
XML

Oinarriak

Xabier Arregi Iparragirre

Xabier Artola Zubillaga

Aurkezpenaren eskema

- Sarrera
 - › Zer da XML
 - › XMLren ezaugarriak
 - Datuen eramangarritasuna
 - Erabilera-testuingurua
 - › Bilakaera
- XMLren oinarriak

Zer da XML

- XML, *eXtensible Markup Language*, dokumentuak *markatzeko* estandarra da.
 - › Sintaxi orokorra definitzen du.
 - Etiketa sinpleak eta irakurgarriak erabiltzean datza
 - Etiketa horiek dokumentuetako edukiak deskribatzen dituzte, eta testuan bertan txertatzen dira
 - Markatzean jarraitu beharreko arau sintaktikoak definitzen dira: nola txertatu etiketak dokumentuetan, nola mugarritu *elementuak*, nola idatzi etiketak, zein izen jarri elementuei...

Zer da XML (II)

- › Formatu estandarra eskaintzen du.
 - Malgua eta irekia:
 - edozein dokumentu marka daiteke
 - etiketa multzoa ez dago aldeztatik mugatua: *eXtensible*
 - Zurruna:
 - oinarrizko arauak errespetatu behar dira, nahitaez
 - XML gramatika

Adibidea

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>
<?xml-stylesheet href="pertsona.css" type="text/css" ?>
<pertsona>
  <izen-abizenak>
    <izena>Pier Paolo</izena>
    <abizena>Pasolini</abizena>
  </izen-abizenak>
  <lanbidea>zine-zuzendaria</lanbidea>
  <lanbidea>idazlea</lanbidea>
</pertsona>
```

XMLren ezaugarriak

- Markatzen dena testua da. Markatze-etiketak testua dira. Dena da testua.
 - › Edozein testu-editore aski da XML dokumentua irakurtzeko eta sortzeko.
- Programek ere XML dokumentuetako datuak maneiatzeko erraztasuna dute.
 - › Liburutegi ugari.
 - › Dokumentuen egitura atzigarri.

XMLren ezaugarriak (II)

- Ez da aurkezpen-lengoaia. Markak ondo erabiliz gero, XMLk dokumentuen egitura eta semantika adierazteko balio du.
- Izatez, XML markatze-metalengoaia da.
 - › Erabiltzaileak defini ditzake bere gustuko etiketak, eta etiketa horien arteko antolaera.
- Dokumentu-mota jakin baterako onartzen den markatzea *eskema* baten bidez adierazten da.
 - › Eskema-lengoiak: DTD, XML Schema (W3C), RELAX NG...

XMLren ezaugarriak (III)

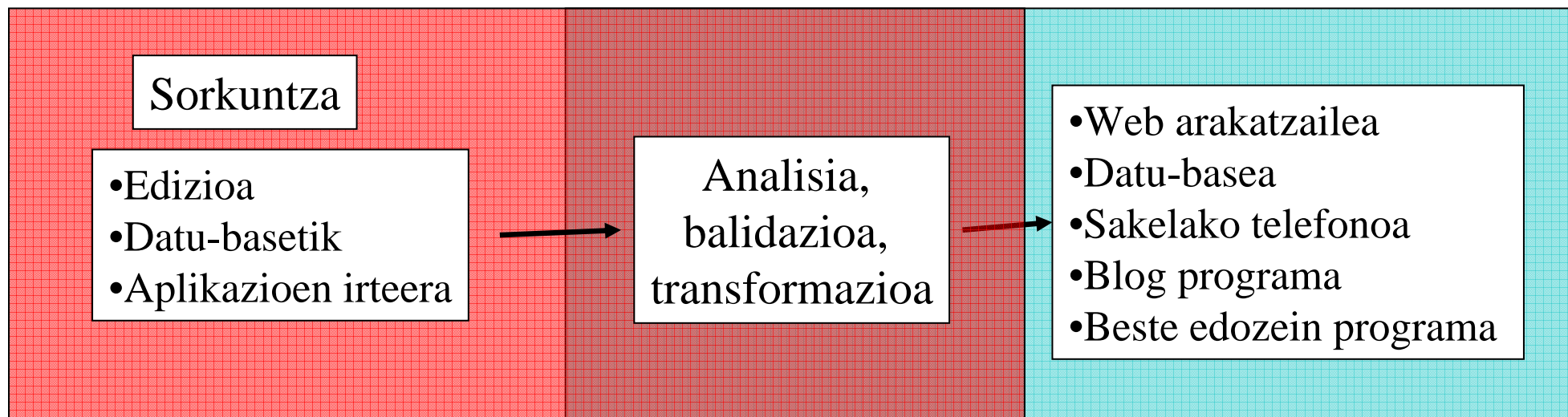
- Dokumentu-motak erabilera zabala duenean, eta erakunde edo talderen batek adostua bada: *XML aplikazioa*.
 - › Adibidez: XSL, XML Schema, ebXML, HL7, RSS, MathML, SMIL...

XMLren ezaugarriak: datuen eramangarritasuna

- Testu soila denez, ez dago arazorik plataforma artean informazioa pasatzeko.
 - › Dokumentuetako edukiak eta egiturazko ezaugarriak esplizituak dira.
- Horregatik, informazioa trukatu, gorde, kudeatu eta argitaratzeko erabiltzen da.

Oinarrizko erabilera- testuingurua

Zerbitzaria:
dokumentuen sortzailea eta banatzailea



Bezeroa: dokumentuen kontsumitzailea

Bilakaera

- SGMLtik eratorria da XML.
- *SGML: Standard Generalized Markup Language.*
 - › 70eko hamarkadan sortua eta 80koan garatua.
