

Mastering XML

XML

从入门到

精通

[美] Ann Navarro
Chuck White 著
Linda Burman

周生炳 宋浩 袁海洋 肖伟 译
周生炳 审校

● XML(eXtensible Markup Language, 可扩展标记语言)是下一个大型Web主导技术,它是一种基于SGML的、简单灵活的语言,提供了比HTML更为广泛的方法来描述文档内容,同时通过一种能够在所有计算机上运行、与平台和操作系统无关的方法,提供描述元数据的机制。



电子工业出版社
Publishing House of Electronics Industry
URL: <http://www.phei.com.cn>

内 容 提 要

XML语言是用于Web开发的一项新技术,利用它我们可以通过Internet或Intranet创建并发布信息。本书首先介绍了XML的基本概念和语法规则,接着介绍如何设计数据及评价不同类型的工具功能,然后介绍了可扩展超文本语言和XML工具箱;最后介绍用于特定功能的各种应用程序,并通过分析实际商业应用案例讨论如何运用XML解决关键任务。本书教授读者如何构建XML数据的技术资源,其围绕技术细节的丰富信息框架在广度和深度上都远非纯计算机著作可比。

本书既适用于XML的新手,也能帮助具有XML经验以及HTML背景的读者拓展知识或进行转换。



Copyright©2000 SYBEX Inc., 1151 Marina Village Parkway, Alameda, CA 94501.
World rights reserved. No part of this publication may be stored in a retrieval system, transmitted, or reproduced in any way, including but not limited to photocopy, photograph, magnetic or other record, without the prior agreement and written permission of the publisher.

本书英文版由美国SYBEX公司出版,SYBEX公司已将中文版独家版权授予中国电子工业出版社及北京美迪亚电子信息有限公司。未经许可,不得以任何形式和手段复制或抄袭本书内容。

图书在版编目(CIP)数据

XML从入门到精通/(美)那瓦罗(Navarro, A.)等著;周生炳等译.-北京:电子工业出版社,2000.8
书名原文:Mastering XML

ISBN 7-5053-6015-9

I. X… II. ①那… ②周… III. 可扩充语言-程序设计 IV. TP312

中国版本图书馆CIP数据核字(2000)第63456号

书 名:XML从入门到精通

著 作 者:[美] Ann Navarro Chuck White Linda Burman

译 者:周生炳 宋 浩 袁海洋 肖 伟

审 校:周生炳

责任编辑:李 莹

印 刷 者:北京天竺颖华印刷厂

装 订 者:三河金马印装有限公司

出版发行:电子工业出版社 URL:<http://www.phei.com.cn>

北京市海淀区万寿路173信箱 邮编:100036 电话:68279077

北京市海淀区翠微东里甲2号 邮编:100036 电话:68207419

经 销:各地新华书店

开 本:787×1092 1/16 印张:33.25 字数:850千字

版 次:2000年8月第1版 2000年8月第1次印刷

书 号:ISBN 7-5053-6015-9

TP·3170

定 价:53.00元

版权贸易合同登记号 图字:01-98-2492

凡购买电子工业出版社的图书,如有缺页、倒页、脱页者,请向购买书店调换,若书店售缺,请与本社发行部联系。

为了Dave——没有你的爱和鼓励我就无法完成这个项目。

——Ann Navarro

献给我的妻子Eileen和我的儿子Lance，虽然我总见不到你们，但我每天都想着你们。

——Charles White

献给Lani——你是我灵感的源泉和向导。

——Linda Burman

致 谢

几乎每本书的致谢页都以这样的话开始，“这么多人为本书的成功作出了贡献……”，这似乎已是陈词滥调，但一部好书确实使一班人成为一个紧密团结的集体。对本书而言，这些人是第二生产工作室（Studio B Productions）的David Rogelberg；Sybex出版公司项目组：开发编辑Kim Goodfriend；合伙出版商Cheryl Applewood；编辑Ben Tomkins、Sally Engelfried；技术编辑Piroz Mohseni；项目组负责人Shannon Murphy；电子出版专家Franz Baumhackl；校对Jennifer Campbell、Richard Ganis、Amey Garber和Carrie Bradley；当然，还有我的合著者 Chuck White和Linda Burman。

