



21世纪高等学校计算机
应用技术规划教材

XML基础 及实践开发教程 (第2版)



◎ 唐琳 刘彩虹 主编
肖紫薇 张坤 副主编

零基础入门

项目式教学

教学资源

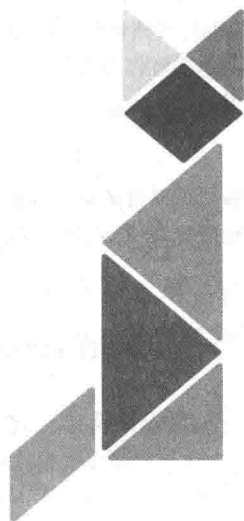
清华大学出版社





21世纪高等学校
应用技术规划教材

XML基础 及实践开发教程 (第2版)



◎ 唐琳 刘彩虹 主编
肖紫薇 张坤 副主编

常州大学图书馆
藏书章

清华大学出版社
北京

内 容 简 介

本书系统地介绍 XML 基本语法及相关技术。

全书共分四部分 13 章：第一部分为 XML 基础，包括 XML 入门、在 XML 中使用 DTD、命名空间、在 XML 中使用 Schema、Schema 高级技术；第二部分为 XML 的显示技术，包括 XML 的显示技术之 CSS、XPath、XSLT；第三部分为 XML 的检索及其应用，包括 XQuery 基础、XQuery 应用；第四部分为基于 Java 的 XML 文档解析技术，包括 DOM、SAX、JDOM 和 DOM4J。

本书可作为高等院校计算机、电子商务以及信息等相关专业的教材，也可作为 IT 从业人员的自学参考书。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目(CIP)数据

XML 基础及实践开发教程/唐琳,刘彩虹主编.—2 版.—北京:清华大学出版社,2018
(21 世纪高等学校计算机应用技术规划教材)
ISBN 978-7-302-47412-8

I. ①X… II. ①唐… ②刘… III. ①可扩展标记语言—程序设计—高等学校—教材 IV. ①TP312

中国版本图书馆 CIP 数据核字(2017)第 127081 号

责任编辑：魏江江 王冰飞

封面设计：刘 键

责任校对：时翠兰

责任印制：沈 露

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>, <http://www.wqbook.com>

地 址：北京清华大学学研大厦 A 座 邮 编：100084

社总机：010-62770175 邮 购：010-62786544

投稿与读者服务：010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载：<http://www.tup.com.cn>, 010-62795954

印 装 者：三河市金元印装有限公司

经 销：全国新华书店

开 本：185mm×260mm 印 张：22.25 字 数：545 千字

版 次：2013 年 10 月第 1 版 2018 年 1 月第 2 版 印 次：2018 年 1 月第 1 次印刷

印 数：5601~7600

定 价：49.50 元

产品编号：067478-01

随着我国改革开放的进一步深化,高等教育也得到了快速发展,各地高校紧密结合地方经济建设发展需要,科学运用市场调节机制,加大了使用信息科学等现代科学技术提升、改造传统学科专业的投入力度,通过教育改革合理调整和配置了教育资源,优化了传统学科专业,积极为地方经济建设输送人才,为我国经济社会的快速、健康和可持续发展以及高等教育自身的改革发展做出了巨大贡献。但是,高等教育质量还需要进一步提高以适应经济社会发展的需要,不少高校的专业设置和结构不尽合理,教师队伍整体素质亟待提高,人才培养模式、教学内容和方法需要进一步转变,学生的实践能力和创新精神亟待加强。

教育部一直十分重视高等教育质量工作。2007年1月,教育部下发了《关于实施高等学校本科教学质量与教学改革工程的意见》,计划实施“高等学校本科教学质量与教学改革工程(简称‘质量工程’)”,通过专业结构调整、课程教材建设、实践教学改革、教学团队建设等多项内容,进一步深化高等学校教学改革,提高人才培养的能力和水平,更好地满足经济社会发展对高素质人才的需要。在贯彻和落实教育部“质量工程”的过程中,各地高校发挥师资力量强、办学经验丰富、教学资源充裕等优势,对其特色专业及特色课程(群)加以规划、整理和总结,更新教学内容、改革课程体系,建设了一大批内容新、体系新、方法新、手段新的特色课程。在此基础上,经教育部相关教学指导委员会专家的指导和建议,清华大学出版社在多个领域精选各高校的特色课程,分别规划出版系列教材,以配合“质量工程”的实施,满足各高校教学质量和教学改革的需要。