 - › markatze-metalengoaia egiturazkoa eta semantikoa.
 - › baina, konplexua...
- HTML: SGML aplikazioa.
 - › web orriak deskribatzeko etiketa multzo finitua.
 - › aurkezpenari begirako etiketak, batez ere.
 - › ez dago beste ezertan erabiltzeko pentsatua.

Bilakaera (II)

- 1996an, SGMLren bertsio sinplifikatu bat egiten saiatzen dira.
 - › Motibazioak:
 - SGMLren estiloa zaindu, baina erabilerrazago bihurtu.
 - HTMLren mugak gainditu.
 - Weberako, SGML baino egokiago eta erabilgarriagoa litzatekeen zerbait egin.
 - › Emaitza: XML 1.0 (1998)

Bilakaera (III)

- Geroztikakoak dira...
 - › *Izen-espazioak XMLn*: XML aplikazio bat baino gehiago dokumentu berean, elkarrekin talka egin gabe, erabiltzeko aukera.
 - › *XSL (eXtensible Stylesheet Language)*: XML dokumentuak transformatzeko lengoia, hasiera batean web arakatzailerik interpreta dezaketen formatura. Gero:
 - *XSLT: XSL Transformations*. XML dokumentuak transformatzeko lengoia orokorra.
 - *XSL-FO: XSL Formatting Objects*. Web orrien edo orri inprimatuen diseinua (aurkezpen, bistaratze kontuak) adierazteko lengoia.

Bilakaera (IV)

- Edukiak erreferentziatzeko eta elkarren artean erlazionatzeko...
 - › XLink (*eXtensible Linking Language*): estekak deskribatzeko lengoia.
 - XLink: dokumentuen arteko estekak.
 - › XPath: dokumentu barruko osagaiak erreferentziatzeko espezifikazioa.
 - › XPointer: dokumentuen osagaiak erreferentziatzeko espezifikazioa, baina XPath baino zehatzagoa.
 - XLink-ekin batera dokumentuen barruko estekak adieraz daitezke.
 - XSLTk eta XPointer-ek XPath darabilte.

Bilakaera (V)

- Programatzeko...
 - › DOM (*Document Object Model*): programazioko objektuen ikuspegia aplikatzen zaie XML dokumentuei. Dokumentuen edukiak atzitzeko eta maneiatzeko interfaze uniformeak (API) da DOM. Datu-egitura gisa hartzen du XML zuhaitza. W3Cren proposamen ofiziala.
 - › SAX (*Simple API for XML*): XML zuhaitza, datu-egitura baino, gertaera-fluxu moduan hartzen da (analizatu ahala sortzen dira gertaerok). Sinpleagoa eta, zenbaitetan, eraginkorragoa.

Bilakaera (VI)

- Dokumentuen eskema kontzeptuala adierazteko, eskema-lengoaiak:
 - › DTD (*Document Type Definition*)
 - › XML Schema (W3C)
 - › RELAX NG, Schematron, NRL schema, DSD

Bilakaera (VII)

- Oraingo eta etorkizunera begirako garabideak...
 - › *XML Query Language*: informazioa erauzteko lengoia. Datu-baseetako moldea, XML munduko SQL modukoa.
 - › *XInclude*: XML dokumentu baliozkoak konbinatuz beste XML dokumentuak osatzeko aukera.
 - › *XML Signatures*: XML dokumentuak digitalki sinatzeko estandarra.
 - › *XML Encryption*: Zifratutako eduki digitalerako XML estandarra.
 - › *XML Native* datu-baseak: Edukiak XMLz antolatzen dituzten datu-baseak. Adibidez: *XIndice*, *eXist*, *Berkeley DB XML*...



HTML *vs.* XML: adibidea

HTML

