



普通高等教育“十一五”国家级规划教材 计算机系列教材

XML基础与Ajax实践教程 (第2版)

王震江 马 宏 编著

清华大学出版社

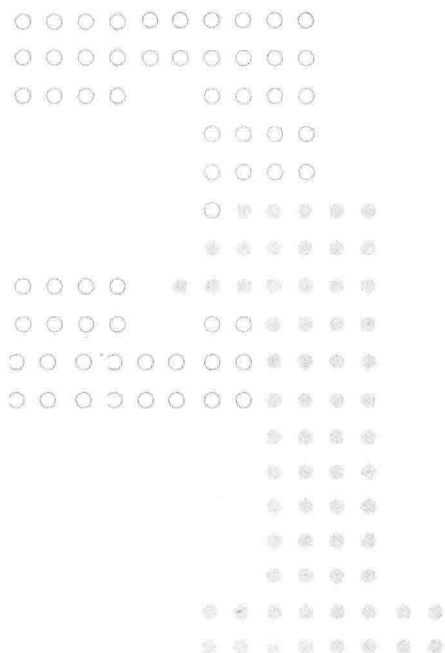




普通高等教育“十一五”国家级规划教材 计算机系列教材

王震江 马 宏 编著

XML基础与Ajax实践教程 (第2版)



清华大学出版社

北京

内 容 简 介

本书共 13 章,介绍了 XML 1.0 语法、XML Schema 2.0、XPath 3.0、XDM 3.0、XQuery Functions 与 Operators 3.0、CSS 2.0、XSL 2.0、XML DOM Level 3、XML 数据库技术、JavaScript 基础等,最后结合书中的各种技术给出了一个 Ajax 开发实例。本书实例丰富,且各章都提供了习题与实验,用来复习本章知识点。

本书适合作为大专院校计算机、软件工程、电子商务、网络工程等相关专业的教材,也可作为各类 IT 技术培训的教材和 IT 从业人员的参考书。

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 基础与 Ajax 实践教程/王震江等编著. —2 版. —北京:清华大学出版社,2016

计算机系列教材

ISBN 978-7-302-42022-4

I. ①X… II. ①王… III. ①可扩展语言—程序设计—高等学校—教材 ②计算机网络—程序设计—高等学校—教材 IV. ①TP312 ②TP393.09

中国版本图书馆 CIP 数据核字(2015)第 263220 号

责任编辑:付弘宇 李 晔

封面设计:常雪影

责任校对:梁 毅

责任印制:王静怡

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:三河市少明印务有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:26.5 字 数:659 千字

版 次:2011 年 9 月第 1 版 2016 年 1 月第 2 版 印 次:2016 年 1 月第 1 次印刷

印 数:1~2000

定 价:49.00 元

产品编号:066719-01

第2版前言

FOREWORD

XML(eXtensible Markup Language) 1.0 技术规范从 1998 年 2 月 10 日问世以来,在 XML 应用技术方面已经发生了翻天覆地的变化,XML 作为新一代的数据交换标准和交换文本已经成为今天网络数据交换的新标准,基本取代了上一代的数据交换标准 EDI(Electronic Data Interchange)。2004 年 10 月 28 日发布的 XML Schema 2.0 技术已经在应用方面日趋成熟,现在在 W3C 发布的新的 XML 各种规范中,出现了大量以 XML Schema 数据模型为基础的描述,标志着 XML 半结构数据库技术日趋成熟。XML 的应用层出不穷,已经在计算机信息处理、非结构化数据表示、异种平台数据交换与传输、Web 编程、网络应用编程、跨平台编程技术等方面得到十分广泛的应用。作为一类文本文件,XML 文档和数据可以在 Internet 上不受任何阻挡,作为各种应用的数据基础承载各种应用。以 XML 为内核的半结构数据库技术,在 Web 和网络应用程序中提供了互操作性好、数据表现灵活、数据类型丰富等特点,并因此受到业界的广泛支持,使得支撑 XML 的各种技术规范和应用(如 Ajax)不断推进和完善,并受到行业的大力支持和软件编程人员的欢迎。

