

Unidade III

XML – DTD XML Schema

PUC Minas
Curso Ciência da Computação
Bancos de Dados
Profº. Palhares – setembro 2017

Referências Bibliográficas

ELMASRI, R.; NAVATHE, S. B. Sistemas de banco de dados: Fundamentos e aplicações. 4ª Edição LTC, 2005. cap. 26.

PUCMinas - Bancos de Dados - Profº. Palhares

1. Introdução

- Um dos objetivos da área de BD é a gerência de dados (inclusão, manutenção, preservar integridades, recuperação dos dados)
- A Web não é um banco de dados mas é um desafio para os pesquisadores da área de banco de dados tratar os dados nela dispostos

PUCMinas - Bancos de Dados - Profº. Palhares

1. Introdução

- Linguagens de marcação de textos sempre existiram
- característica básica a capacidade de descrever marcadores inseridos em um documento de texto, definindo o significado que estes marcadores (tags) conferem a um texto associado
- Um marco importante: a HTML (*HyperText Markup Language*) (1990)

PUCMinas - Bancos de Dados - Profº. Palhares

1. Introdução

- A HTML possibilitou o sucesso da Web e o seu crescimento.
- O desenvolvimento da HTML trouxe consigo um aumento em sua complexidade tornando difícil a representação de conteúdos através de seus recursos
- A XML (**eX**tensible **M**arkup **L**anguage) foi criada no intuito de cobrir essas limitações e, principalmente, proporcionar a troca de dados na Web.
- Ela possibilitou uma flexibilização da descrição de conteúdos em documentos.

PUCMinas - Bancos de Dados - Profº. Palhares

1. Introdução

- Ao contrário da HTML, não existe na XML preocupação nenhuma com a apresentação.
- O foco é o conteúdo dos dados.
- Sua grande contribuição é a flexibilidade permitida na criação da estrutura do documento.

PUCMinas - Bancos de Dados - Profº. Palhares

1. Introdução

- A flexibilização da XML trouxe problemas ligados ao armazenamento e recuperação de dados.
- Existem linguagens de consultas a dados XML (.....)
- Um caminho para vencer tais dificuldades é transformar dados XML em estruturas de um banco de dados relacional .
- o processo inverso também nos é permitido.

PUCMinas - Bancos de Dados - Profº. Palhares

1. Introdução

- A gerência de dados inclui, também, a garantia da integridade.
- Gerenciar dados em ambiente XML é um desafio para a área de bancos de dados.

PUCMinas - Bancos de Dados - Profº. Palhares

2. XML

- A XML (*eXtensible Markup Language*) é um subconjunto da SGML (*Standard Generalized Markup Language*).
- foi criada (1996) com objetivo principal de fornecer aos desenvolvedores da Web maneiras de definir e criar seus próprios marcadores e atributos.

PUCMinas - Bancos de Dados - Profº. Palhares

2. XML - vantagens em relação a HTML

- a HTML trabalha com marcadores definidos pela própria linguagem, a XML possibilita a criação de marcadores conforme a necessidade do usuário.
- a pesquisa na HTML é feita de forma mais bruta, na XML pode-se consultar pelo nome do marcador.
- na HTML a visualização é fixada, na XML um mesmo documento pode ser visualizado de forma diferentes.

PUCMinas - Bancos de Dados - Profº. Palhares

2. XML - vantagens em relação a HTML

- a estrutura do documento XML permite que ferramentas baseadas em banco de dados possam processar e consultar seu conteúdo.
- A XML oferece facilidades tanto para os que definem o documento como para os recebem ou trabalham com ele.

PUCMinas - Bancos de Dados - Profº. Palhares

2. XML - vantagens em relação a HTML

- a XML tende a ser um padrão permanente.
- a XML possibilita troca de dados entre aplicativos, proporcionando maior interoperabilidade.

PUCMinas - Bancos de Dados - Profº. Palhares

2. XML - utilização

- Busca de dados na Web
- 'automação da Web'
- troca de dados
- utilizada em conjunto com a HTML
- catálogo
- automação de grandes redes (bancos, suprimentos, aplicações administrativas, etc.)

