

XML 及 WML 教程

李秋云 编著

- ④ 全面讲解21世纪Web上的“世界语”——XML
- ④ 详细介绍基于XML的无线标记语言——WML
- ④ 结合实例讲述XML与WML的基础知识，相关协议及WAP网页的开发方法

XML 及 WML 教程

李秋云 编著

人民邮电出版社

图书在版编目(CIP)数据

XML 及 WML 教程 / 李秋云编著. — 北京: 人民邮电出版社, 2001. 7

ISBN 7-115-09421-7

I. X... II. 李... III. 可扩充语言, XML、WML — 程序设计 — 教材 IV. TP312

中国版本图书馆 CIP 数据核字(2001)第 036937 号

内 容 提 要

本书是为读者学习 XML 元标记语言和 WML 实例标记语言而准备的教材。书中首先介绍了 XML 基础知识, 包括扩展标记语言 XML 的基础知识及 XML 系列协议中的主要相关协议, 如 Namespaces、XLink、XSLT、Xpointer 等。然后, 结合大量实例全面而详尽地介绍了以下内容: 基于 XML 的实例标记语言 WML (Wireless Markup Language, 无线标记语言) 及其脚本语言 WMLScript; 如何对 WML 文档进行二进制压缩; 如何利用 WML 来设计丰富多彩的 WAP 网页。并对相关的 WAP 开发等问题进行了深入的阐述。

本书可供从事计算机网络工作的高级技术人员参考, 也可供高等院校计算机专业和电子专业的高年级学生或研究生作为 XML 及 WML 的教材使用。

XML 及 WML 教程

◆ 编 著 李秋云

责任编辑 张瑞喜 姚予疆

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号

邮编 100061 电子函件 315@ptph.com.cn

网址 <http://www.ptph.com.cn>

读者热线: 010-67129212 010-67129211(传真)

北京汉魂图文设计有限公司制作

北京顺义振华印刷厂印刷

新华书店总店北京发行所经销

◆ 开本: 787 × 1092 1/16

印张: 23.25

字数: 579 千字

2001 年 7 月第 1 版

印数: 1-5 000 册

2001 年 7 月北京第 1 次印刷

ISBN 7-115-09421-7/TP·2307

定价: 34.00 元

本书如有印装质量问题, 请与本社联系 电话: (010)67129223

前 言

现在,许多在软件业中很有影响力的公司都宣称它们的产品中支持 XML,并促使 XML 成为下一代 Web 的发展方向。甚至有人称,XML 将是 21 世纪 Web 上的“世界语”。XML 到底是什么样一种标记语言,又是如何产生的呢?

近十年来,因特网获得了空前规模的发展,特别是以 Web 为核心的检索业务得到了巨大的发展,而第一代网页设计语言 HTML,又是 Web 技术得以发展的基础,基于 HTML 的 Web 检索技术已经成为被普遍接受的应用模式。在通信领域,HTML 已经成为事实上的表示层句法标准并被广泛使用。近几年来,为了满足不断出现的新的应用需求,HTML 得到了极大的丰富和发展,但同时我们也看到随着网络的进一步发展,HTML 的缺点也逐渐暴露出来,主要包括:

其一,HTML 的标记都是固定的,而实际应用中又常常需要一些特殊标签,要想满足多种应用领域的不同要求,需要在 HTML 中增加大量的标签,显然这是不现实的。

其二,HTML 文档内部结构的条理性较差。为此浏览器的开发者则不得不把大量的精力耗费在文法错误的包容上,这样不但会增加浏览器程序的复杂性,甚至会降低浏览时的效率。

其三,HTML 不利于搜索引擎的搜索。HTML 标签本身,几乎和文档内容没有任何关系,不能给搜索引擎提供足够的线索。如果让搜索引擎在网上的 HTML 文档中搜索某项内容,它返回的条目常常是成千上万,而不能迅速准确地完成搜索任务。

