Tehnologii Web

un model de date pentru Web (I)



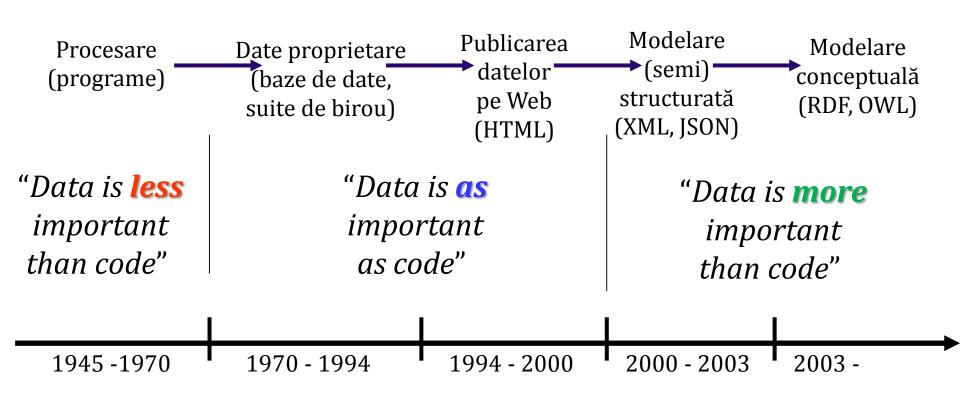
familia XML

"Esențialul în lumânare nu este ceara ce lasă urma, ci lumina."

Antoine de Saint-Exupéry

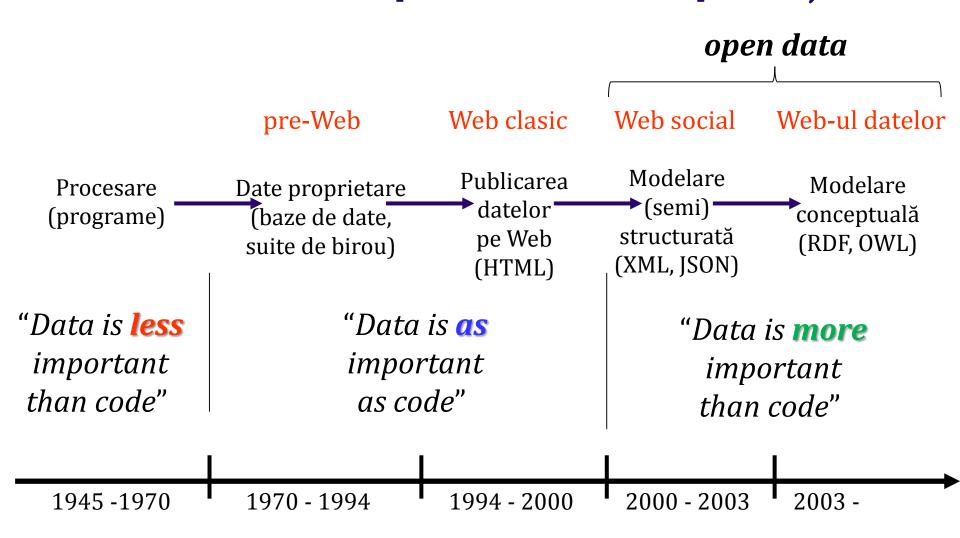
Cum modelăm (reprezentăm) datele?

datele: mai importante ca aplicațiile



evoluția conceptului de "dată" (adaptare după Daconta *et al.*, 2003)

datele: mai importante ca aplicațiile



evoluția conceptului de "dată" (adaptare după Daconta *et al.*, 2003)

Ce model de reprezentare a datelor alegem pentru...

stocarea datelor eterogene ce provin din surse multiple? informații care evoluează în timp? reprezentarea limbajului natural?

Dorim să modelăm & procesăm date privitoare la

antologii de poezii
cataloage de produse ale unui *e-shop*depozite de rețete gastronomice
chestionare
rețele sociale

...