本系列教材立足于计算机公共课程领域,以公共基础课为主、专业基础课为辅,横向满足高校多层次教学的需要。在规划过程中体现了如下一些基本原则和特点。

(1) 面向多层次、多学科专业,强调计算机在各专业中的应用。教材内容坚持基本理论适度,反映各层次对基本理论和原理的需求,同时加强实践和应用环节。

(2) 反映教学需要,促进教学发展。教材要适应多样化的教学需要,正确把握教学内容和课程体系的改革方向,在选择教材内容和编写体系时注意体现素质教育、创新能力与实践能力的培养,为学生的知识、能力、素质协调发展创造条件。

(3) 实施精品战略,突出重点,保证质量。规划教材把重点放在公共基础课和专业基础课的教材建设上;特别注意选择并安排一部分原来基础比较好的优秀教材或讲义修订再版,逐步形成精品教材;提倡并鼓励编写体现教学质量和教学改革成果的教材。

(4) 主张一纲多本,合理配套。基础课和专业基础课教材配套,同一门课程可以有针对不同层次、面向不同专业的多本具有各自内容特点的教材。处理好教材统一性与多样化,基本教材与辅助教材、教学参考书,文字教材与软件教材的关系,实现教材系列资源配套。

(5) 依靠专家,择优选用。在制定教材规划时依靠各课程专家在调查研究本课程教材建设现状的基础上提出规划选题。在落实主编人选时,要引入竞争机制,通过申报、评审确定主题。书稿完成后要真实行审稿程序,确保出书质量。

繁荣教材出版事业,提高教材质量的关键是教师。建立一支高水平教材编写梯队才能保证教材的编写质量和建设力度,希望有志于教材建设的教师能够加入到我们的编写队伍中来。

21世纪高等学校计算机应用技术规划教材
联系人:魏江江 weijj@tup.tsinghua.edu.cn

XML(eXtensible Markup Language,可扩展标记语言)作为一种优秀的数据存储形式、数据交换方式被广泛接受和应用。本书较全面地介绍了 XML 语言及其相关技术,介绍知识的同时更注重实际的应用。

本书共分为四部分 13 章。

第一部分 XML 基础(XML 语法、DTD 和 Schema)

第 1 章 XML 入门:介绍 XML 的基础知识、XML 的语法知识,并介绍 XML 的相关知识。

第 2 章 在 XML 中使用 DTD:介绍 DTD 的基础知识、DTD 的引入方法、DTD 的语法组成,重点讲解 DTD 中元素、属性、实体和符号的语法,最后介绍 DTD 存在的问题。

第 3 章 命名空间:介绍命名空间的基本概念,命名空间的作用和意义,命名空间的声明、作用范围,最后介绍带有命名空间的情况下对 DTD 的影响。

第 4 章 在 XML 中使用 Schema:介绍 Schema 的基本概念,Schema 对于命名空间的支持。通过对 Schema 语法详细的讲解,使读者学会编写和使用 Schema。

第 5 章 Schema 高级技术:对 Schema 的高级特性进行讲解,主要包括三部分内容,分别是 Schema 的高级特性、Schema 的复用和 Schema 的简单实践技巧。

第二部分 XML 显示技术(CSS 和 XSLT)

第 6 章 XML 的显示技术之 CSS:介绍 XML 的显示技术 CSS 和 XSL 的基本概念,重点讲解在 XML 中如何使用 CSS 作为其显示技术,并介绍 CSS 的语法知识。

第 7 章 XPath:介绍 XPath 的基础知识,重点讲解 XPath 的完整路径结构,轴、节点测试和谓词。对实际应用中经常使用的简化路径进行了深入的讲解。

第 8 章 XSLT:介绍 XSLT 的基础知识,掌握 CSS 与 XSLT 的区别。重点讲解 XSLT 的语法知识,最后介绍 XSLT 的复用以及在 XSLT 中使用脚本语言。