XML(可扩展的标记语言)是 W3C 组织于 1998 年 2 月发布的第二代因特网网页设计语言的标准,它的出现为网络应用注入了新的活力。W3C 组织制定 XML 标准的初衷是,定义一种用于在 Internet 上交换数据的标准。W3C 采取了简化 SGML 的策略,在 SGML 基础上,去掉语法定义部分,适当简化 DTD 部分,并增加了部分互联网的特殊成分。因此,XML 也是一种标记语言,基本上是 SGML 的一个子集。因为 XML 也有 DTD,所以 XML 也可以作为派生其他标记语言的元语言。

在 Internet 上,XML 的用途不仅仅是制作网页,它的作用主要有两个,一是作为元标记语言,定义各种实例标记语言;二是作为标准交换语言,担负起描述交换数据的作用。

在没有 XML 的情况下,要想定义一个标记语言并使之推广应用是非常困难的。一方面,相关的组织要对这个标记语言进行多次评定、修改,即使它最终能成为一个正式推荐标准,可能也要等几年的时间。另一方面,为了让这套标记语言得到广泛应用,必须为它开发相应的浏览器,或者去游说各个浏览器厂商支持这种新的标记语言,其艰难是可想而知的。而现在有了 XML,你就可以自由地定制自己的标记语言,而不必通过任何组织的批准;另一方面,浏览器只要支持 XML,就自然会支持这种新的标记语言,而这一点 IE、Netscape 等主流浏览器已经做到了。这样,利用 XML,各个组织、个人就可以创建适合他们自己需要的标记语言,并且,这个标记语言可以迅速地投入使用。

另外,XML 的最大特点是,它把显示格式从文档中分离出去了,另存在样式表中。XML 文档只包含内容。用这种形式存储时,文档内容非常地简单明晰,因为它所携带的信息不包

含对于显示格式的描述，而只是关于文档内容的信息。这样，如果需要改动信息的表现方式时，无需改动信息本身，只要改动样式表文件就够了。同时在 XML 文档中可以进行高效的搜索，因为搜索引擎没必要再去遍历整个 XML 文件，它只需去找一找相关标记下的内容就够了，这样极大程度地提高了搜索的效率。毫不夸张地说，XML 的标记为搜索引擎赋予了智慧！

本书首先介绍了 XML 的起源及其基础知识，然后结合大量实例，系统地介绍了 XML 的文档类型定义和实体，以及 XML 家族的其他重要规范，包括名称空间、扩展链接和 XSL 样式表语言。

前面介绍过，XML 是一种元标记语言，可以用来定义实例标记语言，现在基于 XML 的实例标记语言已经有许多，而 WML(无线标记语言)正是其中最重要的一种。

无线网络和因特网的迅速发展最终促成了 WAP（无线应用协议），也就是利用无线终端检索 Web 信息的标准，除了一系列的通信协议以外，WAP 还包括一种专为无线上网制定的标记语言，即 WML 标记语言。目前 WAP 网站正处于迅速发展的时期，所以在本书的后半部分，详细介绍了 XML 在 WAP 中的应用。

全书结合了大量的实例，以加深读者的理解。

参加本书编写工作的还有郝建国、郝永义、程岩、孙瀚冰、曾凡仔等，同时得到许多其他同志的帮助，在此一并表示衷心的感谢。

由于作者水平有限，书中难免有不妥之处，敬请读者指正。

作 者
2001 年 4 月

目 录

第 1 章 概 述	1
1.1 标记语言	1
1.2 SGML	2
1.2.1 SGML 简史	2
1.2.2 SGML 的应用	5
1.3 HTML	5
1.4 XML	6
1.4.1 认识 XML 文档	8
1.4.2 XML 的由来	9
1.4.3 XML 的应用	10
1.5 WAP 概述	14
1.5.1 WAP 的应用框架	15
1.5.2 WAP 协议栈	16
1.5.3 WAP 的未来发展	17
1.6 WML	18
第 2 章 XML 1.0 规范	19
2.1 绪 论	19
2.1.1 动机和目标	19
2.1.2 术语	20
2.2 文 档	21
2.2.1 结构性 XML 文档	21
2.2.2 字符	21
2.2.3 通用语法结构	22
2.2.4 字符数据和标记	23
2.2.5 注释	23
2.2.6 处理指令	23
2.2.7 CDATA 段	24
2.2.8 序言和文档类型声明	24
2.2.9 独立文档声明	26
2.2.10 空白符处理	27
2.2.11 行尾处理	27
2.2.12 语言标识符	27
2.3 逻辑结构	28
2.3.1 起始标记、结束标记和空标记	29