```
<table border="1">
  <tr>
    <th>titulua</th>
    <th>egilea</th>
    <th>isbn</th>
  </tr>
  <tr>
    <td>XML in Action</td>
    <td>William J.</td>
    <td>0-7564-0562-9</td>
  </tr>
</table>
```

XML

```
<liburua>
  <titulua>XML in Action</titulua>
  <egilea>William J.</egilea>
  <isbn>0-7564-0562-9</isbn>
</liburua>
```

HTMLren bertsioak

- 1992: lehen definizioa
- 1994: HTML 2.0 (estandarra oinarrizko ezaugarrietarako)
- 1997: HTML 4.0 (egitura eta aurkezpena –estilo-orriak– bereizteko saioa)
- **1998: XML**
- 1999: HTML 4.01 (sintaxi-definizio zehatza eta formala)
- 2000: XHTML 1.0 (HTML 4.01en XML bertsioa):
 - › balidatzeko gunea
- 2002: XHTML 2.0



XHTML adibidea

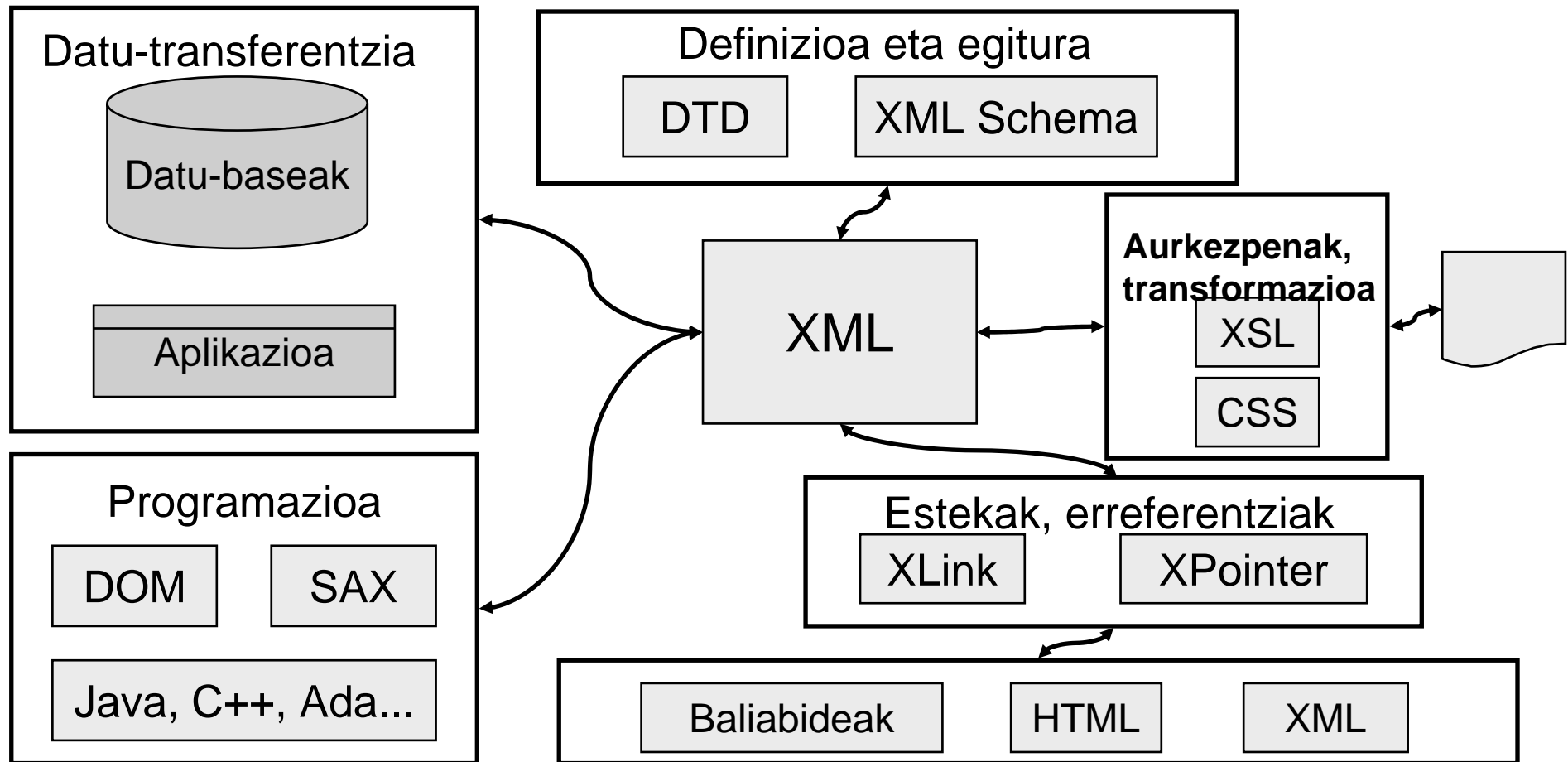
```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html SYSTEM "xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head xml:lang="en">
    <script type="JavaScript"/>
    <title dir="ltr" lang="en"/>
    <style type="das"/>
  </head>
  <body>
    <h1>XHTMLadibidea</h1>
