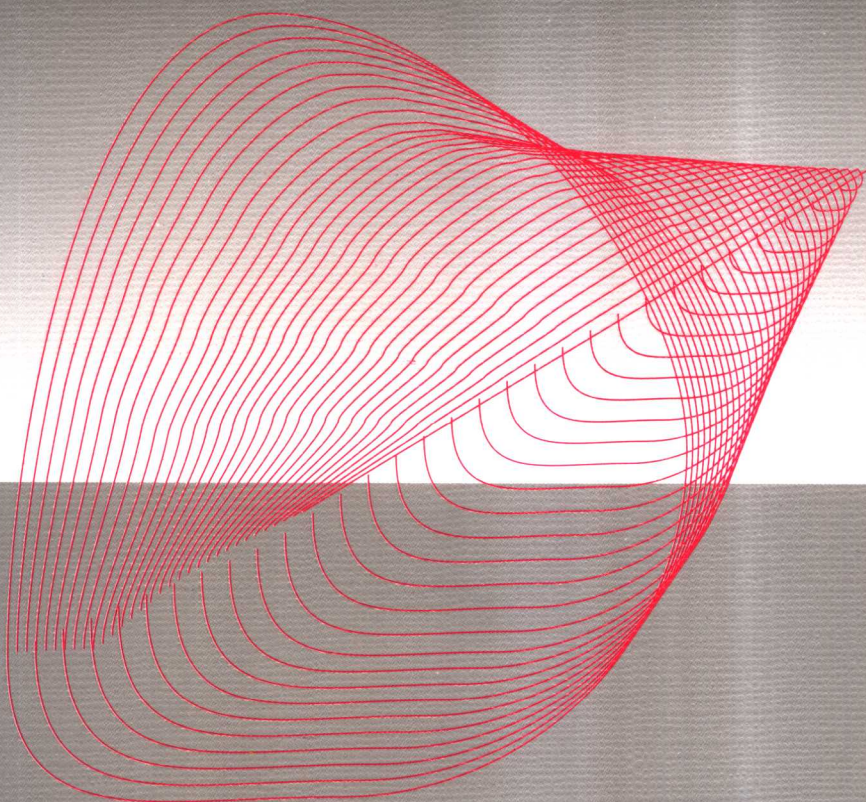


21 世纪高等学校计算机教育实用规划教材

XML 实用技术教程

顾
兵
主
编



清华大学出版社

21

世纪高等学校计算机教育实用规划教材

XML实用技术教程

顾 兵 主编

清华大学出版社
北 京

内 容 简 介

本书共 12 章,分 3 个层次由浅入深,循序渐进地介绍了 XML 技术:第一个层次介绍 XML 基本语法规则及样式显示,这是 XML 的技术基础;第二个层次由 XML 链接技术及 XML 转换技术组成,这是 XML 技术的应用基础;第三个层次介绍 XML 的 DOM 接口、SAX 接口技术与数据交换等内容,这是 XML 应用的重要技术。

本书既可作为高等院校计算机、信息、电子商务等专业相关课程的教材或教学参考书,也可作为广大 XML 技术应用人员和爱好者的参考书。

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 13501256678 13801310933

图书在版编目(CIP)数据

XML 实用技术教程/顾兵主编. —北京:清华大学出版社,2007.1

(21 世纪高等学校计算机教育实用规划教材)