2.3.2	元素类型声明	30
2.3.3	属性表声明	32
2.3.4	条件节	35
2.4	物理结构	36
2.4.1	字符与实体引用	36
2.4.2	实体声明	38
2.4.3	语法分析实体	39
2.4.4	XML 分析器处理实体和引用	40
2.4.5	构造内部实体替代文本	43
2.4.6	预定义的实体	43
2.4.7	表示法声明	44
2.4.8	文档实体	44
2.5	一致性	44
2.5.1	检验和不检验文档有效性的分析器	44
2.5.2	使用 XML 分析器	45
2.6	标 注	45
第 3 章	XML 文档的结构性	47
3.1	XML 文档	47
3.2	XML 文档中的文本	48
3.3	基本语法规则	49
3.3.1	大小写	49
3.3.2	空白符	49
3.3.3	名字(Name)和名字符号(Nmtoken)	50
3.3.4	字符串	50
3.3.5	语法规则	51
3.4	结构性	52
3.5	注 释	52
3.6	几种预定义的字符实体	54
3.7	CDATA	55
3.8	元素标记	56
3.8.1	标记名称	56
3.8.2	空标记	57
3.9	属 性	57
3.10	结构性	58
3.10.1	必须以一个 XML 声明开始	58
3.10.2	起始标记和结束标记必须匹配	61
3.10.3	用“/>”结束空标记	62
3.10.4	根元素完全包含其他元素	62
3.10.5	元素必须是完全嵌套	62
3.10.6	属性值必须用引号引起来	66

3.10.7 使用字符实体引用	66
3.11 URL、URN、URI	66
3.12 xml:lang 属性	67
第 4 章 DTD	69
4.1 结构性和有效性约束	69
4.2 DTD 简介	70
4.3 文档类型声明	70
4.4 元素声明	72
4.5 元素内容声明	73
4.5.1 EMPTY	73
4.5.2 ANY	74
4.5.3 子元素内容	74
4.5.4 混合内容	76
4.6 属 性	77
4.6.1 属性表声明	79
4.6.2 默认值声明	80
4.6.3 属性类型	81
4.7 表示法声明	86
4.8 WML 的文档类型定义	87
第 5 章 实 体	97
5.1 实体的概念	97
5.2 实体声明	98
5.3 实体的分类	99
5.3.1 字符实体	99
5.3.2 语法分析实体和无需语法分析实体	99
5.3.3 内部实体和外部实体	100
5.3.4 通用实体	100
5.3.5 参数实体	105
5.3.6 外部实体	106
5.3.7 实体种类总结	107
5.4 共享 DTD	108
5.4.1 本地共享 DTD	108
5.4.2 远程共享 DTD	109
5.4.3 公共的 DTD	110
5.5 DTD 的内部和外部子集	111
5.5.1 创建公用的外部 DTD	112
5.5.2 连接外部 DTD	113
5.5.3 同时利用内部和外部 DTD	115
5.6 处理指令	116
5.7 表示法和无需语法分析实体	117

5.8	条件节	118
5.9	standalone 属性	119
第 6 章	名称空间(Namespaces)	122
6.1	名称空间的作用	122
6.2	名称空间标识符	126
6.2.1	标识符	126
6.2.2	名称空间实例	127
6.3	名称空间的声明	127
6.3.1	前置字符串	127
6.3.2	名称空间的声明	128
6.4	名称空间的作用域	130
6.5	缺省的名称空间	132
6.6	名称空间的应用	135
第 7 章	WML 基础	137
7.1	了解 User Agent	137
7.2	如何创建 WML 文档	138
7.3	WML 数据类型	138
7.4	字符集	139
7.4.1	Unicode	140
7.4.2	UTF-8	141
7.4.3	UCS-2	143
7.4.4	ISO-8859-1	143
7.4.5	GB2312	143
7.5	WML 文档中的字符	144
7.6	deck/card 结构	145
7.7	其他 WML 语法	146
7.7.1	变量	146
7.7.2	实体	147
7.7.3	元素及其标记	148
7.7.4	属性	149
7.7.5	注释	150
7.7.6	CDATA 段	150
7.7.7	大小写	152
第 8 章	WML 元素	153
8.1	公共属性	153
8.1.1	id 属性	153
8.1.2	class 属性	154
8.1.3	xml:lang 属性	155
8.2	wml 元素	155
8.3	card 元素	156

