

SCC0565 – Sistemas Interativos Web

Prof.: Dr. Rudinei Goularte
(rudinei@icmc.usp.br)

Aula 4 – Introdução a XML

Instituto de Ciências Matemáticas e de Computação - ICMC
Sala 4-229

Baseado no material originalmente produzido por:
Prof. Dra. Renata Pontin e Luciano Pansanato

Roteiro

XML

DTD

2

XML: Linguagem de Marcação Extensível

- W3C: World Wide Web Consortium
– <http://www.w3.org>
- Extensible Markup Language – XML.
 - (XML 1.0) (Fifth Edition), W3C Standard, 26 de Novembro, 2008, <http://www.w3.org/TR/REC-xml/>
 - XML 1.0 fifth edition - 26/11/2008
 - XML 1.1 (Second Edition), W3C Standard, 16/08/2006



3

Documentos e XML



Capítulo 3

Papel de Padrões XML em OHSs

3.1 Considerações Iniciais

As abstrair a entidade "documento", podemos analisá-la sob as seguintes dimensões: conteúdo, estrutura, apresentação, semântica, metadados e malha de hipertexto.¹ Um documento XML representa, basicamente, duas dimensões de informação: o conteúdo propriamente dito e a estrutura organizacional deste conteúdo. A Recomendação XML (Bryer et al., 1998) possibilita a definição de especificações que representam as demais dimensões de informação de um documento.

Paralelamente ao trabalho do W3C (World Wide Web Consortium), desenvolveu-se vários projetos na comunidade de hipertextos aberta (OHSWG, 2001) que visavam entender a infraestrutura da Web com serviços que armazenam as informações de ligação naturalmente ao conteúdo das documentações. Para Case and Hall (1998), OHSs como o OHS Microsoft e o DHM fornecem suporte a dimensões relevantes para a comunidade de hipertextos, aberta sobre serviços de ligação, como:

- Definição das tipos de interface com o usuário para a apresentação de ligações;
 - Variedade de comportamentos relacionados a ligações;
 - Semântica associada a conjuntos de ligações que integram hipertextos, etc.
- ¹ Por exemplo, diferentes tipos de conteúdo: texto, áudio, vídeo, animação, etc. ou a localização física de cada item.
- ² Referências a metadados, índices de busca e filtros, meios de redigir e bilheteiros podem ser considerados como malha de hipertexto sobre documentos, em geral.

- Estrutura
- Apresentação
- Semântica
- Metadados
- Hipertexto

É um documento da Web???

```
<html>
<head>
<title>Exemplo de HTML</title>
<meta name="GENERATOR" content="Microsoft
FrontPage 4.0">
</head>
<body>
<h1>cabecalho</h1></>
<p>parágrafo<b>texto em negrito</b>
<p>texto em novo parágrafo</p>
<a href="http://www.w3.org">W3C</a>
</body>
</html>
```

- Simples ☺
- Fácil aprendizado ☺
- Checagem de sintaxe ☺
- Extensibilidade ☹
- Consciência de conteúdo ☹
- Carência de semântica ☹
- Visa à apresentação ☹
- Intercâmbio de dados ☹
- Conteúdo não-computável ☹
- Pouco reuso ☹

6

Demanda atual de aplicações Web

- Extensibilidade do conjunto de marcadores que
 - ... permitam que autores ou comunidades criem seus **próprios marcadores** para melhor definirem seus documentos de interesse.
 - ... permitam que aplicações possam associar **significado** a dados e campos do documento o que viabilizaria o **processamento** automático dos documentos.
 - ... permitam a construção de aplicações mais apropriadas para dispositivos **portáteis** e de poucos recursos, por exemplo.

7

XML: Linguagem de Marcação Extensível

- É uma linguagem de marcação apropriada à **representação** de dados, de documentos e demais entidades cuja essência se fundamenta na capacidade de **agregar** informações.
- XML é uma **linguagem** ao estabelecer **regras gerais** às quais documentos em **conformidade** com XML devem respeitar.

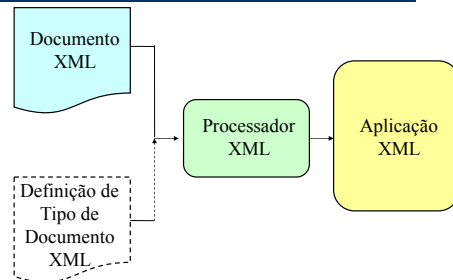
8

Objetivo Principal de XML

- Permitir a especificação de documentos abertos apropriados para intercâmbio entre aplicações na Web.

