



本科电子商务系列教材

XML 及其在 电子商务中的应用

方美琪 编著



清华大学出版社



本科电子商务系列教材

XML 及其在 电子商务中的应用

方美琪 编著



清华大学出版社
北 京

内 容 简 介

本书的第一部分介绍了 XML 的相关技术,包括基础技术和高级技术,由 XML 的基础语法到 Java 的 DOM 编程,由 XML DTD 到 XSL 技术,用大量实例深入浅出地介绍了 XML 的应用技术。

第二部分着重介绍了 XML 在电子商务中的应用,尤其就基于 XML 的电子商务模型和电子商务中的 XML 标准问题进行了深入的介绍,并以大量现实中的实例来分析说明。

本书是《本科电子商务系列教材》丛书之一,适合作为学习 XML 技术的编程与技术人员的教材,也适合关心 XML 的宏观概况的管理与组织人员使用。

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。

图书在版编目(CIP)数据

XML 及其在电子商务中的应用/方美琪编著. 北京:清华大学出版社,2003

(本科电子商务系列教材/陈禹,方美琪主编)

ISBN 7-302-06812-7

I. X… II. 方… III. ①可扩充语言,XML—程序设计—高等学校—教材 ②电子商务—高等学校—教材 IV. ①TP312 ②F713.36

中国版本图书馆 CIP 数据核字(2003)第 050529 号

出 版 者: 清华大学出版社(北京清华大学学研大厦,邮编 100084)

[http:// www. tup. com. cn](http://www.tup.com.cn)

责任编辑: 范素珍

印 刷 者: 北京四季青印刷厂

发 行 者: 新华书店总店北京发行所

开 本: 185×260 印张: 11.25 字数: 255 千字

版 次: 2003 年 8 月第 1 版 2003 年 8 月第 1 次印刷

书 号: ISBN 7-302-06812-7/TP·5063

印 数: 1~5000

定 价: 16.00 元



《本科电子商务系列教材》

编辑委员会名单

主 编 陈 禹 方美琪

编 委 (排名不分先后)

陈 禹 方美琪 陈国青 甘韧初

侯炳辉 邝孔武 盛定宇 左美云

杨坚争 姚立新

策划编辑 范素珍



序 言

近两年来,许多学校陆续开设了电子商务专业。作为一个新兴的试点专业,它已经引起了教育界以及社会各界人士的关注。然而,作为一个新的专业,电子商务专业究竟应当培养什么样的人才;从事电子商务的技术人员和管理人员,究竟应当具备怎样的知识与能力结构;电子商务这个新专业与现有的计算机专业和管理专业究竟应当有什么区别,以及相关的一系列问题还在讨论之中。这些问题的正确回答和处理,对于这个新专业的健康发展是至关重要的。我们编写这样一套教材,其目的不仅仅在于满足目前迫切的应用需求,更是为了在这种研究与讨论中抛砖引玉,研究切磋,以期促使这个新专业健康发展,早日成熟。

电子商务不是突如其来的孤立事物,它是几十年来以电子计算机和现代通信技术的现代信息技术迅速发展和广泛普及的结果,也是人类社会经济发展到新的阶段,生产的社会化程度达到了前所未有的全球一体化的必然产物。从学科发展的角度讲,它又是现代管理科学和新技术革命融合与交叉的最新成果。

事实上,虽然电子商务一词本身是近几年来才流行起来的,但是,实际上这件事情是由来已久的。在美国等技术比较发达的国家,从20世纪50年代中期,就已经出现了把现代信息技术应用于商务(广义的商务——business)的努力。IBM和COBOL这两个词中所包含的business一词就无可争辩地表明,人们当时已经开始考虑和实践电子商务,只不过没有用这个词而已。更有说服力的是在交通运输等领域开展的EDI(电子数据交换),它不仅属于广义的电子商务(EB)的范围,而且也是狭义的、流通领域的电子商务(EC)的内容之一,只不过以前这些内容我们往往把它们列在计算机技术应用、管理信息系统、市场营销等研究名目之下而已。

20世纪90年代初期,互联网开始用于商业用途。这一关键性的转变,使得电子商务的热潮迅速兴起,席卷全球,同样也冲击到了中国大地。现代信息技术的巨大潜力,在这个领域中得到了充分的发挥,使人们不得不刮目相看。也正是在这种热潮的推动下,电子商务这个专业,作为已有的信息管理与信息系统专业的扩展与延伸,登上了高等教育的舞台。