ISBN 978-7-302-14286-7

I. X… II. 顾… III. 可扩充语言,XML—程序设计—教材 IV. TP312

中国版本图书馆 CIP 数据核字(2006)第 147871 号

责任编辑:魏江江

责任校对:梁 毅

责任印制:王秀菊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编:100084

c-service@tup.tsinghua.edu.cn

社 总 机:010-62770175

邮购热线:010-62786544

投稿咨询:010-62772015

客户服务:010-62776969

印 刷 者:北京密云胶印厂

装 订 者:三河市李旗庄少明装订厂

经 销:全国新华书店

开 本:185×260 印 张:17.25 字 数:425 千字

版 次:2007 年 1 月第 1 版 印 次:2007 年 1 月第 1 次印刷

印 数:1~4000

定 价:25.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103 产品编号:022075-01

出版说明

随着我国高等教育规模的扩大以及产业结构调整的进一步完善,社会对高层次应用型人才的需求将更加迫切。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,合理调整和配置教育资源,在改革和改造传统学科专业的基础上,加强工程型和应用型学科专业建设,积极设置主要面向地方支柱产业、高新技术产业、服务业的工程型和应用型学科专业,积极为地方经济建设输送各类应用型人才。各高校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力度,从而实现传统学科专业向工程型和应用型学科专业的发展与转变。在发挥传统学科专业师资力量强、办学经验丰富、教学资源充裕等优势的同时,不断更新其教学内容、改革课程体系,使工程型和应用型学科专业教育与经济建设相适应。计算机课程教学在从传统学科向工程型和应用型学科转变中起着至关重要的作用,工程型和应用型学科专业中的计算机课程设置、内容体系和教学手段及方法等也具有不同于传统学科的鲜明特点。

● 为了配合高校工程型和应用型学科专业的建设和发展,急需出版一批内容新、体系新、方法新、手段新的高水平计算机课程教材。目前,工程型和应用型学科专业计算机课程教材的建设工作仍滞后于教学改革的实践,如现有的计算机教材中有不少内容陈旧(依然用传统专业计算机教材代替工程型和应用型学科专业教材),重理论、轻实践,不能满足新的教学计划、课程设置的需要;一些课程的教材可供选择的品种太少;一些基础课的教材虽然品种较多,但低水平重复严重;有些教材内容庞杂,书越编越厚;专业课教材、教学辅助教材及教学参考书短缺,等等,都不利于学生能力的提高和素质的培养。为此,在教育部相关教学指导委员会专家的指导和建议下,清华大学出版社组织出版本系列教材,以满足工程型和应用型学科专业计算机课程教学的需要。本系列教材在规划过程中体现了如下一些基本原则和特点:

(1) 面向工程型与应用型学科专业,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,反映基本理论和原理的综合应用,强调实践和应用环节。

(2) 反映教学需要,促进教学发展。教材规划以新的工程型和应用型专业目录为依据。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。规划教材建设仍然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工程型和应用型专业教学内容和课程体系改革成果的教材。

(4) 主张一纲多本,合理配套。基础课和专业基础课教材要配套,同一门课程可以有多本具有不同内容特点的教材。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。在制订教材规划时要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、评审确定主编。书稿完成后要认真实行审稿程序,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。建立一支高水平的以老带新的教材编写队伍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教育实用规划教材编委会

联系人: 丁岭 dingl@tup.tsinghua.edu.cn

前言

20 世纪 90 年代后期,随着 Internet 技术及应用不断发展,出现和发展了网络 XML 技术。

在 Internet 飞速发展的过程中,以 HTML 为设计语言的 Web 由于其操作简单、信息获取容易而成为应用的关键。可以说,是 HTML 使 Web 变成了世界的图书馆。

但随着网络应用不断深入以及需求不断上升,HTML 难以满足越来越多的网络数据交换的需求。为此,W3C(World Wide Web Consortium,全球万维网联盟)于 1998 年制定新的网页设计语言 XML。XML 的作用是提供通用数据交换、变革 Web 发布、改造分布式计算。XML 是一项综合性技术,它面向互联网上企业商务应用与数据集成。目前,XML 已成为计算机行业中的热门技术。这一技术得到来自各方的广泛赞誉。IT 界各大厂商、各行业公司都非常看好这一技术,纷纷提供对 XML 标准的支持,以推动以 XML 为基础的应用与开发。

本书从 XML 技术基础入手,分 3 个层次详尽介绍了与 XML 相关的各种技术,其主要内容包括 XML 基础及 XML 编辑工具、DTD 与 XML Schema、XML 样式表显示技术 CSS 和 XSL,XML 链接技术及 XSLT 转换,XML 接口技术、数据岛技术以及 XML 数据交换等应用技术,在此基础上介绍两种 XML 技术的具体应用:同步多媒体合成语言 SMIL 以及可扩展矢量图形语言 SVG,它们在无线通信等领域正发挥着重要的作用。

本书是作者近年讲授这门课程的教学经验总结,内容全面、系统,重点突出,书中大量示例有助于读者准确理解相关概念以及掌握相关内容。本书在编写过程中得到华中科技大学金先级教授、刘本喜教授,湖北省经济管理干部学院郭熙丽教授的大力支持与热情帮助,在此表示诚挚的谢意。