PUCMinas - Bancos de Dados - Prof^o. Palhares

2. XML

```
<?xml version="1.0"?>
<evento>
  <ano>2002</ano>
  <mês>mês</mês>
  <mês-evento>setembro</mês-evento>
  <seminário>Semana de Pós Graduação</seminários>
  <dia>04</dia>
  <sessão>Bancos de Dados</sessão>
  <convidado>
    <trabalho>
      <autor>.....</autor>
      <autor>...</autor>
    </trabalho>
    <trabalho>
      <autor>.....</autor>
      <autor>...</autor>
    </trabalho>
  </convidado>
</evento>
```

PUCMinas - Bancos de Dados - Prof^o. Palhares

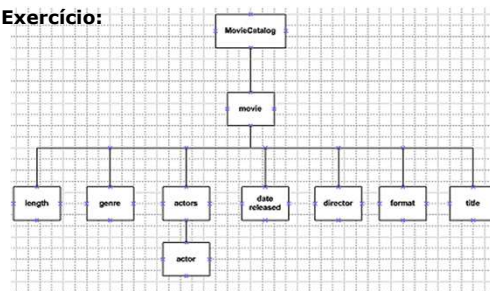
3. Conceitos em XML

- Documento bem formado: obedece ordem de abertura de tags para fechamento das mesmas
- Documento válido: documento bem formado e de acordo com um esquema definido

PUCMinas - Bancos de Dados - Prof^o. Palhares

3. Conceitos em XML

Exercício:



PUCMinas - Bancos de Dados - Prof^o. Palhares

4. DTD

- DTD (*Document Type Definition*) é um documento tipo texto que contém as regras para a elaboração de um documento XML.
- sua função é definir os marcadores que um documento XML pode conter, determinando sua ordem, se são obrigatórios ou opcionais, se são multivalorados ou não, se possuem valor único, etc..
- Podem ser externos ou internos a um documento XML
- Documentos XML são formados, basicamente, por elementos e atributos. Uma DTD suporta a definição desses dois conceitos assim como a relação existente entre os elementos contidos nos documentos e os atributos referentes a esses elementos.

PUCMinas - Bancos de Dados - Prof^o. Palhares

4. DTD

Declarações básicas contidas em uma DTD

Declaração	Significado
ELEMENT	Declaração de um tipo de elemento XML
ATTLIST	Declaração dos atributos associados a um elemento específico, assim como dos valores possíveis para estes atributos
ENTITY	Declaração de conteúdo fixo e reutilizável
NOTATION	Declaração para utilização no processamento do documento XML

PUCMinas - Bancos de Dados - Prof^o. Palhares

4. DTD - Elementos

Declaração de tipo de elemento	
Declaração do elemento	<!ELEMENT S Nome S Especificação de Conteúdo S? '>'
Especificação de Conteúdo	'EMPTY' 'ANY' Mixed children

PUCMinas - Bancos de Dados - Prof^o. Palhares

4. DTD - Elementos

Onde:

- S - significa espaço
- Nome - Corresponde ao nome do elemento. Existem regras para um nome de elemento;
- Especificação de Conteúdo;
- EMPTY (Vazio) - o elemento não possui nem texto nem elementos a ele associados;
- MIXED (Misto) - o elemento possui texto e elementos a ele associados;
- ANY (Qualquer) - o elemento pode conter qualquer combinação de elementos e texto;
- *children* (elementos filhos) - o elemento é composto de outros elementos e não possui texto a ele associado. Refere-se a estes elementos como Subelementos e os elementos que possuem subelementos serão chamados de Grupo.

PUCMinas - Bancos de Dados - Prof^o. Palhares

4. DTD - Elementos

■ Cardinalidades dos elementos

Símbolos de cardinalidade	Significado
?	O elemento contém no máximo uma ocorrência
*	O elemento contém nenhuma ou diversas ocorrências
+	O elemento contém uma ou diversas ocorrências

PUCMinas - Bancos de Dados - Prof^o. Palhares

4. DTD - Elementos

- O símbolo "|" é empregado para simbolizar elementos alternativos.
- A declaração <!ELEMENT a (b | c)>, indica que o elemento *a* é do tipo grupo e que este é formado pelo elemento *b* ou pelo elemento *c*, mas nunca por ambos.
- Uma declaração de conteúdo de um elemento é feita pela combinação de elementos, com ou sem símbolos de cardinalidade e de opcionalidade, aninhados através de parênteses. A ordem como os elementos devem aparecer no documento XML é uma restrição encontrada em XML e é representada na DTD pela vírgula (",") existente entre os elementos.

PUCMinas - Bancos de Dados - Prof^o. Palhares

4. DTD - Atributos

- A declaração de atributos complementa a declaração de um elemento. Um elemento pode ter diversos atributos associados a ele. Um nome de atributo deve ser único para cada elemento.