9

O objetivo principal de XML



10

Dimensões de um documento XML

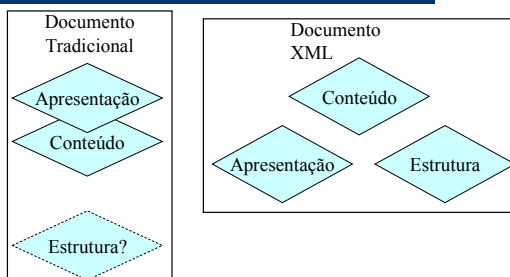
- Conteúdo
- Estrutura

```

<slides>
  <slide name="introdução">
    <p>Introdução AQUI</p>
  </slide>
  <slide name="background">
    <p>background AQUI</p>
  </slide>
  <slide name="resultados">
    <p>resultados AQUI</p>
  </slide>
  <slide name="conclusão">
    <p>Conclusão AQUI</p>
  </slide>
</slides>
  
```

11

Dimensões de um documento XML



12

Dimensões de um documento XML

- Outras dimensões de um documento XML
 - **Estrutura e Semântica**: DTD e Esquema XML
 - **Apresentação**: CSS e XSL (XPath, XSLT e XSL-FO)
 - **Metadados e mais semântica**: RDF e Esquema RDF
 - Estrutura de **hipertexto**: XLink e XPointer
- Processamento de documentos XML
 - Parsers, DOM, SAX, aplicações ..

13 • <http://www.w3.org>

XML: linguagem ou meta-linguagem???

- XML é uma **linguagem**:
 - Estabelece regras gerais às quais documentos em conformidade com XML devem respeitar.
- XML é também uma **meta-linguagem**:
 - Provê recursos para a definição de gramáticas que caracterizam linguagens para classes de documentos específicos com conjunto de elementos, atributos e regras de composição bem determinados.


14

XML como meta-linguagem ...

- **CML**: Linguagem de Marcação Química
 - descrição de fórmulas químicas
- **OFX**: Intercâmbio Financeiro Aberto
 - troca de faturas, recibos, extratos ...
- **WML**: Linguagem de Marcação Sem Fio
 - para dispositivos móveis, como celulares e handhelds
- **MathML**: Linguagem de Marcação Matemática
 - descrição de sentenças e fórmulas matemáticas
- **SVG**: Gráficos Vetoriais Escaláveis
 - ideal para especificação de gráficos vetoriais

15

XML como meta-linguagem ...

- **Esquema XML**: gramática XML com forte tipagem de dados
- **XSLT**: transforma um documento XML em outro documento (XML, HTML, RTF, PS, PDF, ...)
- **RDF**: Framework para Descrição de Recursos
 - descrição de recursos por metadados e **semântica**
- **XHTML**: HTML segundo as regras da especificação XML 
- **XML-RPC**: Chamada Remota a Procedimentos em XML
 - protocolo de comunicação em ambientes distribuídos
- **SOAP**: Protocolo de Acesso Simples a Objetos
 - protocolo de comunicação em ambientes distribuídos
- **SMIL**: Linguagem de Integração de Multimídia Sincronizada
 - especificação de apresentações multimídia

16

Sintaxe e Validação XML

Documentos XML

- **Conceitos básicos**
 - Documentos XML são textos formados por caracteres do conjunto Unicode, contendo caracteres de dado e informações de marcação explicitamente separados
 - **Informações de marcação**
 - Comentários, referências a caracteres e entidades, delimitadores de seção CDATA, elementos, instruções de processamento
 - Definições de Tipo de Documento (DTDs)