本书可作为高等院校计算机、信息、电子商务等专业相关课程的教材或教学参考书,也可作为网络应用相关开发人员的参考书。

由于计算机及网络技术发展迅速,特别是 XML 技术本身也在不断发展和完善中,更由于作者本人水平有限,加之时间仓促,错误在所难免,恳请读者在使用过程中予以指正。

编 者

2007 年 1 月

目 录

第 1 章 XML 技术背景及其发展	1
1.1 标记语言	1
1.1.1 标记语言的历史	1
1.1.2 理解标记语言	2
1.2 通用标记语言	4
1.3 超文本标记语言	4
1.3.1 HTML 语言概述	4
1.3.2 HTML 语言的发展、作用与局限性	5
1.4 XML 现状及其发展	6
1.4.1 XML 技术背景	6
1.4.2 XML 技术作用及影响	8
1.4.3 XML 应用及发展前景	10
1.5 XML 相关的技术	15
1.5.1 XML 技术	15
1.5.2 XML 相关的主要技术综述	15
习题	18
第 2 章 XML 语法基础	19
2.1 XML 工具	19
2.1.1 XML 编辑工具	19
2.1.2 XML 解析工具	22
2.1.3 XML 浏览工具	25
2.2 XML 文档结构	25
2.2.1 XML 文档的逻辑结构	26
2.2.2 XML 文件的物理结构	28
2.3 XML 文档基本语法	30
2.3.1 XML 元素	30
2.3.2 XML 属性	34
2.3.3 XML 语法其他内容	35

2.4 创建格式良好的 XML 文档	37
2.4.1 格式良好的 XML 文档	37
2.4.2 有效的 XML 文档	38
习题	42
第 3 章 在 XML 文档中使用 DTD	44
3.1 DTD 概述	44
3.1.1 XML 文档内部的 DTD	45
3.1.2 XML 文档外部的 DTD	46
3.2 DTD 元素声明	49
3.2.1 DTD 元素声明的语法	50
3.2.2 XML 元素的数据	53
3.3 DTD 属性声明	55
3.3.1 属性声明语法	56
3.3.2 属性值的声明	56
3.3.3 关于属性值的说明	60
3.4 DTD 实体声明	61
3.4.1 通用实体声明的语法	61
3.4.2 参数实体声明的语法	63
3.5 DTD 标记声明	64
3.6 DTD 中的测试开关	65
3.7 DTD 文件存在的问题	66
习题	67
第 4 章 在 XML 文档中使用 Schema	68
4.1 命名空间	68
4.1.1 命名空间概述	68
4.1.2 命名空间的使用	70
4.2 XML Schema	73
4.2.1 XML Schema 的概述	73
4.2.2 XML Schema 的特点	73
4.2.3 XML Schema 基本结构	74
4.2.4 使用 XML Schema	76
4.3 XML Schema 的数据类型	78
4.3.1 基本数据类型	78
4.3.2 内置派生的数据类型	79
4.3.3 自定义数据类型	80

4.4 XML Schema 的元素声明	81
4.4.1 简单类型元素声明	81
4.4.2 复杂类型元素的声明	84
4.5 XML Schema 的属性声明	87
4.6 XML Schema 命名空间	88
习题	90
第 5 章 样式表技术	92
5.1 样式表简述	92
5.1.1 什么是样式表	92
5.1.2 级联式样式表	93
5.1.3 可扩展样式表	93
5.2 使用 CSS 显示 XML	94
5.2.1 CSS 基本语法	94
5.2.2 CSS 常用属性	95
5.2.3 使用 CSS 显示 XML 文档	97
5.3 使用 XSL 显示 XML	102
5.3.1 XSL 语法	103
5.3.2 使用 XSL 显示 XML	108
5.3.3 XSL 其他元素	110
5.3.4 CSS 与 XSL 比较	117
习题	118
第 6 章 XML 链接技术	120
6.1 XLink 语言	120
6.1.1 HTML 的超链接	121
6.1.2 XML 链接	122
6.2 简单链接	128
6.2.1 简单链接实现 HTML 中的超链接	128
6.2.2 简单链接实现的图像嵌入链接	131
6.3 扩展链接	133
习题	135
第 7 章 XSLT 转换	136
7.1 XSLT 概述	136
7.1.1 为什么要使用 XSLT	137
7.1.2 XSLT 的特点	137