Declaração de tipo de atributo	
Declaração do atributo	'<!ATTLIST S NomeElemento EspecificaçãoAtributo S? '>'
Especificação do Atributo	S NomeAtributo S TipoAtributo S ValorPadrão

PUCMinas - Bancos de Dados - Prof^o. Palhares

4. DTD - Atributos

Onde:

- S - significa espaço
- Nome - Corresponde ao nome do elemento ao qual pertence o atributo
- Especificação Atributo
- NomeAtributo - Corresponde ao nome do atributo e deve ser único para cada elemento
- TipoAtributo - Corresponde ao tipo de atributo.

PUCMinas - Bancos de Dados - Prof^o. Palhares

4. DTD - Atributos

- CDATA: corresponde a dados com formato de caracteres que podem ser separados por branco. Não pode conter caractere de marcação.
- ID: corresponde ao identificador do elemento. Possui valor único para um mesmo elemento em todo um documento. Cada elemento pode ter apenas um atributo do tipo ID.
- IDREF: corresponde a uma referência a um atributo do tipo ID no documento, ou seja, o valor do atributo refere-se ao valor de algum atributo do tipo ID no documento e esta referência será a mesma para todas as ocorrências deste atributo.
- IDREFS: corresponde a um conjunto de valores que fazem referência a um atributo do tipo ID no documento, ou seja, cada valor do atributo refere-se a um valor de um mesmo atributo do tipo ID no documento. Esta referência será a mesma para todas as ocorrências deste atributo.

PUCMinas - Bancos de Dados - Prof^o. Palhares

4. DTD - Atributos

- NMTOKEN: corresponde a uma palavra, ou um conjunto de caracteres sem espaço em branco.
- NMTOKENS: corresponde a um conjunto de palavra, ou de caracteres, delimitados por valores em branco.
- Valores Enumerados: corresponde a uma lista de valores possíveis para o atributo. Definem o domínio discreto do atributo.

PUCMinas - Bancos de Dados - Prof^o. Palhares

5. Comparando restrições XML - Relacional

- Restrição de Domínio no XML
 - #REQUIRED, #IMPLIED, #FIXED
 - Valores enumerados
- <!ELEMENT convidado #IMPLIED>
<!ATTLIST mês mês-evento (março | setembro) #REQUIRED>
- Restrição de Domínio no modelo relacional
 - NOT NULL, NULL, default
 - check
- ... convidado char(30),
mês-evento char(8) not null check value in ('março', 'setembro'),
...

PUCMinas - Bancos de Dados - Prof^o. Palhares

5. Comparando restrições XML - Relacional

- <!ELEMENT ARTIGO (título, autor*, referência+, revista?>
- Restrição de Cardinalidade no XML
 - um único valor, '*', '+', '?'
 - (1,1), (0,N), (1,N), (0,1)
 - Restrição de Cardinalidade no modelo relacional
 - sempre um único elemento (NOT NULL)
 - zero ou um elemento (NULL)
 - (0,N) → nova relação, com elemento e sub-elemento
 - (1,N) → nova relação, com elemento e sub-elemento, e trigger para garantir a existência de pelo menos um

PUCMinas - Bancos de Dados - Prof^o. Palhares

5. Comparando restrições XML - Relacional

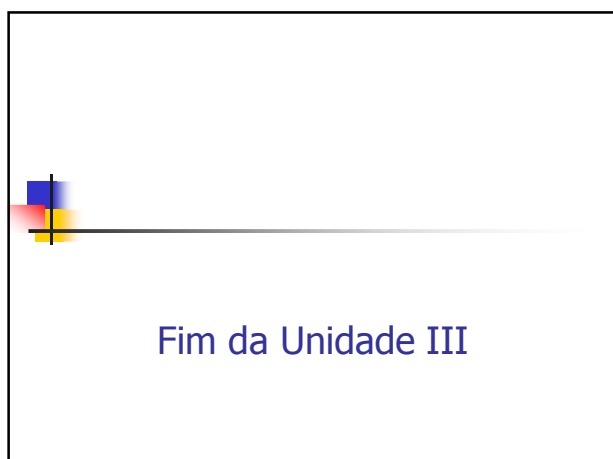
- Restrição de Dependência de inclusão no XML
 - IDREF → ID
 - IDREFS → ID
- Restrição de Integridade Referencial no modelo relacional
 - FK → PK
- Dependência de Inclusão ~ Integridade Referencial

PUCMinas - Bancos de Dados - Prof^o. Palhares

5. Comparando restrições XML - Relacional

- Escolha de subelementos em XML
<!ELEMENT x (y|z)>
- No modelo relacional
Create table x
(y varchar(20),
z varchar(20),
check ((y is NOT NULL and z is NULL) or (y is NULL and z is NOT NULL)))

PUCMinas - Bancos de Dados - Prof^o. Palhares



Fim da Unidade III