

新东方电脑软件程序员系列丛书

# XML 程序设计

新东方电脑教材研发室

栗松涛 编著

清华大学出版社

# (京)新登字 158 号

## 内 容 简 介

本书从程序设计角度讲解了 XML 的系统知识。主要内容包括 XML 概述和基础知识、以及 CSS 和 XML、XSL 和 XML、XML DTD、XML Schema、XML Namespace、XML DSO、XML DOM 等概念及应用,最后是 XML 和 ASP 进行联合编程的技术。全书内容安排由浅入深,通过丰富的教学案例全面讲解了 XML 的实用技术,另外,每章都结合具体工程实践给出一个项目应用例子,介绍一些在具体项目开发中的应用技术。

本书适合初学者或具有中等水平的读者使用。读者只须按书中的讲解,尝试调通书中的每一段程序,相信就可以较为全面地掌握这门新一代的 Internet 语言。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

书 名:XML 程序设计

作 者:新东方电脑教材研发室 栗松涛

出 版 者:清华大学出版社(北京清华大学学研大厦,邮政编码:100084)

<http://www.tup.tsinghua.edu.cn>

印 刷 者: 印刷厂

发 行 者:新华书店总店北京发行所

开 本:787×1092 1/16 印张:24 字数:580 千字

版 次:2001 年 11 月第 1 版 2001 年 11 月第 1 次印刷

书 号:ISBN 7-302-04911-4/TP·2765

印 数:0001~5000

定 价:40.00 元

## 《新东方电脑标准培训教材》策划委员会

主任 王 强

副主任 周怀军 焦金生

委员 (按姓氏笔划为序)

王 强 王文成 包凡一 孙亚刚

杜子华 周怀军 胡 敏 俞敏洪

徐小平 柴文强 温尚书 焦金生

## 《新东方软件程序员》丛书编委会

主 编 王 强

副主编 周怀军 孙亚刚

编 委 栗松涛 冯燕奎 石志国 温尚书

施 松 刘寅宵 姜海鹏 涂 进

# 丛书特点

本丛书是为人们掌握 IT 实用技能，顺利转行进入 IT 行业的需要出版的。

本套丛书有以下几个鲜明的特点：

## 1. 按照北美以及国内流行的 IT 职位需求选材

现在的众多社会培训是面向认证的，可以说是学历教育的翻版，我们认为证书只是进入 IT 行业的敲门砖而已，能否胜任职位工作，要看实际掌握的技能，为此新东方电脑培训率先推出面向职位的就业技能培训；本套丛书作为新东方电脑培训的标准培训教材，与新东方面向职位的就业技能培训课程密切结合。为了让读者更加容易了解 IT 职位的需求，清楚每个职位所需具备的实用技能，丛书按照职位进行划分丛书系列，每个系列面向一个职位，每一本图书针对职位内的一种计算机技能，独立成册。

## 2. 本套丛书目的在教会您一种职业技能，而不止单单宣讲一种技术

每一种学习，都会有一个最佳的学习顺序和学习方法：

- 书中新东方有丰富教学经验的老师将给您一个最佳的学习顺序的指导，这是每本书的写作顺序；
- 没有一种学习方法比通过完整案例，边学边练学得好、学得快，这也是新东方教学方法的经验积累。丛书采用实用易学的案例贯穿始终，凡关键之处必有案例。

## 3. 超越软件版本的限制，分析人们使用软件的规律

软件更新的太快了，我们不盲目赶新潮，要想超越软件版本的限制实际上很简单：以不变应万变，掌握软件的使用规律即可。

最后希望本丛书能够让您从中轻松获益。

# 总 序

新东方学校校长 俞敏洪

当今社会是一个需要技能的社会，新东方最初就是培训学生英语方面的技能，因为良好的英语技能对于工作、前途都起着非常重要的作用。在过去的几年里，新东方的英语培训取得了长足的发展和辉煌的成就，每年培训 10 万学生，并且有大量的新东方学员出国深造。但是在世界上除了英语技能，还有许多其他的技能，电脑技能就是其中之一。在现代化的技术世界中，电脑已经成为人们谋生的一个重要手段。有知识的人离不开英语，但是更离不开电脑，电脑已经成为人们日常学习和工作必不可少的一部分。不可否认，电脑的发展使这个世界更加简单化、条理化、明了化。

