

DTD - XML 构建模块

XML 和 HTML文档的主要的构建模块是元素标签。

XML 文档构建模块

所有的 XML 文档（以及 HTML 文档）均由以下简单的构建模块构成：

- 元素
- 属性
- 实体
- PCDATA
- CDATA

元素

元素是 XML 以及 HTML 文档的**主要构建模块**。

HTML 元素的例子是 "body" 和 "table"。XML 元素的例子是 "note" 和 "message"。元素可包含文本、其他元素或者是空的。空的 HTML 元素的例子是 "hr"、"br" 以及 "img"。

实例:

```
<body>some text</body>

<message>some text</message>
```

属性可提供**有关元素的额外信息**。

属性总是被置于某元素的开始标签中。属性总是以**名称/值**的形式成对出现的。下面的 "img" 元素拥有关于源文件的额外信息：

```

```

元素的名称是 "img"。属性的名称是 "src"。属性的值是 "computer.gif"。由于元素本身为空，它被一个 "/" 关闭。

实体

实体是用来定义普通文本的变量。实体引用是对实体的引用。

大多数同学都了解这个 HTML 实体引用：" ". 这个"无折行空格"实体在 HTML 中被用于在某个文档中插入一个额外的空格。

当文档被 XML 解析器解析时，实体就会被展开。

实体引用	字符
<	<

>	>
&	&
"	"
'	'

PCDATA

PCDATA 的意思是被解析的字符数据 (parsed character data) 。

可把字符数据想象为 XML 元素的开始标签与结束标签之间的文本。

PCDATA 是会被解析器解析的文本。这些文本将被解析器检查实体以及标记。

文本中的标签会被当作标记来处理，而实体会被展开。

不过，被解析的字符数据不应当包含任何 &、< 或者 > 字符；需要使用 &、< 以及 > 实体来分别替换它们。

CDATA

CDATA 的意思是字符数据 (character data) 。

CDATA 是不会被解析器解析的文本。在这些文本中的标签不会被当作标记来对待，其中的实体也不会被展开。

 点我分享笔记