此次改版的主要改动如下:XML 1.0 以第 5 版为基础,第 2、3 章的内容有所增改。XPath 部分有了很大的改动,XPath 采用 3.0 版本,还增加了 XQuery 和 XPath Data Model (XDM) 3.0、XPath 和 XQuery Functions 与 Operators 3.0。XSLT 和 XML DOM Level 3 部分则更改了部分错误。从核心技术上讲,Ajax 不是什么新技术,而是基于 XML 技术的综合应用技术,目前应用已经十分普遍。

本书共分 13 章。第 1 章概述 XML 的基础知识、编辑工具、相关技术及发展。第 2 章讲述 XML 1.0 规范的基本技术要求和规定,对 XML 的元素、属性、实体等重要概念进行描述。第 3 章讲述 XML 文档的元素、元素类型及其定义、XML 1.0 提供的内容模型定义等。第 4 章讲述 XML Schema 2.0,包括定义完整的 XML 数据类型,XML Schema 结构,设计 XML Schema 文档等内容,是 XML 数据库的基础部分。第 5 章讲述用于 XML 文档的查询规范 XPath 3.0、XDM 3.0、XPath 和 XQuery Functions 与 Operators 3.0。第 6 章简要介绍 HTML 4.0 的制表、表单、框架、超级链接、字符控制与多媒体等,为后续的章节提供基础。第 7 章讲述 CSS 2.0 技术规范和要求,用 CSS 转换 XML 文档的方法和技巧。第 8 章讲述专门用于 XML 转换的技术 XSL 2.0,包括样式表、模板规则、样式表设计技术和实现等。第 9 章讲述 XML DOM Level 3,包括 DOM 内核,以及 XML DOM 应用和编程。第 10 章讲述用 ASP 和 ADO 实现 XML 数据库的连接,XML 数据库技术简介。第 11 章,讨论 JavaScript 的数据类型、程序控制(分支、循环)、函数等基础。第 12 章讨论 XML HTTPRequest 对象,如何建立 Web 连接,通过 HTTPRequest 访问 XML 文档。第 13 章给出一个实例,解决如何在 .NET

平台利用 XML、JavaScript、HttpRequest 进行 Ajax 实例设计和编程问题。

为了方便读者学习,第 1 章~第 12 章附有习题,通过练习和实验,可以帮助读者理解 XML 技术及其应用的各个方面。

全书由王震江和马宏编著,由王震江审核。王震江编著第 1~第 12 章,马宏编著第 13 章。本书的实例全部通过上机调试。参与本书研究工作和资料整理的人员有彭嘉凤、方刚、俞锐刚、王玉见、欧晓明、李燕,在此向他们表示感谢。由于编者水平有限,错误和疏漏之处在所难免,衷心希望广大读者给予批评指正。

本书的配套电子课件等资源可以从清华大学出版社网站 www.tup.com.cn 下载,本书及课件的相关问题请联系 fuhy@tup.tsinghua.edu.cn。

编 者

2015 年 6 月

第 1 版前言

FOREWORD

XML(eXtensible Markup Language) 1.0 技术规范已经问世 13 年。伴随 XML 1.0 的问世,相继出现了支持 XML 数据定义、表示、格式化的各种技术规范,从早期的只有简单数据类型的 DTD(Documents Type Definition)发展到可以表示 44 种数据类型的 XML Schema,现在 XML 数据可以表示所有传统关系数据库能表示的数据类型,使得 XML 具有更广泛的应用领域。从使用 CSS 到使用 XSLT 技术对 XML 进行格式化,为 XML 数据的前台表示奠定了坚实的基础,这样对于不考虑数据表现而仅考虑数据定义的 XML 而言,在 Web 应用和网络编程方面就得到快速的发展和广泛的应用。XML 作为新一代的数据交换标准,已经在计算机信息处理、非结构化数据表示、异种平台数据交换与传输、Web 编程、网络应用编程、跨平台编程技术等方面得到广泛应用。作为一类文本文件,XML 文档和数据可以在 Internet 上不受任何阻挡,作为各种应用的数据基础承载各种应用,因而,以 XML 为内核的半结构数据库技术在 Web 和网络应用程序中因为互操作性好、数据表现灵活、数据类型丰富等特点而受到业界的广泛支持,使得支撑 XML 的各种技术规范和应用(如 Ajax)不断推进和完善,并受到行业的大力支持和软件编程人员的欢迎。