新东方一直把为人们谋取更好的工作，更好的生活作为目标，所以在英语培训取得了一定的成就以后，新东方一直在思考是否能在电脑方面为学员提供一点帮助，能够像英语一样，经过一定的培训也能让大家获得更好的电脑技能。出于这样的思考，新东方于 1999 年创立了电脑培训部。经过两年多的发展，电脑培训部由原来只有 20 多台电脑的小小培训部发展到了现在有数百台电脑的培训中心。在发展过程中，电脑培训部得到了微软、联想、Adobe、Macromedia 等世界著名电脑公司的培训授权。迄今为止，新东方电脑培训已经为社会培训了上万名学员，为众多公司及企事业输送了大量的电脑专业人才，有些人经过新东方电脑培训部的培训，在国外的电脑公司也找到了理想的工作。

提到新东方电脑培训部，就不得不提起电脑培训部的创始人新东方学校副校长王强和电脑部的具体业务负责人电脑部主任周怀军。

王强的大名谁都知道，来新东方以前是美国贝尔传讯研究所的软件工程师，回新东方后一直力推电脑部的建设，这里我就不多做介绍了。

周怀军在来电脑培训部负责具体业务之前，已经在加拿大一家计算机公司找到了一份收入不错的工作。但是他有感于新东方的创业精神，有感于自己想要干一番事业的理念，毅然放弃了在加拿大的职位，放弃了在加拿大和家人团聚的机会，从电脑培训部最初一台电脑都没有的情况下做起，一直做到了电脑培训部今天的规模。

岁月流逝，两年的时间匆匆而过，周怀军依然在新东方继续忘我地工作着。周怀军思维敏捷，但语言木讷，尤其在处理学生问题时最能看出他老实、憨厚的个性，这点和会侃、会说、积极主动的新东方主流个性不相吻合。但是周怀军是一个实干的人，刚创办电脑培训部的时候，周怀军有一头浓密乌黑的头发，但是在为新东方电脑部两年多的操劳过程中，他的头发却日渐稀少。但正是由于他这种踏实肯干，对什么事情都勤勤恳恳、认真负责的精神，才使新东方电脑培训部由最初的几十台计算机发展到了现在的几百台计算机，学生也由几十人发展到了现在的几千人。在这个过程中，由于周怀军礼贤下士，两年的时间里在他的周围聚集了大批优秀的电脑专家、电脑教师。

经过两年多的发展，新东方电脑培训部积累了大量的理论和实际操作经验。现在在王强副校长的组织领导下，周怀军老师准备把电脑部积累的这些经验编写成书，把电脑方面从低级到高级的各个领域方面的经验汇集成册，我感到非常高兴。因为新东方在英语方面已经出版了很多的图书，但英语只是人生必需技能的一个方面，电脑也是现代社会人生所需技能的一部分。这些书籍的出版必将对大家尤其是广大电脑爱好者起到强大而明确的指导作用。

其实在当今社会所需要的各种技能之中，不仅仅是英语和电脑，还包括很多其他方面的技能。每个人都在刻画着自己的人生轨迹，每个人都在寻找着自己的奋斗目标，如果一个人连自己的人生目标都不知道，连自己这辈子应该干些什么都不清楚，你的英语和电脑或者其他方面的技能学得再好，学得再精，也只是绣花枕头，没有什么实际作用。无论你的专业是文科还是理科，你的理想是想成为诗人还是想成为物理学家，你是想成为国营公司的总经理还是成为个体老板，这些都是人生目标。一个人只有明确自己的人生目标，在这个人生目标上，你再用英语和电脑把自己武装起来，那你就如虎添翼，在社会中成功的机会就会更高，成功的速度也会更快。英语是使你走向世界的必不可少的工具，电脑是使你管理现代化和接受新思想必不可少的工具。我衷心希望所有看到这篇序言的人，都能够在学习和生活中努力掌握这两项技能，让它们一同伴随着你向人生的终极目标前进！