另外，本书写作期间，很多人提供了想法、建议、解答和其他支持。他们的贡献是宝贵的，尽管在当时看来有些是微不足道的。谢谢你们：Frank Boumphrey（HTML作者指南）帮助我了解DOM和EBNF；Shane McCarron（Open Group）提供普通声卡服务并在深奥的句法问题上与作者有过多次讨论；Tim Berners-Lee（W3C）与作者在MIT召开的HTML WG会议上就名称空间和模式有过一次卓有成效的讨论；Tim Bray（Textuality）对名称空间的深刻洞见；Murray Altheim（太阳微系统公司）对DTD的专业知识和辩论技巧；以及这里未提及的很多人。

最后，感谢Dave——世界上最耐心的男人，居然喜欢有一个像他一样大大咧咧的妻子。

——Ann Navarro

感谢我妻子Eileen，她的耐心超出对她的要求。感谢Denise Santoro和David Wall，他们推荐我参与该计划，而Kim Goodfriend听从了他们的建议。还要感谢Zedak的Barry Goldberg和Jeff Moore，他们对最初为《纽约时报》开发的几个与报纸有关的XML和XSL文件提供了DTD和核心XML文件。感谢网景公司的Dave Hyatt和Mozilla爱好者杂志的Chris Nelson，

前者及时解答了有关XML的问题，后者则提供了屏幕快照。我还想感谢XSL规范的编辑James Clark和Stephen Deach，他们对我提出的公共XSL-List的几个问题迅速作出反应。特别感谢Liam Quinn对关于XSL的几章进行的最后技术编辑。

对我的代理人Waterside的Chris Van Buren的帮助和关怀表示诚挚的谢意。感谢Sybex公司中在本书成书过程中付出心血的人们。最后，感谢我的两位合作者，Ann Navarro和Linda Burman，他们承担了本书大部分工作。

——Chuck White

首先，我想感谢Michael Sperberg-Maqueen、Tim Bray和Jean Paoli，他们是XML规范的编辑；Jon Bosack，他在主持XML工作组时不遗余力地追求“正确地行事”；Dan Connolly，他对W3C“使一切成为可能”发挥了主要作用。

很多人因提供对本书的反馈、洞见、需要的信息以及热情鼓励而帮助过我。Dianne Kennedy除撰写了一章外，还是我们咨询和请教的固定对象；Manish Sharma使本书包括关于SIF的一章；Eric Freeze和Chris Lilley的反馈意见极有价值；Kim Goodfriend不愧她的名字（意为好朋友），她为我们提供了明智的建议和热情的鼓励，Ben Tomkins则表现出极大的耐心；我还想感谢Charles Goldfrab，为他的个人灵感和发明SGML，否则，也许永远不会有XML。

最后，感谢Ophar、Arakan和Rasputin，他们帮助我在写作本书的整个过程中保持幽默感。

——Linda Burman

译 者 序

熟悉HTML的读者都知道，利用HTML能够方便地访问因特网资源，但它不能控制数据，而XML则擅长设计数据。它是一个格式独立、与平台和应用程序无关的语言，只要应用程序都支持XML词汇，就能在应用程序之间无缝地交换数据。因此，XML意在提供Web上的结构化信息交换机制，从而将Web从发布媒体转换为一个应用程序处理环境。