    <p>XMLri buruzko estekak:</p>
    <ul>
      <li>
        <a href="http://www.xml.com/">www.xml.com</a>
      </li>
      <li>
        <a href="http://www.xml.org/">www.xml.org</a>
      </li>
    </ul>
  </body>
</html>
```

Arazoak HTMLrekin

- Hipertestuak deskribatzeko lengoia da:
 - › etiketen multzo finkoa, semantika finkatua;
 - › baina informazioa ez da soilik hipertestua;
- Sintaxia eta semantika nahastu egiten dira:
 - › egiturak aurkezpena baldintzatzen du;
 - › estilo-orrien irtenbidea nola-halakoa da:
- Ez dago egiazko estandarrik:
 - › HTML dokumentu gehienak ez dira baliozkoak;
 - › arakatzailleek, zeinek bere estandarra "definitzen" du;



XML munduaren ikuspegia



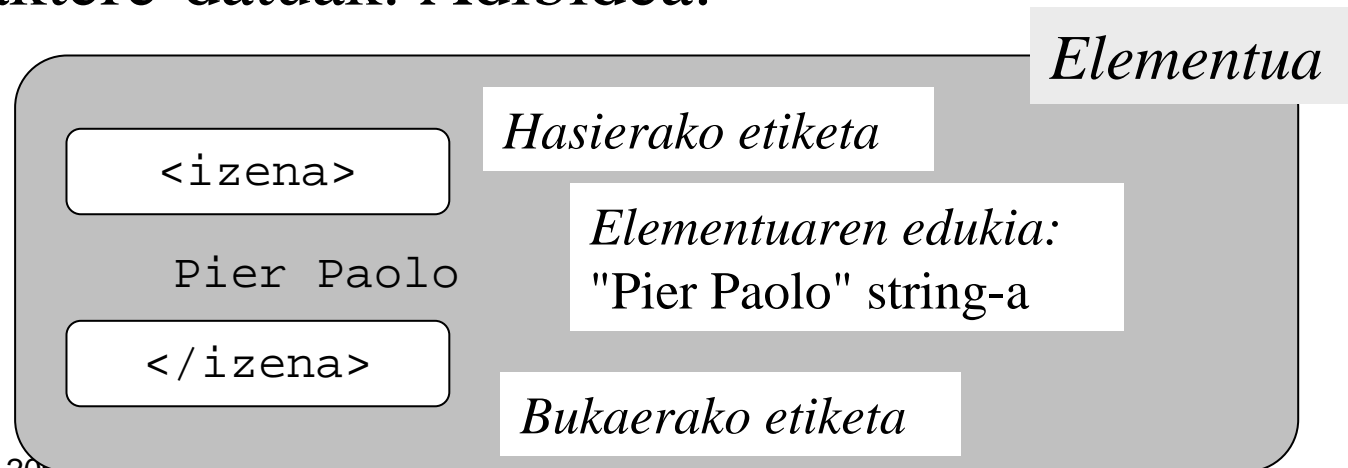
UEU. Iruñea, 2004ko uztaila.

UPV/EHU. Donostiako Informatika Fakultatea. Softwarea garatzeko tresna aurreratuak (2004/07)

XMLren oinarriak.

XML dokumentuak eta elementuak.

- XML dokumentuak prozesagarriak izan daitezzen, sintaxi-arau zorrotzak bete behar dituzte.
 - › Elementuak dira oinarrizko osagaiak.