18

19

Declarações

Instrução de Processamento

DTD

Comentários

Elementos

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE thesis [
  <!ENTITY chapter1 SYSTEM "chap1.xml">
  <!ENTITY chapter7 SYSTEM "chap7.xml">
  <!ELEMENT thesis (abstract,chapter+)>
  <!ATTLIST thesis author CDATA #REQUIRED">
  <!ELEMENT abstract (#PCDATA)>
  <!ELEMENT chapter (title,body)>
  <!ELEMENT title (#PCDATA)>
  <!ELEMENT body (#PCDATA)>
]>
<thesis>
  <abstract>Nesta tese apresentamos ...</abstract>
  <chapter1>
    <!-- falta pouco ---->
  <chapter7>
  </thesis>
```

20

Conjuntos de caracteres aceitos

- ASCII
 - 1 byte, 7 bits → 128 combinações
- ISO 8859-1 Latin-1
 - 1 byte, 8 bits → 256 combinações (ASCII + caracteres para maioria das línguas da Europa Ocidental – **inclusive Português**)
- ISO 8859-(2...15)
 - 1 byte, 8 bits → 256 combinações (ASCII + caracteres para outros conjuntos de línguas)

21

Unicode: conjunto padrão para XML

- 2 bytes → 65534 combinações
- Atualmente pouco mais de 40000 utilizadas
- 0-127 Latin Básico – ASCII (inglês USA)
- 128-255 Latin-1 **português**, inglês UK, francês, italiano, espanhol, etc.
- ...
- 19966-40959 Ideogramas – chinês, japonês, etc.
- 57344-63743 Uso privado por desenvolvedores

22

UTF-8

- Versão compacta do Unicode
- Utiliza apenas 1 byte p/ maioria dos caracteres
- ... ao custo de usar 3 p/ os menos comuns
- p/ inglês → redução de +- 50%

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

23

Regras para documentos XML

- Nomes dos elementos são sensíveis à caixa


```
<mensagem> ≠ <Mensagem> ≠ <MENSAGEM>
```
- Nomes devem começar com letra ou com _ ; o restante pode incluir letras, dígitos, hífens ou o caracter _ , mas **nunca** o caracter de espaço ou com as letras "xml".


```
<slide_1> ☺ <1slide> ☹ <slide1> ☺ <sl de1> ☹
```
- Utilizar atributo **encoding** em instruções de processamento

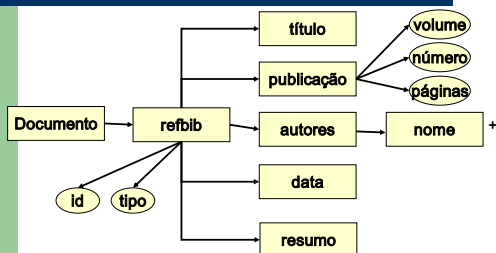

```
<?XML version="1.0" encoding="ISO-8859-1"?>
```

24

Estrutura Física de Documentos XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE bibref SYSTEM "bibref.dtd">
<refbib id="Bul" tipo="Periódico">
  <titulo>Padrões XML e Sistemas Hipermedia Abertos</titulo>
  <publicação volume="2" número="4" páginas="29-66">
    Cadernos de Computação</publicação>
  <autores>
    <nome>R.F. Bulcão Neto</nome>
    <nome>M.G.C. Pimentel</nome>
  </autores>
  <data>16 de Março de 2009</data>
  <resumo>
    A linguagem XML tem papel fundamental nas pesquisas em SHAs ...
  </resumo>
</refbib>
```

Estrutura lógica de documentos XML



25

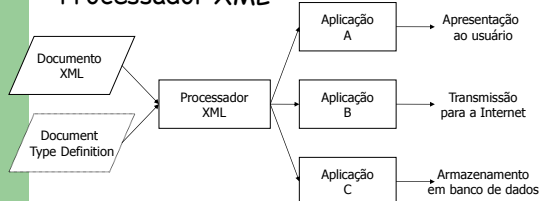
Documentos XML

- Uma aplicação que utilize documentos XML deve **processar** os documentos e **verificar** se seu conteúdo está de acordo com as regras de formação de um documento XML em geral (documento bem formado) e, se for o caso **validar sua estrutura** e conteúdo frente à gramática correspondente, definida na **DTD** ou **Esquema XML** (documento válido).

26

Documentos XML

Processador XML



27

Documentos XML

Processador XML

- Processadores Não-validadores
 - Verificam se um documento é "bem-formado" (se está de acordo com as regras gerais de XML que se aplicam a quaisquer documentos).
- Processadores Validadores
 - Verificam se um documento é bem formado e se o documento está em conformidade com a DTD que define sua gramática.