第三部分 XML 的检索及其应用

第 9 章 XQuery 基础:介绍 XQuery 的基础知识、XQuery 的处理过程,重点讲解 XQuery 的基本语法。

第 10 章 XQuery 应用:讲解 XQuery 的应用,包括在 Java 及数据库中的应用,并以 Oracle Berkeley DB XML 为例进行讲解。

第四部分 基于 Java 的 XML 文档解析技术(DOM、SAX、JDOM 和 DOM4J)

第 11 章 DOM:介绍 XML 文档解析技术,重点讲解 DOM,并基于 JAXP 讲解如何使用 DOM 对文档进行解析。

第 12 章 SAX:介绍 SAX 的基础知识、SAX 的工作机制,重点讲解 JAXP 中的 SAX 接口和类,最后一节对常用的接口和类的使用方法进行实践。

第13章 JDOM 和 DOM4J: 介绍 JDOM 和 DOM4J 的基础知识,后面的章节分别介绍使用 JDOM 和 DOM4J 对 XML 文档进行解析、创建和修改的具体方法。

本书的特点

1. 案例驱动

本书是基于校企合作的教材,每个知识点都通过案例进行说明,帮助读者更深刻地理解知识点。贴近于行业应用,保证了教材内容理论与实际的有机结合,既能符合学校教育的要求,又能够反映企业岗位特征及企业生产技术的特点,注重知识的实用性与实效性,结合市场导向,注重培养学生实践技能。

2. 注重实践能力

通过案例提升读者解决实际问题的能力。

3. 练习充分

每一章均配有习题,帮助读者进一步理解本章的内容。

本书的编写分工如下:刘彩虹负责编写第1、2、4章;张坤负责编写第3章;肖紫薇负责编写第6章;其他章节由唐琳编写。

本书可以作为高等院校计算机、电子商务以及信息等相关专业的教材,也可作为 IT 从业人员的自学参考书。

由于时间仓促,不妥之处欢迎读者批评指正。

编者

2017年10月

第 1 章 XML 入门	1
1.1 了解 XML	1
1.1.1 第一个 XML 文档	1
1.1.2 XML 的发展历史	2
1.1.3 XML 与其他标记语言相比较	3
1.1.4 XML 编辑工具	5
1.2 XML 的语法基础	7
1.2.1 XML 的文档分类	7
1.2.2 XML 的文档组成	11
1.2.3 XML 的基本语法规则	16
1.3 XML 的元素构成	17
1.3.1 元素的形式	17
1.3.2 元素的内容	18
1.4 XML 相关技术及不同用途下的类似技术	21
1.4.1 XML 的相关技术	21
1.4.2 XML 的类似技术	24
1.5 本章小结	30
习题 1	30
第 2 章 在 XML 中使用 DTD	32
2.1 DTD 介绍	32
2.1.1 DTD 概述	32
2.1.2 DTD 的基本语法	33
2.1.3 引入 DTD 的方式	36
2.1.4 使用 XMLSpy 创建 DTD	38
2.2 DTD 中的元素	40
2.2.1 元素定义语法	40
2.2.2 元素类型	41
2.3 DTD 中的属性	47
2.3.1 属性定义语法	47
2.3.2 属性类型	48
2.4 DTD 中的实体和符号	51

2.4.1	实体	51
2.4.2	符号	58
2.5	使用 XMLSpy 做 DTD 与 XML 转换	58
2.5.1	根据 XML 文件产生 DTD	58
2.5.2	根据 DTD 文件产生 XML	61
2.6	DTD 的优缺点	63
2.7	本章小结	64
习题 2	64
第 3 章	命名空间	67
3.1	命名空间概述	67
3.2	命名空间作用域	70
3.3	元素对命名空间的使用	71
3.4	属性对命名空间的使用	72
3.5	DTD 对命名空间的支持	73
3.6	本章小结	74
习题 3	74
第 4 章	在 XML 中使用 Schema	76
4.1	Schema 概述	76
4.1.1	Schema 基础知识	76
4.1.2	第一个 Schema 文件	77
4.2	Schema 的引用方法	78
4.3	Schema 的语法结构	81
4.3.1	元素	81
4.3.2	属性	85
4.3.3	注释	87
4.4	Schema 的数据类型	88
4.4.1	内置数据类型	89
4.4.2	用户自定义数据类型	100
4.5	本章小结	115
习题 4	115
第 5 章	Schema 高级技术	117
5.1	Schema 的高级特性	117
5.1.1	元素的替换	117
5.1.2	抽象元素和抽象类型	118
5.1.3	限制替换元素和限制派生类型	120
5.1.4	限制替换类型	124