Necesități:

un limbaj de marcare explicită a informațiilor

datele dorite a fi modelate pot fi practic nelimitate (*unbound*) și necunoscute

nu există *a-priori* un vocabular/schemă comun(ă)

Necesități: datele trebuie să poată fi auto-explicative (self-explanatory)

ce reprezintă triplul ("Sabin", "Buraga", 30374)?

Necesități:

modelul adoptat să se preteze arhitecturilor navigaționale actuale, bazate pe hipertext

suport pentru specificarea URI-urilor

Documente: formate particulare *vs.* formate generice

Documente: formate particulare *vs.* formate generice

codificare generică ('60): procedurală – apeluri de proceduri bazată pe marcatori (*mark-ups*)

GenCode – Stanley Rice, Norman Scharpf

GML (*Generalized Markup Language*) Charles Goldfarb *et al.* (IBM) definirea formală a tipurilor de documente

SGML (1986) – standard ISO 8879

Marcaj – adnotare, codare, *mark-up*

orice acțiune de a interpreta explicit o porțiune de text (conținut)

Marcaj – adnotare, codare, *mark-up*

exemple:

semnele de punctuație pentru limbile scrise, deliminatorii folosiți în codul-sursă

```
// Un program JavaScript care implementeaza un client pentru serviciul de salut
var http = require ('http'); // folosim 'http', un modul predefinit
http.get ('http://127.0.0.1:8080/', // emite o cerere HTTP
  function (raspuns) {
    console.log ('Am primit raspuns de la server -- cod HTTP: '
                 + raspuns.statusCode); // statusCode: 200, 404,...
  })
 // tratam diverse evenimente
.on ('error', // eroare
  function (e) { console.log ('Eroare: ' + e.message); })
.on ('response', // receptare raspuns de la server
  function (raspuns) { // exista date de procesat
                                                                  marcaje
    raspuns.on ('data', function (date) { ←
                                                                  speciale
    console.log ('Continut receptionat: ' + date);
  });
  console.log ('Campuri-antet HTTP primite: ' <</pre>
                + JSON.stringify(raspuns.headers));
```

Limbaj de specificare (de adnotare, de marcare) set de convenții de marcare utilizate pentru codificarea datelor

Limbaj de specificare (de adnotare, de marcare) set de convenții de marcare utilizate pentru codificarea datelor

definește mulțimea de marcaje obligatorii, modul de identificare și de structurare a marcajelor pe baza unei gramatici

xml

Extensible Markup Language

meta-limbaj de marcare

descendent simplificat al SGML destinat utilizării în Internet (1996—prezent)

xml

Extensible Markup Language

standard W3C (1998, 2000, 2004, 2006, 2008)

http://www.w3.org/TR/xml/

xml

Extensible Markup Language

o tehnologie + o familie de limbaje

XML TECHNOLOGY



On this page → technology topics

news •

upcoming events and talks

XML Technologies including XML, XML Namespaces, XML Schema, XSLT, Efficient XML Interchange (EXI), and other related standards.

XML Essentials

XML is shouldered by a set of essential technologies such as the infoset and namespaces. They address issues when using XML in specific applications contexts.

Transformation

Very frequently one wants to transform XML content into other formats (including other XML formats). XSLT and XPath are very powerful tools for creating different representations of XML content.

Processing

A processing model defines what operations should be performed in what order on an XML document.

Schema

Formal descriptions of vocabularies create flexibility in authoring environments and quality control chains. W3C's XML Schema, SML, and data binding technologies provide the tools for quality control of XML data.

Query

XQuery (supported by XPath) is a query language for XML to extract data, similar to the role of SQL for databases, or SPARQL for the Semantic Web.

Internationalization

W3C has worked with the community on the internationalization of XML, for instance for specifying the language of XML content.

Security

Manipulating data with XML requires sometimes integrity, authentication and privacy. XML signature, encryption, and xkms can help create a secure environment for XML.