这个新兴的专业具有十分鲜明的综合性与实践性。

所谓综合性是指它在传统的学科结构中处于交叉学科的位置。商务活动属于管理科学的领域,它是电子商务的业务主体;而现代信息技术属于工学领域,它是电子商务的手段和技术基础,前者是“的”,后者是“矢”。从事电子商务的人不但要深入了解“的”,熟练运用“矢”,而且要能够做到“有的放矢”。这就决定了这个学科的交叉和综合性质。有一

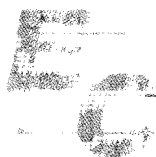
种相当普遍存在的误解,就是以为只要把计算机专业的若干门课加上经济管理的若干门课,就成了信息管理或电子商务。20 年来的经验已经表明,这种“拼盘式”的教学方案是不能真正培养出综合型的人才的。综合不等于化合,把“的”和“矢”简单地堆砌在一起,并不会自动地实现“有的放矢”。这种“化合”必须通过有特色的教材才能体现出来。

另一个特点是实践性。由于电子商务这个领域本身还在迅速地发展和变化之中,理论上还不成熟,因而实践就显得更为重要。正如人们常说的,这种人才是在实际工作中干出来的,而不是课堂上教出来的。因此,这种专业的教材需要大量实例,而且更新极快,实践环节必须大大加强。无疑,这对于教材和教师提出了更高的要求。

总之,社会对于电子商务的人才需求是非常迫切的,技术的发展也非常迅速,现在的问题和困难在于教材和老师。我们编写这样一套教材,目的在于努力探索把现代信息技术、管理科学、系统科学有机地结合起来,为开设这一专业的学校和教师提供参考和帮助。我们希望通过抛砖引玉,提出一家之言,与各界各校的同行共同推进这个新兴的专业。由于水平和经验的限制,不当之处和错误肯定不少,我们迫切希望得到广大同行的批评指正。

丛书主编 陈禹 方美琪

2002.8



前言

自从 XML 1.0 于 1998 年通过 W3C 审查以来,XML 的应用日益增多,范围越来越广,如网络、无线通信、数据库等都能看到 XML 技术的应用。微软公司提出的下一代操作系统也将完全以 XML 作为数据格式。

具体来看,在电子商务领域,XML 的诞生为电子数据交换提供了新的思路,EDI 的缺陷正是 XML 应用的长处。因为 XML 可以跨越所有边界,要获得真正独立的、跨平台的数据传送格式,XML 可能是惟一的选择。XML 的出现,突破了 EDI 的发展瓶颈,它已经开始逐渐取代传统的 EDI,成为企业之间信息交换的有力工具。原来的基于 EDI 的电子商务模型也发生了一定的变化,企业开始建立自己的基于 XML 的电子商务模型。在本书中,将会介绍几个目前基于 XML 的企业电子商务模型以及电子商务模型更前沿的发展。

随着 XML 在电子商务中的应用发展到一定阶段,出现了大量的平台互通、互联、互操作需求时,为了达到某种程度的统一而出现的对数据交换和流程交易的规范定制。在本书中就目前出现的一些电子商务标准进行分类介绍,同时选取其中发展比较完善的微软的 BizTalk 框架、ebxml 以及国内制定的 cnXML 这 3 个标准进行分析。其中还重点介绍了 Microsoft 根据 BizTalk 框架所设计出来的一项产品 Microsoft BizTalk Server 2000。

本书的第一篇介绍了 XML 的相关技术,包括基础技术和高级技术,由 XML 的基础语法到 Java 的 DOM 编程,由 XML DTD 到 XSL 技术,并用大量实例深入浅出地介绍了 XML 应用的技术。

第二篇着重介绍了 XML 在电子商务中的应用,尤其就基于 XML 的电子商务模型和电子商务中的 XML 标准问题进行了深入的介绍,以大量现实中的实例来分析说明。

本书适用的读者

本书面对的读者大致可以分为两部分。一部分是对于 XML 的技术知识不太了解的编程与技术人员。这部分读者迫切想了解 XML 的情况,包括由来发展应用状况,并且希望通过本书的学习掌握 XML 的技术,同时了解 XML 技术在电子商务平台中的应用。这部分读者应该仔细地阅读本书的第一篇。因为第一篇详细介绍了 XML 的各方面的技术,运用了比较通俗的语言和大量的实例,剖析了 XML 技术之所以被广泛应用的本质原因。在第一篇深入学习的基础上,再结合第二篇的讨论,认识 XML 在电子商务平台上应