5.1.5 元素和属性的约束	126
5.2 Schema 的复用	128
5.2.1 使用 include 元素复用 Schema	128
5.2.2 使用 redefine 元素复用 Schema	131
5.2.3 使用 import 元素复用 Schema	133
5.3 Schema 实践技巧——空元素的表示	135
5.4 本章小结	136
习题 5	137
第 6 章 XML 的显示技术之 CSS	138
6.1 XML 的显示技术	138
6.2 在 XML 中引入 CSS	139
6.3 CSS 的基本语法	141
6.3.1 CSS 语法	141
6.3.2 CSS 属性	142
6.3.3 CSS 单位	144
6.3.4 CSS 选择器	145
6.3.5 CSS 实践	145
6.4 本章小结	147
习题 6	148
第 7 章 XPath	149
7.1 XPath 概述	149
7.2 XPath 结点	149
7.3 XPath 路径	151
7.3.1 轴	153
7.3.2 XPath 结点测试	156
7.3.3 谓词	160
7.3.4 简化路径	162
7.4 XPath 运算符	164
7.5 XPath 函数	165
7.6 表达式	171
7.7 本章小结	173
习题 7	173
第 8 章 XSLT	175
8.1 XSLT 概述	175
8.1.1 XSLT 的基本概念	175
8.1.2 使用 XMLSpy 工具创建 XSLT	176

8.1.3	第一个 XSLT	178
8.2	在 XML 中引用 XSLT	181
8.3	XSLT 的转换模式	182
8.4	XSLT 的基本语法	184
8.4.1	XSLT 文档结构	184
8.4.2	output 标签	185
8.4.3	模板及模板调用	187
8.4.4	转换为 HTML 文档常用标记	197
8.4.5	转换为 XML 文档常用标记	202
8.5	XSLT 的复用	206
8.6	XSLT 进阶	210
8.6.1	多 XML 文档输入	211
8.6.2	多 XML 文档输出	214
8.6.3	自定义函数	215
8.6.4	分组重排	217
8.6.5	字符串处理	221
8.6.6	XSLT 其他常用标记	223
8.7	本章小结	226
	习题 8	226
第 9 章	XQuery 基础	228
9.1	XQuery 介绍	228
9.2	第一个 XQuery	229
9.2.1	路径表达式	229
9.2.2	FLWOR 表达式	230
9.3	XQuery 的处理过程	232
9.4	XQuery 基本语法	237
9.4.1	基本表达式	237
9.4.2	比较表达式	237
9.4.3	条件表达式	240
9.4.4	逻辑表达式	241
9.4.5	构造器	241
9.4.6	FLWOR	243
9.4.7	量化表达式	245
9.4.8	序列表达式及其操作	246
9.4.9	类型相关表达式	247
9.4.10	运算表达式	250
	习题 9	251

第 10 章 XQuery 应用	253
10.1 在 Java 中使用 XQuery	253
10.1.1 XQJ 介绍	253
10.1.2 使用 Saxon 编程	254
10.2 XQuery 在 XML 数据库中的应用	260
10.2.1 XML 数据库介绍	260
10.2.2 原生 XML 数据库中的 BDB XML 介绍	260
10.2.3 XQuery 在 BDB XML 中的应用实例	262
习题 10	265
第 11 章 DOM	266
11.1 XML 文档解析技术	266
11.1.1 XML 文档解析技术概述	266
11.1.2 DOM 与 SAX 相比较	267
11.1.3 JAXP	268
11.2 使用 DOM 解析 XML 文档	268
11.3 DOM 接口及其应用	271
11.3.1 DOM 的核心概念——结点	271
11.3.2 使用 JAXP 通过 DOM 解析 XML 文档	273
11.3.3 使用 JAXP 通过 DOM 输出 XML 文档	278
11.3.4 使用 JAXP 通过 DOM 修改 XML 文档	281
11.4 本章小结	284
习题 11	285
第 12 章 SAX	286
12.1 SAX 概述	286
12.1.1 SAX 基础知识	286
12.1.2 第一个 SAX 程序	288
12.2 使用 SAX 解析 XML 文档	291
12.2.1 XMLReader 和 XMLReaderFactory	291
12.2.2 SAXParser 和 SAXParserFactory	294
12.3 SAX 接口及其应用	294
12.3.1 ContentHandler 接口	294
12.3.2 Attributes 和 Attributes2 接口	302
12.3.3 ErrorHandler 接口	304
12.3.4 DTDHandler 和 DeclHandler 接口	311
12.3.5 EntityResolver 和 EntityResolver2 接口	311
12.3.6 LexicalHandler 接口	313