2001年3月

# 序 言

如果说 20 世纪是计算机的时代，那 21 世纪则是 Internet 和 Web 的时代。与此相关的技术也得到了飞速发展。硬件方面已从原来的 8086 和 8088 发展到目前具备 1000 兆赫兹处理速度的中央处理器，其发展速度之快远非几年前人们所能预料。在软件方面更是这样，从 Internet 和 Web 的出现到 1992 年发布 HTML 的第一个版本，也只仅仅几年的时间。尽管 HTML 的第一个版本十分简陋，但它却从此改变了计算机的发展方向，此后各种与此相关的技术如雨后春笋般层出不穷，Java、CSS、DOM、ActiveX、COM 等技术相继形成并逐渐开始发展成熟。尽管所有这些技术的出现都曾给 Web 技术带来过一定的震动，但从没有哪项技术能像 XML 一样如此轰动整个 Web 世界了。XML 出现不过几年时间就开始影响并变革整个 Internet 的发展趋势。

XML 的全称是 eXtensible Markup Language（可扩展的标记语言）。为什么 XML 语言具有如此的震动效应呢？这里要首先回顾一下整个 Web 的发展历史。

谈到 Web 的发展首先应该是 HTML 标记语言的发展。HTML 技术的发展几乎伴随整个 Internet 的发展。正是 HTML 这种简单的标记语言使几乎所有人都体会到了 Internet 的脉搏的跳动，它对于计算机的发展几乎可以和 BASIC 的作用相提并论，但 HTML 技术本身却存在诸多缺陷。首先 HTML 是一种样式语言，在目前 Internet 上它只是充当了数据表示的主要角色。而这种不协调在 Internet 发展的初期还没有什么影响，但随着 Internet 上信息量的增多，HTML 变得越来越难以胜任；其次 HTML 对浏览器的过度依赖性也形成了 HTML 标准的严重不统一，从而导致许多信息表示只能由某种特定的浏览器来解释。HTML 的这些不足导致人们重新思考 HTML 在 Internet 上的角色并开始研究一门新的语言来弥补 HTML 的缺陷，XML 的产生就正是这种重新思考的最终结果。

XML 是一种专门为 Internet 所设计的标记语言，它的重点是管理信息的数据本身，而不是数据的样式，数据的显示则交给另外的技术来解决。这种明确的分工导致的将是更高效的 Web 程序设计、更快的搜索引擎的出现、更统一的数据表示和更方便的数据交流的出现。XML 是一种定义语言的语言，它克服了 HTML 标记有限的缺点，给了程序员更大的程序设计的自由空间。使用 XML DTD 可以根据具体的项目背景设计出适合于自身的标记语言，而不是所有的信息都是使用 HTML 标记来表示。

XML 是 SGML 的一个子集，本质上是一种特殊的 SGML 标记语言，它继承了 SGML 的优点剔除了 SGML 的不适合 Internet 的部分，并对 SGML 进行了大刀阔斧的改革。目前 XML 已经是一门十分友好的语言。在语言的简单性方面可以和 HTML 相媲美，是 Internet 的下一代语言。

本书内容安排由浅入深，适合于多种层次的读者学习。本书的另一个特点是采用案例教学方式，书中几乎所有的概念都有具体而简单的例子程序来演示。针对具体的项目开发是本书的另一个特点，几乎每章最后都提供了一个 XML 在具体项目中应用的例子。本书

的第三个特点是重点面向程序设计，这里使用了像 VBScript，HTML，Javascript，ASP 等技术，但应用都是很初步的，读者只要参考一下相关方面的书籍就能够正常地阅读本书。

本套丛书是在新东方电脑部《网页编程班》讲义的基础上，在新东方电脑培训部主任周怀军和清华大学出版社副总编焦金生老师的大力支持下，由孙亚刚老师组织完成。本书由栗松涛老师编写，在写作过程中，冯燕奎、石志国、梁如军、温尚书、施松、刘寅宵老师也为本书的大纲制定和编写提供了很多宝贵意见和建议。

本书经过紧张的策划、设计和创作，能够在这么短的时间内与广大读者见面，这与清华大学出版社的柴文强副编审和田在儒编辑的辛勤努力是分不开的，在此我们表示深深的谢意。

由于编写时间短促，本书中可能会有一些疏漏，希望广大读者以及新东方教师和学员给予指正，可以发 E-mail 至：[books@getjob.com.cn](mailto:books@getjob.com.cn)。

如果您对我们的详细课程感兴趣，可以访问：

新东方学校网站：[www.neworiental.org](http://www.neworiental.org)，[www.tol24.com](http://www.tol24.com)

新东方电脑部网站：[www.getjob.com.cn](http://www.getjob.com.cn)

