

Documentos XML bem-formatados

José Carlos L. Ramalho
jcr@di.uminho.pt
gEPL - DI - UM



Um documento XML

- Conteúdo = Dados + Anotações
- Dados = blocos de texto
- Anotações:
 - marcas de início de elementos
 - marcas de fim de elementos
 - marcas de elementos vazios
 - referências a entidades
 - comentários
 - limitadores de secções especiais de texto
 - declarações de tipo de documento
 - instruções de processamento

O exemplo tradicional

```
<?xml version="1.0" encoding="UTF-8"?>
<doc>
  Hello World!!!
</doc>
```

A declaração XML

- Anotação especial que deve iniciar todos os documentos XML

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

version - obrigatório, valores possíveis: 1.0
standalone - opcional, valores possíveis: yes, no;
encoding - opcional, para o português o valor deverá ser: iso-8859-1

Comentários

- Podem aparecer em qualquer ponto dum documento XML.
- Começam pela marca: <!--
- e terminam com a marca: -->.

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!-- Isto é um comentário no início -->
<doc>
  Olá Mundo!!!
</doc>
```

Comentários (2)

- Existem algumas restrições à utilização de comentários:
 - Não podem aparecer antes da declaração.
 - Não podem aparecer dentro duma anotação.
 - Não se pode utilizar a sequência de caracteres "--" dentro dum comentário.

Comentários (3)

- Podem ainda ser utilizados para remover temporariamente partes do documento, desde que essas partes não contenham comentários.

```
<RECEITAS>
<TITULO> O Meu Livro de Receitas </TITULO>
<RECEITA ORIGEM="Portugal">
<TITULO> Bolo </TITULO>
<!--
<INGREDIENTE> 500g de farinha </INGREDIENTE>
-->
<INGREDIENTE> 200g de açúcar </INGREDIENTE>
<INGREDIENTE> 300g de manteiga
</INGREDIENTE>
</RECEITA>
</RECEITAS>
```

Documentos XML bem-formatados - jcr - UM - 2002.03.08

7

Instruções de Processamento

- As instruções de processamento são uma reminiscência da anotação procedimental
- Uma instrução de processamento não faz parte do conteúdo do documento.
- É uma indicação directa de que algo deve ser executado naquele ponto.
- Uma instrução de processamento começa por: `<?id-processador`
- e termina por: `?>`
- Exemplo: a declaração XML

Documentos XML bem-formatados - jcr - UM - 2002.03.08

8

Instruções de Processamento 2

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<agenda>
<?html action="hr"?>
  <entrada id="e1" tipo="pessoa">
    <nome>José Carlos Ramalho</nome>
    <email>jcr@di.uminho.pt</email>
    <telefone>253 604479</telefone>
  </entrada>
<?html action="hr"?>
  ...
</agenda>
```

Documentos XML bem-formatados - jcr - UM - 2002.03.08

9

Elementos

- Blocos lógicos em que um documento pode ser decomposto
- Exemplo:
 - Vais ver o espectáculo a `<lugar>Braga</lugar>?`
- Uma anotação de início começa por `<` e termina por `>`,
- e uma anotação de fim começa por `</` e termina por `>`.
- Uma anotação contém o nome do elemento que inicia ou que termina, respectivamente.

Documentos XML bem-formatados - jcr - UM - 2002.03.08

10

Caracteres reservados

- No conteúdo dum elemento, nunca deverão ser usados os caracteres `'<'` e `'>'` pois são os caracteres que limitam as anotações.
- Em lugar deles devem-se usar, respectivamente, as entidades do tipo carácter `'<'` e `'>'`.
- Qualquer processador ou editor de XML fará a substituição automática daquelas entidades pelos caracteres correspondentes.

Documentos XML bem-formatados - jcr - UM - 2002.03.08

11

Caracteres reservados 2

Se, no documento, estivesse o seguinte texto:

A anotação `<nome>` é usada para anotar nomes.

Um editor mostraria o mesmo texto da seguinte maneira:

A anotação `<nome>` é usada para anotar nomes.