本书共分 13 章。第 1 章概述 XML 的基础知识、编辑工具、相关技术及其发展、Ajax 技术简介。第 2 章讲述 XML 1.0 规范的基本技术要求和规定,对 XML 的元素、属性、实体等重要概念进行描述。第 3 章讲述 XML 文档的元素、元素类型及其属性定义、XML 1.0 提供的内容模型定义等。第 4 章讲述 XML Schema 2.0,包括定义完整的 XML 数据类型、XML Schema 结构、设计 XML Schema 文档等内容,是 XML 数据库的基础部分。第 5 章讲述用于 XML 文档的查询规范 XPath 2.0。第 6 章简要介绍 HTML 4.0 的制表、表单、框架、超链接、字符控制与多媒体等,为后续的章节提供基础。第 7 章讲述 CSS 2.0 技术规范和要求,用 CSS 转换 XML 文档的方法和技巧。第 8 章讲述专门用于 XML 转换的技术 XSL 2.0,包括样式表、模板规则、样式表设计技术和实现等。第 9 章讲述 XML DOM Level 3,包括 DOM 内核,以及 XML DOM 应用和编程。第 10 章讲述用 ASP、ADO 技术实现 XML 与数据库连接,以及 XML 数据库技术。第 11 章讨论 JavaScript 的数据类型、程序控制(分支、循环)、函数等基础,为 Ajax 编程做准备。第 12 章讨论 XML HttpRequest 对象,如何建立 Web 连接,通过 HttpRequest 访问 XML 文档。第 13 章给出一个实例,解决如何在 .NET 平台下利用 XML、JavaScript、HttpRequest 进行 Ajax 实例设计和编程问题。

为了方便读者学习,每章都附有习题,通过练习和实验,可以帮助读者理解 XML 技术及其应用的各个方面。

本书第 1 章~第 12 章由王震江编著,第 13 章由马宏编著。本书的实例全部通过上机调

试。全书由王震江统稿和审核。参与本书研究工作和资料整理的人员有彭嘉凤、王武、方刚、俞锐刚、马宏、王玉见、欧晓明、李燕,在此向他们表示感谢。由于编者水平有限,错误和疏漏之处在所难免,衷心希望广大读者给予批评指正。

本书的课件及书中提到的网络资源可以从清华大学出版社网站(www.tup.com.cn)下载,本书和课件等资源的使用中如有问题,请联系 fuhy@tup.tsinghua.edu.cn。

编 者

2011 年 8 月

第 1 章 概述	1
1.1 XML 技术简介	1
1.1.1 XML 的历史	1
1.1.2 XML 与 HTML 的比较	2
1.1.3 XML 的编辑工具	4
1.1.4 浏览 XML 文档	5
1.1.5 XML 的相关技术及其发展	7
1.2 Ajax 技术简介	11
1.2.1 什么是 Ajax	11
1.2.2 Ajax 运行模式	11
1.2.3 Ajax 所使用的技术	14
1.2.4 使用 Ajax 可以干什么	14
习题 1	15
第 2 章 XML 的语法基础	16
2.1 XML 的语法基础	16
2.1.1 结构良好性	17
2.1.2 XML 声明	17
2.1.3 XML 的元素	19
2.1.4 属性	21
2.1.5 标记与字符数据	22
2.1.6 实体	22
2.1.7 处理指令	27
2.1.8 CDATA 节	28
2.1.9 注释	30
2.2 XML 的文档结构	31
2.2.1 XML 的序言	31
2.2.2 XML 数据	32
2.3 名称空间	34
2.3.1 名称空间的概念	34

