

Tutorial

Introdução ao XML e tecnologias associadas

Sessão 4

Aplicações e dialectos XML

Miguel Ferreira

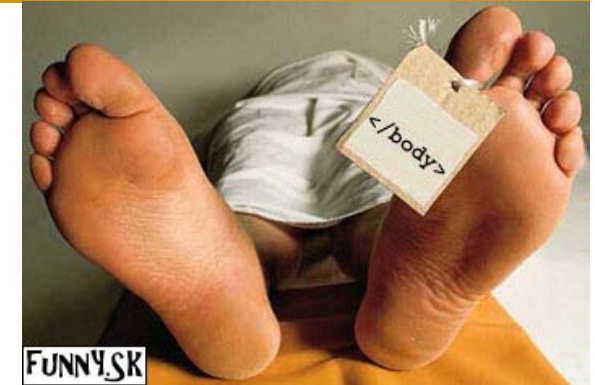
mferreira@dsi.uminho.pt

Departamento de Sistemas de Informação
Universidade do Minho

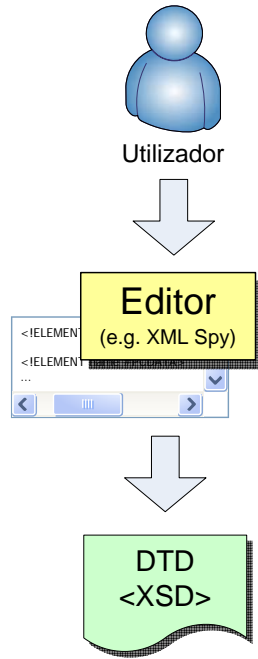
- Vantagens do XML
- Aplicações do XML
- Armazenamento de XML
- Processamento de XML
- Transmissão de XML
- Dialectos
 - SVG, GML, MusicML, DocBook, Dialectos para metainformação, ChessML
- Repositórios de DTD/Schemas
- Exercício e discussão

Vantagens do XML

- **É apenas texto!**
 - Interoperabilidade/portabilidade
- Não tem dono, **é de toda a gente!**
 - Norma aberta (W3C)
- Feito para ser **lido por máquinas, mas também por humanos!**
 - Auto-descritivo, **simples** de compreender e processar
- E se eu quiser manipular XML?
 - **Grande nº de ferramentas** e bibliotecas para todas as linguagens de programação
- Pode ser usado em que contextos?!
 - Todos! É **extensível!**

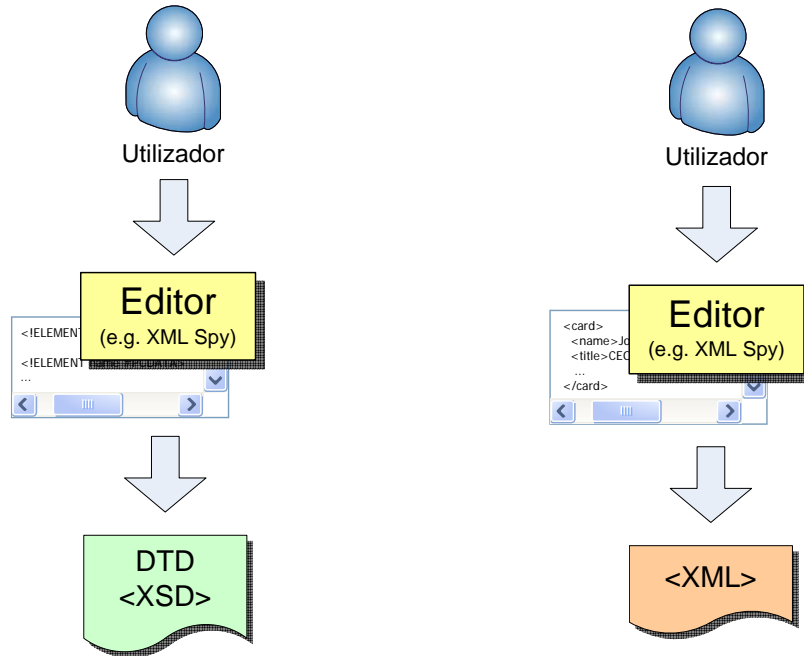


- **Publicação electrónica**
 - Uma representação neutra, múltiplas saídas!
- **Comércio electrónico**
 - Intercâmbio entre diferentes aplicações
- **Metainformação**
 - Grande expressividade e estruturação
- **Computação ubíqua**
 - Mesmos dados, múltiplos dispositivos
- **Aplicações em geral**
 - Representação, armazenamento e transmissão de informação