7.1.3 XSLT 工作机理	138
7.2 XSLT 基本架构	139
7.2.1 XSLT 处理器架构	139
7.2.2 XSLT 的基本架构	140
7.3 Xpath 语言	141
7.3.1 Xpath 数据模型	141
7.3.2 Xpath 位置路径	142
7.3.3 节点之间关系	143
7.3.4 Xpath 的函数	144
7.4 XSLT 元素	144
7.4.1 定义模板规则的元素	144
7.4.2 XML 元素的转换	146
7.5 XSLT 与脚本语言	149
7.6 转换程序	152
习题	154
第 8 章 DOM 接口技术	155
8.1 DOM 接口概述	155
8.2 DOM 的结构	156
8.3 DOM 对象	158
8.3.1 DOM 基本接口	158
8.3.2 DOM 对象	160
8.4 DOM 的使用	161
8.4.1 创建 DOM 对象	161
8.4.2 加载 XML 文档	161
8.4.3 遍历 XML 文档	162
8.4.4 DOM 接口应用	164
8.5 DOM 对文档的操作	165
8.5.1 动态创建 XML 文档	165
8.5.2 添加子元素	166
8.5.3 添加或修改元素内容	167
8.5.4 删除子元素	168
8.5.5 异常处理	169
8.5.6 动态转换 XML 文档	170
8.6 DOM 应用实例	172
8.6.1 DOM 的 ParseError 对象应用实例	172
8.6.2 DOM 节点对象应用实例	174

8.6.3 服务器端 ASP 应用实例	177
习题	180
第 9 章 SAX 接口	182
9.1 SAX 接口概述	182
9.1.1 SAX 接口的特点	182
9.1.2 SAX 工作机制	184
9.2 SAX 接口及其方法	184
9.2.1 SAX 应用程序接口	185
9.2.2 SAX 接口中类及方法	188
9.3 SAX 接口应用示例	191
9.3.1 Java+SAX 接口读取 XML 文档内容	191
9.3.2 用 C/C++ 实现 SAX 接口及其应用	198
9.4 SAX 接口与 DOM 接口比较	208
习题	209
第 10 章 XML 数据岛	210
10.1 数据岛基本概念	210
10.1.1 什么是数据岛?	210
10.1.2 与数据岛相关的术语	211
10.2 在 HTML 中显示 XML 数据	211
10.2.1 数据岛形式	212
10.2.2 数据岛中 XML 元素的显示	215
10.2.3 XML 标记属性的处理	216
10.3 数据岛对象	217
10.3.1 数据岛对象的访问	218
10.3.2 数据岛对象的深入研究	221
习题	229
第 11 章 XML 数据交换	230
11.1 数据交换基本概念	230
11.1.1 数据交换类型	230
11.1.2 数据存取机制	232
11.2 XML 与数据库	233
11.2.1 XML 与关系数据库	234
11.2.2 XML 与面向对象的数据库	234
11.2.3 SQL Server 2000 对 XML 的支持	235

11.3 XML 与 SQL Server 的数据交换	236
11.3.1 配置 IIS 服务器	236
11.3.2 通过 HTTP 方式访问 SQL Server	239
11.4 使用 HTTP 访问数据库对象	240
11.5 使用 ADO 实现数据交换	242
习题	245
第 12 章 XML 应用	246
12.1 SMIL 语言及应用	246
12.1.1 SMIL 语言结构	246
12.1.2 SMIL 中常用标记和属性	247
12.1.3 SMIL 应用领域	249
12.2 可扩展矢量图形语言 SVG	249
12.2.1 SVG 的特点	250
12.2.2 SVG 语言	251
12.2.3 SVG 主要应用领域	257
习题	257
参考文献	259

第 1 章

XML 技术背景及其发展

文本处理是计算机科学的一个分支,标记语言则是伴随着文本处理系统的发展过程而出现和不断发展的一种语言。GML、SGML、HTML 以及随后出现的 XML 都是在文本处理及网络应用中有着重要作用的标记语言。其中可扩展标记语言 XML 是随着网络应用的不断深入以及网络数据交换的需要而出现的一种重要的标记语言。本章首先介绍标记语言的发展历史,在对有关标记语言比较的基础上,引出 XML 语言,然后对 XML 语言的特点、作用以及与之相关的技术进行简要的介绍。

