链滴

【XML】《XML 实用教程》笔记

作者：ZephyrJung
原文链接：https://ld246.com/article/1478855900498
来源网站：链滴
许可协议：署名-相同方式共享 4.0 国际 (CC BY-SA 4.0)

XML是eXtensible Markup Language的缩写，称之为可扩展标记语言。

符合W3C制定的基本语法规则的XML文件称为规范的XML文件，规范的XML文件如果再符合额外的些约束就称之为有效的XML文件。

一个规范的XML文件应当满足如下语法规则：

- XML文件用“XML声明”开始。
- XML文件有且仅有一个根标记
- XML的非根标记都必须封装在根标记中。
- 非空标记必须由“开始标记”与“结束标记”构成。
- XML文件的标记必须形成树形结构，即标记不允许出现交叉。
- 空标记没有“开始标记”和“结束标记”

XML声明

一个规范的XML文件应当以XML声明作为文件的第1行，在其前面不能有空白、其他处理指令或注释如：

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

还可以添加属性：standalone，指为yes或no，表明XML文件是否是完全自包含的，即是否引用了外“实体”。

标记

一个标记包含的内容由两部分构成：文本数据部分和子标记部分。一个标记包含的文本数据部分可以普通字符、CDATA段和实体引用。

根标记：每个XML文件有且仅有一个根标记，其他标记都必须封装在根标记中。

空标记：即不包含子标记或文本内容的标记，如：<name arg1 arg2.../>

非空标记：<name arg1 arg2...>...</name>

标记命名规则：可以由字母、数字、下划线、点或连字符组成，但必须以字母或下划线开头。标记名区分大小写。

属性

属性指标记的附加信息，命名规则与标记相同，且区分大小写。属性不体现数据结构，不要因属性的繁使用破坏XML的数据结构。

特殊字符

< <
> >
' '
" "
& &

CDATA段

CDATA段中的内容可以包含任意的字符。但CDATA段不能相互嵌套，如：

```
<![CDATA[
boolean boo=true&&false
<三国演义>
]]>
```

注释

```
<!--与HTML文件相同—>
```

注释不可以在XML声明的前面。

名称空间

XML允许自定义标记，那么不同的XML文件以及同一XML文件就可能出现名字相同的标记。想要区这些标记，就需要使用名称空间。

有前缀的名称空间语法如下：

xmlns:前缀=名称空间的名字，如xmlns:person="China.dalian"

无前缀的名称空间语法如下：

xmlns=名称空间的名字，如xmlns=www.yahoo.com

名称空间的名字决定了相互之间的区分，前缀只是为了方便标记引用。

名称空间的作用域是声明该命名空间的标记及其所有子孙标记。尽管子标记可以通过名称空间的前缀引用父标记声明名称空间，表明自己隶属于该名称空间，但子标记也可以重新声明名称空间，如：

```
<?xml version="1.0" encoding="UTF-8" ?>
<people xmlns="public.of.china">
 <p1:Lisi xmlnsp1="Liaoning">
  在建筑公司担任工程师
  <usa:Lichuguo xmlns:usa="American">
   在美国学医英语
  </usa:Lichuguo>
 </p1:Lisi>
 <Zhangsan xmlns="Shanghai">
  在上海中心医院医师
  <Zhangchuguo xmlns="France">
   在法国学习法语
  </Zhangchuguo>
 </Zhangsan>
</people>
```

名称空间的名字

W3C推荐使用统一资源标识符（Uniform Resource Identifier, URI）作为名称空间的名字。在XML，URI不必是有效的，仅仅用作名称空间的名字，往往使用URL作为URI。

有效的XML文件

规范的XML表示合乎XML文件语法，有效的XML文件表示组织符合需要的逻辑。

对XML的数据结构进行约束的方式有两种：使用文档类型定义（Document Type Definition，DTD和XML Schema模式。

一个规范的XML文件如果和某个DTD文件相关联，并且遵守该DTD文件规定的约束条件，就称之为有的XML文件。

DTD文件

在DTD文件中，用关键字ELEMENT来定义一个元素，格式如下：

<!ELEMENT 标记名称 标记的约束条件>

DTD文件的扩展名必须是“.dtd”，保存时所选择的编码必须与其关联的XML文件一致。

<!ELEMENT 手机用户表 (用户*)>
<!ELEMENT 用户 (号码，姓名)>
<!ELEMENT 号码 (#PCDATA)>
<!ELEMENT 姓名 (#PCDATA)>

在XML文件中使用“文档类型声明”与一个DTD文件相关联。有两种形式的关联：SYSTEM和PUBLI，SYSTEM关联表明所关联的DTD文件由个人或工作小组所定义且认可，PUBLIC关联表明所关联的DD文件已经得到某一领域的认可，是经过许多人讨论得到认可的DTD文件：

