

# Introdução à XML

Érica Saito

Serviço de Biblioteca da EEFE-USP



# Pauta

---

O que é XML e sua aplicação

Um pouco de história

Estrutura de um documento XML

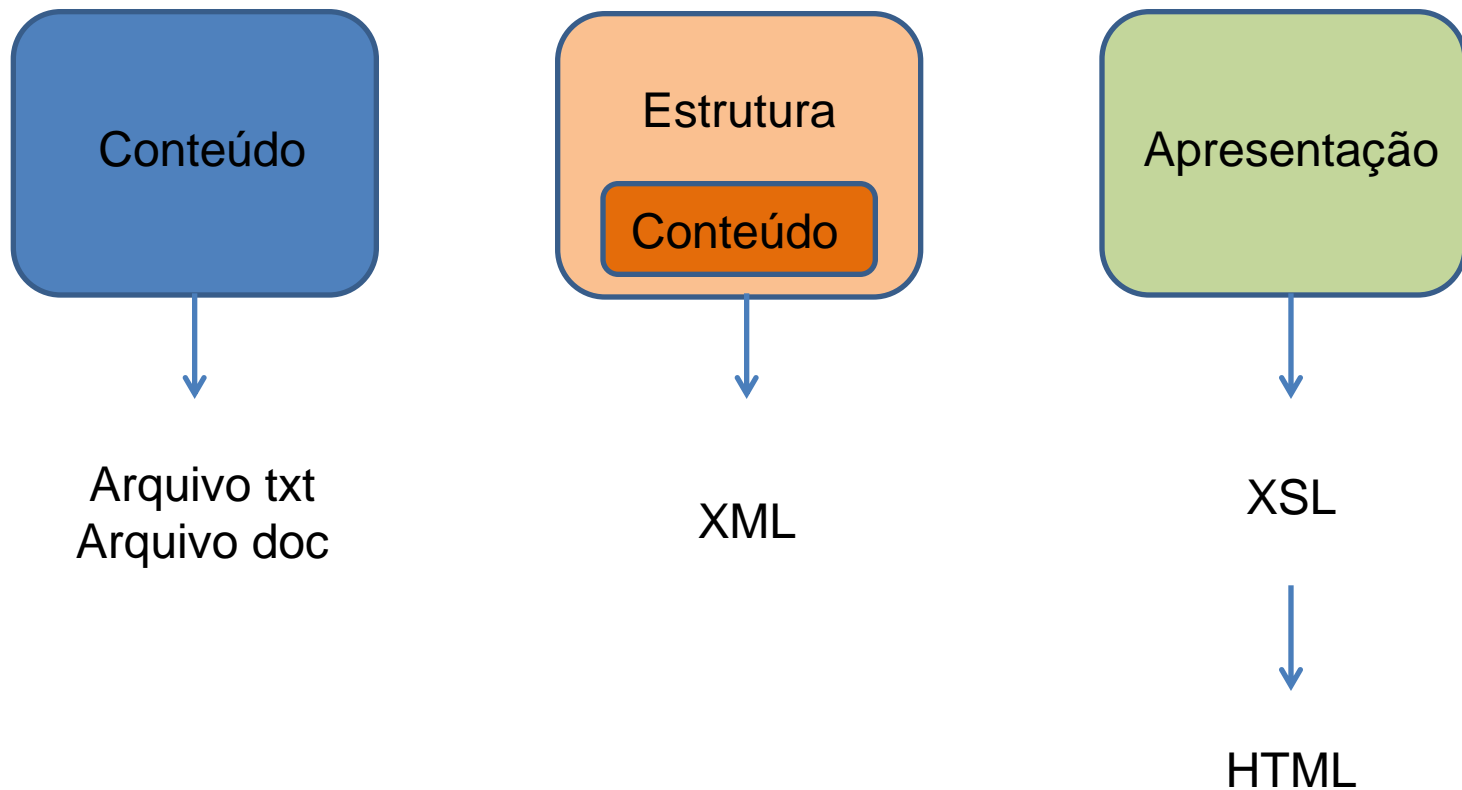
DTD

XML Schema

## O que é XML (EXtensible Markup Language)?

- É uma metalinguagem de marcação;
  - É uma recomendação do W3C;
- Foi desenhado para ser auto-descritivo.