你面前的这本书不仅全面系统地介绍了XML语言的技术细节，而且提供了利用XML解决实际问题的丰富案例。

正如本书作者在前言中所说的那样，这本书几乎适应于有兴趣了解XML的一切读者，不管他是初学者，还是XML老手；不管他是否熟悉HTML；不管他有没有计算背景。对初学者，本书从最基本的XML元素入手，循序渐进地介绍XML文档创建、XML数据设计、名称空间、链接以及显示XML文档的层叠样式表（CSS）和可扩展样式表语言。对具有HTML背景的读者，本书详细介绍了如何将HTML文档转换为XML文档，以及W3C的一个新的建议，可扩展超文本语言（XHTML）。另外，本书还介绍了目前流行的XML开发工具、具备XML能力的浏览器、服务器以及XML与Java和JavaScript的关系。鉴于XML尚未完全标准化，本书还介绍了XML标准化过程，我们可从中了解到主流IT企业和组织（如微软公司、IBM公司、网景公司、W3C等）在XML领域所作的大量工作和激烈竞争。本书最后用了差不多四分之一的篇幅介绍几个利用XML实际问题（包括电子商务）的案例，这将对国内企业提供有益的启发。

综观全书，我们不仅从中获得全面系统的XML技术介绍，而且对XML语言的现状、各大公司和组织对XML标准和产品的开发情况、XML的应用情况都有清楚的了解，这是本书与纯计算机著作的主要区别。

本书由周生炳、宋浩、袁海洋、肖伟译，周生炳统稿。翻译过程中得到王宴民、李满卫、刘晋伟、陈黎、鲍明涛、陈艳超、陈豫新、王欣、王芳、姜苗苗、吴传芝、孙岩、沈明华、黄志颖、许蕾、吕娟、扬凡、莫琳琪、李晓燕、刘笑凯等同志的协作，对他们的帮助，我们深表感谢。由于本书涉及范围广，内容新，语言不太规范，译者学识所限，错误在所难免，希望读者不吝指教。

译 者

引 言

尽管市面上有许多关于XML的著作，但本书与它们有很大差别。《XML从入门到精通》提供了教授你如何构建XML数据的技术资源，其围绕技术细节的丰富信息框架在广度和深度上都远非纯计算机技术著作可比。因此，本书是你案头的必备书，不管你是否已经拥有XML方面的其他书籍。

本书适合你吗？

如果你是一名XML新手，希望了解关于XML的一切新奇事物，那么是不会失望的。读完本书，你将熟练掌握XML的关键要素，能够亲手编写XML代码，了解XML家族的所有成员，并对如何开发标准以及如何在实际环境中使用XML有所领悟。

如果你已经具备XML经验，那也能在本书中找到丰富的有用信息。它将改进你的XML技能，并将你带到XML世界的最前沿。本书将拓展你的基础知识，使你从容面对产生DTD和模式、开发你自己的名称空间、将XML应用程序集成到你的指令表中的挑战。

对具有HTML背景的读者，将学习如何转换你的知识并了解为什么HTML适合显示而XML适合数据。如果你已经熟悉SGML，但想将XML添加到你的公文包中，将发现本书价值非凡，因为它告诉你SGML对Web所作的方方面面的贡献。

阅读《XML从入门到精通》无需计算背景。它将为你提供探索XML奥妙所必需的知识。

本书的组织方式

本书分为八大部分，共30章。

第一部分：XML初阶 介绍XML激动人心的发展史。我们将了解到XML究竟是什么、它擅长什么（商务）、什么组织在持之以恒地开发它们以及未来可能的趋势。该部分还包含一份XML Web资源指南。

第二部分：XML文档创建的基本组件 指导你循序渐进地创建XML文档。该部分提供所有XML文档的公共基本句法及XML语法规则。我们将学习如何创建自己的元素、属性和实体，以及什么时候使用哪一个以及它们如何组成文档对象模型。

第三部分：设计数据 集中介绍创建XML数据或文档之前必须作出的设计决策。我们将学习如何评估为特定目的而需要或交换的信息，并决定应如何定义它以便于他人使用。还会看到DTD和模式之间的区别，了解什么时候选择哪一种来约束数据。