- Oinarrizko markatze-eredua. Elementuak, etiketak eta karaktere-datuak. Adibidea:



Etiketen sintaxia

- Hasierako etiketak: `<izena>`
- Bukaerako etiketak: `</izena>`
 - Beti jarri behar dira.
- Etiketa-izenak edukiaren izaera adierazten du.
- Elementu hutsak, edukirik ez dutenak:
`<ezkontza-urtea/>`
- Maiuskulak eta minuskulak bereizi egiten dira.
- Elementuen habiaketa: elementu baten hasierako etiketa beste baten baitan badago, bukaerakoak ere halaxe egon behar du.

UEU. Iruñea, 2004ko uztaila.

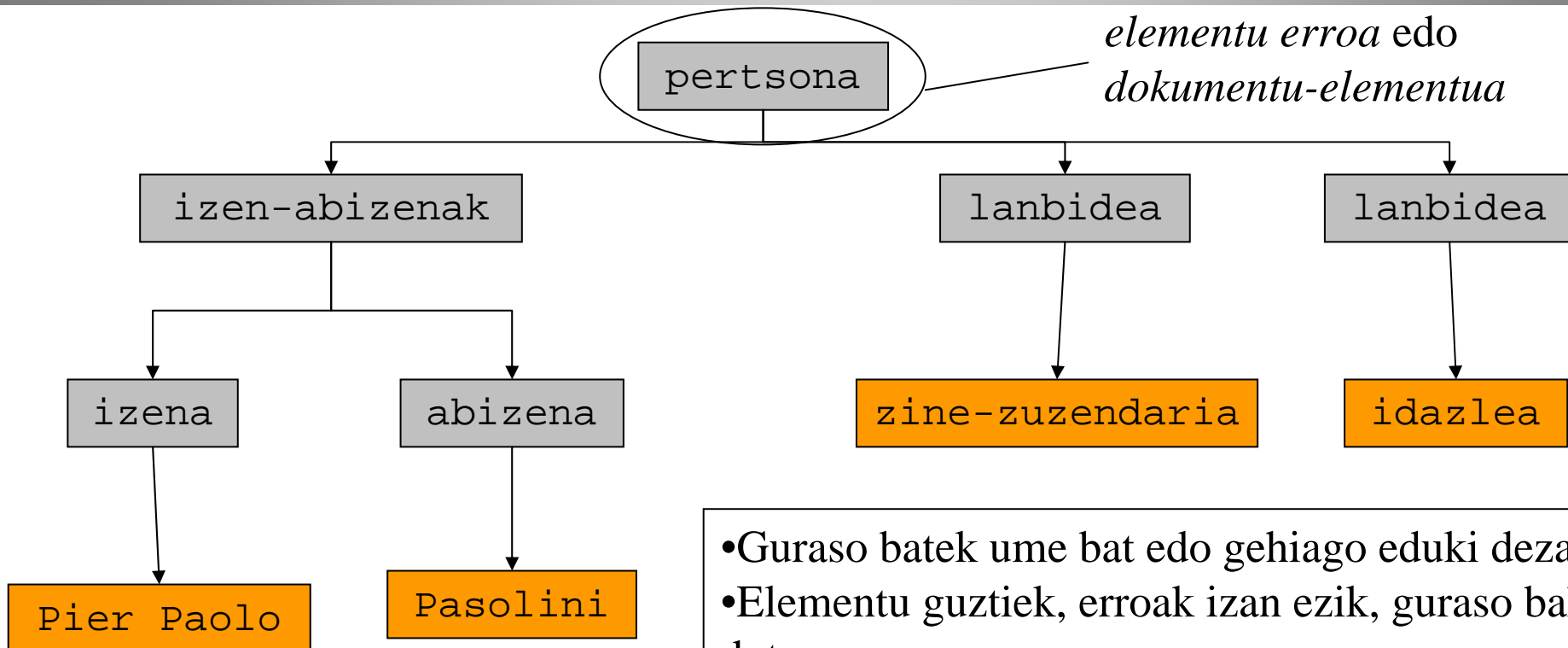
UPV/EHU. Donostiako Informatika Fakultatea. Softwarea garatzeko tresna aurreratuak (2004/07)

XML zuhaitzak

- XML dokumentuak zuhaitz-erako datu-egituratzat har daitezke.
- Har dezagun honako adibide hau:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>
<pertsona>
  <izen-abizenak>
    <izena>Pier Paolo</izena>
    <abizena>Pasolini</abizena>
  </izen-abizenak>
  <lanbidea>zine-zuzendaria</lanbidea>
  <lanbidea>idazlea</lanbidea>
</pertsona>
```