Components

The XML ecosystem is using additional tools to create a richer environment for using and manipulating XML documents. These components include style sheets, xlink xml:id, xinclude, xpointer, xforms, xml fragments, and events

Publishing

XML grew out of the technical publication community. Use XSL-FO to publish even large or complex multilingual XML documents to HTML, PDF or other formats; include SVG diagrams and MathML formulas in the output.

www.w3.org/standards/xml/

xml: caracterizare

Marcaje descriptive

<para> <response> <Person> <tag>

case sensitive

xml: caracterizare

Tipuri de documente

detalii în alt curs

Document Type Definition (DTD)

specificare formală a tipurilor de documente (constituienți + structură)

folosește la verificarea corectitudinii sintactice

xml: caracterizare

Independența datelor

suport pe orice platformă hardware/software

extinderea marcajelor

translatarea/transformarea documentelor

Ușor de implementat, cu suport pentru Web

procesoare XML disponibile pentru toate limbajele de programare

Suport pentru utilizarea internațională

folosirea Unicode

independent de codificare/limbă

Meta-limbaj

definire de alte limbaje

portabil

Soluție pentru reprezentarea conținutului resurselor Web identificate de URI/IRI

asigurarea inter-operabilității (lingua franca)

xml: constituienți

Prolog (preambul)
Elemente
Atribute
Entități
Secțiuni de marcare
Instrucțiuni de procesare

xml: prolog

Declarație care specifică versiunea și codificarea documentului

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

xml: prolog

Declarație care specifică versiunea și codificarea documentului

trebuie să apară o singură dată la începutul documentului

Element = componentă structurală (unitate-text)

Element = componentă structurală (unitate-text)

nume – identifică un element sintaxă similară cu cea a identificatorilor de variabile

produs

Sintactic, un element este specificat via marcatori (*tag*-uri) – de început și de sfârșit

odus>Ping Uinix

Case sensitive

<marcaj> \neq <Marcaj> \neq <MARCAJ>

xml: elemente

Un element poate avea conținut vid

```
cprodus>
```

cprodus />

Model structural

desemnează relațiile dintre elemente: secvență, ierarhie, grupare, incluziune

Elemente imbricate în alte elemente (pot conține un text și/sau alte elemente)

Elemente imbricate în alte elemente (pot conține un text și/sau alte elemente)

Elementele trebuie să fie închise și imbricate corect

```
<?xml version="1.0" ?>
<antologie>
                                    preambul
 <poem>
  <titlu>...</titlu>
    <strofa>
     <vers>...</vers>
     <vers>...</vers>
   </strofa>
 </poem>
 <poem>
   <!-- mai multe poeme... (acesta e un comentariu) -->
 </poem>
</antologie>
```

document XML modelând o antologie de poezii

```
<game>
              <title>Bad Piggies</title>
              <platform>Android</platform>
              <pla><platform>iOS</platform>
              <url>...</url>
              <player>
                            <identity>
                                                 <first-name>Sabin</first-name>
                                                <a href="mailto:square">| <a href="mailto:last-name">| <a href="mailto:las
                                                <!-- eventual, și alte informații -->
                             </identity>
                             <points>30374</points>
              </player>
</game>
```

date XML referitoare la un joc electronic

```
oduse>
  codus>
   <nume>Ping Uinix</nume>
   <ofertant>http://www.pinguin.info</ofertant>
   o> Mascota lunii
 </produs>
 codus>
   <!-- un soi de portocale albastre -->
   <nume>Blue Ory</nume>
   <descriere />
 </produs>
 codus>
   <nume>Ceas "inteligent" cu gust de măr</nume>
 </produs>
</produse>
    un posibil catalog de produse în cadrul unui e-shop
```

Atribut

descrie o anumită proprietate (caracteristică) a unei apariții particulare a unui element

Atributele apar doar în marcajul de început

```
<antologie stare= "ciorna" data="2015-03-29">
...
</antologie>
```

```
<student matricol="TuPi33">
  <nume initiala= "I">Tuxy Pinguinescu</nume>
  </student>
```