新东方电脑教材研发室

2001 年 7 月

# 目 录

第 1 章 XML 引言	1
1.1 XML 的发展历史	1
1.1.1 标记语言的出现	1
1.1.2 标记语言的特点	3
1.1.3 RTF 标记语言	4
1.1.4 HTML 标记语言	8
1.1.5 标准通用标记语言 (SGML)	13
1.1.6 可扩展标记语言 (XML)	14
1.2 XML 的一般概念	18
1.2.1 XML 是什么	18
1.2.2 XML 的设计目标	19
1.2.3 XML 的样式语言	20
1.2.4 XML 的链接语言	26
1.2.5 XML 的名称空间	27
1.2.6 XML 的文档对象模型 (DOM)	27
1.2.7 XML 的文档类型定义 (DTD)	27
1.2.8 XML 的文档类型模式	29
1.3 XML 的应用	30
1.3.1 应用在科学和数学领域中的 XML	30
1.3.2 应用在化学领域中的 XML	32
1.3.3 矢量标记语言 (VML)	33
1.3.4 音乐标记语言 (MusicML)	34
1.3.5 资源描述框架 (RDF)	35
1.3.6 人力资源标记语言 (HRML)	35
1.3.7 开放的金融交换标准 (OFX)	36
1.3.8 可扩展表单描述语言 (XFDL)	36
1.3.9 无线标记语言 (WML)	36
1.3.10 同步多媒体合成语言 (SMIL)	37
小结	38
第 2 章 XML 应用程序	39
2.1 良构的 XML 文档	39
2.1.1 问题的描述	39

2.1.2	使用记事本创建 XML 文档	40
2.1.3	结构良好的 XML 文档	42
2.2	XML 文档中的中文编码问题	52
2.3	XML 的解析器	54
2.4	文件的结构	58
2.4.1	处理指令	59
2.4.2	文件格式定义	59
2.4.3	XML 文档的树状结构	60
2.4.4	标记的命名规则	61
2.4.5	XML 文档的属性	62
2.4.6	XML 文档中的注释	63
2.4.7	程序中的文字 CDATA 标记	64
2.5	项目开发	67
	小结	72
<b>第 3 章</b>	<b>使用 CSS 格式化 XML 页面</b>	<b>73</b>
3.1	CSS 技术简介	73
3.2	CSS 文档的基本语法	74
3.2.1	样式的直接定义法	74
3.2.2	成组选择符	75
3.2.3	利用 CLASS 来定义样式	76
3.2.4	在 CSS 文件中使用 ID	77
3.2.5	STYLE 属性的使用	79
3.2.6	CSS 中的继承特性	79
3.2.7	使用 @import 结合样式表单	80
3.2.8	!important 声明	81
3.3	将 CSS 和 XML 结合的方法	82
3.3.1	外部样式文件	82
3.3.2	将样式语句嵌入到 XML 文件当中	84
3.3.3	混合的方法指定样式	86
3.3.4	使用多个样式文件	88
3.4	CSS 格式化 XML 文档	90
3.4.1	与文字相关的属性	90
3.4.2	与定位相关的属性	93
3.4.3	与版式相关的属性	101
3.5	在 XML 中使用 HTML 标记	108
3.5.1	在 XML 文档中显示图形	109
3.5.2	在 XML 文档中使用超链接	110

3.5.3 XML 中的换行分段标记	111
3.5.4 在 XML 文档中实现表格	112
3.5.5 在 XML 文档中显示表单	113
3.6 在 XML 文档中添加脚本语言	114
3.7 项目设计	115
3.7.1 学生信息样式的设计	117
3.7.2 课程信息的显示	119
3.7.3 教师信息的显示样式	121
小结	124
<b>第 4 章 使用 XSL 显示 XML 页面</b>	<b>125</b>
4.1 XSL 的特点	125
4.2 使用 XSL 格式化 XML 页面的例子	126
4.3 XSL 的模式语言	130
4.3.1 相对定位和绝对定位	131
4.3.2 XSL 的路径匹配模式	135
4.3.3 XSL 的控制语句	151
4.4 在 XSL 当中使用脚本语句	159
4.5 多模板 XSL 文件格式	160
4.5.1 多模板 XSL 程序的作用机理	163
4.5.2 多模板程序设计的方法	165
4.6 项目设计一	167
4.6.1 绘制整个系统的 XML 结构图	168
4.6.2 设计学生信息的 XSL 模板	169
4.6.3 设计课程信息的 XSL 模板	171
4.6.4 设计教师信息的 XSL 模板	173
小结	174
<b>第 5 章 XML 文档类型定义</b>	<b>175</b>
5.1 文档类型定义	175
5.2 文档类型声明	175
5.2.1 内部 DTD 的声明	175
5.2.2 外部 DTD 的声明	177
5.2.3 外部 DTD 和内部 DTD 联合使用	178
5.3 元素声明的语法	179
5.3.1 #PCDATA 型数据	180
5.3.2 ANY	180
5.3.3 声明子标记的一般语法	182