用的大致框架结构和目前应用的情况。

本书所面对的另一部分读者是对于 XML 感兴趣,但是并不急于掌握 XML 的具体技术细节,而更关心其宏观概况的管理与组织人员。这部分读者可能更想知道 XML 在目前的电子商务中到底是如何运用的,目前的软件厂商对于在电子商务中利用 XML 提供了怎样的支持,XML 到底为电子商务带来了怎样的革命,等等。对于这部分读者,我们建议首先阅读本书的第 1 章,了解什么是 XML,以及 XML 在目前情况下应用的现状及前景,然后重点阅读第二篇。对于第二篇中出现的 XML 相关技术,可以查阅第一篇的对应章节。

常燕杰参加了第 1 章、第 2 章和第 5 章的编写;关琳参加了第 3 章、第 4 章的写作;甘涛参加了第 6 章的编写;洪祖强、陈起、兰兴伟参加了第 7 章、第 8 章和第 9 章的编写工作;程华和郝亮参加了第 10 章至第 13 章内容的编写。方美琪对全书进行了统校。

编 者

2003.4



目 录

第一篇 XML 基础技术

第 1 章 XML 绪论	3
1.1 什么是置标语言	3
1.2 什么是 XML	5
1.3 为何采用 XML	9
1.4 目前 XML 的应用情况概览	10
小结	12
第 2 章 XML 基本语法	13
2.1 什么是“格式良好的”XML 文件和有效的 XML 文件	13
2.2 XML 文档格式如何排错	15
2.3 XML 的逻辑结构	17
2.3.1 XML 的整体结构	17
2.3.2 XML 文档的有效对象	18
2.3.3 CDATA	22
小结	23
第 3 章 DTD	24
3.1 什么是 DTD	24
3.2 为什么引入 DTD	24
3.3 DTD 的声明	25
3.3.1 内部 DTD 声明	25
3.3.2 外部 DTD 声明	25
3.4 元素的声明	27
3.4.1 空元素的声明	27
3.4.2 文本元素的声明	27
3.4.3 无限制元素的声明	27
3.4.4 包含子元素的元素声明	28
3.5 实体的声明	28

3.5.1	什么是实体	28
3.5.2	内部实体和外部实体	29
3.5.3	可分解实体和不可分解实体	30
3.5.4	一般型实体和参数型实体	31
3.5.5	多个外部 DTD 同时使用	32
3.6	属性的声明	33
3.6.1	属性类型	33
3.6.2	属性默认值类型	34
小结		34
第 4 章	XML Schema	35
4.1	为什么引入 XML Schema	35
4.2	W3C 的 XML Schema 和 Microsoft 的 XML Schema	35
4.3	XML Schema 基本语法	36
4.3.1	为 XML 文档声明 XML Schema	36
4.3.2	XML Schema 的基本框架	36
4.3.3	XML Schema 声明	37
小结		41
第 5 章	XML 名称空间	42
5.1	为什么引入名称空间	42
5.2	如何使用名称空间	44
5.3	名称空间的使用范例	45
小结		45
第 6 章	XSL	47
6.1	什么是 XSL	47
6.2	XSL 总体结构	50
6.3	XSL 语法	52
6.3.1	在 XSL 中表现 XML 数据	52
6.3.2	在 XSL 输出中添加 XML 数据	63
6.3.3	在 XSL 中实现条件控制	71
6.3.4	XSL 元素的结构关系	75
6.4	XPath 路径表示法	76
6.4.1	XPath 概述	76
6.4.2	XPath 数据模型	76
6.4.3	基本表达式	77
6.4.4	样本 XML 文档的逻辑图表	77

6.4.5	上下文	79
6.4.6	轴(axis)与结点测试(node test)	79
6.4.7	位置路径与缩写	80
6.4.8	谓词	81
6.4.9	函数	81
小结		82
第7章	XML 的链接语言和指针语言	83
7.1	概述:XLink 和 XPointer	83
7.2	XLink	83
7.3	XPointer	87
7.3.1	XPointer 轴	88
7.3.2	XPointer 结点测试	88
7.3.3	XPointer 谓词断言	89
7.3.4	XPointer 位置单元集函数	89
7.4	XLink 和 XPointer 的应用实例	90
小结		92
第8章	DOM	93
8.1	DOM 概述	93
8.1.1	DOM 对象与树状结构	93
8.1.2	DOM 树的结点	94
8.1.3	不同级别的 DOM 模型	94
8.1.4	什么时候用 DOM	95
8.2	JAVA 与 XML DOM	95
8.2.1	Java 环境的准备	95
8.2.2	库文件的引入	95
8.3	JAVA DOM API 简介	96
8.3.1	Document 对象	96
8.3.2	Node 对象	97
8.3.3	NodeList 对象	97
8.3.4	Element 对象	98
8.3.5	Attr 对象	98
8.4	DOM 的实际使用	99
8.4.1	DOM 编程初探	99
8.4.2	从 XML 文档中获取结点的文本内容	101
8.4.3	从 XML 文档中获取属性值	102
8.4.4	修改 XML 文档的文本值和属性值	104