Atributele pot fi specificate în orice ordine

Numele atributelor este case sensitive

Valorile atributelor trebuie să fie delimitate obligatoriu de ghilimele sau apostrofuri

Atributele fără valoare nu sunt acceptate

... este eronat

```
<!-- Meta-date asociate unui program, folosite de http://www.pcjs.org/ -->
<manifest>
  <title>VisiCalc</title>
  <version>VC-176Y2-IBM-TEST</version>
  <type>Application</type>
  <category>Productivity</category>
  <company>Software Arts/company>
  <releaseDate>December 16, 1981/releaseDate>
  <machine
    href="/devices/pc/machine/5150/mda/64kb/machine.xml"
    state="/apps/pc/1981/visicalc/state.json"/>
  <disk id="disk" dir="/apps/pc/1981/visicalc/bin/">
    <file>VC.COM</file>
    <file dir="../">README.md</file>
    <link href="http://www.bricklin.com/history/vclicense.htm">
       VisiCalc License
    </link>
  </disk>
</manifest>
```

xml: referințe la entități

Scop:
codificare și referentiere a unei părți de document

sintaxă:

&identificator;

sau

&#număr;

xml: referințe la entități

Entități predefinite:

< > & "

Entități (referințe) de tip caracter:

k; **ă**; **❀**;

xml: secțiuni

Anumite părți din documente necesită procesări speciale

CDATA – inhibă procesarea XML

xml: secțiuni

```
<script type="application/javascript">
```

```
if (vizite < 10) { // nu e vizitator fidel
    $("#mesaj").html ("<p>Salut!");
}
```

Eroare la procesarea XML

</script>

xml: secțiuni

Nu dorim ca procesorul XML să interpreteze codul JavaScript

xml: instrucțiuni de procesare

Includ informații privitoare la aplicațiile (externe) care urmează a fi executate pentru procesarea conținutului

<?processing-instruction ... ?>

xml: instrucțiuni de procesare

```
<script>

echo "Salut!\n";
?>
</script>
```

Procesorul XML ar putea invoca interpretorul PHP pentru execuția programului

xml: instrucțiuni de procesare

Exemplificare:

asocierea unei foi de stiluri CSS pentru redarea conținutului unui document XML

<?xml-stylesheet type="text/css" href="stiluri.css" ?>

xml: familia

XML (Extensible Markup Language) sintaxa XML Information Set - Infoset modelul de date XML (abstract) XLL (Extensible Linking Language) XLink – legături între documente XPointer – localizare relativă a resurselor XSL (Extensible Stylesheet Language) transformări și formatări XQuery (împreună cu XPath) interogarea datelor XML

Structurarea/formatarea conţinutului (formate de prezentare a datelor)

în navigatorul Web: (X)HTML (*Extensible HTML*) formulare electronice: XForms

grafică vectorială: SVG (Scalable Vector Graphics)

lumi tridimensionale: X3D (Extensible 3D Graphics)

în medii fără fir: XHTML-MP (Mobile Profile), Tiny SVG

Reprezentarea diferitelor tipuri de continut

expresii matematice: MathML

date multimedia sincronizate:

SMIL (Synchronized Multimedia Integration Language)

informații vocale: VoiceXML

componente ale interfeței-utilizator:

XUL (Extensible User-interface Language)

XAML (Extensible Application Markup Language)

informații cartografice: KML (Keyhole Markup Language)

Reprezentarea diferitelor tipuri de conținut

documentații: DocBook (Documentation Book)
informații prelucrate de suite de birou – e.g., Open Office:

ODF (Open Document Format)
mediatizare (syndication) – fluxuri de știri:
RSS (Really Simple Syndication), Atom
reguli de specificare a proceselor de afaceri: reguli de specificare a proceselor de afaceri: BRML (Business Rules Markup Language)