8.4	p 元素	160
8.5	br 元素	164
8.6	文本显示格式	165
8.7	table 元素	168
8.7.1	table 元素	168
8.7.2	tr 元素	168
8.7.3	td 元素	169
8.8	img 元素	170
8.9	input 元素	172
8.10	select 元素	175
8.10.1	select 元素	176
8.10.2	option 元素	178
8.10.3	optgroup 元素	180
8.11	fieldset 元素	182
第 9 章	事件及导航	184
9.1	历史记录	185
9.2	事 件	186
9.2.1	do 元素	186
9.2.2	内在事件	192
9.2.3	onevent 元素	198
9.2.4	timer 元素	201
9.3	template 元素	203
9.4	deck/card 任务覆盖	205
9.5	postfield 元素	208
9.6	setvar 元素	208
9.7	WML 任务	209
9.8	anchor 元素	214
9.9	a 元素	215
9.10	head 元素	216
9.11	access 元素	216
9.12	meta 元素	217
第 10 章	WMLScript	218
10.1	WMLScript 的作用	218
10.2	WMLScript 字节码解释器	218
10.3	语法规则	221
10.3.1	大小写	221
10.3.2	空格	221
10.3.3	分号	221
10.3.4	注释	221
10.3.5	常量	221

10.3.6	标识符	223
10.3.7	保留字	223
10.3.8	名字空间	223
10.4	变量和数据类型	224
10.4.1	变量声明	224
10.4.2	变量的生存周期	224
10.4.3	变量存取	224
10.4.4	变量类型	225
10.4.5	L-Values	225
10.4.6	自动类型转换	225
10.5	操作符和表达式	225
10.5.1	赋值运算符	225
10.5.2	算术运算符	226
10.5.3	逻辑运算符	226
10.5.4	字符串操作	227
10.5.5	比较操作	227
10.5.6	数组操作	228
10.5.7	逗号运算符	228
10.5.8	条件运算符	228
10.5.9	typeof 运算符	228
10.5.10	isvalid 操作	229
10.5.11	表达式	229
10.5.12	运算符优先级	229
10.6	函 数	230
10.6.1	函数声明	230
10.6.2	函数调用	231
10.6.3	函数返回值	232
10.7	语 句	232
10.7.1	空语句	232
10.7.2	表达式语句	233
10.7.3	块语句	233
10.7.4	变量语句	233
10.7.5	if 语句	234
10.7.6	while 语句	235
10.7.7	for 语句	235
10.7.8	break 语句	236
10.7.9	continue 语句	236
10.7.10	reture 语句	237
10.8	标准库函数	237
10.9	pragmas	238

10.9.1	外部编译单元	238
10.9.2	访问控制 pragma	239
10.9.3	元 pragma	239
10.10	自动类型转换	240
10.10.1	基本规则	240
10.10.2	操作符和数据类型转换	242
10.10.3	操作符总结	243
第 11 章	WMLScript 标准库函数	246
11.1	Lang 库函数	246
11.2	Float 库函数	250
11.3	String 库函数	252
11.4	URL 库函数	260
11.5	WMLBrowser 库函数	265
11.6	Dialogs 库函数	267
第 12 章	WML 网页实例	269
12.1	表 格	269
12.2	WMLScript 应用实例	279
第 13 章	二进制 XML 文档	290
13.1	二进制 XML 内容结构	290
13.1.1	多字节整数	290
13.1.2	字符编码	291
13.1.3	文档结构	292
13.1.4	版本号	293
13.1.5	文档公共标识符	293
13.1.6	字符集	294
13.1.7	字符串表 (String Table)	295
13.2	标记符结构 (Token Structure)	295
13.2.1	解释器状态机	296
13.2.2	tag 码空间	296
13.2.3	Attribute 码空间	297
13.2.4	通用标记符	298
13.3	对 XML 文档进行编码	301
13.4	通用标记符	302
13.5	XML 二进制文档编码实例	303
13.6	WML 文档的二进制格式	307
第 14 章	XSLT	314
14.1	XSL	314
14.2	XSLT 和 XSL 的关系	315
14.3	XSLT 简介	316
14.4	如何在网络上应用 XSLT	317