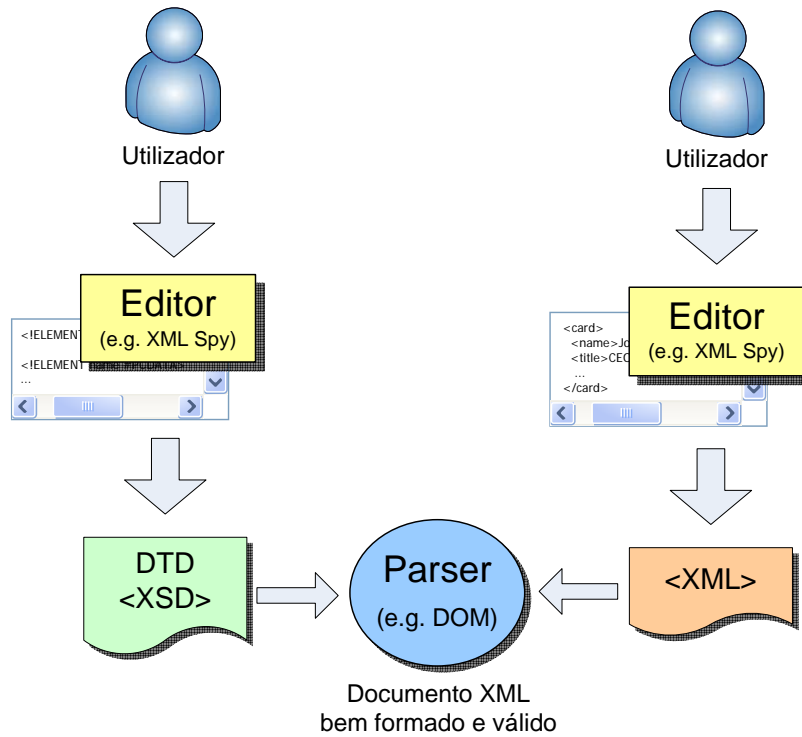
Ciclo de vida de um XML

XATA'06
8-10 Fevereiro



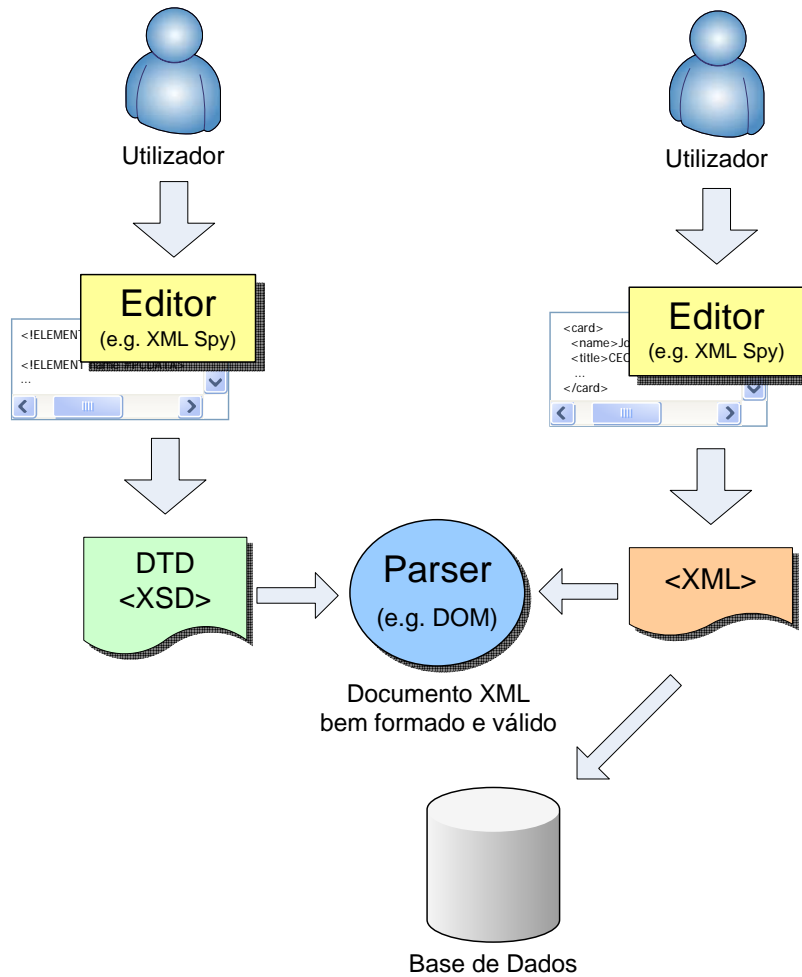
Ciclo de vida de um XML

XATA'06
8-10 Fevereiro



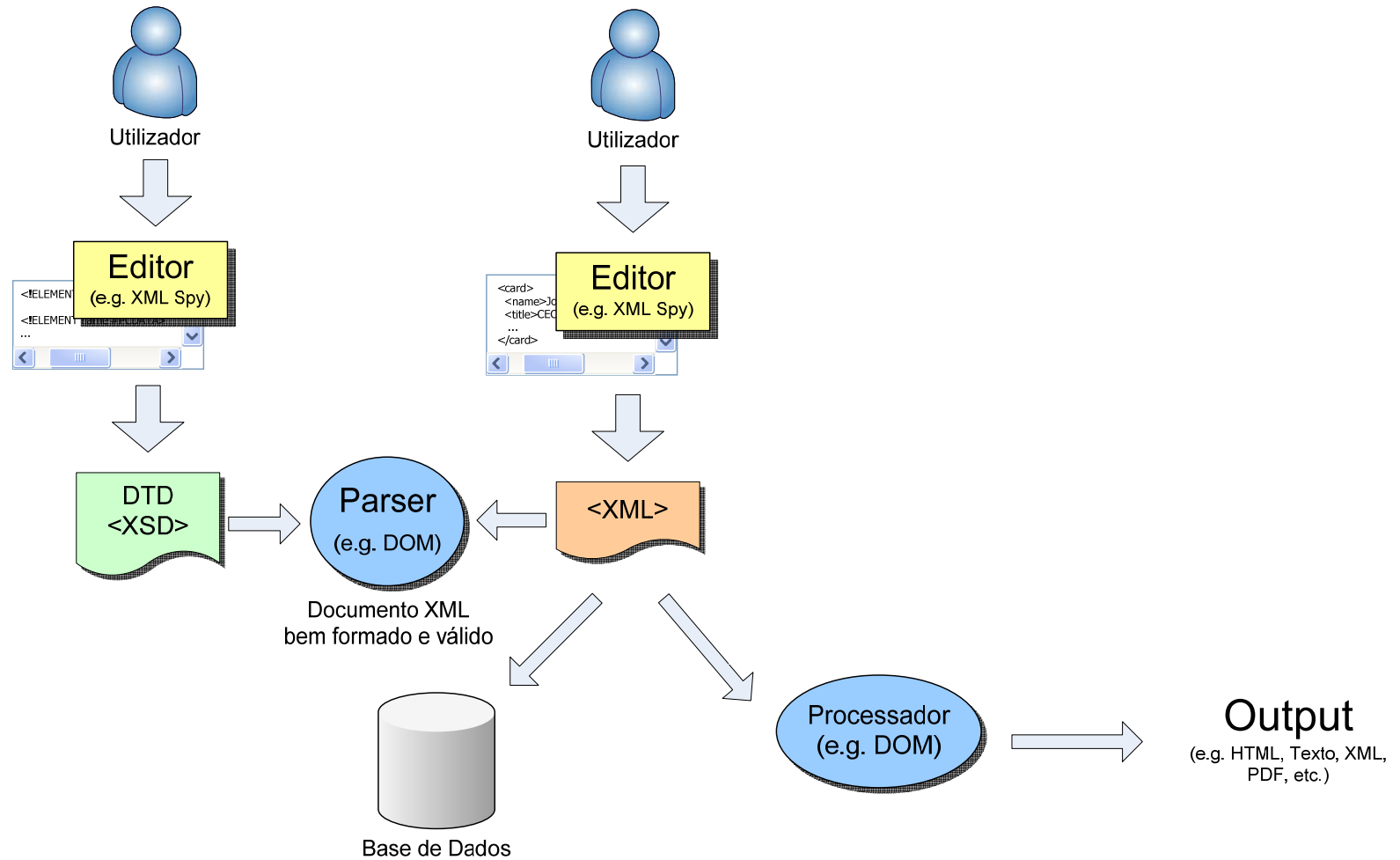
Ciclo de vida de um XML

XATA'06
8-10 Fevereiro



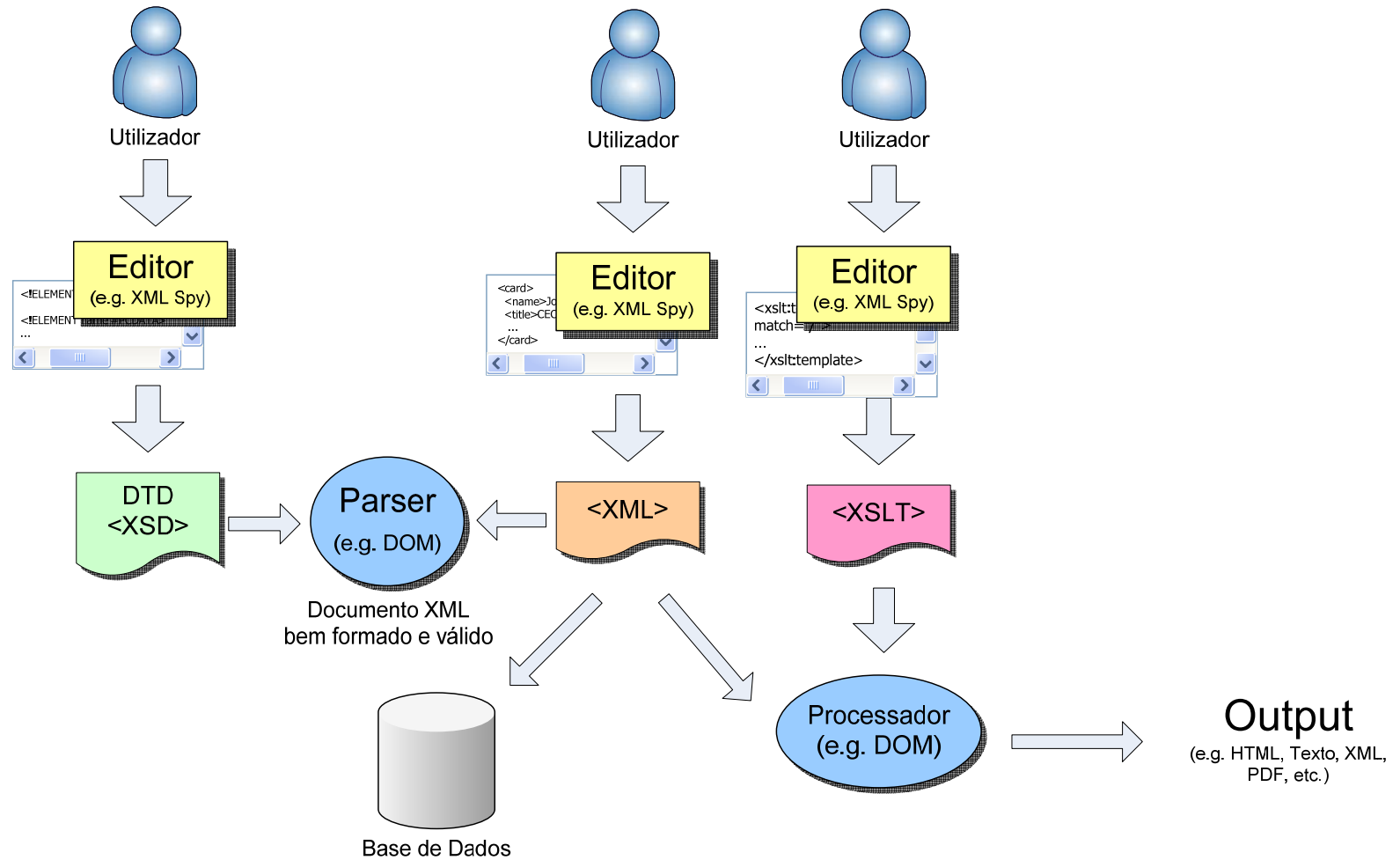
Ciclo de vida de um XML

XATA'06
8-10 Fevereiro



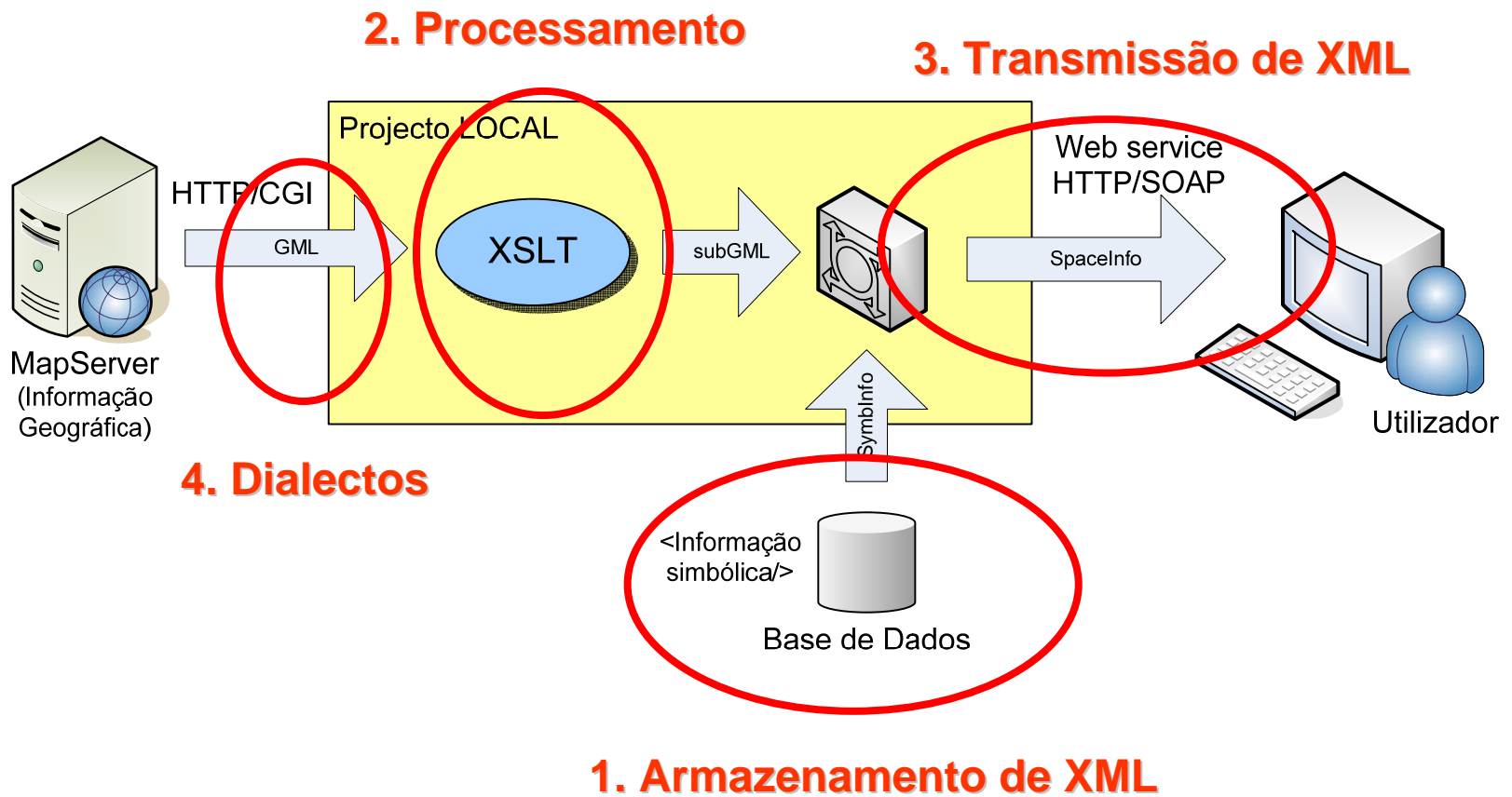
Ciclo de vida de um XML

XATA'06
8-10 Fevereiro

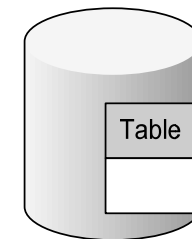
