoc abr="DOM" src="http://www.w3.org/DOM">DOM</webdoc>
  <webdoc abr="KIF"
src="http://logic.stanford.edu/kif/kif.html">Knowledge Interchange
Format</webdoc>
  <webdoc abr="KQML" src="http://www.cs.umbc.edu/kqml">KQML</webdoc>
  - <webdoc abr="RDF" src="http://www.w3.org/RDF">
    RDF Specification
    <an>1999</an>
  </webdoc>
  <webdoc abr="RDFS" src="http://www.w3.org/TR/WD-rdf-schema">RDF
Schema Specification</webdoc>
  - <articol abr="Sho93">
    <autor>Y. Shoham, Agent-oriented programming</autor>
  - <revista nr="60">
    <titlu>Artificial Intelligence</titlu>
    <an>1993</an>
    <pp>51-92</pp>
  </revista>
</articol>
  <webdoc abr="XML" src="http://www.w3.org/XML">XML</webdoc>
  - <carte abr="XMLc" isbn="1-861003-4-12">
    <autor>Cagle, K.</autor>
    <autor>Gibbons, D.</autor>
    <autor>Hunter, D.</autor>
    <autor>Ozu, N.</autor>
    <autor>Pinnock, J.</autor>
    <autor>Spencer, P.</autor>
    <titlu>Beginning XML</titlu>
    <editura>Wrox Press,</editura>
    <an>2000</an>
  </carte>
</bib>

```

XHTML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
    "DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
  <head>
    <title>Virtual Library</title>
  </head>
  <body>
    <p>Moved to <a href="http://vlib.org/">vlib.org</a>.
    </p>
  </body>
</html>
```

Documente XML valide

- Un document XML bine format devine un **document XML valid**, dacă, în plus față de restricțiile de bună formare, are asociată și o definiție de tip de document (**Document Type Definition** - DTD), pe care o respectă.