# Entendendo o fluxo



# Importância do XML para as Bibliotecas

---

- Catalogação
- Migração e transferência de dados
- Construção de bases de dados

# Importância do XML para as Revistas

---

- As revistas foram uma das primeiras áreas a utilizarem tecnologias de marcação com sucesso;
- Facilita a disseminação e apresentação das Revistas em formato eletrônico;
- Neutralidade e reutilização de conteúdo.

# **Não confunda...**

XML – Dados estruturados  
Compreensão por máquinas e pessoas

HTML – Exibição do documento

## Um pouco de história

XML é uma versão abreviada de SGML

Jon Bosak (Sun) liderou um grupo entre os anos de 1996-1998 na W3C e desenvolveu o XML.

Algumas linguagens baseadas em XML são:

XHTML

MARXML

EAD Version 2002



# Exemplo XML

---

```
<catalogo>
  <livro>
    <autor> Rowling, Joanne K.</autor>
    <titulo>Harry Potter e as relíquias da morte</titulo>
  </livro>
</catalogo>
```

# TAGs

---

Utilizar os sinais < >

Toda tag precisa ser **fechada**

Forma inicial: <elemento>

Forma final: </elemento>

Forma abreviada: <elemento/>

# Exemplo XML

---

```
<catalogo>
  <livro>
    <autor> Rowling, Joanne K.</autor>
    <titulo>Harry Potter e as relíquias da morte</titulo>
  </livro>
</catalogo>
```

# Elementos

---

Elemento no XML é tudo, desde a tag inicial até a tag final.

O elemento pode conter:

- Outros elementos
- Texto
- Atributos
- Uma mistura das três alternativas anteriores.

# Elementos

---

<raiz>

<pai>

<filho1>Texto</filho1>

<filho2>Texto</filho2>

<filho3>Texto</filho3>

<filho4>Texto</filho4>

</pai>

</raiz>

# Elementos

---

## Exemplo

<nota>

<para>Camila</para>

<de>Erica</de>

<cabecalho>Lembrete</cabecalho>

<corpo>Feijoada no sábado</corpo>

</nota>

# Atributos

---

Os atributos em XML são usados para descrever os elementos XML ou para fornecer uma informação adicional sobre os elementos.

```
<aviso data="12/09/13">  
    <cabecalho>Lembre-se</cabecalho>  
    <corpo>Prova de Matemática</corpo>  
</aviso>
```

# Atributos

---

Exemplo com mais de 1 atributo:

```
<cachorro nome="Haru" id="001"  
tamanho="medio" cor="branco e marrom"  
inteligencia="esperta"/>
```



# Atributos

---

- Sempre utilizar sinal de aspas (simples ou dupla)
- Um elemento pode ter um ou mais atributos
- Se houver mais que um atributo, ele não poderá ser repetido no mesmo elemento
- Podem ser expressos em qualquer ordem dentro do elemento

# Nome dos elementos e atributos

---



- São sensíveis a maiúscula e minúscula (case sensitive);
- Preferencialmente de fácil entendimento humano;
- Nomes podem conter letras, números, e outros caracteres.

## Nome dos elementos e atributos

---



- Nomes não devem começar com um número ou caractere de pontuação;
- Nomes não devem começar com as letras xml (ou XML ou Xml).
- Aconselhamos não usar “.” e “-”;
- Nomes não podem conter espaço.