14.5 XSLT 的转换流程及工作原理	318
14.5.1 应用实例	319
14.5.2 XSLT 详解	322
14.5.3 输出字符码设定	330
第 15 章 XML 链接语言 (XLink)	331
15.1 XLink 的基础	331
15.2 基本概念	333
15.3 XLink 的应用意义	334
15.4 XLink 标记	335
15.5 链接元素及其属性	337
15.5.1 扩展链接	338
15.5.2 简单链接	352
15.5.3 XLink 属性	354

第1章 概 述

HTML 是一种基于 SGML 的实例标记语言，它的特点是简单、易用，这使它迅速地在 Web 上推广应用，也为 Web 的发展壮大起了巨大的推动作用。

几年来，为了满足不断更新的应用需求，HTML 得到了极大的丰富和发展，但同时面对不断涌现的新的需求，HTML 的缺点也逐渐暴露出来。为此，W3C 制定了新的元标记语言 XML（扩展标记语言），XML 是 SGML 的一种简化形式。

WAP 作为因特网和无线网络的融合，为人们提供了更快捷、方便的信息服务。而 WAP 网页则是用 WML(无线标记语言)创建的。WML 是 XML 的一种实例标记语言。

本章将分别对以上内容进行介绍。

1.1 标记语言

我们日常书写所用的语言，称为书面自然语言。如果为了标识某些信息，在书面自然语言中加入一些标记，这种书面自然语言就可以称为标记语言（Markup Language）。比如在一段书面语言中，为了说明某一句话的重要性，可以在这句话下面划上一条下划线，在这里，下划线就可以算是一种标记。不妨这样说，如果将一篇手稿交给排版人员排印，其中没有任何格式，甚至连段落和不同的字体都没有，而只是一整篇连续的文字。手稿中用手写方式插入了一些用来规定打印格式的标注，诸如“这里用黑体，加下划线”、“这里用斜体”等，这就是“格式化标记”，可以说是一种比较原始的标记。

我们在这里要介绍的标记语言，实际上是一种为了方便计算机处理而设计的标记语言，也就是说，是为了让计算机看懂，而不只是让人看懂。其中所用到的标记，往往是代表一定含义的文字或数字。通常的做法是，根据需要先定义一套标记，然后将这套标记添加到书面语言的合适位置中去，使书面语言变成标记语言。

例如，我们为了让计算机“了解”在一段书面语言中，哪一部分是标题，哪一部分是作者的名字，哪一部分是正文。我们定义了如下一套标记：

<标题> </标题>

<作者> </作者>

<正文> </正文>

那么，利用上述标记，我们就可以写出下面这段标记语言。

<标题>什么是标记语言？ </标题>

<作者>张三</作者>

<正文>我们日常书写的语言，被称为书面自然语言。……</正文>

标记语言不同于一般的控制程序设计语言，它基本上可以被视为是一种文档结构描

述语言。在计算机处理过程中，标记语言的标记既可以作为数据，也可以作为控制语句来使用。在计算机程序语言中，一般把一种语言的定义整体称为某某程序设计语言，而把用这种语言编写的文字段落，称为某某语言程序。套用这个惯例，我们把标记语言的定义整体称为某某标记语言，而把用这种标记语言编写的文字段落，称为某某语言文档。如果我们把上面定义的那套标记叫做“文章标记语言”的话，那么下面那段标记语言，就可称为“文章标记语言文档”。