第四部分：处理与显示XML 描述分析和处理XML的每个阶段。考察不同类型的工具功能，这使我们能够评价哪些系统和方法最适合自己。该部分将介绍更高级的特征，如利用XLink和XPointer链接文档和信息，以及利用CSS或XSL控制XML的显示。

第五部分：转换为XML 展示W3C的一个令人兴奋的建议：可扩展超文本语言（eXtensible HyperText Language, XHTML）。XHTML是一种一致的XML词汇，因此把这两个标准汇聚在一起。该部分描述XHTML的模块化和扩展的表处理等高级特征。

第六部分：XML工具箱 介绍XML开发工具、支持XML的浏览器和XML服务器技术。通过简要的实例，了解当前大型和小型企业如何在Web上使用XML。

第七部分：XML应用程序 解释“标准”的含义，并在标准组织世界内考察一番。然后提供这些应用和词汇的若干实例——常用于特定目的的XML实现，如在MathML中显示数学计算、使SVG提供可伸缩向量图，等等。

第八部分：使用XML解决企业的实际问题 强调基于XML的解决方案来解决任务中的关键商业问题。XML解决方案的开发者展示了XML多种多样的可能性，并讨论实际XML商业应用的案例。

此外，书末还有两个附录。附录A提供XML属性及其取值的全面资源——XML文档的样式和显示须知。附录B列出用XHTML建立可扩展文档所使用的抽象模块的细节。最后，本书选配光盘提供了宝贵的XML开发实用程序，它们有助于实现你的XML开发计划。

本书使用的约定

本书使用若干约定帮助你尽快找到需要的信息。如下所示的提示、说明和警告用于突出要点。

提示：在这里，你将发现内行的提示、诀窍或有趣的信息。

说明：表示作者希望引起你注意的细节。

警告：指出处理当前论题时应格外谨慎的事项。决策可能对成功有重大影响。

这是一个附栏

比说明更值得注意的有趣但有点离题的信息将在附栏中出现。

在附栏中，可找到正文中提到的公司或技术的背景信息或值得一提的某些论题的进一步介绍。

准备好了吗？

引进HTML时，我们觉得再没有什么比它更令人惊奇的了，但我们很快意识到，尽管我们能够随意去到美妙的地方并看到大量新信息，但用HTML不能控制数据。

随着XML的诞生，Web的一个新纪元开始了。你准备加入吗？翻开下一页，让我们踏上XML之路。

目 录

第一部分 XML初阶	1
第1章 通向XML之路——标记语言的演化	1
早期文字处理系统：开始，我们在纸上书写	1
Web诞生记	5
从HTML到XML	7
下章简介	9
第2章 认识XML	10
XML是什么？	10
XML的设计目标	10
介绍XML及其近亲——一族标准	12
W3C标准产生过程：如何发展XML家族	14
XML的基本部件与特征：概述	17
为什么选择XML：商业案例	19
XML的未来	21
XML资源	22
下章简介	23
第二部分 XML文档创建的基本组件	25
第3章 创建XML文档	25
掌握基本句法	25
XML语法规则	32
名称空间简介	35
下章简介	37
第4章 了解和创建元素	38
什么是元素？	38
构造元素标注	39
创建你的第一个元素	41
元素在DOM中的作用	49
撰写有效的元素标记	51
从数据库中得到元素	54
下章简介	55

第5章 了解并创建属性	56
什么是属性?	56
了解属性句法	57
了解属性类型	58
综合运用	64
下章简介	68
第6章 了解并创建实体	69
实体基础	69
如何使用实体	70
确保实体标记合法	70
一般实体	71
参数实体	77
利用实体的力量	78
下章简介	81
第三部分 设计数据	83
第7章 XML设计原理	83
选择设计方法	83
功能规范	86
设计会议	86
使用结构图	87
文档分析	87
XML可用性	93
捎带一提DTD!	95
下章简介	95
第8章 创建DTD	96
DTD: XML的语法	96
元素属性	99
建立实体	103
综合运用	106
建立DTD的工具	111
下章简介	114
第9章 XML数据模式	115
什么是模式?	115
比较模式与DTD	115
数据类型	116
模式句法	118
下章简介	122