XML zuhaitzak (II)

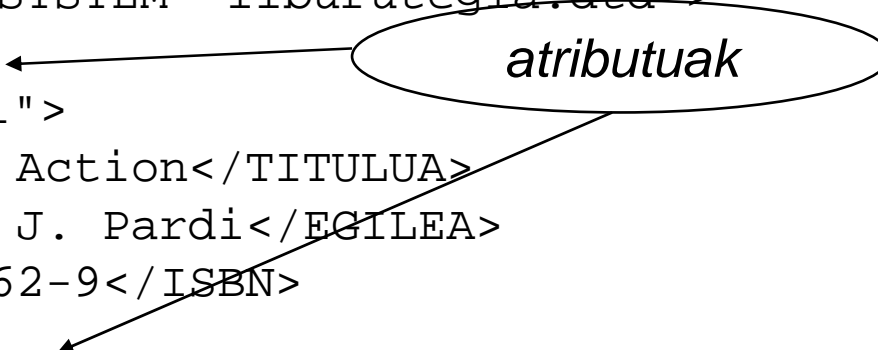


- Guraso batek ume bat edo gehiago eduki dezake.
- Elementu guztiek, erroak izan ezik, guraso bakarra dute.
- Erro bakarra eduki behar dute dokumentuek.

Beste adibide bat

liburutegia.xml

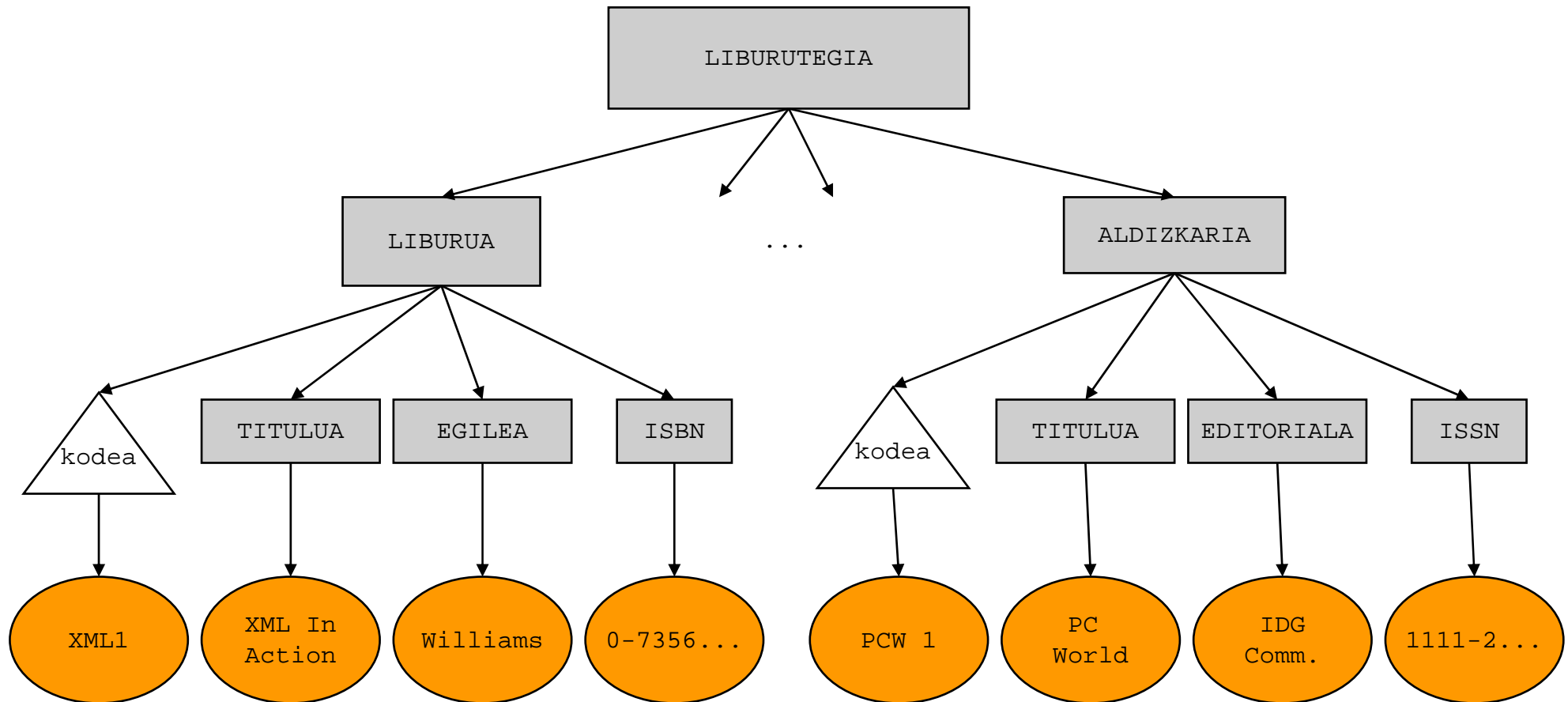
```
<?xml version="1.0" ?>
<?xml:stylesheet type="text/xsl" href="liburutegia.xsl" ?>
<!DOCTYPE liburutegia SYSTEM "liburutegia.dtd">
<LIBURUTEGIA>
  <LIBURUA kodea="XML1">
    <TITULUA>XML in Action</TITULUA>
    <EGILEA>William J. Pardi</EGILEA>
    <ISBN>0-7356-0562-9</ISBN>
  </LIBURUA>
  <ALDIZKARIA kodea="PCW1">
    <TITULUA>PC World</TITULUA>
    <EDITORIALA>IDG Communications</EDITORIALA>
    <ISSN>1111-2222</ISSN >
  </ALDIZKARIA>
</LIBURUTEGIA>
```



UPV/EHU. Irutegia, 2004ko uztaia.