## Nome dos elementos e atributos

---

<nome completo> </nome completo> ✗

<1comentario> </1comentario> ✗

<xmlfile> </xmlfile> ✗

**Não confunda...**

<L**iv**ro> é diferente de <li**v**ro>

# Documento XML

---

Formado por:

- Prólogo

- ❖Declaração XML

- ❖Instrução de processamento

- ❖Declaração de tipo de documento

- Instância do Documento

# Documento XML

## Exemplo 1 de Documento XML

Prólogo  
Instância do Documento

```
<?xml version="1.0" encoding="UTF-8"?>      → Declaração XML
<?xml-stylesheet type="text/xml" href="catalogo.xsl"?> → Instrução de processamento
<!DOCTYPE catalogo SYSTEM "catalogo.dtd">      → Declaração de Tipo de Documento
<catalogo>
  <libro>
    <autor>Goldfarb, Charles F.</autor>
    <autor>Prescod, Paul</autor>
    <titulo>Charles F. Goldfarb's <ênfasis>XML</ênfasis> Handbook</titulo>
    <edicion numedic="4">
      <lugar>Upper Saddle River (New Jersey)</lugar>
      <editorial>Prentice Hall PTR</editorial>
      <anyo>2002</anyo>
    </edicion>
    <materia>XML</materia>
    <isbn>0-13-065198-2</isbn>
    <notas>Buen estado.<lb/>&noprest;</notas>
  </libro>
  <libro>...</libro>
</catalogo>
```

# Declaração XML

---

É obrigatório:

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



# Instrução de Processamento

---

Por exemplo:

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

# Declaração de tipo de documento

---

Por exemplo:

```
<!DOCTYPE catalogo SYSTEM  
    "catalogo.dtd">
```

```
<!DOCTYPE article SYSTEM "JATS-  
journalpublishing1.dtd" PUBLIC "-  
//NLM//DTD JATS (Z39.96) Journal  
Publishing DTD v1.0 20120330//EN">
```

# DTD

---

DTD (Document Type Definition) é o conjunto de regras que define quais tipos de dados e entidades farão parte de um documento XML.

# Tipos de DTD

---

Pode ser: externa, interna ou mista

Para documentos grandes,  
recomendamos a utilização da DTD  
externa, que é o caso da DTD Pubmed.

# Declaração de tipo de documento

---

Por exemplo:

```
<!DOCTYPE catalogo SYSTEM  
    "catalogo.dtd">
```

```
<!DOCTYPE article SYSTEM "JATS-  
journalpublishing1.dtd" PUBLIC "-  
//NLM//DTD JATS (Z39.96) Journal  
Publishing DTD v1.0 20120330//EN">
```

# DTD PMC

```
JATS-journalpublishing1 - Bloco de notas
Arquivo  Editar  Formatar  Exibir  Ajuda

<!-- ===== -->
<!-- MODULE:   Journal Publishing DTD -->
<!-- VERSION:  ANSI/NISO JATS Version 1.0 (Z39.96-2012) -->
<!-- DATE:    March 2012 -->
<!-- ===== -->

<!-- ===== -->
<!-- PUBLIC DOCUMENT TYPE DEFINITION -->
<!-- TYPICAL INVOCATION -->
<!-- ===== -->
"-//NLM//DTD JATS (Z39.96) Journal Publishing DTD v1.0 20120330//EN"
  Delivered as file "JATS-journalpublishing1.dtd"
  Available at:
  http://jats.nlm.nih.gov/publishing/1.0/JATS-journalpublishing1.dtd
<!-- ===== -->

<!-- ===== -->
<!-- SYSTEM:   Journal Publishing DTD of the JATS DTD Suite -->
<!-- ===== -->
<!-- PURPOSE:  DTD for creation of new journal articles -->
<!-- ===== -->
<!-- The Journal Publishing DTD is an application of -->
<!-- the ANSI/NISO Z39.96 Journal Publishing Tag Set. -->
<!-- It is a subset of the Journal Archiving and -->
<!-- Interchange DTD that is optimized for the -->
<!-- creation or publishing of existing journal -->
<!-- articles and some non-article material such as -->
<!-- product and book reviews in XML. It describes -->
<!-- both the metadata for a journal article and -->
<!-- the full content of the article. -->
<!-- ===== -->
<!-- This DTD was constructed using the modules in the -->
<!-- JATS DTD Suite. -->
<!-- ===== -->
<!-- TAG SET SPONSOR -->
<!-- National Center for Biotechnology -->
<!-- Information (NCBI) -->
<!-- National Library of Medicine (NLM) -->
<!-- ===== -->
```

<http://jats.nlm.nih.gov/publishing/1.0/dtd.html>

# DTD

---

Na DTD encontramos as declarações de:

- Elementos
- Atributos
- Entidades
- Notação de tipos

# DTD

---

- PCDATA  
(Parsed Character Data)

- CDATA  
(Character Data)



# Declaração de (tipos de) elementos

---

Vazio

```
<!ELEMENT element-name EMPTY>
```

**Exemplo**

**DTD**

```
<!ELEMENT capa EMPTY>
```

**Documento XML**

```
<capa href="foto001.jpg"> </capa>
```

# Declaração de (tipos de) elementos

---

Algun conteúdo

```
<!ELEMENT element-name ANY>
```

**Exemplo**

**DTD**

```
<!ELEMENT pessoa ANY>
```

**Documento XML**

```
<pessoa>
```

```
  <c>texto</c>
```

```
  <a>alguma coisa</a>
```

```
</pessoa>
```

# Declaração de (tipos de) elementos

---

Com caracteres analisáveis

```
<!ELEMENT element-name (#PCDATA)>
```

**Exemplo**

**DTD**

```
<!ELEMENT ingrediente (#PCDATA)>
```

**Documento XML**

```
<ingrediente>Farinha</ingrediente>
```

# Declaração de (tipos de) elementos

---

Com elementos filhos

```
<!ELEMENT element-name  
(#PCDATA|elemento|elemento|...)>
```

**Exemplo**

**DTD**

```
<!ELEMENT lista (ingrediente)>
```

**Documento XML**

```
<lista>
```

```
    <ingrediente>Café</ingrediente>
```

```
</lista>
```

# Declaração de atributos > Atributo obrigatório

```
<!ATTLIST element-name attribute-name attribute-type  
#REQUIRED>
```

## **Exemplo**

### **DTD**

```
<!ATTLIST pessoa numero CDATA #REQUIRED>
```

### **XML Válido**

```
<pessoa numero="100" />
```

### **XML Inválido**

```
<pessoa />
```

## Declaração de atributos > Atributo opcional

---

<!ATTLIST autor e-mail CDATA #IMPLIED>

### **Exemplo**

#### **DTD**

<!ATTLIST contato email CDATA #IMPLIED>

#### **XML Válido**

<contato email="contato@contato.com.br" />

#### **XML Válido**

<contato />

## Declaração de atributos > Atributo fixo

---

<!ATTLIST element-name attribute-name attribute-type #FIXED "value">

### **Exemplo**

#### **DTD**

<!ATTLIST nome instituicao CDATA #FIXED "USP">

#### **XML Válido**

<nome instituicao="USP" />

#### **XML Inválido**

<nome instituicao="UNICAMP" />

# Comentários

---

É possível incluir comentários com a utilização de

“<!-- texto -->”

Os comentários não fazem parte do documento.

Exemplo:

<!-- Atencao: Este documento apresenta uma variacao estrutural que nao esta contemplada na DTD -->



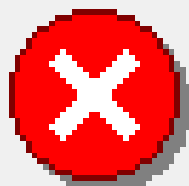
# Validação

---

XML bem formado ou well-formed é aquele que possui a sintaxe correta

Lembrando: precisa ter todas as tags fechadas, elemento raiz, os atributos precisam ter aspas e não confundir minúscula com maiúscula já que as tags são case sensitive

Exemplo:



This file is not well-formed:  
Lengua closing element name expected.

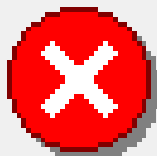
# Validação

---

XML válido tem:

- Sintaxe do XML correta
- Foi validado com a DTD

Exemplo:

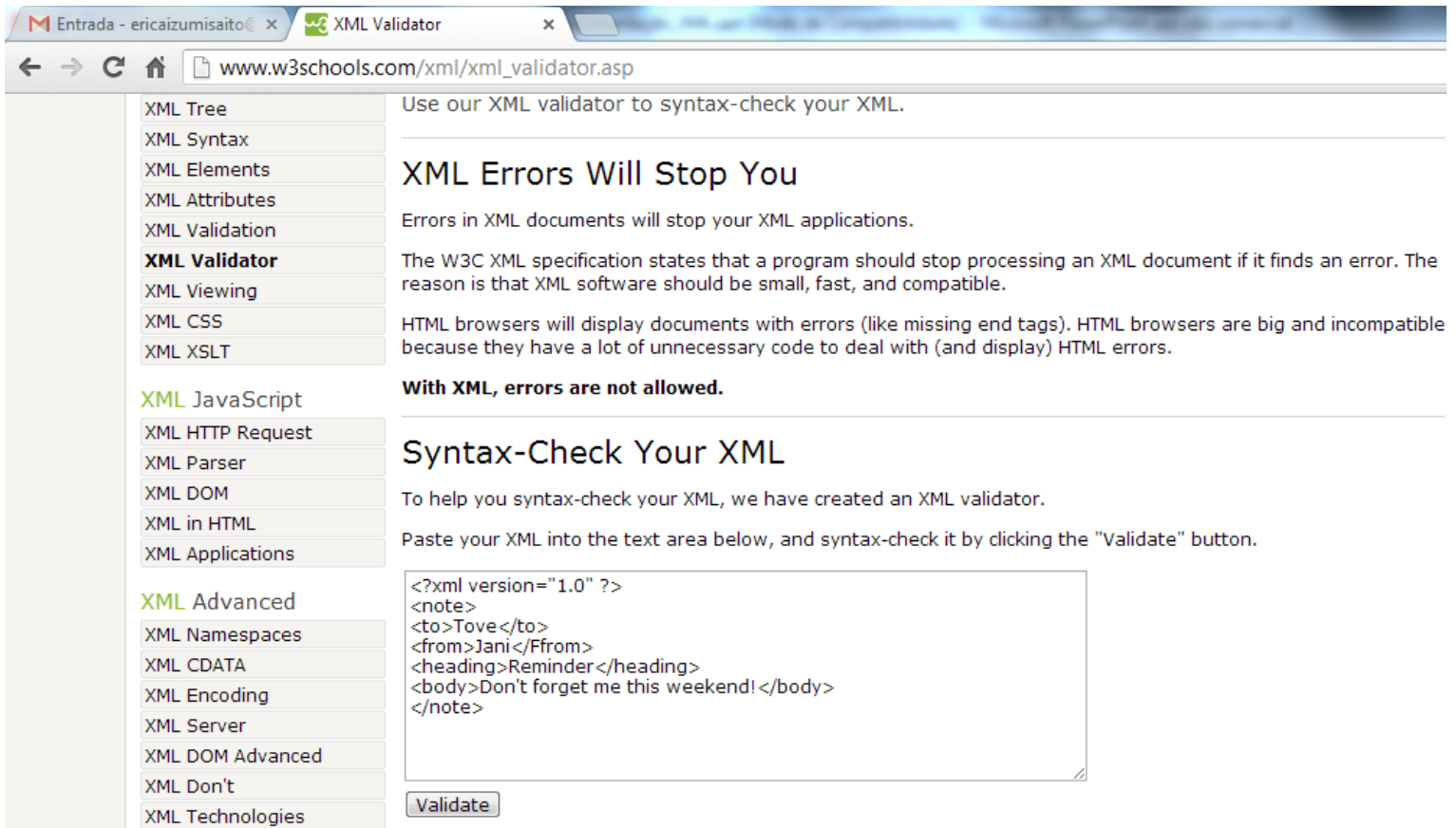


This file is not valid:

Mandatory element 'Lengua' expected in place of 'Lingua'

# XML Validator

[http://www.w3schools.com/xml/xml\\_validator.asp](http://www.w3schools.com/xml/xml_validator.asp)



The screenshot shows a web browser window with the URL `www.w3schools.com/xml/xml_validator.asp`. The page has a left sidebar with a navigation menu and a main content area on the right.

**Navigation Menu (Left Sidebar):**

- XML Tree
- XML Syntax
- XML Elements
- XML Attributes
- XML Validation
- XML Validator** (highlighted)
- XML Viewing
- XML CSS
- XML XSLT
- XML JavaScript
- XML HTTP Request
- XML Parser
- XML DOM
- XML in HTML
- XML Applications
- XML Advanced
- XML Namespaces
- XML CDATA
- XML Encoding
- XML Server
- XML DOM Advanced
- XML Don't
- XML Technologies

**Main Content Area:**

Use our XML validator to syntax-check your XML.

