

A Linguagem XML: Conceitos Básicos e Características

A XML – **eXtensible Markup Language** – é uma linguagem de marcadores como a HTML e foi desenhada para descrever dados.

A sua grande vantagem é que ela é extensível, ou seja, você não está limitado a um certo número de tags, e pode criar as suas próprias tags, assim sendo ela é uma linguagem auto definível.

Para descrever os dados a XML usa a DTD – **Document Type Definition**.

Diferenças entre XML e HTML

A linguagem XML não é uma linguagem que veio para substituir a HTML pois, XML foi criada com um objetivo diferente da HTML.

Enquanto HTML foi criada para **exibir** dados e ela se preocupa como os dados serão exibidos, a XML foi criada para **descrever** dados e ela se preocupa com o que os dados são.

HTML está relacionada com exibir dados enquanto que XML está relacionada em descrever dados.

As tags usadas em documentos HTML e a estrutura dos documentos HTML são predefinidas e o autor de um documento HTML pode usar somente as tags que estão definidas em HTML padrão.

Já a XML permite que o autor defina as suas próprias tags e a própria estrutura do seu documento.

A utilização de XML vem crescendo dia a dia, mas não devemos pensar que ela será a substituta da HTML, ela é um complemento à HTML e será usada para descrever e estruturar os dados enquanto HTML será usada para formatar e exibir os mesmos dados.

XML foi desenvolvido para estruturar, armazenar e enviar os dados.

As Regras de Sintaxe do XML

1) Todos os documentos XML devem conter a “*declaração XML*”

- Define a versão do XML e a codificação de caracteres usada no documento.
- `<?xml version="1.0" encoding="ISO-8858-1"?>`
- A declaração XML não é parte do documento XML, não é um elemento do documento XML, e por isso não precisa de uma *tag* de fechamento.

2) Todos os documentos XML devem conter um, e apenas um, elemento raiz

- `<raiz> ... </raiz>`
- Entre as *tags* do elemento raiz que estarão todas as outras *tags* do seu documento XML
- Elemento que “diz” o que é o documento

3) Todos os elementos devem ter uma *tag* de fechamento

- Existem duas construções válidas:
- `<teste>Teste 1</teste>` o elemento teste tem conteúdo
- `<teste />` o elemento teste é vazio
- Diferente do HTML que pode ter tags que não são fechadas

4) XML é *case sensitive*

- `<teste>` é diferente de `<Teste>` que é diferente de `<TESTE>`

5) Os elementos XML devem estar corretamente aninhados

- Correto: `<b><i>texto em negrito e em itálico</b></i>`
- Incorreto: `<b><i>texto em negrito e em itálico</i></b>`

6) Os valores dos atributos devem estar entre aspas

- Tanto faz se são aspas duplas ou simples
- Correto: `<data="22/12/1981"></data>`
- Incorreto: `<data=22/12/1981></data>`

7) Os nomes dos elementos devem seguir as regras:

- Nomes podem possuir letras, números e outros caracteres
- Nomes não podem começar com número ou caractere de pontuação
- Nomes não podem começar com as letras XML e suas variações
- Nomes não podem conter espaços

8) Sintaxe dos comentários:

- `<!--Comentário-->`

Um documento XML que siga essas regras de sintaxe é um documento XML Bem Formado

Um documento XML é dito válido se, além de ser bem formado, ele estiver de acordo com um *DTD* ou *XML Schema*.

Vejamos um exemplo de documento XML (*vamos dar a ele o nome de `aviso.xml`*), e a seguir vamos analisar cada linha de código:

```
<?xml version="1.0"?>
<aviso>
  <para>Janice data="01/04/2000"</para>
  <de>Jefferson</de>
  <cabecalho>Lembre-se</cabecalho>
  <corpo>Amanha voce tem prova de matematica</corpo>
</aviso>
```

Agora vamos analisar linha a linha...

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

Esta primeira linha do documento é uma declaração XML e deve sempre ser incluída pois define a versão XML do documento. Neste caso estamos especificando a versão 1.0 da XML. O Internet Explorer vem com um analisador XML que atende o padrão 1.0.

```
<aviso>
```

Esta linha define o primeiro elemento do documento – o elemento raiz.(nó raiz)

```
<para>Janice data="01/04/2000"</para>  
<de>Jefferson</de>  
<cabecalho>Lembre-se</cabecalho>  
<corpo>Amanha voce tem prova de matematica</corpo>
```

Estas quatro linhas definem 4 elementos filhos da raiz (*para, de, cabecalho e corpo*).

```
</aviso>
```

A última linha define o fim do elemento raiz.

As tags usadas no documento foram criadas pelo autor e por isso receberam esses nomes e somente a primeira linha é obrigatória pois define a XML para o analisador do navegador.

Vejamos as características principais de um documento XML:

- Perceba que cada tag inicial tem um tag final, ou seja, as tags são usadas sempre em pares. Você não pode usar: **<de>** Jefferson
- Todo documento XML possui um elemento raiz. Os demais elementos devem estar aninhados dentro de um elemento raiz.

- O elemento possui um atributo, a data de criação do documento; da mesma forma os elementos podem possuir atributos aos pares **nome/valor**, e, o atributo deve sempre vir entre aspas.
- As tags usadas em XML são *case sensitive*, ou seja, você deve escrever uma tag de fechamento da mesma maneira que escreveu uma tag de início. Assim **<para>** e **</Para>** irá ocasionar um erro no documento XML.

Questões para fixação do conteúdo

1. O que significa XML e qual a sua utilização?
2. Cite uma vantagem no uso da XML.
3. Cite quais são as diferenças entre XML e HTML.
4. Escreva o comando que é obrigatório a todo documento XML e explique sua função.
5. Cite 5 particularidades que devem ser seguidas na formação de um documento XML.
6. Escreva qual é o comando utilizado para inserir um comentário em um documento XML
7. Represente como deve ser uma estrutura básica de um documento XML.