5.3.4	可选择的子标记	184
5.3.5	复合标记类型定义	185
5.3.6	混合型标记内容	187
5.3.7	空标记的定义方法	190
5.3.8	DTD 中注释的使用语法	191
5.4	实体的声明和使用	192
5.4.1	内部一般实体	192
5.4.2	外部一般实体	195
5.4.3	内部参数实体	196
5.4.4	外部参数实体	199
5.5	属性的声明	200
5.5.1	属性的声明	201
5.5.2	属性的内容类型	205
5.6	项目设计 (设计 SCLML 学生选课系统标记语言)	213
5.6.1	设计系统的 DTD 片段	214
5.6.2	设计系统 DTD 文档	215
5.6.3	系统的 XML 数据片段	215
5.6.4	系统的所有数据	217
	小结	218
<b>第 6 章 XML Schema</b>		219
6.1	逻辑 XML Schema 的文档结构	219
6.2	元素的定义	220
6.2.1	元素内容类型的指定	222
6.2.2	标记的子标记的出现次数	223
6.2.3	标记的子标记出现的次序	225
6.2.4	子标记的分组	226
6.2.5	元素的开放性和封闭性	228
6.3	Schema 中的注释	233
6.4	项目一 (一个完整 Schema 应用程序)	235
6.4.1	程序源代码	235
6.4.2	样式文件代码分析	240
6.4.3	Schema 文件代码分析	246
	小结	247
<b>第 7 章 XML 中的名域空间</b>		248
7.1	名域空间的基本概念	248
7.2	名域的宣告	248

7.3 名域的范畴	251
7.4 默认的名域	252
7.5 使用名域的属性	254
小结	255
<b>第 8 章 XML 的数据源对象 (数据岛)</b>	<b>256</b>
8.1 数据岛的一般概念	256
8.1.1 DSO 的数据显示方法	256
8.1.2 内嵌的数据岛	257
8.1.3 外部调用数据的数据岛	259
8.2 数据岛的数据显示	261
8.2.1 通过简单表格来显示数据岛数据	261
8.2.2 通过嵌套表格来显示数据岛数据	262
8.3 通过记录集管理数据岛	268
8.3.1 创建一个记录集	268
8.3.2 通过记录集访问数据岛中各个标记的方法	270
8.3.3 操作记录集的方法	272
8.3.4 遍历记录集	274
8.3.5 对记录集进行分页	277
8.3.6 把记录取到数组中	280
8.4 数据源对象的属性和方法	282
8.4.1 数据岛文档的根	282
8.4.2 数据岛中的节点	283
8.4.3 带有数据类型的 XML 数据岛	285
8.4.4 节点数据类型的访问和设定	288
8.5 项目开发 (使用数据岛实现数据库查询的分页显示)	290
8.5.1 项目背景	290
8.5.2 项目的实现	291
8.5.3 项目的实现代码	294
小结	296
<b>第 9 章 使用 DOM 管理 XML 文档</b>	<b>297</b>
9.1 声明一个 DOM 对象	297
9.2 使用 DOM 技术访问 XML 文档中的所有节点	300
9.3 节点对象的属性	302
9.3.1 和标记内容相关的属性	302
9.3.2 和标记及标记数据类型相关的属性	304
9.4 节点和属性的创建	308