Descrierea – conceptuală – a resurselor Web

```
cadrul general: RDF (Resource Description Framework)
exprimarea vocabularelor de meta-date:
DCMI (Dublin Core Metadata Initiative),
FOAF (Friend Of A Friend),
DOAP (Description Of A Project),...
exprimarea modelelor conceptuale – ontologii:
OWL (Web Ontology Language)
```

într-un curs viitor

Descrierea serviciilor Web.

serializarea datelor transmise conform RPC: XML-RPC (XML Remote Procedure Calls) descrierea serviciilor:

WSDL (*Web Service Description Language*) exprimarea protocolului de transfer: SOAP modelarea proceselor de afaceri via servicii Web: BPEL4WS

(Business Process Execution Language for Web Services)

Aşadar, până la urmă, ce este XML?

xml: privire de ansamblu

xml: privire de ansamblu

XML se pretează pentru stocarea (reprezentarea) datelor semi-structurate

sistemele tradiționale (*e.g.*, baze de date relaționale) necesită specificarea unei scheme rigide în vederea stocării/accesării datelor

punct de vedere în general centralizat

xml: privire de ansamblu

XML se pretează pentru stocarea (reprezentarea)
datelor semi-structurate

datele de pe Web sunt descentralizate, eterogene și pot fi, deseori, accesate – chiar modificate – de către oricine
resursele disponibile la un moment dat
nu sunt a-priori cunoscute

nu sunt *a-priori* cunoscute

xml: direcții importante

Comunicații/interacțiune

person-to-person – instant messaging (e.g., XMPP)

person-to-computer - syndication (formatul Atom)

computer-to-computer – servicii Web ▶ mash-up-uri

xml: direcții importante

Publicare deschisă a datelor

modelare (semi-)structurată

procesare independentă de platforma/limbaj

prezentări multiple ale conținutului, în diverse formate și pe sisteme eterogene

xml: direcții importante

Căutare și regăsire – *knowledge retrieval*

meta-date

tagging

navigare

filtrare

import/export de date ▶ inter-operabilitate

...

în contextul aplicațiilor Web sociale

xml: direcții importante

Modelarea proceselor (de afaceri, mai ales)

dataflow

workflow

PRESENTATIONS FROM XML LONDON 2014





DR. MICHAEL KAY & DR. DEBBIE LOCKETT

Benchmarking XSLT Performance

xmllondon.com www.xmlprague.cz

ABEL BRAAKSMA

Streaming Design Patterns or: How I Learned to Stop Worrying and Love the Stream

10:10

10:40

12:40

15:10



MATT KOHL & SANDRO CIRULLI

From monolithic XML for print/web to lean XML for data: realising linked data for dictionaries



WILLIAM NARMONTAS

XML Processing in Scala

GEORGE BINA

XML Authoring On Mobile Devices



DR. ELIAS WEINGÄRTNER

Engineering a XML-based Content



CELINA HUANG

A Visual Comparison Approach to A



STEVEN PEMBERTON

Live XML Data

Hans-Juergen Rennau (Traveltainmen GmbH) 11:10 Coffee break

Native XML Databases: Death or Coming of Age 11:40

Mansi Sheth (Veracode Inc)

Building Security Analytics solution using Native XML Database

Xavier Franc (Qualcomm Technologies Inc) and Michael Paddon (Qualcomm Technologies Inc) A Unified Approach to Design and Implement data-centric and document-centric XML Web 12:10

Node search preceding node construction - XQuery inviting non-XML technologies

Applications

Christine Vanoirbeek (Epfl), Stéphane Sire (oppidoc) and Houda Chabbi (hefr)

Graphical User Interface Tool for Designing Model-Based User Interfaces with UIML

Anne Brüggemann-Klein, Lyuben Dimitrov and Marouane Sayih (Technische Universität München)

13:10 Lunch

Survey State Model (SSM) - XML Authoring of electronic questionnaires Jose Lloret (Robert Gordon University) and Nirmalie Wiratunga (Robert Gordon University)

Schematron for Information Architects

xml: instrumente

Analizoare (procesoare, parsers)
Apache Xerces, Expat, libxml, MSXML,...