UPV/EHU. Donostiako Informatika Fakultatea. Softwarea garatzeko tresna aurreratuak (2004/07)

XML zuhaitza



UEU. Iruñea, 2004ko uztaila.

UPV/EHU. Donostiako Informatika Fakultatea. Softwarea garatzeko tresna aurreratuak (2004/07)

Eduki mistoa

- Karaktere-datuak (testu laua) eta elementu umeak nahastu egiten dira zenbaitetan:

```
<biografia>
<izen-abizenak><izena>Bernardo</izena>
<abizena>Bertolucci</abizena></izen-abizenak> <lanbidea>zine-
zuzendari</lanbidea> italiarra, hasieran <izen-abizenak><izena>Pier
Paolo</izena> <abizena>Pasolini</abizena></izen-abizenak> maisuaren
laguntzailea izan zen eta haren eragina nabari zaio...
</biografia>
```

- *Data-centric XML*: datuei orientatua (DBak)
- *Document-centric XML*: "narratiboagoa" (testu markatuak)

UEU. Iruñea, 2004ko uztaila.

UPV/EHU. Donostiako Informatika Fakultatea. Softwarea garatzeko tresna aurreratuak (2004/07)

Atributuak

- Elementuek atributuak eduki ditzakete.
- Atributua izen-balio bikotea da, elementuaren hasierako etiketan doana.
- Balioak, beti, komatxoan artean jartzen dira.

```
<pertsona id="G1900000" info-jatorria="wikipedia">  
  <izen-deiturak>  
    <izena>Pier Paolo</izena>  
    <deitura>Pasolini</deitura>  
  </izen-deiturak>  
  <jaioa>1922</jaioa>  
  <hila>1975</hila>  
</pertsona>
```

Atributuak (II)

```
<pertsona info-jatorria="wikipedia">  
  <jaiotza-urtea>1922</jaiotza-urtea>  
  <izen-abizenak>  
    <izena>Pier Paolo</izena>  
    <abizena>Pasolini</abizena>  
  </izen-abizenak>  
  <lanbidea mota="nagusia">zine-zuzendaria</lanbidea>  
  <lanbidea mota="bigarrena">idazlea</lanbidea>  
</pertsona>
```

- Noiz erabili elementu (ume)ak? Noiz atributuak?
 - Irizpide bat:
 - atributuak meta-informaziorako erabili
 - elementu (ume)ak informaziorako
 - Kontuan hartzekoa:
 - elementu batek ez du onartzen atributu bat baino gehiago izen berarekin
 - atributuen balioek ez dute egiturarik onartzen.

XML izenak

- XML espezifikazioak arau berberak ezarri ditu edozein *XML izen* idazteko, berdin dio elementuak diren, atributuak diren, edota bestelako izenak diren.
- Arauen laburpena:
 - › Edozein karaktere alfanumeriko onartzen da.
 - Hiru puntuazio-karaktere onartzen dira: _ - .
 - Gainontzeko puntuazio-karakterek ez dira onartzen.
 - Zuriunerik ez.
 - › Hasi, edozein letrarekin edo "_" karakterearekin has daitezke.

Entitate-erreferentziak

- Arazoa sortzen da "<" karakterea elementu baten karaktere-datuetan jarri nahi denean, beti etiketa-hasiera gisa interpretatzen baita.
- Irtenbidea: *entitate-erreferentzia* erabiltzea.
 - "<" erabili "<" karakterea adierazteko.
- Era berean, "&" karakterea ere nahasgarria da, eta horrela beste batzuk ere bai.

Entitate-erreferentziak (II)

| | |
|---|---|
| <code>&lt;</code> (<code>&#60;</code>) | < |
| <code>&amp;</code> (<code>&#38;</code>) | & |
| <code>&gt;</code> | > |
| <code>&quot;</code> | " |
| <code>&apos;</code> | ' |

- `<` eta `&` derrigorrez erabili behar dira, gainontzekoak ez beti.
- Aurredefinitutako bost entitate-erreferentzia hauez gain, DTDan beste batzuk ere defini daitezke.

