

XML, cuceritorul



Sabin-Corneliu Buraga

Facultatea de Informatica

Universitatea "Al.I.Cuza" Iasi

busaco@infoiasi.ro

<http://www.infoiasi.ro/~busaco>



XML, cuceritorul

"Un punct care ieri era nevazut, astazi a ajuns ținta și mîine va fi punctul de plecare."

(T.Macaulay)

Cuprins

- Ce este XML
- Familia XML
- Trăsături XML
- Scheme XML
- Aplicații ale XML
- De la HTML la XML
- Viitorul XML





Ce este XML

- XML (Extensible Markup Language)
- Meta-limbaj de adnotare (marcare)
 - Set de conventii de marcare utilizate pentru codificarea informatiilor
 - Specifica multimea de marcaje (tag-uri) obligatorii, identificarea si semantica marcajelor
- SGML simplificat, pentru Web



Ce este XML

- Marcaje descriptive
- Tipuri de documente:
 - Specificarea formala a partilor & structurii
 - **DTD (Document Type Definition)**
 - **XML Schema**
- Independenta datelor
- Independent de platforma hard/soft
- Suport pentru uz international



Ce este XML

- Recomandare a Consorțiului Web:
XML 1.0 (updated) oct.2000
- Exemplu:

```
<?xml version="1.0" ?>  
<web data="05 mai 2001" loc="Iasi">  
  <tutorial>XML, cuceritorul</tutorial>  
  <autor email="busaco@infoiasi.ro">  
    Sabin-Corneliu Buraga  
  </autor>  
</web>
```

Instr.procesare

Atribut

Tag



Familia XML

- **XML (Extensible Markup Language)**
meta-limbajul propriu-zis
- **XLL (Extensible Linking Language)**
 - XLink - hiper-legaturi intre documente
 - XPointer - localizarea relativa a resurselor
- **XSL**
(Extensible Stylesheet Language)
transformare/formatare date XML