2.3.2 声明名称空间	35
2.4 语言标识	37
2.4.1 国家代码和语言代码	37
2.4.2 覆盖	39
习题 2	40
第 3 章 文档类型定义	41
3.1 概述	41
3.2 元素	42
3.2.1 元素声明	42
3.2.2 元素类型	45
3.2.3 元素取值	46
3.3 属性	46
3.3.1 基本概念	46
3.3.2 属性	48
3.3.3 条件节	58
3.4 内部和外部 DTD	60
3.4.1 内部 DTD	60
3.4.2 外部 DTD	61
习题 3	62
第 4 章 XML Schema	63
4.1 概述	63
4.2 一个 XML Schema 文档示例	64
4.3 XML Schema 文档的结构	66
4.3.1 模式的基本概念	68
4.3.2 复杂类型和简单类型	69
4.3.3 元素声明	75
4.3.4 属性声明	79
4.3.5 其他	81
4.4 XML Schema 的数据类型	87
4.4.1 数据类型体系	87
4.4.2 数据类型二分法	92
4.4.3 内置数据类型	94
4.5 XML Schema 的名称空间	109
4.5.1 不同的名称空间	109
4.5.2 名称空间表示	110
4.5.3 在 Web 上定位模式文件	110
4.5.4 一致性	111
4.6 XML Schema 设计	111

4.6.1	定义局部和全局成分	111
4.6.2	定义名称空间	115
4.6.3	模式组装	117
习题 4		121
第 5 章 XPath		122
5.1	概述	122
5.2	XPath	122
5.2.1	XPath 应用举例	124
5.2.2	XPath	126
5.2.3	表达式上下文	127
5.2.4	路径表达式	128
5.2.5	序列表达式	133
5.2.6	过滤表达式	133
5.3	XPath Data Model	134
5.3.1	数据模型构建	134
5.3.2	访问程序	135
5.3.3	结点类型	136
5.4	XPath 和 XQuery Functions 与 Operators	142
5.4.1	数值运算和函数	142
5.4.2	字符串函数	145
5.4.3	逻辑运算和函数	148
5.4.4	日期时间运算及函数	149
5.4.5	上下文函数	152
5.5	应用举例	153
习题 5		154
第 6 章 HTML 技术		155
6.1	概述	155
6.2	提出问题	155
6.3	HTML 的语法基础	157
6.3.1	HTML 的语法基础	157
6.3.2	简单网页设计	158
6.4	HTML 的制表	167
6.4.1	简单表格	168
6.4.2	制表元素	169
6.4.3	表格的高级技巧	170
6.5	表单	176
6.5.1	表单的定义	177
6.5.2	表单中常用的元素	179

6.6	框架和超链接	181
6.6.1	框架的概念	181
6.6.2	框架的元素	184
6.6.3	超链接	185
6.7	字符控制	187
6.7.1	字体	187
6.7.2	列表控制	189
6.8	多媒体	191
6.8.1	多媒体技术	191
6.8.2	文本的滚动	192
6.9	XHTML 简介	194
6.9.1	关于元素的规定	194
6.9.2	关于属性	195
习题 6		196
第 7 章	CSS 技术	197
7.1	概述	197
7.2	问题的引入	197
7.3	显示属性	199
7.3.1	可视区和包含块	200
7.3.2	显示属性及其值	201
7.3.3	表格属性	204
7.3.4	none 值	205
7.3.5	程序优化	205
7.4	颜色与背景	206
7.4.1	颜色	206
7.4.2	背景	208
7.5	字体与文本	209
7.5.1	尺寸设置	209
7.5.2	字体属性	210
7.5.3	文本属性	214
7.6	边界设置	215
7.6.1	认识边界	215
7.6.2	边界的概念	216
7.7	元素筛选	220
7.8	用 CSS 对 XML 进行转换	222
7.8.1	以数据为主的 XML 文档转换	222
7.8.2	以文本为主的 XML 文档转换	228
7.8.3	内部和外部 CSS	231
习题 7		232