CDATA sekzioak

- Sekzio osoak txerta daitezke, sekzioko karaktereak banaka kodetu gabe.
- Adibidez:

```
<azalpena> <kodea> xmlns </kodea> atributua erabil daiteke  
<kodea> svg </kodea> aurrizkia elementu guztiei jarri gabe;  
horrela:
```

```
  <![CDATA[
```

```
    <svg xmlns="http://www.w3.org/2000/svg"
```

```
      width="12cm" height="10cm">
```

```
        <ellipse rx="110" ry="150" />
```

```
      </svg>
```

```
    ]]>
```

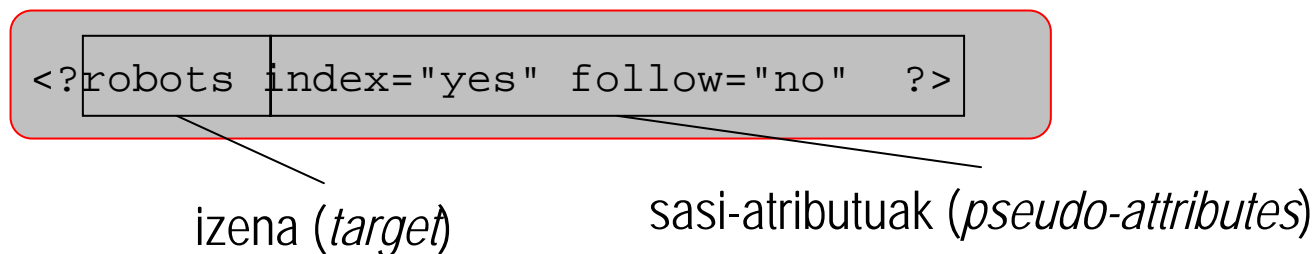
```
</azalpena>
```

Iruzkinak

- XML dokumentuetan iruzkinak jar daitezke.
- Ez dira elementuak.
- Non jarri:
 - › Karaktere-datuen barruko edozein tokitan.
 - › Elementu erroaren aurretik edo ondoren.
- Nola:
 - › "`<!--`" eta "`-->`" string-en artean.

Prozesatze-aginduak

- Dokumentua prozesatuko duten aplikazioei informazioa pasatzeko modua da *prozesatze-agindua*.
- Ez dira elementuak.
- Non: etiketa baten barruan izan ezik, edonon.
- Nola: " <? " eta " ?> " ikurren artean.
- Adibideak:



Prozesatze-aginduak (II)

```
<?xml-stylesheet href="pertsona.css" type="text/css" ?>
<pertsona>
    Pier Paolo Pasolini
</pertsona>
```

- Adibide honetan, estilo-orri bat atxikitzen zaio XML dokumentuari. Ohikoa da.
- Elementu erroa baino lehenago jartzen da.
- Arakatzaleei esaten zaie *pertsona.css* estilo-orria aplikatu behar dela dokumentua erakusteko.

XML erazagupena

- XML dokumentuek *XML erazagupen* batekin hasi behar lukete, baina ez da derrigorrezkoa.
- Itxuraz, XML erazagupena prozesatze-agindu baten modukoa da, baina teknikoki ez da hala.

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>  
<pertsona>  
    Pier Paolo Pasolini  
</pertsona>
```

XML erazagupena (II)

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>
```

- encoding:
 - › Hautazko atributua da.
 - › Atributu honen bitartez adierazten da zein karaktere-kodeketa erabiltzen den dokumentuan.
 - › Besterik ezean Unicode-ko UTF-8 kodeketa hartzen da.
- standalone:
 - › Hautazko atributua da. Besterik ezean, "no" balioa.
 - › "no" balioa badu, kanpoko DTD baten premia dago, dokumentua zuzenki tratatzeko.
 - › "yes" balioa badu, ez da beharrezkoa kanpoko informaziorik irakurtzea.

Ongi eratutako dokumentuak

- XML dokumentu guztiek, salbuespenik gabe, "ongi eratuta" egon behar dute erabilgarriak izango badira (*well-formed*).
- Edozein analizatzailek egiaztatu egingo du ea dokumentua ongi eratuta dagoen. Erroreren bat aurkituz gero, ez du analisia burutuko.
- Beraz, XML dokumentua publiko egin aurretik, analizatu egin behar da.

Ongi eratutako dokumentuak: arauak

1. Hasierako etiketa guztiek bukaerako bat behar dute
2. Elementuak habia daitezke, baina ez teilakatu.
3. Elementu erro bat eta bakarra egon behar du.
4. Atributu-balioak komatxoan artean jarri behar dira.
5. Elementu batek ezin dezake izen bereko atributu bat baino gehiago eduki.
6. Iruzkinak eta prozesatze-aginduak ezin dira jarri etiketen barruan.
7. Elementu baten edo atributu baten karaktere-datuetan ezin dira zuzenean jarri "<" eta "&" karaktereak.

Ongi eratutako dokumentuak: nola egiaztatu

- Web arakatzaille batean kargatuz, ongi eratua badago, erakutsi egingo du; bestela, errore-mezua azalduko du.
- XML analizatzailearen bat erabiliz.
- XML inguruneetan, edizio aukeraz gain, analisia egiteko aukera ere eskaintzen da.