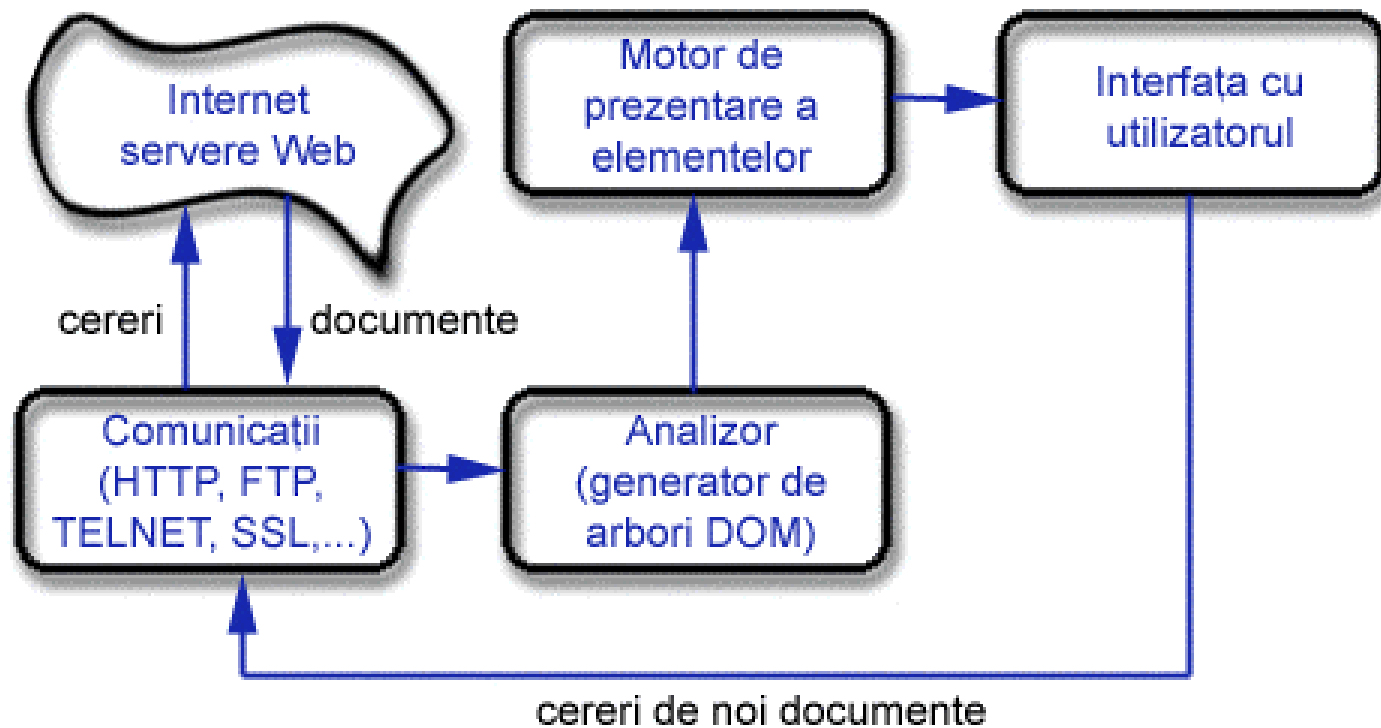
XML vs. SGML

- Declaratia unui document: `<?xml ...?>`
- DTD optional, XML Schema optionala
- Tag-uri de sfirsit obligatorii:
`` sau ``
- Case-sensitive: `<tag>` \neq `<Tag>` \neq `<TAG>`
- Instructiuni de procesare: `<?instr...?>`
- Numele incepind cu XML, xml sunt rezervate



XML

■ Arhitectura unui browser XML





Spatii de nume XML

- **Spatiu de nume** - vocabular utilizat pentru calificarea elementelor in mod unic

```
<tutorial id="03">  
  <title>XML, cuceritorul</title>  
  <year>2001</year>  
</tutorial>
```

```
<student>  
  <name>Stefan Tanasa</name>  
  <year>4</year>  
</student>
```

Conflict!



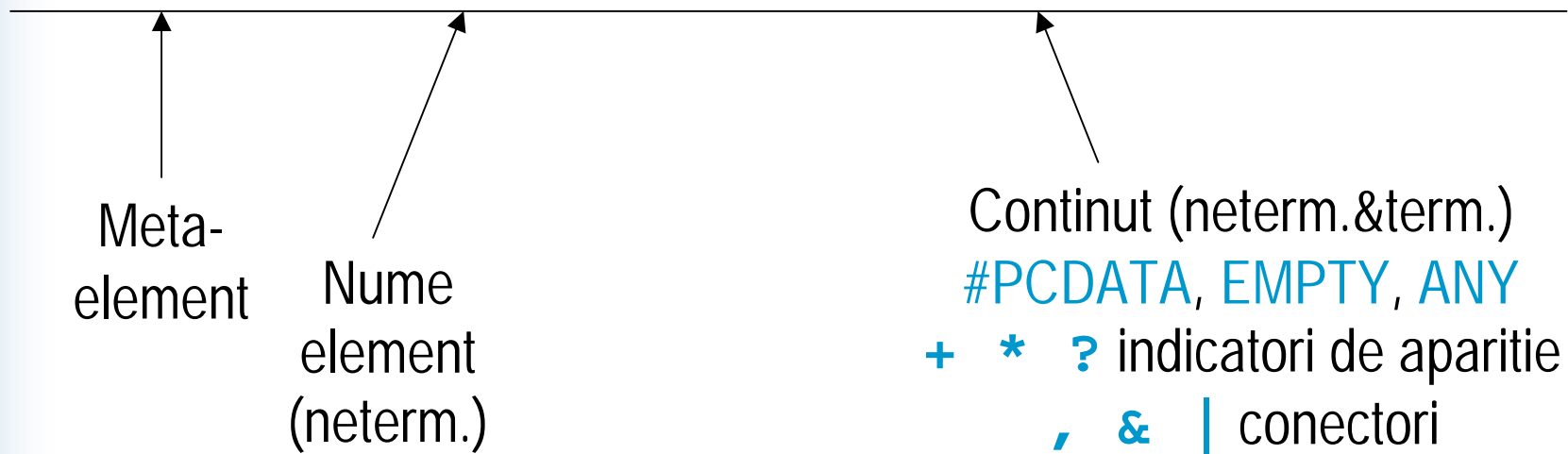
Spatii de nume XML

```
<?xml version="1.0" ?>
<web xmlns:b="urn:infoiasi.ro:busaco-ns">
  <b:tutorial b:ident="03">
    <title>XML, cuceritorul</title>
    <b:year>2001</b:year>
    <b:desc>
      <h2
xmlns="http://www.w3.org/TR/REC-html40">
        Un <i>tutorial</i> despre XML</h2>
    </b:desc>
  </b:tutorial>
</web>
```