Vizualizatoare și editoare structurale Firefox, Open Office, <oXygen />, MS Visual Studio etc.

Formatatoare – *e.g.*, procesoare XSLT/XSL-FO FOP, Saxon, Xalan, XEP

Sisteme de gestiune a bazelor de date XML eXist, Mark Logic's CIS, Sedna etc.

xml: demo



Dacă alegem nume de marcatori/atribute deja definite de alte limbaje XML?

```
<event uri="http://sophia.estec.esa.int/socis2015/faq">
        <name>ESA Summer of Code in Space</name>
        <year>2015</year>
    </event>
```

```
<participant>
    <name uri="mailto:tux@info.uaic.ro">
         Tuxy Pinguinescu</name>
    <year kind="Bachelor">2</year>
    </participant>
```

```
<event uri="http://sophia.estec.esa.int/socis2015/faq">
  <name>ESA Summer of Code in Space</name>
  <year>2015
</event>
                                          conflict!
<participant>
  <name uri="mailto:tux@info.uaic.ro">
     Tuxy Pinguinescu</name>
  <year kind="Bachelor">2</year>
</participant>
```

Spațiu de nume

desemnează un vocabular utilizat pentru calificarea – în mod unic – a elementelor/atributelor XML

Vocabularul definit – colecție de nume de elemente și attribute, plus maniera lor de structurare – poate fi desemnat de un URI

Vocabularul definit poate fi desemnat de un URI

atributul xmlns specifică acest URI, atașând opțional un identificator unic fiecărui vocabular folosit

specificație W3C (2009): http://www.w3.org/TR/xml-names/

```
<?xml version="1.0"?>
<c:calendars xmlns:c="http://www.calendar.info">
 <e:participant xmlns:s="http://www.info.uaic.ro/Students/"
               xmlns:e="http://www.info.uaic.ro/Events/">
    <s:name>Tuxy Pinguinescu</s:name>
    <s:year s:kind="Bachelor">2</s:year>
    <c:calendar>
      <e:event e:ident="SOCIS">
       <e:name>Code in Space</e:name>
                                                      fără
       <e:year>2015</e:year>
                                                   conflicte!
     </e:event>
     <e:event e:ident="StagiiPeBune" />
    </c:calendar>
  </e:participant>
</c:calendars>
```

xml: spații de nume - exemple

Vocabularul XHTML/HTML5: http://www.w3.org/1999/xhtm

Vocabularul Atom: http://www.w3.org/2005/Atom

Vocabularul modelului conceptual oferit de DBpedia:

http://dbpedia.org/ontology/

Vocabularul JSP (Java Server Pages):

http://java.sun.com/JSP/Page

Vocabularul XUL (Extensible User-interface Language):

http://www.mozilla.org/keymaster/gatekeeper/there.is.only.xul

Studiu de caz:

includerea în documentele (X)HTML a unor construcții provenite din alte limbaje XML

exemplificare:

SVG (Scalable Vector Graphics) - http://www.w3.org/TR/SVG/

```
<!DOCTYPE html>
<a href="http://www.w3.org/1999/xhtml">
<head><title>HTML + SVG</title></head>
<body>
 <h1>O ilustrație SVG (grafică vectorială) inclusă într-o pagină Web</h1>
 <!-- Elemente și atribute SVG specificate în documentul HTML,
     recunoscute pe baza spațiului de nume SVG -->
 <svg width="500" height="200" xmlns="http://www.w3.org/2000/svg">
  <!-- o zonă rectangulară cu colțuri rotunjite -->
  <rect x="50" y="50" rx="7" ry="7" width="450" height="150"
        style="fill: #6699FF; stroke: #3333CC;"/>
  <!-- continut textual -->
  <text x="70" y="90" style="stroke: navy; fill: white; font-size: 32pt;">
       SVG direct în browser...</text>
  <!-- un cerc galben -->
  <circle cx="400" cy="150" r="33" style="fill: yellow; stroke: red;" />
 </svg>
 De utilizat un navigator Web care oferă suport nativ pentru SVG.
</body>
</html>
```