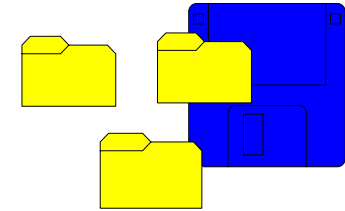


Exemplo de uma aplicação XML

XATA'06
8-10 Fevereiro

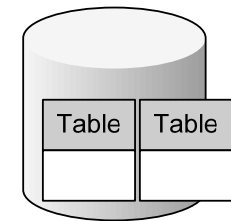


- Sistema de **ficheiros**
 - ❑ Um ficheiro por cada documento
 - ❑ Dificuldade em **encontrar** informação
 - ❑ Necessidade de **índices auxiliares** (e.g. RBDMS)
- Bases de dados **relacionais**
(uma tabela com uma coluna)
 - ❑ Bloco de texto
 - **CLOB** (Character Large Object)
 - Varchar, Text
 - ❑ Não é possível **questionar** o XML
 - ❑ Não permite **actualizações** parciais



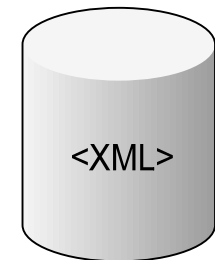
BD Relacional

- Bases de dados **relacionais**
(múltiplas tabelas com múltiplas colunas)
 - ❑ **tabelas** construídas a partir de um **DTD**
 - ❑ Não suporta documentos sem **conhecimento** prévio do seu **DTD**
 - ❑ Permite **actualizações parciais**
 - ❑ Não preserva a **ordem** dos elementos
 - Não pode ser aplicada na EIPub
 - ❑ Solução baseada em **middleware**
 - XML-DBMS - <http://www.rpbouret.com/xmldbms/>
 - ❑ As BD mais recentes já suportam o tipo de dados XML