1.1 标记语言

标记语言(markup language),特指用一系列约定好的标记来对电子文档进行标记,以实现对其语义、结构及格式的定义。这些标记必须能够很容易地与内容相区分,并且易于识别。标记语言必须定义什么样的标记是允许的,什么样的标记是必需的,标记是如何与文档的内容相区分的,以及标记的含义是什么。比如,HTML(hypertext markup language,超文本标记语言)就是目前网络中网页文档格式定义的一种重要的标记语言,它与 SGML 和 XML 都属于一个大家族——标记语言家族,标记语言的家族及发展如图 1-1 所示。

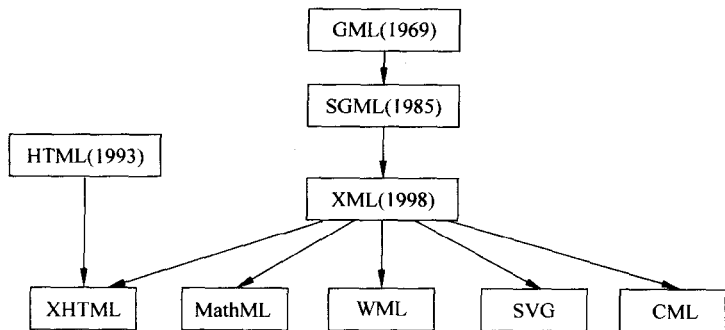


图 1-1 标记语言的家族

1.1.1 标记语言的历史

了解标记语言发展的历史,对于了解 HTML、SGML、XML 之间的联系和正确理解标记语言这一技术更有帮助。

为了促进数据的交换和操作,在 20 世纪 60 年代,通过 IBM 公司研究人员的杰出工作,得出了重要的结论:要提高系统的可移植性,必须采用一种通用的文档格式,这种文档的格式化必须遵守特定的规则。这也就是创建 GML(Generalized Markup Language,通用标记语言)的指导原则。从人们所产生的将文件结构化为标准的格式的动机出发,IBM 公司创建了 GML,并在其出版系统内部予以实现。IBM 公司采用 GML,以使用单一类型的源文件来维护书籍、报表以及其他各类文件。

IBM 公司的研究人员 Charles Goldfarb 带领的开发团队继续完善着 GML,在对这种标记语言的概念达成共识的基础上,将其称为 SGML(Standard Generalized Markup Language,标准通用标记语言),这是第一个标准化的信息结构化技术。SGML 成为了 IBM 内部格式化和维护合法文件的手段。SGML 后来被扩展和修改,作为一种全面的信息标准以适应工业范围的广泛应用。1986 年,SGML 被国际标准化组织(ISO)所采纳。

SGML 的功能非常强大,它为描述网络电子文档提供了一套必要的通用框架,并且允许建立新的一致性规则,由 SGML 所提供的功能已被美国一些大型企业所采用,如汽车公司和电信公司,同时作为一种文档标准而被美国政府机构所采用。但是它非常复杂,需要许多昂贵的软件配合运行,因而在很长一段时间没有得到推广。

1989 年,欧洲粒子物理实验室(CERN)的研究员 Tim Berners-Lee 和 Anders Berglund 共同创建了一种基于标记的语言,为在 Internet 上共享的文档做标记,于是产生了 HTML。HTML 可看作是 SGML 的简化应用,它的产生推动了信息产业的大发展。

HTML 当初仅仅提供了一种对静态文本信息显示的方法,但显然并不能适应越来越多的需要。因此就有越来越多的标签产生,比如用于描述图片,<script>通过为网页加入脚本支持,提供动态网页内容等。两大浏览器厂商微软和网景公司,甚至创建了自己的产品兼容的标签,使得 HTML 变得愈来愈臃肿,兼容性不好,失去了严谨的结构化。

1996 年,人们开始致力于描述一个新的标记语言,它是一种在 Web 中应用 SGML 的灵活性和强大功能的方法,W3C(World Wide Web Consortium,全球万维网联盟——领导万维网,制定其公共的协议,促进万维网的发展并确保其互操作性的国际组织)专门成立了专家小组以从事这项工作。1998 年 2 月,W3C 批准了 XML 1.0 规范。标志着一种崭新的、重要的语言 XML(eXtensible Markup Language,可扩展的标记语言)的产生。XML 具备了 SGML 的核心特性,但又非常简洁,XML 规范的内容甚至不到 SGML 的十分之一。随即 XML 的应用如雨后春笋般地展开,XML 被看作是 IT 行业中一颗闪亮的新星。