相应地，用“XML（标记）语言”标识的文档称为“XML（标记语言）文档”。

1.2 SGML

SGML（Standard General Markup Language，标准通用标记语言）是一种元标记语言，实际上它是一种通用的文档结构描述标记语言，主要用来定义文献模型的逻辑结构和物理结构。SGML 是 ISO 组织于 1986 年发布的 ISO 8879 国际标准。

一个 SGML 文档由三部分组成，即：

- 语法定义。
- 文档类型定义(简称 DTD——Definition Type Document)。
- 文档实例。

语法定义，规定了文档类型定义和文档实例所必须遵循的语法结构；文档类型定义，规定了文档实例的结构和组成结构的元素类型；文档实例是 SGML 文档的主体部分，用来存放文档的具体内容。

SGML 是一种元语言，在实际应用中可以派生出其他的语言。也就是说，每一个特定的 DTD 都定义了一类文档。例如，所有的数学论文都可以使用同一个 DTD。我们通常把具有某一特定 DTD 的 SGML 语言，称为某某标记语言。

我们在 Web 上常用的 HTML 语言就是由 SGML 派生出来的标记语言。HTML 是基于 SGML 的一种实例标记语言，也就是说，它的 DTD 作为标准被固定下来，用户不能修改它，而只能用它来创建文档实例。因此，HTML 不能作为定义其他标记语言的元语言。

XML 也是 SGML 的一个子集。严格地讲，XML 是一种简化的 SGML。与 HTML 不同的是，XML 可以定义标记语言的 DTD，因而也可以像 SGML 那样，作为元语言，来定义其他标记语言。

如果把标记语言分为元标记语言和实例标记语言的话，SGML 和 XML 都是元标记语言，而 HTML 和由 XML 派生的 XHTML 都是实例标记语言。

SGML 已经存在了很长的时间，并且在过去数十年间的上千个工程中的应用也已经证明了它的正确性。实际上，由于 XML 是 SGML 的一个子集，因此可以从 SGML 的发展历史来了解 XML 的由来。

1.2.1 SGML 简史

计算机科学的一个重要应用是文本处理，其研究内容主要是如何通过计算机系统，使

文档编辑和发布的部分过程自动化。自动化文本处理的一个重要应用方面是计算机排版。我们将文档输入到计算机中，然后计算机根据我们对文档样式的描述将文档打印出来。

前面所说的这种过程只是将文档打印出来。但在后来，人们又希望对文档做更多、更复杂的处理，例如，自动化的管理、搜索、排版等。在这个时候，IBM 成立了专门的研究小组（以 Charles Goldfarb 领导），试图来建立一个能够进行存储、查找、管理和发布法律文档的系统。

这个研究小组发现，IBM 内部存在很多种系统，他们之间无法互相通信，不同的系统对信息采用了不同的表示形式，即文件格式，而且每个系统使用一种不同的命令语言。所以它们无法互相读取文件。

就这个研究课题和发现的以上问题，Goldfarb 和其他两个 IBM 研究人员，Ed Mosher 和 Ray Lorie，经过认真研究，得出了一些重要结论。这些结论继而演变为 SGML 的基本思想，它们具体包括：

（1）有必要创建一种公共的标记语言，用来创建文档。

试想在实际生活中，如果大家都讲方言，互相交换信息时都要进行翻译，那么这将是一个难以沟通和交流的世界，所以，我们需要一种“世界语”，例如在学术会议中一般都采用英语。同样，如果各个工具软件都采用各自的语言，它们就无法共同工作，因此它们也需要一种公共的语言。

也就是说，在文本处理领域中，有必要采用一种公共的标记语言作为文档表示形式，这种公共的标记语言称为通用标记语言。

（2）对于不同的应用来说，要使用特定的文档类型。针对前面的任务，即“建立一个能够存储、查找、管理和发布法律文档的系统”，则要求所用的标记必须是专用于法律文档的。

律师们和科学家们都可以使用同一种标记语言，但他们并不使用同样的术语（也就是文档类型）。这类似于他们都使用英语，但是他们分别有各自的专业术语。

当我们向计算机描述文档时，这些领域的专用术语就显得更重要了。如果用一种简单的，类似于标准的字处理软件，而不是专门用于某个特定的应用领域的软件，则这种表示形式将无法实现对文档进行复杂处理的要求。