28

Notação e Especificação XML

- Notação básica e estruturas lógicas
 - Um documento XML é constituído de marcações (*markup*) e de *character data*
 - *Markup*: *start-tags*, *end-tags*, *emph-element tags*, referências a entidades, comentários, seções *CDATA*, declarações de tipo de documento e instruções de processamento.
 - *Character data*: todo texto que não é marcação constitui o *character data* do documento.

29

Notação e Especificação XML

- Elemento
 - Construtor básico de um documento.
 - De maneira geral, um elemento é qualquer cadeia de caracteres que aparece entre os caracteres delimitadores `<` e `>`, desde que não esteja contido em um comentário ou em uma seção *CDATA*.
 - Pode conter outros elementos, outras marcações (comment, PI, entity references, etc).

30

Notação e Especificação XML

• Elemento

- Cada elemento tem um tipo, identificado por um nome, e pode ter um conjunto de especificações de atributos associados.

Diagram illustrating XML element structure:

```

    start-tag  conteúdo  end-tag
    <compositor>Noel Rosa</compositor>
    <version status="draft" target="private"/>
  
```

Labels: empty-element tag, start-tag, conteúdo, end-tag.

31

Notação e Especificação XML

• Elementos XML são extensíveis.

- Documentos XML podem ser estendidos para carregarem mais informações.

```

<note>
  <to> Tover </to>
  <from> Jani </from>
  <body> Don't forget me this weekend</body>
  <heading> Reminder </heading>
</note>
  
```

32

Notação e Especificação XML

Diagram illustrating XML element structure:

```

<book>
  <title> My first XML </title>
  <prod id="123" media="paper"/>
  <chapter> Introduction to XML
    <para> What is XML? </para>
  </chapter>
</book>
  
```

Annotations: Book é o elemento pai de title, prod e chapter. title, prod e chapter são elementos irmãos.

33

Notação e Especificação XML

• Elementos XML possuem conteúdo.

- Conteúdo: elemento, texto, ou texto.
- Conteúdo: simple ou text
- Conteúdo: empty
- Conteúdo: mixed

```

<book>
  <title> My first XML </title>
  <prod id="123" media="paper"/>
  <chapter> Introduction to XML
    <para> What is XML? </para>
  </chapter>
</book>
  
```

34

Notação e Especificação XML

• Atributo

- Um atributo é um par (nome, valor) presente na start-tag do elemento, logo após seu nome

Diagram illustrating XML attribute structure:

```

  atributo  atributo  atributo
  <publication volume="2" number="4" pages="29-66">
    World Wide Web Journal</publication>
  
```

Os valores dos atributos devem estar entre aspas.
Um atributo não pode aparecer mais de uma vez no mesmo elemento.

35

Notação e Especificação XML

• Entidades

- Uma referência a uma entidade é da forma *&nome da entidade*
- Entidade externa: &chapter1;
- Entidade interna: *Graça* representa Graça
- *x <*; *y* representa *x < y*

&	&
<	<
>	>

36

Notação e Especificação XML

• CDATA

- Uma seção CDATA permite a inclusão de trechos que devem ser interpretados como caracteres e não como elementos de marcação

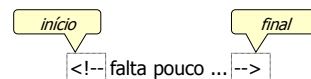
```
<![CDATA[ *p = &q; b = (i <= 3); ]]>
```

37

Notação e Especificação XML

• Comentários

- Não fazem parte do conteúdo do documento



38

Notação e Especificação XML

• Declaração XML

- Documentos XML podem, e devem, começar com uma declaração XML

```
<?xml version = "1.0" ?>
```

39

Notação e Especificação XML

• Instruções de processamento

- Os parsers XML apenas repassam essas informações para a aplicação



40

Notação e Especificação XML

• Documento bem formado

- A instrução de processamento deve literalmente iniciar o documento.
- Todo documento deve ter um elemento (raiz) que inclui todos os demais.
- O documento XML deve ter pelo menos 1 elemento.
- Não é permitido desrespeitar a estrutura de aninhamento de elementos.
- Elementos devem ter tags inicial e final.
- As tags são case-sensitive.
- Os atributos devem ser valorados e os valores devem estar entre aspas.
- Utilize < e & apenas para iniciar uma tag e entidade.