1.1.2 理解标记语言

从以上标记语言的发展历史可以看出,标记语言是由定义网络中网页文档格式的一些代码或者控制标记按一定规则组成的。这些代码和控制标记本身没有意义,它们与文档信息结合后形成电子文档,而应用软件通过解读电子文档中的标记语言,并通过标记语言的意义来对电子文档进行特定的处理,例如输出到屏幕上或进行信息的交换等。

由此可知,标记语言主要功能是使电子文档具有结构性。正是这种结构性使得应用

软件能对电子文档进行正确的解读和应用。标记语言可以分为特殊用途的标记语言与通用标记语言。所谓特殊用途的标记语言是指专为某种应用软件而特别设计和定制的,这类语言的代表就是 HTML,它是专门为了应用在 Web 上的网页设计的。

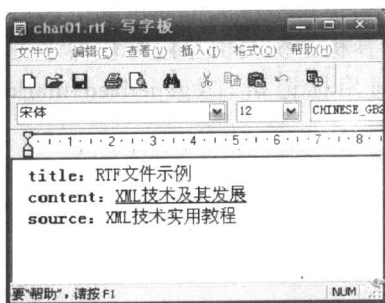


图 1-2 在写字板中输入文本内容

RTF 也是一种特殊用途的标记语言,实际上,它是一种最简单的标记语言,Windows 系统中的写字板就可以支持(解读)这种标记语言的控制标记。

建立一个简单的 RTF 文件示例:

在写字板中输入文字,如图 1-2 所示。将新建的文档取名为 char01,并将其保存为 .rtf 文件格式。RTF 格式的文件本身也是一种文本文件,因此可用记事本打开并阅读。所看到的代码如图 1-3 所示。

在图 1-3 中可以看出,除了在先前创建 RTF 文件中输入的内容之外,还有一些其他代码,这些代码就是 RTF 的控制标记。写字板可以解读这些标记,并按照这些标记的含义将所输入的文字正确地显示出来。

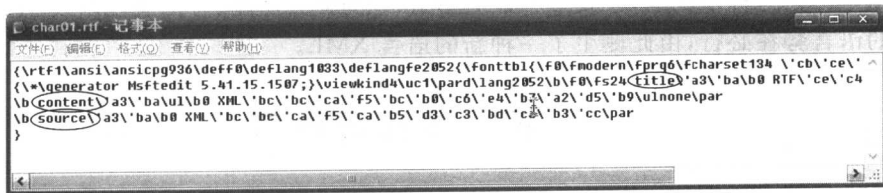


图 1-3 记事本中查看 RTF 文件

实际上微软公司的 Word 也可以正确解读 RTF 格式的文件,通过 Office 组件中的 Word 程序加载前面创建的 char01.rtf 文件,其显示结果如图 1-4 所示。

通过 Office 组件中的 Word 程序所建立的 .doc 格式文件,也是一种特殊用途的标记语言。特殊用途的标记语言实质上是有着固定格式的置标语言,它不具备扩展性。

另一类标记语言则以 SGML (standard generalized markup language, 标准通用标记语言)为代表,它功能很强,可用于创建成千上万的置标语言,因此也称为元语言,即用于创建其他标记语言的语言,而且具有极好的扩展性。

SGML 和 HTML 之间、XML 和 HTML 之间的差别是巨大的。HTML 可以看作是 SGML 的一种具体应用,SGML、XML 则作为元语言,可以定义出无数新的标记语言,如 WML、SMIL、SVG 即是一批用 XML 定义的新置标语言。SGML 和 XML 之间的差别则是微小的,XML 作为 SGML 的子集,继承了 SGML 的优点,如可扩展性、结构化和有效性等。

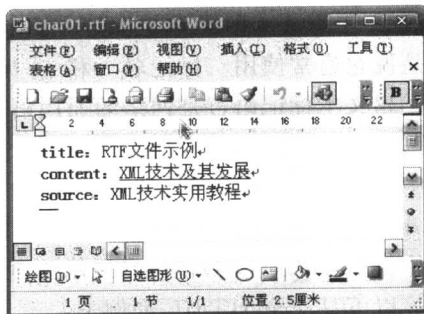


图 1-4 在 Word 程序中查看 RTF 文件

1.2 通用标记语言

IBM 公司于 20 世纪 60 年代开发了一种文档描述语言,用来解决不同系统中文档格式不同的问题。这样,文档的显示和打印需要不同的计算机系统提供专门的软件来显示和打印文档。IBM 公司把自己这种标识语言称为通用标记语言(genelized markup language,GML)。