Documentos XML bem-formatados - jcr - UM - 2002.03.08

12

Tipos de Conteúdo

- elemento com conteúdo textual
 - `<lugar>Braga</lugar>`
 - `<INGREDIENTE>Meia dúzia de ovos</INGREDIENTE>`
 - `<data>(1922)</data>`
- elemento com conteúdo misto
 - `<verso>Olha, <nome>Daisy</nome>: quando ...</verso> <p>Vais ver o espectáculo a <lugar>Braga</lugar?</p>`

Tipos de Conteúdo (2)

- elementos com conteúdo vazio: são normalmente utilizados pelo seu significado posicional - referências, pontos de inserção de imagens, ...
 - Como será discutido num capítulo mais à frente (`<ref id="cap5"/>`) ...
 - São representados por uma única anotação que é iniciada por '`<`' e termina em '`>`', que é a forma abreviada de escrever "`<elem-ident></elem-ident>`".

Atributos

- Um elemento pode ter um ou mais atributos que, por sua vez, podem ser opcionais ou obrigatórios.
- Visam qualificar o elemento a que estão associados.
- Não há limite para o número de atributos que podem estar associados a um elemento.
- Aparecem sempre na anotação que marca o início dum elemento, uma vez que vão qualificar o conteúdo que se segue.

Atributos (2)

- Um atributo é definido por um par constituído por um nome e um valor:
 - o nome e o valor devem estar separados pelo sinal '=' e
 - o valor deverá estar colocado dentro de aspas simples ou duplas.
 - Exemplo:
 - `<ref destino="exemplo5"/>`
 - `<imagem path="figs/img3.gif"/>`

Elemento versus Atributo

- Não existe uma fronteira entre os dois e muitas vezes a escolha não é simples.

Informação nos elementos

```
<agenda>
<entrada id="e1" tipo="pessoa">
<nome>José Carlos Ramalho</nome>
<email>jcr@di.uminho.pt</email>
<telefone>253 604479</telefone>
</entrada>
```

Informação nos atributos

```
... <agenda>
<age <entrada id="e1" tipo="pessoa" nome="José Carlos Ramalho"
email="jcr@di.uminho.pt" telefone="253 604479"/>
...
</agenda>
```

Atributos reservados

- `xml:lang`
 - Pode: `<para xml:lang="en">Hello</para>` indica elemento `<para xml:lang="pt">Olá</para>`
 - `<para xml:lang="fr">Bonjour</para>`
- `xml:space`
 - Serve para indicar se o espaço branco no conteúdo do elemento em causa é ou não relevante: **preserve** ou **default**.

Secções Marcadas



- Úteis para incluir exemplos de XML

```
<![CDATA [
The <p> tag is used for paragraphs
]]>
```

- Ou para texto com muitos caracteres reservados:

```
Prima a tecla <<<ENTER>>>.
Prima a tecla &lt;&lt;&lt;ENTER&gt;&gt;&gt;.
<![CDATA[Prima a tecla <<<ENTER>>>.]]>
```



Regras de bem-formação



- Um documento XML deve ter sempre uma declaração XML no início
- O documento deve incluir um ou mais elementos
- Todos os elementos têm anotações de início e fecho (excepto os vazios)
- Os elementos deverão estar aninhados correctamente
- Os valores de atributos têm de estar dentro de aspas



```
<?xml version="1.0" encoding="ISO-8859-1"?>
<sumarios>
  <disciplina> Processamento Estruturado de
    Documentos</disciplina>
  <professor>
    <nome>José Carlos Ramalho</nome>
    <email>jcr@di.uminho.pt</email>
    <url>http://www.di.uminho.pt/~jcr</url>
  </professor>
  <aula tipo="T">
    <data>2000.10.02</data>
    <sumario>
      <p>
        Anotação de Documentos: um pouco de
        história.</p>
      <p>
        Linguagens de Anotação como meta-linguagens:
        o SGML e o XML.</p>
      <p>
        Anotação Descritiva. Ciclo de vida dos
        documentos estruturados.</p>
    </sumario>
  </aula>
  ...
</sumarios>
```

Documento bem- formado



NameSpaces



To be continued ...