第四部分 处理与显示XML	123
第10章 分析和处理XML	123
显示XML面临的挑战	123
Java与XML	128
XML编程基础	128
下章简介	133
第11章 了解名称空间	134
为什么需要名称空间?	134
什么是名称空间?	134
名称空间表示法	135
名称空间作用域和缺省	138
下章简介	139
第12章 XML链接	140
定位器句法	140
XPath	148
下章简介	151
第13章 显示XML: 使用层叠样式表	152
CSS是什么?	152
在XML中使用样式表	153
了解样式规则	153
块级格式化	156
考察CSS格式化性质	159
考察其他CSS媒体类型	175
CSS3一瞥	175
下章简介	176
第14章 显示XML: XSL介绍	177
XSL简介	177
创建XSL文档	182
了解转换	183
XSL格式化	199
掌握模板	215
下章简介	229
第15章 显示XML: 高级XSL	230
使用XSL工具	230
使用高级元素句法	231
XSL变量	245
使用表达式	248

XSL函数	254
关于扩展	257
格式化对象	258
下章简介	265
第五部分 转换为XML	267
第16章 从HTML变换到XML	267
为什么要转向XML?	267
HTML与XHTML之间的差异	268
由HTML进入XHTML	273
设计用于当前Web浏览器的XHTML文档	276
下章简介	277
第17章 使用XHTML	278
XHTML的模块化	278
XHTML的文档和客户机配置文件	281
XHTML展望	283
下章简介	283
第六部分 XML工具箱	285
第18章 XML开发工具	285
开发工具	285
用IDE管理XML应用程序	293
分析器	295
IBM的alphaWorks工具	296
其他工具	298
下章简介	310
第19章 XML浏览器	311
文档对象模型	311
使用XSL FO浏览器InDelv	316
使用XML和Internet Explorer	316
使用XML与Netscape Navigator	323
下章简介	334
第20章 XML服务器及服务器技术	335
什么是服务器?	335
是什么使服务器支持XML?	335
DynaBase的动态Web发布	336
Enhydra Java/XML应用程序服务器	338

XML服务器技术	340
SAIC/Bellcore: 通用商业语言与协议 (UCLP) 以及MISTI应用程序	340
Adobe System公司等: 信息与内容交换协议 (ICE)	343
下章简介	345
第21章 XML与Java	346
为什么XML需要Java	346
了解XML应用程序体系结构	347
XML应用程序体系结构中的Java	347
考察分析器	349
Java的规则表示	354
小结	355
下章简介	355
第22章 用JavaScript提取XML数据	356
在XML文档之间切换	356
列出XML元素	364
Tag Lister做什么	364
选择要显示的XML元素	369
小结	377
下章简介	377
第七部分 XML应用程序	379
第23章 标准、应用程序及词汇	379
标准与标准实体	379
标准化过程	382
词汇	385
应用程序	389
确定标准、应用程序和词汇的位置	389
下章简介	390
第24章 通道定义格式探秘	391
实现CDF	391
高级通道选项	396
微软的CDF Generator	403
通道设计提示	406
下章简介	407
第25章 使用XML应用程序	408
Web上的数学计算: MathML规范	408
SMIL: 同步多媒体集成语言	412