Validarea: DTD

```
<!ELEMENT web      (tutorial+)>
<!ELEMENT tutorial (title,author*)>
<!ELEMENT title    (#PCDATA)>
<!ELEMENT author   (#PCDATA)>
<!ATTLIST web      data NUMBER #REQUIRED>
```





Validarea: XML Schema

- **Schema** = specificatie formală a gramaticii asociate unui document XML
- Validarea documentelor XML, folosind sintaxa XML

- Recomandare W3C - 02 mai 2001

```
<ElementType name="title" content="textOnly" />
<ElementType name="author" content="textOnly" />
<ElementType name="tutorial" order="many">
  <element type="title" minOccurs="1" maxOccurs="1" />
  <element type="author" maxOccurs="*" />
  <AttributeType name="ident" required="yes" type="int" />
  <attribute type="ident" />
</ElementType>
```

05 mai 2001

Sabin-Corneliu Buraga
<busaco@infoiasi.ro>

13



XML Schema vs. DTD

- Moduri diferite de specificare formală a conținutului documentelor XML
- Schemele pot fi deschise, mai flexibile decât DTD-urile
- Tipuri de date XSchema variate și complexe
- Declarațiile spațiilor de nume nu pot fi exprimate în DTD
- Schemele sunt mai ușor de validat și procesat

Aplicatii

■ MathML

■ Amaya

The screenshot shows the 'Editing Mathematics' application window. The main window displays a list of available constructs for insertion into a formula. The constructs listed are:

- Root with index, mroot in MathML: $\sqrt[3]{x + 1}$
- Square root, msqrt: $\sqrt{x + 1}$
- Enclose, menclose: $\boxed{1234}$
- Fraction, mfrac: $\frac{1}{x + 1}$

A smaller, overlapping window titled 'Editing Mathematics' provides a detailed view of the 'mfrac' construct. It shows a tree structure of the MathML code for a fraction:

```
li
├── Fraction,
│   ├── code
│   └── mfrac
│       ├── math
│       └── mfrac
│           ├── mn 1
│           ├── mrow
│           └── mi x
│               ├── mo +
│               └── mn 1
```

The main window also contains text explaining the use of these constructs, such as 'Constructs within a formula. If the caret is not in a formula, a match element is created first to accept the new construct. The available constructs are:' and 'The last you can... Enter... You are express...'. On the right side, there are partial views of mathematical expressions like $x_i^n + 1$ or \int_0^∞ and a summation symbol $\sum_{i=1}^n$.



Aplicatii XML

■ RDF

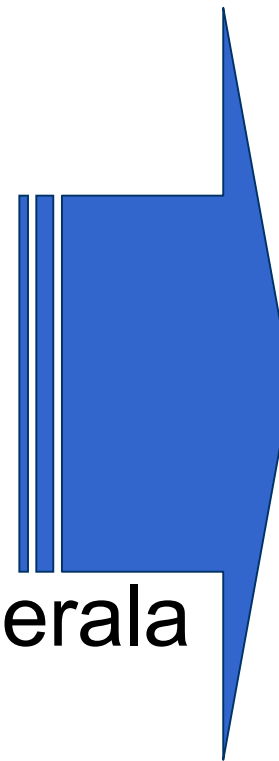
(Resource Description Framework)

- Reprezentarea metadatelor:
resurse, proprietati, declaratii
{subiect, predicat, obiect}
- **Scheme RDF** - declara proprietatile unei
resurse si semantica asociata



RDF - utilizari

- Inspectarea resurselor
- Catalogarea datelor
- Agenti inteligenti
- Descrierea drepturilor de proprietate intelectuala
- Securitate personala/generala (semnaturi digitale)



Structurare
inteligenta

Reprezen-
tare a cunos-
tintelor

Standar-
dizare Web

Semantic
Web



RDF - exemplu

"Stefan, Cristian si Alexandrina sunt participanti la atelierul de tehnologii Web."

```
<rdf:RDF>
  <rdf:Description about="http://www.infoiasi.ro/~web/">
    <s:Participanti>
      <rdf:Bag>
        <rdf:li resource="http://www3.infoiasi.ro/~stanasa" />
        <rdf:li resource="http://www3.infoiasi.ro/~colaru" />
        <rdf:li resource="http://www3.infoiasi.ro/~alexo" />
      </rdf:Bag>
    </s:Participanti>
  </rdf:Description>
</rdf:RDF>
```

A diagram consisting of a black arrow pointing from the 'Bag' tag in the RDF code to a rectangular box with a black border. Inside the box, the text 'Bag | Seq | Alt' is displayed, representing the possible cardinality constraints for the 'Bag' property.