BD Relacional

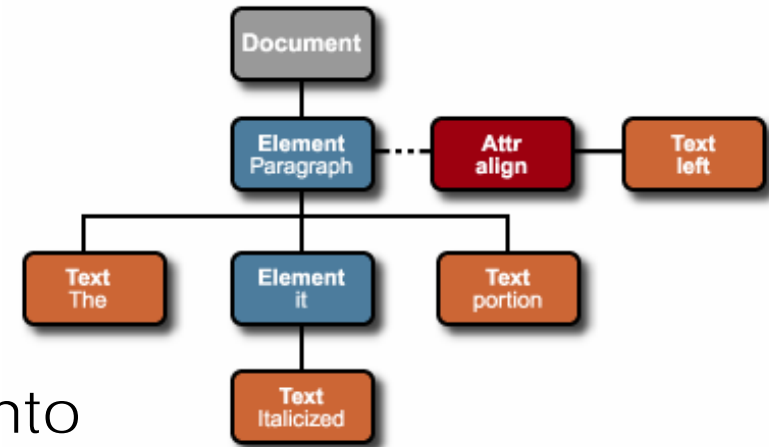
- Bases de dados **XML-Nativas**
 - Suporta **qualquer** tipo de XML
 - Sistema de **interrogação** integrado
 - XPath, XQuery, XUpdate
 - **Indexação** pesada, por vezes lenta
 - **Pesquisas** muito **rápidas**
 - Algumas bases de dados
 - Xindice (<http://xml.apache.org/xindice/>)
 - eXist (<http://exist.sourceforge.net/>)
 - dbXML (<http://www.dbxml.com>)
 - Tamino (<http://www.softwareag.com>)



BD XML-Nativa

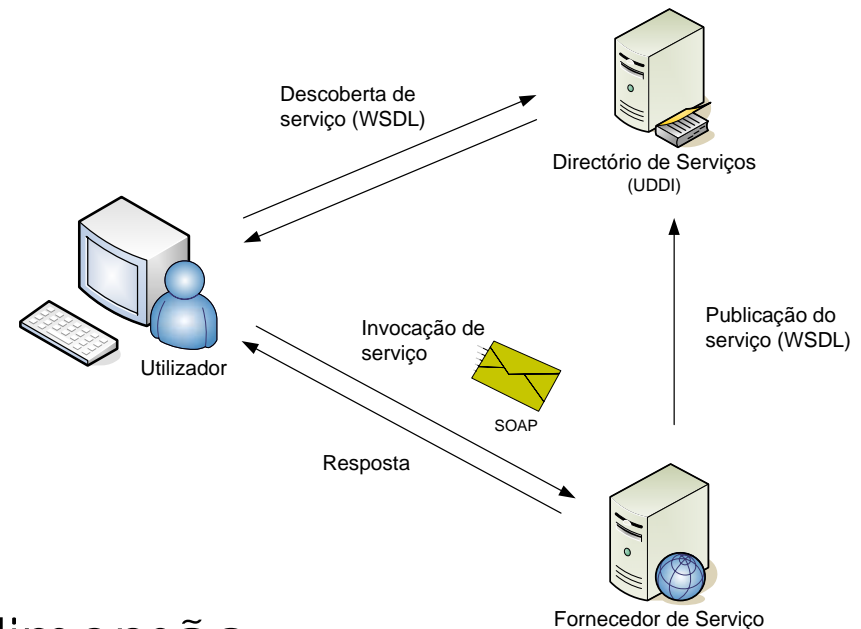
- Document Object Model (DOM)

- Norma da W3C
- Representa um documento XML como uma **árvore** de objectos
- Permite **ler e escrever** XML
- Suporta **validação** (Schema)
- Suporta **transformações XSLT**
- Suporta **XPath**
- Implementações em (quase) **todas as linguagens** de programação
 - Java, .Net, Perl, Python, PHP, Javascript, etc.



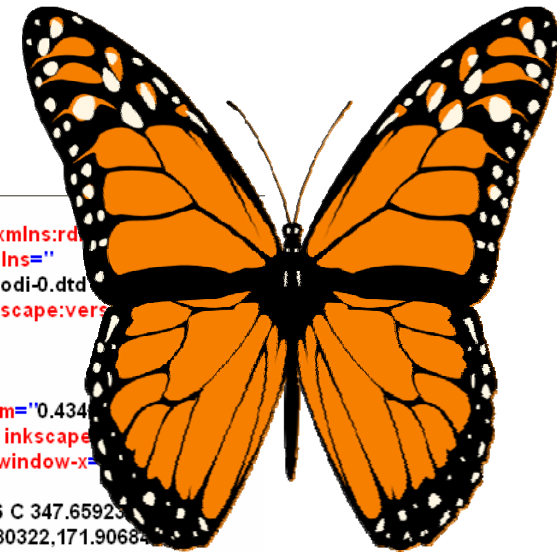
- Simple API for XML (**SAX**)
 - Processamento conduzido por **eventos**
 - e.g. abrir etiqueta, fechar etiqueta, instrução de processamento, ...
 - Processamento muito **rápido**
 - Útil no processamento de **documentos XML grandes** (i.e. da ordem dos Megabytes)
 - **Complexo** de programar
 - Suporta **XSLT**

- Desenvolvimento de **aplicações distribuídas** sobre protocolos Internet
- **Independente da plataforma**
 - e.g. Windows/Linux, Java/Perl, telemóvel/mainframe
- **Protocolos XML abertos**
 - WSDL, SOAP, UDDI, XML
- Funciona sobre **HTTP, ...**
 - Sem reconfiguração de firewalls
- **Reutilização** de serviços
- **Mensagens de grande dimensão**
 - Poderá não ser adequado a todos os cenários de utilização



- Linguagens de anotação para **domínios de aplicação específicos**
- Procuram resolver a **generalidade** dos **problemas** nesse domínio
- Podem ser **grandes** e complexas
 - Dependendo da complexidade da aplicação e do domínio
- Geralmente é necessário "**podar**" a linguagem antes de a utilizar