RDF: 资源描述框架	415
P3P: 保密选择平台句法规范	417
下章简介	421
第八部分 使用XML解决企业的实际问题	423
第26章 实例简介	423
存在真正的XML实现吗?	423
如何在现实世界中使用XML?	423
现实世界方案	426
小结	430
第27章 XML用于企业间数据集成: webMethods案例研究	431
关于Dun & Bradstreet	431
关于webMethods	433
最终产品: D&B Global Access	435
发布Global Access	446
XML开发提示	446
小结	447
第28章 WDDX: Allaire案例研究	448
企业问题: 透明的数据传输与集成	448
小结	456
第29章 利用XML管理内容: Dell案例研究	457
Dell怎样开始使用XML: 确定企业问题	458
HTML、SQL、XML与因特网信息结构	461
利用XML管理内容的技术问题	466
小结	482
第30章 XML与学校框架内的信息共享: SIF案例研究	484
学校互操作框架简介	484
SIF组件	487
SIF功能	489
探讨框架组件规范	492
小结	498
附录A 可扩展样式表语言的格式特性	499
附录B XHTML模块化中定义的语义模块	508

第一部分 XML初阶

第1章 通向XML之路——标记语言的演化

如果你最近读过很多关于Web的读物，那么无疑听说过XML（这也许就是你购买本书的原因之一）。在本章中，我们将介绍XML（Extensible Markup Language，可扩展标记语言），它最适合作为构造数据的工具。XML提供了将文本和其他媒体安置到构件中的规则，并允许你管理和操纵得到的结果。

在全面了解XML是什么以及为什么开发它之前，了解它的来龙去脉是有益的。为此，我们追溯到一般标记语言的根源，从计算机时代早期的文字处理解决方案开始。

早期文字处理系统：开始，我们在纸上书写

很早就出现了文字。不只是一个字，而是很多——数以千计、百万计、亿万计。如何处理它们很快成为一个问题。如何处理它们才能使其本来含义清晰无误？世界上第一个文字处理系统牵涉到用纸和笔（或者可能是凿子和岩石）记录文字。在计算机时代，文字处理不仅包括初始文档产生，还包括文档存储、格式编排、出版和其他分析、润色和修改文本的工具。

过去十年在字处理和桌面出版系统方面的进展掩盖了文本格式编排任务的复杂性。用流行的字处理程序（如Microsoft Word或Corel的WordPerfect）编写的文档在屏幕和在打印纸上相同，这产生了一个拗口的首字母缩写：WYSIWYG，即“所见即所得（What You See Is What You Get）”。

尽管确实很方便，但WYSIWYG文档处理工具也具有相当大的欺骗性。用户不知道产生页面外观和版面布局的计算机指令。在字处理的这些技术出现之前，这种格式编排是由遵循书面格式指示的人类印刷工人进行的。从字面上看，“标记”这个术语来自在打印文稿上书写格式指示的作法——“把它标出来”。今天，同样的标记过程出现在Word和其他文字处理环境中，包括XML。

通向XML之路也从纸和笔开始

基于纸和笔的标记与今天电子标记之间转换的过程经过了若干一般性阶段。当标记过程第一次转向计算机时，提供同样的指示，例如：“将这一节加粗”、“字体样式设置为12磅加拉蒙字体”，这一系列指令称为过程——产生文档的代理程序将要完成的过程，因而产生“过程标记”这个术语。

遗憾的是，过程标记仍具有很多与手写标记相同的隐患。它的注重点是表示而不是结构。两段具有相同视觉处理的文本不一定在结构上相关：一段文本可能是一个地址，另一段可能是脚注。更改为预想的表示要求一个附属标记过程。此外，因为标记过程注重视觉结果，所以它常常依赖于打印机、计算机屏幕和纸张的能力。

宏命令的引入

标记技术的下一阶段是宏的引进。宏是一组命令和/或一系列键盘击键，它由一个唯一的名称或标注符号来标识，该标注被插入到文档中应执行该指令集的地方。然后，在最终显示之前，这个类属代码传给一个附属格式编排应用程序。