12.4	DefaultHandler 和 DefaultHandler2 类开发实践	318
12.5	本章小结	324
	习题 12	324
第 13 章	JDOM 和 DOM4J	326
13.1	JDOM 和 DOM4J 概述	326
13.1.1	JDOM 基础知识	326
13.1.2	DOM4J 基础知识	327
13.1.3	DOM4J 与 JDOM 相比较	328
13.2	使用 JDOM 对 XML 文档进行操作	328
13.2.1	使用 JDOM 解析 XML 文档	328
13.2.2	使用 JDOM 创建 XML 文档	330
13.2.3	使用 JDOM 修改 XML 文档	332
13.3	使用 DOM4J 对 XML 文档进行操作	334
13.3.1	使用 DOM4J 解析 XML 文档	334
13.3.2	使用 DOM4J 创建 XML 文档	338
13.3.3	使用 DOM4J 修改 XML 文档	340
13.4	本章小结	342
	习题 13	342
	参考文献	343

本章学习目标

- 了解 XML 的发展历史
- 掌握 XML 的文档结构
- 熟练掌握 XML 的基本语法
- 掌握 XML 的用途及相关技术

本章先向读者介绍 XML 的基础知识,重点讲解 XML 的语法知识,最后介绍 XML 的其他相关的知识。

1.1 了解 XML

1.1.1 第一个 XML 文档

XML 的全称是 eXtensible Markup Language,即可扩展标记语言,它提供了一套跨平台、跨网络、跨程序语言的数据描述方式。XML 是一个简单的、基于文本格式表示的结构化信息(包括文件、数据、配置、书籍、交易、发票等)。XML 文档本身是非常简单、易于理解的。下面来看一个 XML 文档实例,见代码 1-1。

代码 1-1 简单的 XML 文档

```
<?xml version = "1.0" encoding = "UTF - 8"?>
<个人信息>
  <姓名>田诗琪</姓名>
  <生日>2011 - 04 - 11 </生日>
  <性别>女</性别>
  <身高>83cm </身高>
</个人信息>
```

将该文档保存为扩展名为.xml的文件,可以使用浏览器查看该文件。浏览器显示 XML 文档的原理,实际上是将该文档转换为文档树结构并显示出来。在此使用 IE 浏览器显示,如图 1-1 所示。

虽然我们目前还没有学习 XML,但是查看上面的文档即可了解到,该文档包含了一个人的个人基本信息,通过标记将内容信息有条理地保存下来。代码 1-1 主要包括两个部分,即文档声明和文档内容。

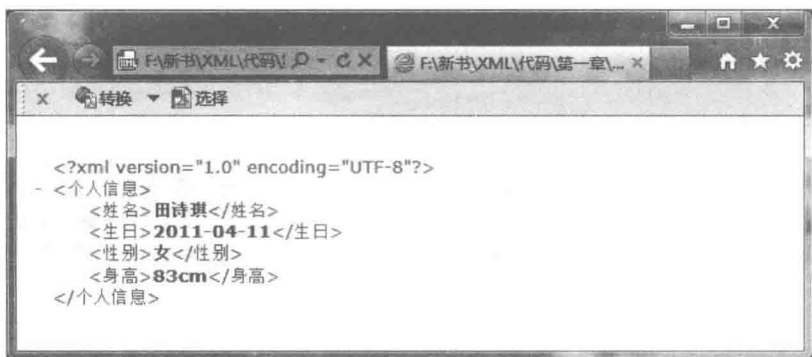


图 1-1 XML 文档显示结果页面

文档声明是一个 XML 文档所必需的,必须位于文档的第一行,形如<?xml version = "1.0" encoding = "UTF-8"?>。这一行代码会告诉解析器或浏览器,这个文件应该按照 XML 规则进行解析。当前文档声明信息包括使用的是 XML 1.0 版本,所使用的编码方式为 UTF-8。