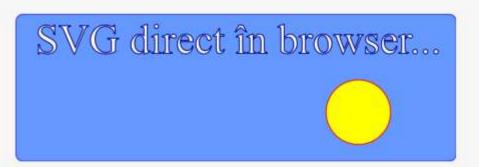
```
HTML *
<!DOCTYPE html>
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
 <title>HTML + SVG</title>
</head>
<body>
 <h1>0 ilustratie SVG (grafica vectoriala)
      inclusa intr-o pagina Web</hl>
  <!-- Elemente si atribute SVG incluse
       in documentul HTML,
       recunoscute pe baza spatiului de nume SVG -->
  <svg width="500" height="200"
       xmlns="http://www.w3.org/2000/svg">
    <!-- zona rectangulara cu colturi rotunjite -->
    <rect x="50" y="50" rx="7" ry="7"</pre>
      width="450" height="150"
       style="fill: #6699FF; stroke: #3333CC;"/>
    <!-- continut textual -->
    <text x="70" y="90" style="stroke: navy;</pre>
          fill: white; font-size: 32pt;">
       SVG direct în browser...
      </text>
    <!-- un cerc galben -->
    <circle cx="400" cy="150" r="33"</pre>
            style="fill: yellow; stroke: red;" />
  </svg>
 De utilizat un navigator Web
    care ofera suport nativ pentru SVG.
</body>
</html>
```

Output

Run with JS

Auto-run JS V

O ilustratie SVG (grafica vectoriala) inclusa intr-o pagina Web



De utilizat un navigator Web care ofera suport nativ pentru SVG.

xml: spații de nume - demo

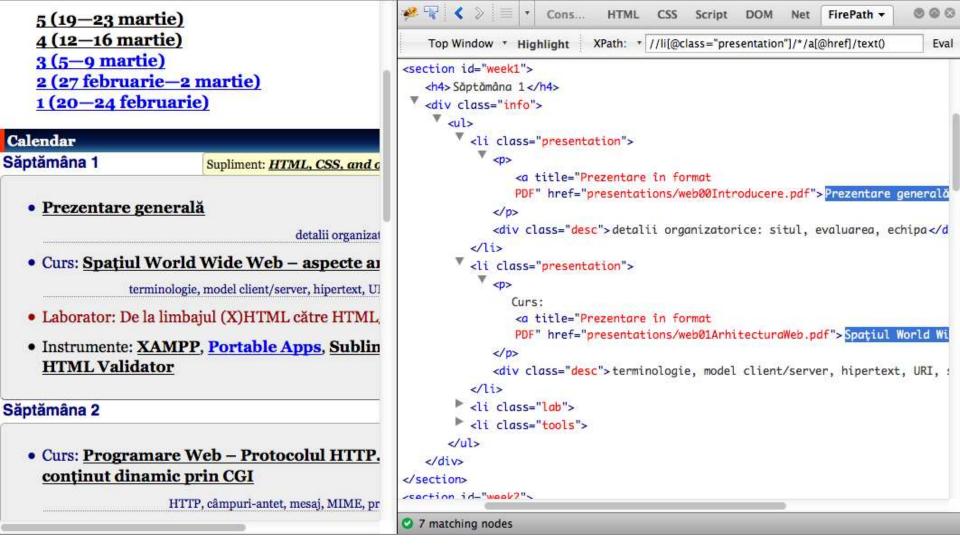


rezumat

modelarea datelor prin XML



caracterizare & constituienți, aplicații și utilizări, spații de nume XML



episodul viitor:

extragerea datelor cu XPath validarea documentelor XML prin DTD