---

9.4.1	创建一个标记	308
9.4.2	创建一个属性	310
9.4.3	生成复杂节点的方法	312
9.4.4	复制某一个子节点	313
9.5	节点的删除和编辑	316
9.6	节点的查询	318
9.7	项目开发 ( 客户端分页技术 )	321
	小结	327
<b>第 10 章</b>	<b>XML 在 ASP 中的应用</b>	<b>328</b>
10.1	使用 DOM 创建一个 XML 文档	328
10.1.1	在 ASP 文档中调用 DOM 的方法	328
10.1.2	由字符串生成 XML 文档	329
10.1.3	从一个 XML 文档生成另一个 XML 文档	330
10.2	在 XML 文件中定位的方法	331
10.3	使用 DOM 在 XML 文档中添加标记和属性	334
10.3.1	在 ASP 中生成新的标记	334
10.3.2	生成新的属性	334
10.4	在 XML 文件中进行查询	336
10.5	从数据库中提取 XML 文件	338
10.6	ADO 的 XML 特性	343
10.6.1	Recordset 对象	343
10.6.2	Response 对象	346
10.6.3	Stream 对象	347
10.7	XMLHTTP 对象	349
10.8	XML 的服务器端样式	351
10.9	项目开发 ( 一个客户端的 BBS )	353
	小结	365

# 第 1 章 XML 引言

在学习使用 XML 以前,先让我们来学习什么是 XML,并了解相关的一些概念,例如什么是规范、规范的实现、标记语言、专用标记语言、通用标记语言等,尤其是标记语言的概念,希望读者对此有一定了解,因为这是学习 XML 语言的基础。在此基础之上,本章又重点介绍了 XML 设计目标、XML 样式语言、XML 超链接语言、DOM(文档对象模型)、DSO(数据源对象)、DTD(文档类型定义)等基本概念,但这些概念在本书的其他章节中还有重点讲解,此处只需要有一个初步的认识。除此之外,本章最后还介绍了 XML 在当前的应用情况及其发展前景,但该章内容不是重点,读者可以略过或进行粗略阅读。

## 1.1 XML 的发展历史

XML 的全称是 eXtensible Markup Language(扩展标注语言)。什么是标注语言、XML 的扩展性?这在以后的章节中会做进一步叙述,读者这里可以简单地理解 XML 就是一种计算机语言。尽管 XML 的应用已经有好几年的历史,但是直到 1998 年 2 月,W3C(全球信息网协会)才正式制定出统一的标准来规范 XML 的使用,这就是我们所说的 XML 1.0 规范。

什么是规范呢?规范就是计算机业界制定的一种协议,它使某种软件的应用在全世界范围内达到统一,以方便人们的使用和相互交流。而和规范密切相关的另一个概念是规范的实现,所谓规范的实现就是软件开发商开发出合乎规范的软件,但这个实现也需要有计算机厂商的支持。从确定规范到实现规范,这中间的时间很短。对厂商而言,是否能够获得市场并抓住机遇,就看如何利用好这段时间了。

计算机软件的发展在很大程度上依赖于规范的制定。道理很简单,因为只有合乎规范的软件才是真正实用和有效的软件,才能被全世界统一接受,换句话说,才是能帮助厂商赚钱的软件。因此在规范形成之前,任何厂商都会持观望态度,不敢轻举妄动,因为生怕自己的行动和将来制定出的规范相抵触,从而不被大家所接受。由此可见,制定并且遵守统一的规范对于计算机业的发展是多么重要。

目前和 XML 相关的许多规范还没有被正式通过,但就其目前的状况而言,该语言的发展方向已经基本确定。

### 1.1.1 标记语言的出现

什么是标记语言呢?标记语言就是使用某种“记号”来表示特殊信息的语言,例如用

一种“记号”来表示格式信息或表示数据信息。

什么是格式？格式就是一段文字将被显示成的样式，例如字体的大小、颜色、字型等。CSS 就是专门用来定义 HTML 文档格式的一种语言。而什么是数据呢？数据就是能够表达某种意思的一段文字，例如“中国人民”这四个字，无论用什么格式表示，都传达了“中国人民”这样的信息，因此这四个字，我们就称之为一段数据。

下面是一个具体的例子，如图 1-1 所示。

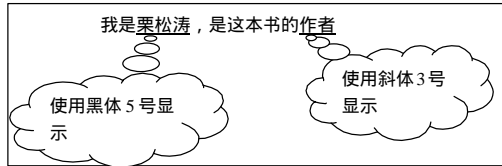


图 1-1 标记语言示意图

如果上面这段文字要送给打字小姐排版，那该如何通知打字小姐有关文字格式的信息呢？这里使用了加下划线然后在旁边进行说明的办法来解决，如图 1-1 所示，这就是一种通俗的标记语言。