第 8 章 XSL 转换 XML	233
8.1 概述	233
8.2 一个 XSL 文档的讨论	234
8.3 XSL 样式表	236
8.3.1 样式表元素	236
8.3.2 样式表名称空间	239
8.3.3 样式表模块组合	240
8.4 模板规则	241
8.4.1 模板	241
8.4.2 模式	249
8.4.3 序列构造器	251
8.4.4 条件	254
8.4.5 循环	255
8.5 样式表设计	257
8.5.1 排序	259
8.5.2 求和	259
8.5.3 彩色效果	262
8.5.4 自动编号	266
8.5.5 创建元素和属性	270
8.5.6 变量和参数使用	272
习题 8	277
第 9 章 XML DOM 技术	278
9.1 概述	278
9.2 一个 DOM 示例	278
9.3 DOM 基础知识	280
9.3.1 DOM 的结构模型	280
9.3.2 核心模块基础知识	281
9.4 DOM 常用接口	283
9.4.1 Document 接口	283
9.4.2 Node 接口	289
9.4.3 Element 接口	290
9.4.4 Attr 接口	291
9.4.5 Text 接口	292
9.4.6 Comment 接口	292
9.4.7 DocumentFragment 接口	293
9.4.8 DOMImplementation 接口	293
9.4.9 NodeList 接口	293
9.4.10 NamedNodeMap 接口	293

9.4.11	CharacterData 接口	294
9.4.12	NameList 接口	294
9.5	XML DOM 设计	294
9.5.1	创建和浏览 DOM 对象	295
9.5.2	修改 DOM 树结构	301
9.5.3	实现 XSLT 转换	305
9.5.4	出错判断	307
9.5.5	浏览 DOM 树	308
习题 9		311
第 10 章	XML 数据库技术	313
10.1	概述	313
10.2	数据岛	313
10.2.1	基本概念	313
10.2.2	简单 XML 文档的处理	313
10.2.3	多级 XML 文档的处理	316
10.3	XML 数据源对象	319
10.3.1	基本概念	319
10.3.2	处理数据源对象	319
10.4	数据集操作	321
10.4.1	数据集的概念与操作	321
10.4.2	移动指针	323
10.4.3	ASP 对象	325
10.4.4	ADO	330
10.4.5	连接数据库	331
10.4.6	通过 ADO	332
10.5	XML 数据库	334
10.5.1	XML 数据库概念	335
10.5.2	访问 XML 数据库	335
习题 10		336
第 11 章	JavaScript 基础	337
11.1	概述	337
11.2	JavaScript 基础	337
11.2.1	构建 JavaScript 编程环境	337
11.2.2	入门示例	338
11.2.3	JavaScript 的交互方法	340
11.2.4	变量、数据类型和表达式	342
11.3	控制语句	346
11.3.1	条件控制	346

11.3.2	循环	347
11.4	函数与对象	348
11.4.1	函数定义	348
11.4.2	函数调用和参数传递	349
11.4.3	对象	351
11.5	其他	353
11.5.1	其他错误捕获和处理	353
11.5.2	内部 JavaScript 和外部 JavaScript	356
11.5.3	注释	357
习题 11		357
第 12 章 XML HttpRequest		358
12.1	概述	358
12.2	XML HttpRequest 对象	358
12.2.1	操作 HttpRequest 对象	359
12.2.2	HttpRequest 请求检测	366
12.3	通过 ASP 访问 XML	369
习题 12		372
第 13 章 Ajax 实例		373
13.1	概述	373
13.1.1	简要需求	373
13.1.2	功能分析	373
13.1.3	技术分析	374
13.2	建立数据库及项目	374
13.2.1	数据建模	374
13.2.2	建立项目	378
13.3	年级班级管理功能模块	380
13.3.1	数据访问层	380
13.3.2	显示层及页面后台代码	381
13.4	教师信息管理功能模块	397
13.4.1	数据访问层	397
13.4.2	显示层及页面后台代码	397
13.5	学生信息管理功能模块	398
13.5.1	数据访问层	398
13.5.2	显示层及页面后台代码	398
13.6	其他	398
13.6.1	一些常见问题	398
13.6.2	JSON 简介	405
参考文献		407