文档内容根据实际内容的不同而不同,内容是通过标记有序地展现出来的,标记是用来描述内容结构的元数据,不能包含内容。

(1) 每个标记都必须包括开始和结束标记,开始标记以“<”开头,并以“>”结束,中间为元素名称;结束标记以“</”开头,并以“>”结束,开始标记和结束标记之间的内容为标记内容。

(2) 开始标记和结束标记中间的内容为元素名称,名称大小写敏感。对于文档中标记的名称,用户可以根据实际需要自己定义。例如<姓名>这个标记的名字,就是由编者来定义的,在编写的时候也可以使用<名字>或< name >等,但开始标记和结束标记的元素名称必须相同,如图 1-2 所示。

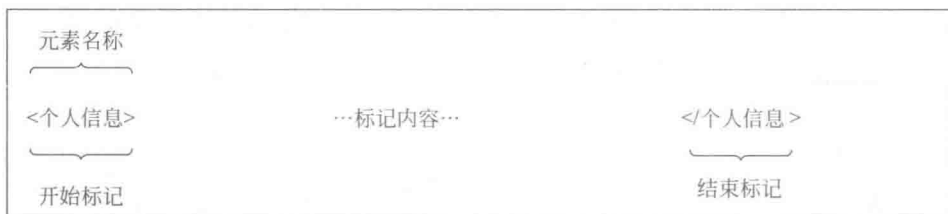


图 1-2 XML 文档元素

(3) 一个 XML 文档有且只能有一个根标记,其他所有标记必须嵌套在当前标记内部,本文档的根标记为<个人信息>。标记必须合理嵌套,一个 XML 文档中所有的标记会形成文档树结构,本文档的文档树结构如图 1-3 所示。

1.1.2 XML 的发展历史

在 20 世纪 80 年代早期,IBM 公司提出在各文档之间共享一些相似的属性,例如字体大小和版面。IBM 设计了一种文档系统,文档中能够明确地将标识和内容分开。所有文件



图 1-3 XML 文档树结构

的标识使用同一种方法,IBM把这种标识语言称为通用标记语言(Generalized Markup Language),即GML。

经过若干年的发展,1984年国际标准化组织(ISO)开始对其提案进行讨论,并于1986年正式发布了为标准化文档而定义的标记语言标准(ISO 8879),成为新的标准通用标记语言,即SGML。SGML的功能非常强大,是可以定义标记语言的元语言。但是它过于严谨的文件描述法,导致其难以理解和学习,影响了它的推广和应用。

一个非常成功的SGML应用就是HTML。HTML作为SGML的子集,关注的是数据的显示,而不是数据本身。HTML限制定义新的标签,显然已经背离了SGML的思想,但是在实际应用中取得了成功。HTML使因特网迅速走出了实验室,成为人人皆可使用的工具,使Web几乎成为因特网的全部内涵。随着因特网的日趋成熟,新的需求已经形成,需要一种语言能满足更复杂和更大规模的用途。

1998年2月,W3C发布了XML 1.0标准,后来又修订了4次,2007年11月发布了XML 1.0第5版。其目的是在Web上能以现有的超文本标记语言(HTML)的使用方式提供、接收和处理通用的SGML。XML是SGML的一个简化子集,它以一种开放的、自我描述的方式定义了数据结构,在描述数据内容的同时能突出对结构的描述,从而体现出数据与数据之间的关系。W3C组织于2004年2月4日发布了XML 1.1的推荐标准,后来又于2006年8月发布了第2版,这是最新的XML版本。不过,目前大多数的应用还是基于XML 1.0的推荐标准。

XML正式规范1.0中阐述了XML的10个设计目标:

- (1) XML应该可以直接用于因特网。
- (2) XML应该支持广泛的、多样化的应用程序。
- (3) XML应该与SGML兼容。
- (4) XML文档的处理程序应该容易编写。
- (5) XML中的可选项应尽可能少,在理想状态下应为零。
- (6) XML文档应该清晰明了、可读性强。
- (7) XML应易于设计。
- (8) XML的设计应该正式而且简洁。
- (9) XML文档应该易于创建。
- (10) XML标记的间接性应在保证前面所有目标要求的前提下予以重视。