---

## XML Errors Will Stop You

Errors in XML documents will stop your XML applications.

The W3C XML specification states that a program should stop processing an XML document if it finds an error. The reason is that XML software should be small, fast, and compatible.

HTML browsers will display documents with errors (like missing end tags). HTML browsers are big and incompatible because they have a lot of unnecessary code to deal with (and display) HTML errors.

**With XML, errors are not allowed.**

---

## Syntax-Check Your XML

To help you syntax-check your XML, we have created an XML validator.

Paste your XML into the text area below, and syntax-check it by clicking the "Validate" button.

```
<?xml version="1.0" ?>
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```

# XML Schema

---

XML Schema descreve a estrutura de um documento XML. É como se fosse a DTD.

Define:

- Elementos e atributos do documento XML

# XML Schema

---

Exemplo:

```
<?xml version="1.0"?>
<xs:schema
xmlns:xs="http://www.w3.org/2001/XMLSchema">
<xs:element name="message">
<xs:complexType>
<xs:sequence>
<xs:element name="To" type="xs:string" />
<xs:element name="Note" type="xs:string" />
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:schema>
```

## Alguns software comerciais para trabalhar com XML

---

Altova XML Spy  
[www.altova.com](http://www.altova.com)

Liquid XML Studio  
<http://www.liquid-technologies.com/xml-studio.aspx>

<oXygen/> XML Editor  
[www.oxygenxml.com](http://www.oxygenxml.com)

## Alguns software comerciais para trabalhar com XML

---

XMLwriter

<http://xmlwriter.net>

eXtiles

<http://www.inera.com/extyles-products>

## Alguns software gratuitos para trabalhar com XML

---

XML Pad 3

[www.wmhelp.com/xmlpad3.htm](http://www.wmhelp.com/xmlpad3.htm)

Notepad++

<http://notepad-plus-plus.org/>

Eclipse

<http://www.eclipse.org/>



# Revisão e Comparações de Editores de XML

---

## Revisão de Editores de XML

<http://www.cmsreview.com/XML/Editors/>

## Comparações de Editores XML

[http://en.wikipedia.org/wiki/Comparison\\_of\\_XML\\_editors](http://en.wikipedia.org/wiki/Comparison_of_XML_editors)

# Quiz sobre XML

---

20 questões

[http://www.w3schools.com/xml/xml\\_quiz.asp](http://www.w3schools.com/xml/xml_quiz.asp)



**Dúvidas**

**&**

**Comentários**

## Para saber mais

---

BROWN, Alex. XML in serial publishing: past, present and future. **OCCLC Systems & Services**, v. 19, n. 4, p.149-154, 2003.

Disponível em:

<<http://www.emeraldinsight.com/journals.htm?articleid=863213>>.

Acesso em: 25 jul. 2013.

MILLER, Dick R. XML: libraries' strategic opportunity. **Library Journal: net connect**, v. 125, n. 10, 2000. Disponível em: <<http://cds.cern.ch/record/450652/files/ext-2000-173.pdf>>. Acesso em: 25 jul. 2013.

TENNANT, Roy. **XML in libraries**. New York: Neal-Schuman Publishers, 2002.



@bibeefe  
@ reveefe  
@ rbefe

**SAVE THE DATE**

**15/10/2013**

Palestra sobre a  
Internacionalização da  
Formação do Profissional  
Bibliotecário

# Presença Confirmada



## Elisangela Alves Silva

Bibliotecária com graduação pela ECA/USP, mestre em Ciência da Informação pela mesma instituição e especialização em gestão pública. Atualmente trabalha como supervisora de Acervo da Biblioteca Mário de Andrade.

# **Uma Escola, um livro**

1998 - If All of Seattle Read the  
Same Book"

Seattle Public Library

**O Alienista**



# Contato

---

Érica Saito  
[esaito@usp.br](mailto:esaito@usp.br)