第 1 章 概 述

1.1 XML 技术简介

XML 技术是在 Internet 的广泛应用之后,传统的 Web 技术 HTML 的可扩展性、结构化和灵活性已经不能满足应用需要,并已经影响到 Internet 应用的发展的背景下提出的。1998 年 2 月,W3C 提出了 XML 技术的第一个规范 XML 1.0,目标是创建一种标记语言,并同时具备定义严格、语法明确、表示方便、结构良好、适用于所有行业的新的标记定义等。用来彻底解决在 Internet 应用中存在的问题。XML 技术源自 SGML,它既具备 SGML 的核心特征,又有 HTML 的简单性。目前,XML 技术已经开始在 Web、新型数据库系统中广泛应用,在计算机网络应用、网络编程、跨平台编程、移动互联网、物联网技术中发挥越来越重要的作用。

1.1.1 XML 的历史

1. SGML

1979 年美国国家标准化组织 ISO,设立了一个文本处理小组,开始开发一种基于 GML (Generalized Markup Language)的文本描述语言。1980 年,这项工作导致了第一个 SGML 工作草案的出版,1983 年,这个文本描述语言最终演化成 SGML(Standard GML),这是一个标准化的信息结构化技术,后来 SGML 扩展和修改成为一种全面适应工业范围的信息标准。1986 年,国际标准化组织 ISO 采纳了 SGML。

SGML 语言庞大,功能强,体系严密,同时技术比较复杂,价格昂贵,需要大量的软件来支持它,导致运营成本较高。20 世纪 80 年代主要用于电子产品交易、科技文献分类等方面。

2. HTML

HTML(HyperText Markup Language),意为超文本标记语言。1989 年,在欧洲核子物理实验室问世,这个技术采用超文本传输协议(HyperText Transfer Protocol,HTTP)。

HTML 的出现给 Internet 的爆炸性发展产生了积极的作用,WWW 成了人类了解信息,了解世界的一种全新的概念和模式。HTML 的巨大成功,使 HTML 迅速从 1.0 发展到 4.0。在发展的过程中,给 HTML 赋予了比最初设想要复杂得多的功能,目的是使 HTML 完成所有来自于商业应用、科学研究、信息发布的任务,使得 HTML 的语言失去了最初的简单性。并且在使用 HTML 时出现随意性、不规范和不严格等问题。

另一方面,HTML 的专用词表有限,用户无法自由增加新标记并进行有效性验证,用 HTML 来完成不同行业内的数据定义、数据表示,以及行业之间的数据交换很不方便,随着 Internet 的广泛应用,这种交换又是必需的、大量的、十分广泛的,这使得 HTML 捉襟见肘,无法满足这种要求。

3. XML

为了解决 HTML 在 Internet 应用中的局限性,1996 年,W3C(World Wide Web

Consortium)开始寻找在 Web 中使用 SGML 的方法。因为,SGML 具备 HTML 所没有的三种优势:可扩展性、结构化和灵活性。其目标是创建一种标记语言,这种语言既要具备 SGML 的核心特征,又要有 HTML 的简单性。同时具备许多新的特征,如定义严格,语法明确,表示方便,结构良好,适用于所有行业的新的标记定义等。1998 年 2 月 10 日,W3C 发布了 XML 1.0 规范。这就是 XML(Extensible Markup Language),可扩展标记语言。

XML 是 Web 发展到一定阶段的必然产物。W3C 在 XML 1.0 规范中是这样定义 XML 的:“可扩展标记语言(缩写为 XML)是用来描述一种称为 XML 的文件的数据对象,同时也部分地描述了处理这些数据对象的计算机程序的行为。XML 是 SGML 在应用上的一个子集,或为 SGML 的某种限制形式。根据指定规格的定义,XML 文件是符合规格的 SGML 文件。”

根据 XML 1.0(第 5 版)规范^①,XML 的设计目标是:

- (1) XML 应能直接用于 Internet;
- (2) XML 应支持广泛的应用;
- (3) XML 应与 SGML 兼容;
- (4) 处理 XML 文档的程序应该容易编写;
- (5) XML 的可选择性特征保持绝对小,理想情况下为零;
- (6) XML 文档应该是人易读的且合理清晰的;
- (7) XML 设计应该是可很快准备好的;
- (8) XML 文档的设计应该正规和简洁;
- (9) XML 文档应容易创建;
- (10) 在 XML 标记(markup)中简洁性可以忽略不计。

在 XML 1.0 规范中,包含三个主要的部分,分别是 XML 文档内容、文档的逻辑结构、文档的物理结构。

XML 文档是由被称为实体的存储单元组成的。实体或者是解析的,或者是不可解析的。解析的数据由字符组成,其中一部分形成字符数据,一部分形成标记(markup)。标记对描述 XML 文档的存储布局 and 逻辑结构进行编码。XML 提供了把各种约束强加在该存储布局和逻辑结构上的机制。

XML 文档内容中,包括结构良好性、字符集、通用语法结构、字符数据与标记、注释、处理指令、CDATA 节、序言及文档类型定义、文档的独立性声明、空白符处理、行结束控制和语言标识等十余项内容。对书写 XML 文档的数据给出了详细的规定。

在 XML 文档的逻辑结构部分,定义了 XML 文档的标记书写的规则、元素类型声明、属性表声明和条件节等内容。从而规定了 XML 元素和属性的定义规范。

在 XML 文档的物理结构中,对字符和实体参考、实体声明、实体解析、XML 处理器对实体和参考的处理、构造内部实体替代文本、预定义实体、表示法声明和文档实体等内容。对 XML 文档中实体的定义、引用、内部和外部实体进行了统一规定。

1.1.2 XML 与 HTML 的比较

对于 XML 的语法,将在后续章节中详细讨论。为了便于理解 XML 的语法、标记、文件格

^① Tim Bray, et al. Extensible Markup Language (XML) 1.0 (Fifth Edition) [EB/OL]. <http://www.w3.org/TR/2008/REC-xml-20081126/>.

式,把 HTML 与 XML 进行简单比较讨论,是一件有意义的事情,可以帮助我们理解 XML。

1. HTML 文档

HTML 文件,是一系列用“<”和“>”符号,把具有特定含义的英文字符串括起来,构造成为标记的元素来描述语法的。一个简单网页的 HTML 程序如下:

例 1.1 一个简单的网页文件。

```
<html>
  <head><title>我的第一个网页</title></head>
  <body bgcolor="#c0c0c0">
    <h1 align="center">学习 HTML,设计自己的网页!</h1>
    <p align="center">这是用 HTML 语言编写的第一个主页。</p>
    <p align="center">
      
    </p>
  </body>
</html>
```

在 HTML 文件中出现的标记都是由 HTML 规范规定的,如<html>、<head>、<body>、<p>、<h1>等,每个标记应该有一个结束标记,结束标记由“</”和“>”包括标记名构成,如</body>、</p>。一个 HTML 文件必须在第一行写上<html>,最后一行写上</html>来包含所有内容。浏览器判断一个文件是否是 HTML 文件,则根据文件是否包含在<html>和</html>这一对标记中。但是也有几个元素没有结束标记,如、
、<hr>等。目前使用的是 W3C 在 1998 年 4 月 24 日推出的 HTML 4.0 规范。

2. XML 文档

一个 XML 文档是由用户自行定义的标记组成,这些标记与 HTML 一样,都使用“<>”来包含标记名。下面来考查一组数据的表示,如表 1.1 所示。

表 1.1 学生成绩表

学号	姓名	性别	专业	成 绩		
				高等数学	程序设计	电路基础
200811010201	于丹	女	软件工程	89	73	92
...	...	...	...	...	...	...

表 1.1 中数据的 XML 文档表示如下。

例 1.2 根据表 1.1 建立 XML 文档。

```
<?xml version="1.0" encoding="GB2312"?>
<studentlist>
<student id="200811010201">
  <name>于丹</name>
  <sex>女</sex>
  <major>软件工程</major>
  <score>
    <mathematics>89</mathematics>
    <programming>73</programming>
    <circuit>92</circuit>
  </score>
</student>
```