我们知道，计算机不会真正理解文档中的文字内容。就像前面说的搜索引擎一样，例如，我们要搜索有关“XML”的书籍，则在挨个查找许多无关的文档后，它可能返回许多只是包含上述单词并且和书籍无关的文档。

解决的办法就是告诉计算机有关文档的尽可能多的信息。即定义一组有特定含义的标签，计算机通过标签就可以“理解”其中的内容。这样看来，不同领域需要的标签是不一样的，我们没有必要创建一种包罗万象的标记语言，而只要为不同的应用领域分别创建一组标签，即一种特定的文档类型就可以了。

特定的文档类型不但有利于自动搜索，而且有利于自动排版。因为计算机“理解”了文档的内容，所以我们能够通过编程，让计算机来为我们重新排版，即更改样式表文档。

实际上，标准通用标记语言（SGML）文档的表现形式是在与文档内容分开的样式表中定义的。这是 SGML 的一个重要特点，XML 也继承了 SGML 的这一特性。而在 HTML 文档中，内容和格式是混合在一个文档中的，因此，格式的改变意味着要重新编辑所有的 HTML 文档。如果需要频繁地改变文档的表现形式，其工作量更是繁重。另外，如果愿意，通用标

记语言的用户也可以不对文档的表现形式作任何规定，而由计算机根据某些约定格式，将文档打印出来，或显示在屏幕上，或生成语音信息。

另一方面，我们也可以为文档准备两个或多个不同样式表，在显示时，计算机可以根据用户的有关资料或用户的选择，决定采用哪一种样式表，也就是说将同一个文档格式化为不同的样式。

不同的文档类型对应不同的实例标记语言，在文档类型定义（DTD）中，我们只要定义文档的内容结构。在 SGML 中，我们用“元素”来表示文档的各个组成部分。元素由标记和内容组成，其中标记（包括起始标记和结束标记）由尖括号括起来。在文档实例中，起始标记和结束标记之间是元素的内容，内容包括文本内容和子元素。例如下面这段例子。

```
<书>
  <摘要>
    这是一本世界著名童话故事书
  </摘要>
  <书名>
    安徒生童话故事选
  </书名>
  <作者>
    安徒生
  </作者>
  <定价>
    ¥28
  </定价>
</书>
```

这段例子中有 5 种类型的元素，元素是在 DTD 中定义的，上面看到的只是元素实例。例如“书”元素实例，由起始标记“<书>”开始，到结束标记“</书>”结束，中间是元素内容，即 4 种子元素。其中每一种子元素又分别包括起始标记、字符内容和结束标记。

这样在进行计算机自动化处理时，例如要查找一本名叫“XML 的历史”的书的价格，搜索引擎只需查找这样一个“书”元素，它的“书名”子元素的内容为“XML 的历史”，然后将此“书”元素的“定价”子元素中的内容返回即可。而在搜索时还可以加以限制，例如要查找的书的作者必须为某某，或者返回价钱的同时还要同时返回书的摘要，这些在 SGML 或 XML 文档中都是很容易实现的。

你也许已经看出来了，通用标记文档和数据库很相像。确实，它在很多方面都与传统的数据库数据相似，它的很多特性与数据库也是一样的。例如它们都可以保存抽象数据，并把它们和显示格式信息分离开来。

在通用标记文档中，可以包含各种各样的数据，并且用元素来表示数据间的结构。我们可以说 SGML（或 XML）文档不是传统意义上的文档，而是一种数据库化的文档。这一特性使它可以应用于 Web 数据库和电子数据交换(EDI)中。

(3) 在文档类型定义中，要严格定义文档结构所必须遵循的规则，只有这样，才能使计算机系统能够有根据地检查文档的合法性，以及可靠地处理文档。

例如，对于上面的例子，我们可以这样规定文档类型：关于一本书，必须要描述的信