注意:W3C是万维网联盟(World Wide Web Consortium)英文的缩写,万维网联盟成立于1994年10月,以开放论坛的方式来促进开发互通技术(包括规格、指南、软件和工具),开发网络的全部潜能。万维网联盟(W3C)从1994年成立以来,已发布了90多份Web技术规范,领导着Web技术向前发展。W3C认为自身不是官方组织,因此将它正式发布的规范称为推荐(建议)标准,意思是进一步标准化的建议,但是由于该组织自身的权威性往往成为事实上的标准。

1.1.3 XML与其他标记语言相比较

在将XML与其他标记语言比较之前,我们先来了解一下什么是标记语言。标记语言

(Markup Language)指用一系列约定好的标记对电子文档进行标记,以实现电子文档的语义、结构及格式的定义。标记语言包括很多种语言,它们的发展及相互关联如图 1-4 所示。

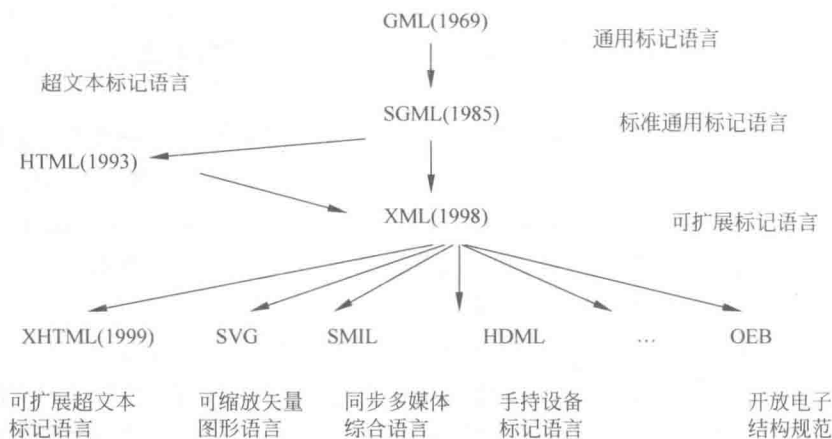


图 1-4 标记语言的发展

XML 与 HTML 是标记语言体系中最成功的两种标记语言,因为它们的结构有些相似,常被人们拿来作比较。但是它们处于不同层次,对于新闻、网络日记、论坛留言等大部分短期数据,HTML 仍是在 Web 上快速“出版”数据的最简单的方法。如果数据要长期使用,并且需要更多的一些结构,我们推荐使用 XML。

HTML 与 XML 的主要区别如表 1-1 所示。

表 1-1 HTML 与 XML 的区别

	HTML	XML
功能	用于数据显示	用于描述数据和保存数据
标记	标记是固定不变的	没有固定标记,根据实际需要进行定义,XML 本身是一种元语言
语法	语法要求很宽松	有严格的语法要求

SGML、HTML 和 XML 之间又是什么关系呢? SGML 在 Web 发明之前就早已存在,它是一种定义标记语言的元语言。HTML 和 XML 都是从 SGML 发展而来的标记语言,因此,它们有一些共同点,例如相似的语法和标记的使用。不过 HTML 是在 SGML 定义下的一种描述性语言,是 SGML 的一个应用,其 DTD 作为标准被固定下来,而 XML 是 SGML 的一个简化版本,是 SGML 的一个子集。从严格意义上来说,XML 仍然是 SGML。

HTML 不能用来定义新的应用,而 XML 可以。例如 RDF 和 CDF 都是使用 XML 定义的应用,事实上 XML 和 SGML 是兼容的,但又没有 SGML 那么复杂。XML 规范的定义者之一 Tim Bray 说,XML 的设计出发点是取 SGML 的优点,取出复杂的部分使其保持轻巧,可以在 Web 上工作。

HTML、SGML 和 XML 将继续用于其适合的地方,它们的任何一个都不会使其他一个废弃。不同于 HTML 和 XML,SGML 可能永远不会在因特网上被广泛接受,因为它不是为某个网络协议而设计的,也从来没有为某个网络协议的需求而优化过。对于高端的、复杂