• Bem formado x válido

41

DTD

• Document Type Definition

- Corresponde à declaração que define a gramática do documento

• Palavras reservadas:

- As palavras reservadas devem aparecer sempre em letra maiúscula

- ☐ ELEMENT
- ☐ ATTLIST
- ☐ CDATA
- ☐ PCDATA
- ☐ NOTATION
- ☐ SYSTEM

42

DTD

- **Document Type Definition**
- **ELEMENT**
- **ATTLIST**
- **PCDATA**
- **CDATA**
- **NOTATION**
- **SYSTEM**
- os nomes de elementos, atributos, entidades e notações
 - devem começar com uma letra, underscore ("_") ou dois-pontos (":"); seguido de letras, dígitos ou dos sinais "_", "-", ".", mas não podem começar com as letras "xml"

43

DTD

- **Declaração de Tipo de Documento**
- Indica que o documento deve seguir um DTD
- DTDs internos

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE HELLO [ <!ELEMENT hello (#PCDATA)>]>
<hello>
  Olá XML!
</hello>
```

44

DTD

- **Declaração de Tipo de Documento**
- Indica que o documento deve seguir um DTD
- DTDs externos

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE HELLO SYSTEM "hello.dtd">

<hello>
  Olá XML!
</hello>
```

45

DTD

- **Declaração de elementos**
- Identifica o nome do elemento e a natureza de seu conteúdo
- **#PCDATA**
 - *Parseable Character Data*
- **EMPTY**

	seqüência (and)
	escolha (ou)
?	opcional
*	zero ou mais filhas
+	um ou mais filhas
()	agrupamento

```
<!ELEMENT thesis (abstract, chapter+)>
<!ELEMENT abstract (#PCDATA)>
<!ELEMENT chapter (#PCDATA)>
```

46

DTD

- **Declaração de atributos**
- Nome do atributo, tipo de dados e valor padrão
- Tipo de dados:

- Valor

#REQUIRED	obrigatório
#IMPLIED	opcional
#FIXED <value>	fixo
<default value>	padrão

CDATA	Texto (Character Data)
ID	Um nome exclusivo no documento
IDREF	Uma referência para outro elemento
IDREFS	Diversas referências separadas por espaços
ENTITY	Uma entidade declarada
ENTITIES	Diversas entidades separadas por espaços
NMTOKEN	Um XML name token
NMTOKENS	Diversos XML name tokens
NOTATION	Um nome de notação declarada

```
<!ATTLIST thesis author CDATA #REQUIRED>
```

47

DTD

Exemplo (na Lousa)

- DTD

```
<!ELEMENT film (title, director, actor+)>
<!ATTLIST film code ID #REQUIRED>
<!ATTLIST film type (DVD|VHS) #REQUIRED>
<!ELEMENT title (#PCDATA)>
<!ELEMENT director (#PCDATA)>
<!ELEMENT actor (#PCDATA)>
```


DTD

```
<film code="A010" type="DVD">
  <title>The Lord of the Rings - The Two Towers</title>
  <director>Peter Jackson</director>
  <actor>Elijah Wood</actor>
  <actor>Ian McKellen</actor>
</film>
```

DTD

- **Limitações**
- A verificação de que os valores contidos no documento correspondem ao esperado é responsabilidade da aplicação
- Por exemplo, não existem mecanismos para especificar que:
 - Um elemento <data> deve conter apenas valores válidos para dia-mês-ano
 - Valores de <custo> devem conter apenas valores positivos em alguma moeda corrente
 - Um elemento <titulo> deve conter entre 1 e 255 caracteres

50

DTD

- **Limitações**
- Utiliza uma sintaxe especial que não é parecida com a sintaxe XML usual
- Extensibilidade e escalabilidade
- Compatibilidade com namespaces
- Outras limitações, ??

51

Bibliografia

- W3C:
 - <http://www.w3.org/TR/2006/REC-xml11-20060816/>
 - <http://www.w3schools.com/xml/default.asp>
- "XML: How to Program". Harvey M. Deitel, Paul J. Deitel, Tem R. Nieto, Ted Lin, Praveen Sadhu.

Tarefa

- Individual: desenvolver um arquivo XML para representar/estruturar/armazenar as informações do formulário HTML para cadastro e compra de passagem aérea.
- DTD para validar o XML.
- A modelagem é importante!

Por enquanto é isso!!

Esses slides têm sido elaborados conjuntamente pelos profs. do grupo de pesquisa "Sistemas Web e Multimídia Interativos"