

Aula 03 - Principais componentes do XML

Vantagens XML

- Simplifica a troca de dados entre sistemas
- Possibilita codificação mais fácil
- Possibilita buscas mais inteligentes

TAG XML

É o texto que está entre "<" e ">"

Pode ser de abertura "<tag>" ou de fechamento "</tag>"

ELEMENTO XML

É composto pelas tags de abertura, de fechamento e tudo que esta entre elas.

ATRIBUTO XML

É um par nome-valor dentro de uma tag de abertura de um elemento

Regras de Documentos XML

A especificação XML requer que um parser rejeite documentos que não sigam as regras básicas. Existem 3 tipos de documentos XML:

- **Inválidos:** Não segue as regras básicas de documentos XML ou não segue as regras definidas no DTD ou schema.
- **Válidos:** Seguem da XML e as regras definidas no DTD ou schema
- **Bem formatados:** Seguem as regras da XML mas não tem um DTD ou schema associado

Elemento Raiz (root)

Um documento XML deve estar contido dentro de um único elemento, esse elemento é chamado *elemento raiz* e contem todos os outros elementos do documento

Regras básicas de Documentos XML

- Todo documento deve conter um elemento raiz
- Elementos XML não podem se sobrepor
- Tags de fechamento são obrigatórias
- Elementos são *case sensitive*
- Atributos devem ter valores entre aspas (Se o valor contem aspas, podem ser usadas entidades **"** para aspas duplas e **'** para aspas simples)

Declarações XML

A maioria dos arquivos XML começa com uma declaração XML que fornece ao *parser* informações básicas sobre o documento, a declaração pode conter até 3 pares nome-valor.

- **version:** é a versão do XML sendo usada (atualmente 1.0)
- **encoding:** é o conjunto de caracteres usados no XML (ISO-8859-1 , UTF-8)
- **standalone:** especifica se o XML pode ser processado sozinho ou se ele depende de outro arquivo pra ser processado (pode ser "yes" ou "no")

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
```

Coisas Adicionais em Documentos XML

Comentários: Podem estar em qualquer lugar do documento, começam com "`<!--`" e terminam com "`-->`", não pode conter dois hifens "`--`" dentro do comentário, tirando essa situação pode conter qualquer coisa.

Instruções de Processamento: É um tipo de instrução para ser executada por um pedaço em especial de código.

`<? cocoon-process type="sql" ?>` (Cocoon é um framework de processamento de XML, da Apache)

Entidades: Entidades são strings pré determinadas, sempre que o parser encontrar uma entidade ao ler o XML ele substituirá pelo texto na declaração da entidade:

`<!ENTITY dwa "Desenvolvimento Web Avançado">`

Quando o parser encontrar `&dwa;` ele substituirá por *Desenvolvimento Web Avançado*

A especificação XML define 5 entidades:

1. **<** para `<`
2. **>** para `>`
3. **"** para `"`
4. **'** para `'`
5. **&** para `&`

NAMESPACES XML

Utilizado para definir para que serve as tags XML de determinado documento. Para utilizar um namespace é preciso definir um prefixo de namespace

```
<acervo
  xmlns:livros="http://www.biblio.com/livros"
  xmlns:filmes="http://www.biblio.com/filmes"
>
```

```
  <livros:copia>
    <titulo>Usando XML</titulo>
    <tam>340</tam>
  </livros:copia>
  <filme:copia>
    <titulo>Rambo II</titulo>
    <tam>105</tam>
  </filme:copia>
</acervo>
```