- **Imagens** vectoriais 2D
- Norma aberta da **W3C**
- Suporta **animação**



```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<svg xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:cc="http://web.resource.org/cc/" xmlns:rd
http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:svg="http://www.w3.org/2000/svg" xmlns="
http://www.w3.org/2000/svg" xmlns:sodipodi="http://inkscape.sourceforge.net/DTD/sodipodi-0.dtd
="http://www.inkscape.org/namespaces/inkscape" id="svg2" sodipodi:version="0.32" inkscape:vers
width="210mm" height="297mm" sodipodi:docname="butterfly.svg" sodipodi:docbase="
/home/knill/web1/Harvard/computing/svg">
  <defs id="defs11">
    <sodipodi:namedview inkscape:cy="467.29273" inkscape:cx="305.25953" inkscape:zoom="0.434
inkscape:window-height="785" inkscape:window-width="983" inkscape:pageshadow="2" inkscape:
0.0" borderopacity="1.0" bordercolor="#666666" pagecolor="#ffffff" id="base" inkscape:window-x=
inkscape:window-y="393" inkscape:current-layer="svg2"/>
    <path style="fill:#f67f00;stroke:none;stroke-width:1.0000000" d="M 362.08873,345.70536 C 347.65923
366.32740,325.14287 366.14703,324.43887 C 340.44448,220.30643 298.99576,174.34148 295.80322,171.90684
285.95509,164.39757 281.19336,165.10156 282.94294,161.31759 C 284.67447,157.53363 296.05575,169.38419
297.60691,171.17351 C 343.15001,223.97306 367.95072,323.70554 367.95072,323.70554 C 367.95072,323.70554
371.10717,326.63886 371.55809,324.43887 C 389.59496,237.90628 425.66870,192.44000 437.84359,174.10682 C
450.01848,155.77365 456.42156,153.98433 439.64727,174.84015 C 387.79127,239.37294 372.92889,323.41221
373.81269,323.70554 C 394.55509,331.03882 374.71454,347.17202 375.61638,345.70536 C 381.06352,336.84677
381.02745,363.30521 381.02745,363.30521 C 381.02745,363.30521 411.69013,260.63942 531.63532,130.10720 C
539.03044,122.06993 797.67918,-107.49078 707.49482,172.64016 C 702.82328,187.18938 671.42108,245.97288
671.42108,245.97288 C 671.42108,245.97288 674.12661,270.90599 663.30448,291.43915 C 661.44669,294.95913
656.99158,329.57217 646.16945,348.63867 C 643.12123,354.03596 646.13338,344.00404 646.16945,348.63867 C
647.07130,436.63792 602.75470,434.70194 605.58650,435.17127 C 622.72153,438.10458 625.17454,474.27227
625.42706,474.77093 C 635.34734,493.83743 631.73996,498.23740 635.34734,507.03733 C 654.82716,554.52758
634.28316,551.68228 637.15102,559.83687 C 644.36577,580.37003 635.63592,588.84729 638.05287,595.03657 C
644.36577,611.16977 633.39936,621.46568 633.54365,625.83631 C 634.44550,650.76943 624.23662,641.67618
624.52521,652.23609 C 625.42706,684.50247 618.21231,671.30259 613.70309,683.03582 C 600.17543,718.23553
594.99885,705.18230 593.86253,709.43560 C 588.45147,729.96876 581.20065,725.04079 578.53119,729.96876 C
572.21829,741.70199 565.90538,741.70199 563.19985,746.10196 C 550.57404,766.63512 543.41340,752.14457
541.55560,753.43522 C 512.69661,773.96838 499.16895,757.36586 494.65974,757.83519 C 481.13208,759.30184
```

GML - Geography Markup Language

XATA'06

8-10 Fevereiro

-
-
-
-

```
<?xml ver  
<wfs:Featu  
<gml:bo  
<gml:  
<gr  
</gml:  
<gml:bc  
<gml:fe  
<my:n  
<gr  
<  
</g  
<my  
<my  
<my  
<my  
<my  
<my  
<my  
<my
```



- Formato XML para partilha de **pautas musicais**

```

<?xml version="1.0" encoding="UTF-8"?>
<score-partwise version="1.1">
  <movement-title>Elite Syncopations</movement-title>
  <identification>
    <creator type="composer"> Scott Joplin</creator>
    <encoding>
      <software>Finale 2005 for Windows</software>
      <software>Dolet 3.0.2 for Finale</software>
      <encoding-date>2005-11-16</encoding-date>
    </encoding>
  </identification>
  <defaults>
    <scaling>
      <millimeters>7.112</millimeters>
      <tenths>40</tenths>
    </scaling>
    <page-layout>
      <page-height>1591</page-height>
      <page-width>1246</page-width>
      <page-margins type="both">
        <left-margin>55</left-margin>
        <right-margin>98</right-margin>
        <top-margin>50</top-margin>
        <bottom-margin>60</bottom-margin>
      </page-margins>
    </page-layout>
    <system-layout>
      <system-margins>
        <left-margin>0</left-margin>