Aplicatii XML

■ XUL

(Extensible User Interface Language)

- Descrie interfete grafice in Mozilla/Netscape 6 indiferent de platforma hard/soft
- Bazat pe XML 1.0, HTML 4.01, CSS 1&2
- Controale (widgets):
 - » Ferestre: `<window>`
 - » Cutii de dialog: `<box>`, `<text>`, `<textfield>`
 - » Meniuri: `<menu>`, `<toolbar>`, `<tab>`
 - » Butoane: `<radio>`, `<checkbox>`, `<button>`
 - » Complexe: `<grid>`, `<tree>`, `<stack>`



De la HTML la XML

■ Motive

- Extensibilitatea
- Modularizarea
- Profile ale documentelor
- Profile ale dispozitivelor
- Transformarea marcajelor



De la HTML la XML

■ Solutia: **XHTML 1.0** (2000)

– Aplicatie a XML: `text/html` sau `text/xml`

– Tag-uri & attribute scrise cu minuscule (case-sensitive)

– Tag-uri de sfirsit obligatorii

– Tag-urile vide scrise: `<tag />` sau `<tag></tag>`



De la HTML la XML

■ Module XHTML:

- » Modul de baza (`html`, `head`, `body`,...)
- » Modul tranzitional (`font`, `center`,...)
- » Modul de stiluri (`style`, `link`)
- » Modul script (`script`, `noscript`)
- » Modul de fonturi (`tt`, `b`, `i`, `big`, `small`)
- » Modul de frazare (`abbr`, `cite`, `code`,...)
- » Modul de inflexiune (`pre`, `sub`, `sup`, `hr`,...)



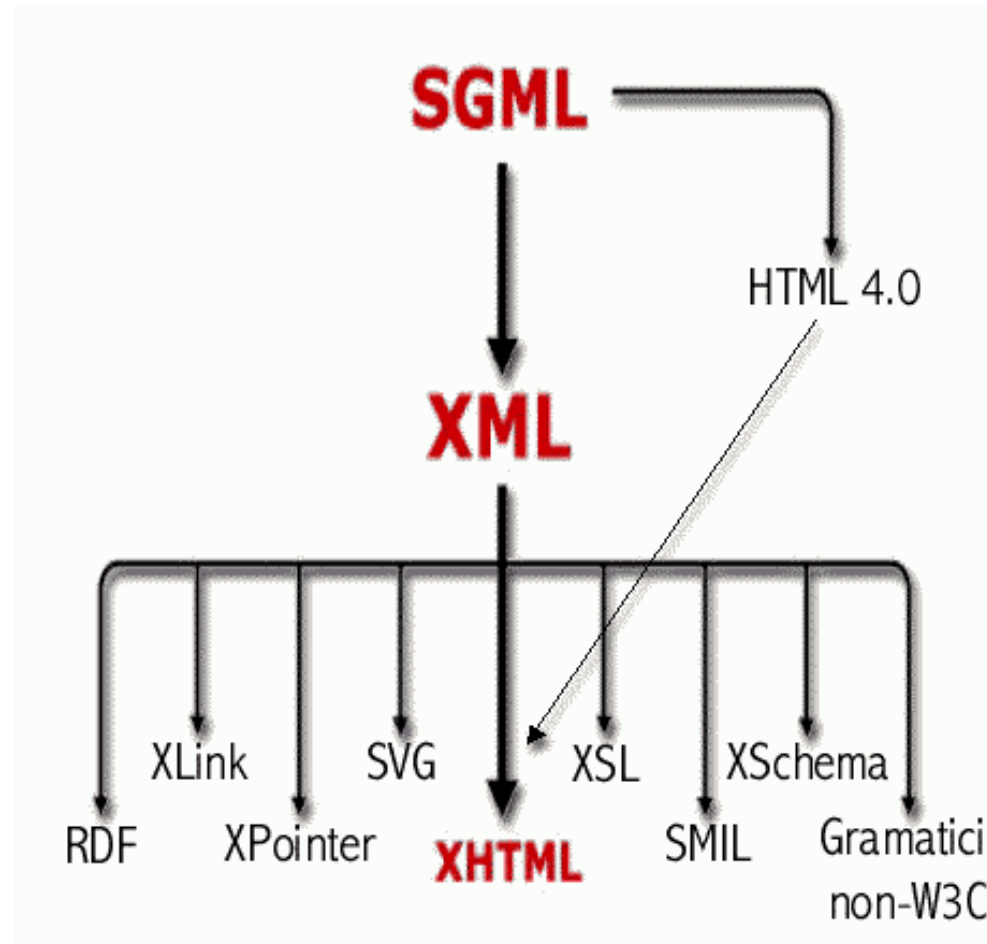
De la HTML la XML

■ Module XHTML - continuare:

- » Modul de editare (`del`, `ins`,...)
- » Modul liste (`ul`, `ol`, `li`,...)
- » Modul formulare (`form`, `button`, `input`,...)
- » Modul de tabele (`table`, `tr`, `td`,...)
- » Modul harti de imagine (`area`, `map`)
- » Modul applet-uri (`applet`)
- » Modul obiecte (`object`)
- » Modul cadre (`frameset`, `frame`, `iframe`,...)

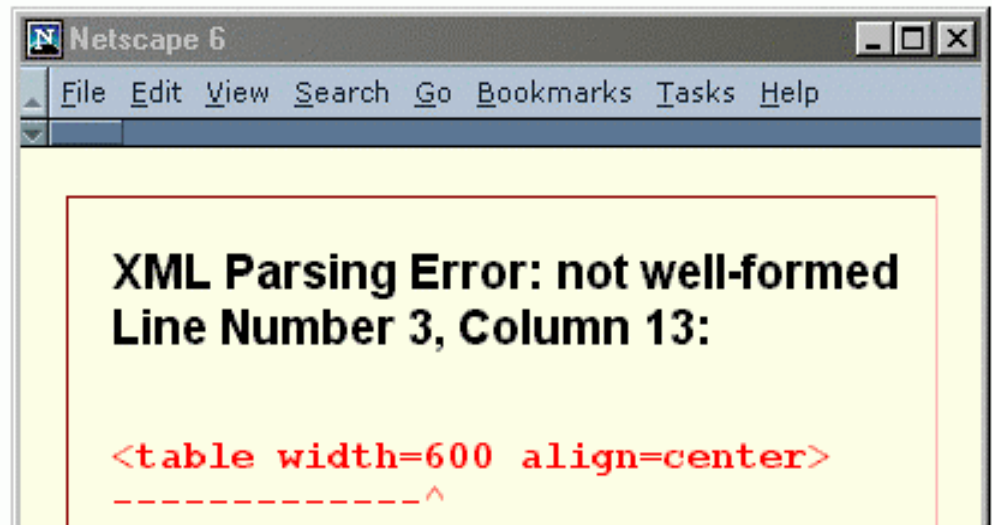


De la HTML la XML



Arborele
genealogic
al XHTML

XHTML validare



05 mai 2001



Beneficii ale XML

- Format flexibil de transmitere a informatiilor pe Web
- Standard deschis, independent de platforma
- Oferă suport pentru cautare inteligenta si eficienta pe Web
- Oferă scalabilitate (virtual) nelimitata
- Suport pentru hipermedia
- Suport pentru dezvoltarea de aplicatii Web
- Usor de inteles si de programat



Proiecte majore

- RDF → **Semantic Web**
(T.Berners-Lee 2001)
- SOAP (Simple Object Application Protocol) =
RPC + HTTP + XML → **Microsoft .NET**
- Agenti Web, motoare de cautare,
biblioteci digitale
- Marcarea documentelor hipermedia:
**SMIL (Synchronized Multimedia
Integration Language)**
- Situri Web mai flexibile si interactive
- XML EDI (Electronic Document Interchange)



Unelte XML

- **Editoare:** Visual XML, Xeena, XML Spy, Xpose, xmlTool
- **Procesoare:** expat, XML4C, SAX, XML2Beans, Perl libxml, Lark, Xerces
- **Browsere:** Netscape, IE, DocZilla, Jumbo, eXeMel, Opera
- **Utilitare pentru RDF:** Mozilla, SiRPAC, Protégé, RDF for XML, InterDataWorking
- **Biblioteci:** xml-lib, JDOM, SAX, XAF



Adrese

- Consortiul Web: www.w3.org
- Extensibility: www.extensibility.com
- Textuality: www.textuality.com
- XML.org: www.xml.org
- James Clark: www.JClark.com
- XHTML.org: www.xhtml.org
- Semantic Web: www.SemanticWeb.org
- Oasis: www.oasis-open.org

XML, cuceritorul

- Ce este XML
- Familia XML
- Trăsături XML
- Scheme XML
- Aplicații ale XML
- De la HTML la XML
- Viitorul XML



Mulumiri pentru
atenția acordată!