8.4.5 往 XML 文档中添加结点并保存	105
小结	108
第 9 章 SAX 介绍	109
9.1 SAX 的简介	109
9.1.1 什么是 SAX	109
9.1.2 SAX 和 DOM	110
9.2 应用 SAX	112
9.2.1 一个简单的 SAX 程序	112
9.2.2 SAX 的主要类和接口	114
小结	116

第二篇 XML 在电子商务中的应用

第 10 章 XML 对电子商务发展的影响	119
10.1 传统的基于 EDI 形式的电子商务	119
10.2 XML 在电子商务中应用的发展	121
10.3 相关技术的发展促进了基于 XML 的电子商务的发展	123
小结	124
第 11 章 基于 XML 的电子商务模型	125
11.1 3 种基于 XML 的企业电子商务模型	125
11.2 电子商务模型中的关键问题	126
11.3 动态的电子商务模型	127
11.3.1 Web 服务概述	127
11.3.2 Web 服务与电子商务的结合——动态电子商务	131
小结	132
第 12 章 电子商务中的 XML 标准分析	133
12.1 电子商务中的 XML 标准分类	133
12.1.1 标准是怎样产生的	133
12.1.2 目前电子商务的标准分类	134
12.1.3 关系体系结构的标准	134
12.1.4 功能相关的标准	135
12.1.5 业界相关的标准	137
12.2 微软的 BizTalk 框架概述	137
12.2.1 BizTalk 的框架介绍	137
12.2.2 BizTalk 框架中的术语	138
12.2.3 BizTalk 框架的体系结构	140

12.2.4	BizTalk 的消息结构	141
12.3	ebxml 框架介绍	142
12.3.1	ebxml 框架概述	142
12.3.2	ebxml 的体系结构	143
12.3.3	ebxml 在电子商务中的应用方式	145
12.3.4	ebxml 的消息结构	146
12.4	cnXML 框架介绍	146
12.4.1	cnXML 发展概述	147
12.4.2	cnXML 的消息格式	148
12.4.3	cnXML 消息规范——通信方式	153
	小结	154
第 13 章	BizTalk Server 2000	156
13.1	BizTalk Messaging Service	156
13.1.1	管理交易伙伴	157
13.1.2	转换文件格式	157
13.1.3	多样化的传输方式	159
13.1.4	信息追踪	159
13.1.5	信息交换的完成	160
13.2	BizTalk Orchestration Service	161
13.2.1	BizTalk Orchestration Service 整体结构	162
13.2.2	BizTalk Orchestration Designer	164
	小结	165
	参考文献	166

第
一
篇

XML 基础技术

本科电子商务系列教材

XML 及其在电子商务中的应用

欢迎学习 XML(eXtensible Markup Language,可扩展置标语言)。

本章将讲述 XML 发展的过程,XML 可以订制自己的置标语言,因此它超越了超文本置标语言(HTML)等一系列置标语言。但是,实际上,XML 和 HTML 并不是并列的,XML 实际上是一种元语言,因为通过它可以创建自己的置标语言。

如果读者对于 XML 几乎没有了解,那也没有关系,本书会从它的由来、它的语法、它的发展等各个方面给出具体的介绍,解决读者的每一个问题。本章主要是对 XML 进行一个概要性的讲解,包括和 XML 相关的其他知识,比如说,XML 是置标语言,那么到底什么是置标语言,XML 为什么是置标语言;既然 XML 解决的是和 HTML 相同的问题,为什么必须要用 XML,XML 的优势到底在哪里;都说 XML 有力地支持了电子商务,那么它对于电子商务的意义到底在哪里?