```



Elite Syncopations

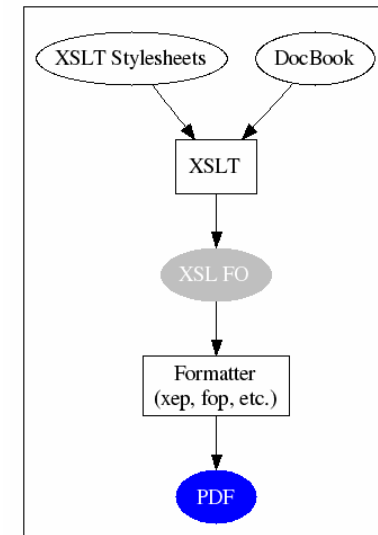
Not fast. Scott Joplin

Piano

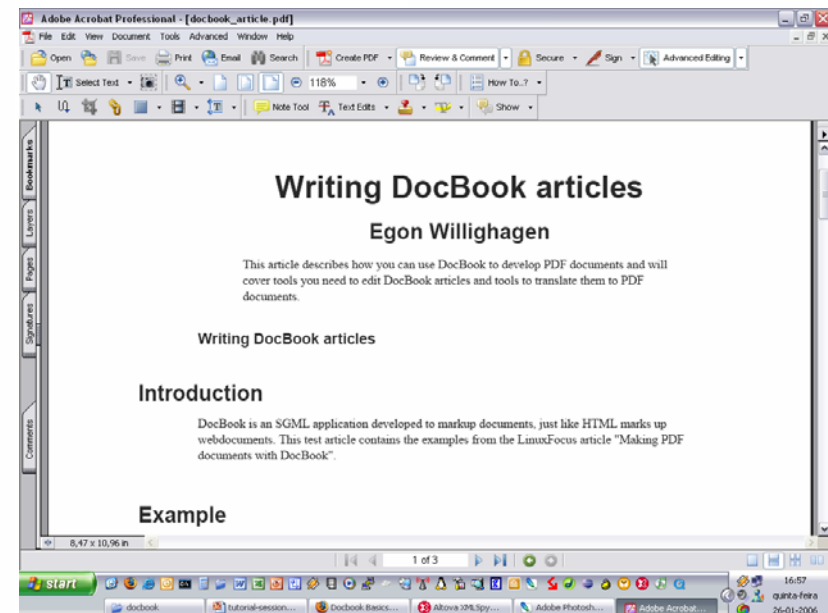
- Dialecto XML para descrever documentos de texto
- Livros, artigos, etc.

```

<?xml version="1.0"?>
<article>
  <title>Writing DocBook articles</title>
  <artheader>
    <abstract>
      This article describes how you can use DocBook to develop
      PDF documents and will cover tools you need to edit DocBook
      articles and tools to translate them to PDF documents.
    </abstract>
    <author>
      <firstname>Egon</firstname>
      <surname>Willighagen</surname>
    </author>
    <date>
    </date>
  </artheader>
  <section>
    <title>Introduction</title>
    <para>
      DocBook is an SGML application
      developed to markup documents, just like HTML marks up webdocuments.
      This test article contains the examples from the LinuxFocus
      article "Making PDF documents with DocBook".
    </para>
  </section>
  <section>
    <title>Example</title>
    <para>Make sure to notice that this document makes recursive use
    of the section element.</para>
  </section>
  <title>Adding examples</title>
  
```



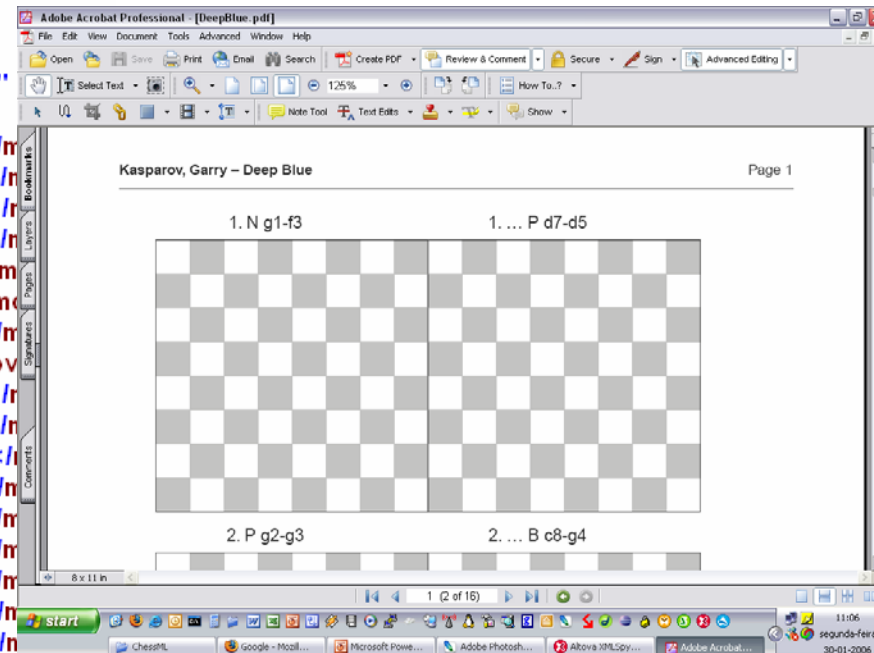
Publishing HTML



- **EAD**
 - Descrição de fundos documentais (arquivos)
- **Dublin Core**
 - Descrição de recursos
- **XMARC**
 - Catálogos bibliográficos
- **METS**
 - Estruturação e transmissão de objectos digitais
- **RDF/OWL e Topic Maps**
 - Mapas de conceitos (i.e. ontologias)
- Muitos **outros...**