1986 年国际标准化组织(ISO)认为 IBM 公司提出的概念很好,并发布了为生成标准化文档而定义的标记语言标准 ISO 8879,称为新的语言 SGML,即标准通用标记语言。SGML 是结构化的、可扩展的语言,这些特点使它在很多公司受到欢迎,被用来创建、处理和发布大量的文本信息。

SGML 是一种通用的文档结构描述置标语言,为语法置标提供了异常强大的工具,同时具有极好的扩展性,因此在数据分类和索引中非常有用。但 SGML 强大功能的背后是它的复杂度太高,不适合网络的日常应用,另外,SGML 价格昂贵,导致开发成本高,还有一个关键因素是不被主流浏览器厂商所支持,这些原因使得 SGML 在 Web 上的推广受到了阻碍。在这种情况下,开发一种兼具 SGML 的强大功能、可扩展性以及 HTML 的简单性的语言势在必行,由此诞生了一种新的语言 XML。

1.3 超文本标记语言

1.3.1 HTML 语言概述

HTML(超级文本标记语言),起源于 SGML(标准通用标记语言),是一种国际标准文本排版标记语言。由于 HTML 是专门用于网络超文本描述的语言,限于网络传输的弊端及网络安全等因素,它必须尽可能得简单,是一种完全基于 ASCII 码的文档。HTML 中最基本的元素是标记。比如像<html>和</html>代码就是一个标记对,也就是说它通常使用<标记名>和</标记名>来表示标记的开始和结束。

超文本标记语言是网络的通用语言,它通过标签(tag)来描述将在网页上显示的信息(例如文字、图像、声音、动画等各种资源),且只需在一个简单的文字编辑器或专用的编辑器中按 HTML 的固定格式编写所需内容即可,所建立的文本文件的扩展名必须为.html 或.htm。用它编写的 HTML 文件可以在网络上被任何人浏览,浏览器通过解释的方式来执行 HTML 代码,将信息展示给浏览者。

以下是一个 HTML 的例子:

```
<HTML>
  <HEAD>
    <TITLE>超文本标记语言(HTML)示例</TITLE>
  </HEAD>
  <BODY>
    <H1 ALIGN="center">XML 技术发展</H1>
```

```

<HR>
<P ALIGN="center">欢迎使用 XML 技术实用教程</P>
</BODY>
</HTML>

```

把上面的例子保存成 HTML 格式,然后用浏览器(IE)打开,即可看到如图 1-5 所示的画面。



图 1-5 应用程序解读 HTML 的示例

1.3.2 HTML 语言的发展、作用与局限性

从 20 世纪 80 年代以来,Internet 飞速发展,在这个发展过程中,Web 由于其操作简单、信息获取容易而成为网络应用的关键。在这当中,HTML 的作用功不可没。正是由于 HTML 的出现,迎合了人们只需要一个简单程序,就可以看到图文并茂的网页这种需求,因此在 HTML 出现后,网页成了人们接近网络、了解网络信息、发布消息的一个主要渠道。可以说,是 HTML 使 Web 变成了世界的图书馆。

HTML 是一种特定的 SGML 文档类型,由于其学习和实现非常容易,而且免费提供源代码,所以很早就得到各个 Web 浏览器厂商的支持。

HTML 最初由 GERN 在 1990 年进行设计,1993 年由 Berners-Lee 等人完成 HTML 1.0 标准,后来,W3C 承担了 HTML 的开发和标准化工作,经过不断完善,现在已发布了 HTML 4.0 标准。

在 W3C 所建议使用的 HTML 4.0 中,所有的控制标记都是固定的,且数目有限。所谓固定是指其控制标记的名称是固定不变的,因而其提供的功能与使用的属性也是固定的。因此,HTML 不允许网页设计者自行创造控制标记。也就是说,由于 HTML 不是一种元语言,所以它不能创建某一特定领域的标记集。

随着网络应用的深入,特别是电子商务的应用,HTML 过于简单的缺点很快凸现出来。因为标记语言的出现,使建立需要存储大量电子文档的数据仓库变得更加方便,通常,这样的电子文档由 3 个部分组成。

- 文档结构: 提供如何书写文档的基本框架。
- 文档内容: 标记出文档所包含的内容。