<!DOCTYPE 根标记名称 SYSTEM "DTD文件的URI">
<!DOCTYPE 根标记名称 PUBLIC "正式公用标识符" "DTD文件的URI">

URI如果是一个文件的名字，该文件必须和当前XML文件在同一目录中；如果URI是一个URL，该URL须是可以访问的。

DTD文档类型声明应该写在XML声明的后面。

约束标记的子标记

标记的约束条件是用小括号括起来的子标记列表，子标记列表可以约束XML文件中的标记可以包含哪子标记以及这些子标记出现的先后顺序。

子标记列表中的每个子标记可以尾加模式限定符来限定该子标记出现的次数，不加限定符的子标记必出现且只能出现一次。限定符有一下三种：

- +：必须出现一次或多次
- *：可以出现零次或多次
- ？：可以出现零次或一次

子标记列表的分项也可以是几个标记的“或运算”，并用括号括起来，如：<!ELEMENT 应聘者 (姓名,专科|本科|研究生)?,(男|女)>

约束(#PCDATA)表示文本数据，可以单独使用，也可以与子标记或运算使用，但与子标记或运算时，定符必须为*

约束EMPTY表示标记是一个空标记，或者只能包含有空字符的非空标记。

约束ANY表示不准备对该标记有任何约束

DTD的完整性

即满足以下两个条件：

1. 不允许无穷嵌套
2. XML文件中的每个标记都必须在DTD中有相应元素对其进行约束

DTD中的属性约束列表

格式如下：

```
<!ATTLIST 标记名称
  属性名称 属性类型 默认值
  属性名称 属性类型 默认值
  .....
>
```

其中，默认值有如下几种情况：

- 字符串：标记必须有该属性，且有默认值
- #IMPLIED：标记可以没有该属性，没有默认值
- #REQUIRED：标记必须有该属性，没有默认值
- #FIXED：标记可以没有改属性，但是如果有该属性，那么属性指固定不变

常用属性类型如下：

- CDATA：属性值可以为任意字符串，但不能包含左右尖括号、与符号、单双引号。如果想用这些字，可以使用实体引用。
- Enumerated：属性值只可以使枚举值，其并非关键字，而是用小括号括起来，使用竖线分隔的枚值组。
- NMTOKEN：属性值可以由字母、数字、下划线、点或连字符组成，属性值中不能含有空格，可以数字、点或连字符开头。
- NMTOKENS：即用空格分隔的多个NMTOKEN值
- ID：如果希望某个属性的属性值具有专用性，即不允许其他类型相同的属性再取这个属性值，就可将属性的类型取为ID类型。可以由字母、数字、下划线、点或连字符组成，但必须以字母或下划线开。其默认值必须为#REQUIRED或#IMPLIED。
- IDREF：属性值为已有的某个ID类型的属性的值
- IDREFS：属性值为若干个其他ID类型的属性值的组合，可以用空格分隔。

内部DTD

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE 列车时刻表 [
 <!ELEMENT 列车时刻表 (T28次,T226次)>
 <!ELEMENT T28次 (开车时间,终到时间)>
 <!ELEMENT T266次 (开车时间,终到时间)>
 <!ELEMENT 开车时间 (hour,minute)>
 <!ELEMENT 终到时间 (hour,minute)>
 <!ELEMENT hour (#PCDATA)>
 <!ELEMENT minute (#PCDATA)>
]>
<列车时刻表>
 <T28次>
  <开车时间>
   <hour>21点</hour>
   <minute>12分</minute>
  </开车时间>
  <终到时间>
   <hour>23点</hour>
```

```
  <minute>25分</minute>
 </终到时间>
</T28次>
<T226次>
 <开车时间>
  <hour>08点</hour>
  <minute>45分</minute>
 </开车时间>
 <终到时间>
 <hour>19点</hour>
 <minute>36分</minute>
 </终到时间>
</T28次>
</列车时刻表>
```

DTD调试

```
IGNORE：忽略某些约束条件
<![ IGNORE [
DTD中的某些约束条件
]]>
INCLUDE：包含某些约束条件
<![INCLUDE [
DTD中的某些约束条件
]]>
```

DTD与命名空间

如果有效的XML文件中的标记想定义名称空间，那么必须保证DTD中有相应的约束。格式如下：

```
<!ATTLIST 前缀:标记名称 xmlns:前缀 类型 默认值情况>
```

如：

```
<!ATTLIST ya:张三 xmlns:ya CDATA #FIXED "www.yahoo.com">
```

名称空间的目的是有效的区分名字相同的标记，否则不允许对一个标记约束两次。

XML关联CSS

```
<?xml-stylesheet href="样式表的URI" type="text/css" ?>
```