- XML para descrever jogadas de xadrez

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<chessgame event="IBM Kasparov vs. Deep Blue Rematch"
  site="New York, NY USA"
  date="1997.05.03"
  round="1"
  white="Kasparov, Garry"
  black="Deep Blue"
  opening="Reti - King's Indian attack, Keres variation"
  result="1-0">
  <move> <white> N g1-f3</white> <black> P d7-d5</black></move>
  <move> <white> P g2-g3</white> <black> B c8-g4</black></move>
  <move> <white> P b2-b3</white> <black> N b8-d7</black></move>
  <move> <white> B c1-b2</white> <black> P e7-e6</black></move>
  <move> <white> B f1-g2</white> <black> N g8-f6</black></move>
  <move> <white> O-O</white> <black> P c7-c6</black></move>
  <move> <white> P d2-d3</white> <black> B f8-d6</black></move>
  <move> <white> N b1-d2</white> <black> O-O</black></move>
  <move> <white> P h2-h3</white> <black> B g4-h5</black></move>
  <move> <white> P e2-e3</white> <black> P h7-h6</black></move>
  <move> <white> Q d1-e1</white> <black> Q d8-a5</black></move>
  <move> <white> P a2-a3</white> <black> B d6-c7</black></move>
  <move> <white> N f3-h4</white> <black> P g7-g5</black></move>
  <move> <white> N h4-f3</white> <black> P e6-e5</black></move>
  <move> <white> P e3-e4</white> <black> R f8-e8</black></move>
  <move> <white> N f3-h2</white> <black> Q a5-b6</black></move>
  <move> <white> Q e1-c1</white> <black> P a7-a5</black></move>
  <move> <white> R f1-e1</white> <black> B c7-d6</black></move>
  <move> <white> N d2-f1</white> <black> P d5-e4 x</black></move>
```



Repositórios de DTD/Schemas

- XML.ORG Registry (<http://www.xml.org/xml/registry.jsp>)
- Repositório de especificações XML (i.e DTD/Schemas)
- Vasto conjunto de normas
- Qualquer pessoa pode publicar uma especificação
- Procurar sempre utilizar DTD já existentes!

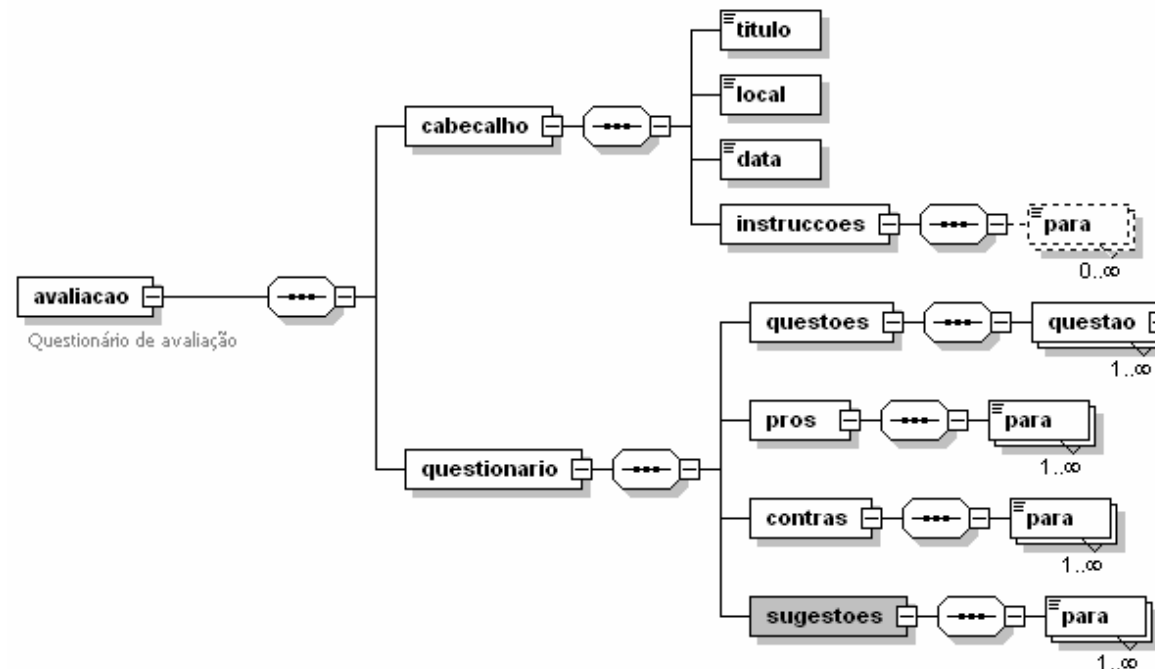
Registered Submissions By Industry		
Accounting (9)	EDI (2)	Public Service (1)
Aerospace (4)	ERP (1)	Publishing/Print (3)
Agriculture (1)	Education (9)	Real Estate (3)
Arts/Entertainment (6)	Energy/Utilities (28)	Robotics/AI (2)
Astronomy (4)	Financial Service (22)	Science (12)
Automotive (4)	Food Services (2)	Security (1)
Banking (1)	Geography (3)	Social Sciences (3)
Biology (7)	Healthcare (9)	Software (11)
Business Reporting (1)	Human Resources (9)	Supply Chain (5)
Business Services (1)	Industrial Control (2)	Telecommunications (8)
Catalogs (1)	Internet/Web (9)	Transportation (4)
Chemistry (4)	Manufacturing (5)	Travel (1)
Construction (2)	Marketing/PR (1)	Weather (1)
Customer Relation (2)	Multimedia (8)	XML Technologies (20)
Databases (2)	News (5)	
E-Commerce (9)	Other Industry (5)	

Alguma questão?

<nota> *antes de passarmos ao exercício?! </nota>*



- Criação de uma interface para o **Authentic** para editar um documento XML.
- Geração de **derivadas** HTML e PDF do documento



■ Armazenamento XML

- Storing XML in Relational Databases
 - <http://www.xml.com/pub/a/2001/06/20/databases.html>
- Introduction to Native XML Databases
 - <http://www.xml.com/pub/a/2001/10/31/nativexmldb.html>
- Ronald Bourret
 - Consulting, writing, and research in XML and databases
 - <http://www.rpbouret.com>

- **Processamento XML**
 - DOM - <http://www.w3.org/DOM/>
 - SAX - <http://www.saxproject.org/>
- **Web services**
 - A Web Services Primer
 - <http://webservices.xml.com/pub/a/ws/2001/04/04/webservices/index.html>
 - W3C Web Services Activity
 - <http://www.w3.org/2002/ws/>

■ Dialectos

- ❑ SVG: <http://www.w3.org/TR/SVG/>
- ❑ GML: <http://opengis.net/gml/>
- ❑ Music & XML: <http://xml.coverpages.org/xmlMusic.html>
- ❑ DocBook: <http://www.docbook.org/>
- ❑ Metainformação
 - Dublin Core: <http://www.dublincore.org>
 - EAD: <http://www.loc.gov/ead>
 - MARC 21:
<http://lcweb.loc.gov/marc/marcdocz.html>
 - MARCXML: <http://www.loc.gov/marc/marcxml.html>
 - METS: <http://www.loc.gov/standards/mets>
 - RDF: <http://www.w3.org/RDF>
- ❑ ChessML: <http://xml.coverpages.org/chessML.html>