这是一种很直观的语言，简单明了，但这种“简单明了”是相对于打字小姐而言的，对于计算机则非常复杂，因为计算机根本没有办法理解“栗松涛”要“使用黑体 5 号显示”的解释，也没法完成“作者”要“使用斜体 3 号显示”的要求，因为这种方法太人性化了。但尽管如此，它已经是一种标记语言了，只是本身还不能被计算机理解并执行。

那么如果要想让计算机理解并按我们的要求显示上面这段文字，也就是说如何将这些格式和数据信息通知计算机呢？很简单，我们可以用一种计算机可以“看懂”的语言来解决这个问题，具体请看例子如下：

```
我是<bold>栗松涛</bold>，是这本书的<italic>作者</italic>
```

上面文字信息使用了一种标记语言，这种标记语言的详细定义如下。

### ● 标记定义

具体的标记定义情况如表 1-1 所示。

表 1-1 标记作用对照表

标记名字	标记作用
<bold></bold>	该标记开始和结束之间的内容使用黑体 5 号显示
<italic></italic>	该标记开始和结束之间的内容使用斜体 3 号显示

### ● 语法定义

如<bold>，我们称为 bold 标记的开始标志；前面加一个反斜杠，如</bold>，则表示该标记的结束标志，而该标记的内容指的就是该标记开始标志和结束标志之间的那些内容。

这种计算机可以“看懂”的语言，也称为标记语言，但和刚开始提到的标记语言相比，其特点是所有记号都是统一定义的，并且在语法书写上有固定要求。

提示：

与最初提到的标记语言相比，该种语言明显具有计算机语言的特点，尽管不很直观，但对照着语言定义，人们还是能明白语言本身所要表达的意思。它相对于最初提到的那种标记语言有一个根本性的改变——剔除了最初标记语言中人性化的特点，使计算机能够理解并按人们的要求进行处理。

这种演变体现了标记语言的本质——用计算机能够理解的“记号”在数据文本上做标记，而有关“记号”的定义和用法的规定就是该种标记语言的核心。

### 1.1.2 标记语言的特点

标记语言是由一些格式标记（codes）或控制标记（tags）所组成的，这些标记决定了信息的显示格式或数据的意义，但这些标记不能单独存在，只是用来修饰它所包括的文本信息。例如下面的代码：

```
我是<bold>栗松涛</bold>,是这本书的<italic>作者</italic>
```

`<bold>`、`</bold>`、`<italic>`、`</italic>`是我们所使用的标记，“我是栗松涛，是这本书的作者”就是它所修饰的信息。就这些标记而言，没有任何意义，但标记结合了文本信息，例如：“我是栗松涛，是这本书的作者”以后，就表示了一种具有特定格式的文本。

通俗地讲，标记语言就是有关使用什么样的记号，和如何使用这些记号的一种语言，通过它可以和计算机进行自由的交流。像所有其他计算机语言一样，它也有一定的词汇，但这些词汇数量不多，例如有名的 HTML 标记语言，总的词汇（标记）量也不超过 100 个。除此之外，它还有一定的语法，这些语法用来控制这些词汇如何使用。

目前标记语言的种类很多，根据具体的使用范围可以分为“专用标记语言”和“通用标记语言”两种。

#### 专用标记语言

专用标记语言是用来表示某种特殊的信息，通常被某一种或几种应用软件所支持，这是终于特殊用途的需要而发展起来的。例如 HTML 语言就是一种有特殊用途的标记语言，该种标记语言是为了应用在 WWW 上的网页而专门设计的，其重点在于信息的显示，是 IE 或 Netscape 等浏览器都普遍支持的一种语言（本书后面会对 HTML 语言做讲解）。

其他的例如 Word、WPS、写字板等应用软件，也都是通过标记语言中的标记信息来处理文本中所要求的特殊格式的，例如字体大小和颜色，而它们所支持的标记语言也属于专用标记语言的范畴。

所谓专用标记语言就类似于日常生活中的“行话”。“行话”是某一个行业或领域中的人进行交流的专用语言，同样专用标记语言也是有特殊用途的一类软件相互之间共同使用的语言，可以称之为此类应用软件之间的“行话”。