1.1 什么是置标语言

XML 是由 W3C(World Wide Web Consortium,互联网联合组织)于 1998 年 2 月发布的一种标准,XML 同 HTML 一样是 SGML(Standard Generalized Markup Language,标准通用置标语言)的一个简化子集。鉴于此,我们有必要在具体地讲解 XML 的细节之前,对于置标语言,以及在本书中将会频繁出现的 XML、HTML 和 SGML 作简要的介绍。

什么是置标语言?我们日常书写的语言,称为书面自然语言。如果在书面自然语言中为了标识某些信息,而加入一些标记,这种书面自然语言就可称为置标语言(英语为 Markup Language)。比如在一段书面语言中,为了说明某一句话的重要,在这句话下面画上下划线。但是,我们在这里所关注的置标语言,实际上是一种为了计算机处理而设计的置标语言,其中所用到的标记,往往使用代表一定含义的文字或数字表示。通常的做法是,根据需要,先定义一套表示描述语义和格式等信息的标记,然后将这套标记添加到书面语言中去,使书面语言变成置标语言。例如,在我们所熟悉的置标语言 HTML 中,为了让计算机了解,一段书面语言中某一部分的字体是粗体,我们可能看到的是这样一点文字 `<B> Hello! </B>`,其中 `<B></B>` 就是描述相关段落“Hello!”这部分的格式信息的标记,我们将在 HTML 和 XML 中看到,这些标记用尖括号“`<`”和“`>`”括起来,分别置于他们所描述的段落的开始和结尾部分,结尾部分增添“`/`”。

虽然 XML 在 1998 年才被首次发布,而 HTML 也不过是近十几年的产物,但他们的元语言 SGML 说起来却比较“古老”了。早在 20 世纪 60 年代,IBM 公司的三位学者为了解决由不同专用格式创建的法律文件不能在公司不同部门间相互移植的问题,提出了任

何可互操作系统的 3 个主要要求：

- 文件处理程序需要支持一个公共文档格式；
- 这个公共文档格式对特定的领域是专用的；
- 为了达到高度可靠性文档格式必须遵守特殊规则。

这个文档格式编排系统成为通用置标语言 GML，在后来的几年里经过反复的研究和探讨，对 GML 进行充分地丰富和发展，建立了标准通用置标语言 SGML，并于 1986 年被国际标准化组织 ISO 批准采用，发布为 SGML（通用置标语言标准 ISO 8879:1986）。

可以这样理解，SGML 就是对于通用置标语言 GML 的一个简化。

SGML 提供了描述文档和创建新的一致性衡量准则所必需的公共框架，它是一种指示文档标准语言或标记集的标准。这种说明本身就是一种文档标准定义。SGML 本身并不是一个文档语言，但它描述了一种文档语言，因此它是一种元语言。SGML 的思想是基于文档应该有结构和语义结构而设计的。它不关心这些语义元素如何显示，而只关心如何组织它。避免了其他文档格式共同的表示特征，而集中于信息的结构。它不提倡一种特殊的文档结构，而是要求定制数据，因此显示的方式可能有所不同。SGML 的优点可概括为：

- 不提倡一种特殊的文档结构；
- 不存在必须使用的有限标注集；
- 不限制创建新文档的标准的潜力。

但是由于 SGML 的诞生先于网络飞速发展的年代，因此它在很多方面是针对于书面而非网络传输。而且由于 SGML 当初的设计过于考虑完备性，而使最终的 SGML 文档过于冗余，仅语法说明部分就有 500 页之多，这不但影响到它的实施，更为后来实现一个完整而成本低廉的支持该标准的软件系统设置了障碍，20 世纪 80 年代只有少数大公司内部实现了 SGML 的推广。总的来说，因为 SGML 十分庞大，既不容易学，又不容易使用，在计算机上实现也十分困难，所以并没有真正地推广。

HTML 是 SGML 的一个实例，它的标签集合和文档格式作为标准固定下来。因此，HTML 不能作为定义其他置标语言的元语言。众所周知，它是 Tim Berners-Lee 设计构想的。1989 年，他对所在的研究中心的大量研究资料的不可移植性和不可兼容性深感失望，并构想了一个世界：在其中存取数据是一项简单的任务，以一种一致的方式进行，不考虑使用什么终端和程序，包含任何地点的任何计算机应能只用一个简单的常用的程序存取数据。他意识到传统的线性思路不能达到这种跨系统表达信息的要求，于是产生了超级链接概念，并最终诞生了以超链接为显著特征的超文本置标语言 HTML（HyperText Markup Language）。

也可以这样理解，HTML 是 SGML 的一个实例，相对于 SGML 针对于书面，HTML 是针对于网络，它的重点是在网络上很清晰地组织信息。

HTML 的出现为网络带来的蓬勃发展，这里自不用赘述，但是随着 Web 应用的日益广泛，曾经给网络发展带来勃勃生机的 HTML，反过来在很多方面限制着网络的进一步深入发展。一方面由于 HTML 只是 SGML 的一个实例，它的标签集是固定不可扩充的，因而缺乏扩展性，尽管 HTML 推出了一个又一个的新版本，但始终不能满足不断增长的