提示：格式编排应用程序不一定是一个独立的附属程序，它常常只是更大程序的另一个组件，就像大多数文字处理软件包一样。

当作者打算改变处理文档结构的方式时，类属代码可能大大节省时间。如要更改一级标题的表示方式，只要在一个地方编辑指令即可：即，只需更改为一级标题所定义的宏，而无需更改一个一级标题在文档中的每次出现。因为对一个宏标注，只定义一次格式编排指令，所以它们也起到写作速记的作用。

除了辅助写作外，类属代码还提供一种使得自动过程与文档交互的手段。可以命令编辑程序查找所有项目列表，并将它重新编排为编号列表。更复杂的过程可用于提取和编辑大型文档或其他邮件写作处理任务的概要。

GML：按结构表示

六十年代末，IBM公司的三位研究人员Charles Goldfarb、Ed Mosher和Ray Lorie开始着手解决如何应付数以千计的法律文件的这一全球性问题，这些文件是在不同系统中使用专用格式创建的。因为这些约束，它们不可能在整个公司间移植。他们的研究揭示了任何可互操作系统的三个主要要求：

- 文件处理程序需要支持一个公共文档格式。
- 这个公共格式对他们的领域（即法律文件）应是专用的。
- 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称为广义标记语言（Generalized Markup Language, GML）。

说明：GML还代表三位研究人员Goldfarb、Mosher和Lorie的首字母。

一个公共文档格式

GML的第一个要求是涉及的系统必须支持一个公共文档格式。当时的文档创建包含文本格式指令，所以它是将GML建立在文档标记模型基础上的一个逻辑步骤。然而，这种标记的语义在第一次实现时应是统一的。

与领域有关的格式编排

Goldfarb、Mosher和Lorie尝试解决的问题是处理法律文件。任何接触过法律契约的人都知道它们整个自成体系。拉丁词组和极其特殊和晦涩的术语随处可见。如果打算创造一种通

用语言，那么该语言的特征表达不了法律文件的独特需要。因此，专门化是一个好东西，它成为他们的第二个要求：他们的语言专门应用于他们的需要。

基于规则的格式编排

最后一个要求认识到，对于成功的计算机文档处理，必须拥有某些手段来检验标记与公共文档格式是否一致以及是否遵守任何与领域有关的格式约束。

这个最终要求可能不如另两个显而易见。新系统应是可移植的，要解决的特殊问题是清晰无误的。但考虑一下如何才能满足这两个要求，最重要的是，如何应对与领域相关的问题。

为了更好地理解这个概念，考虑备忘录这种简单的文档样式，看看它包含的常见元素，如图1.1所示。

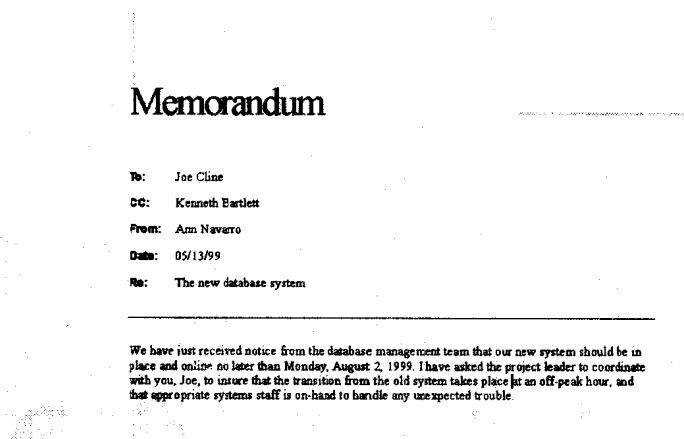


图1.1 一份简单的办公室备忘录

你所关心的格式是文档的结构格式编排，而不是视觉形式。对该文档系统而言，使用30磅New Roman字表达标题“Memorandum”是无关紧要的，同样，它包含水平线之类视觉元素也没什么关系。重要的是每份备忘录中包含的六种元素：

- 该文件提交给谁（“To:”域）
- 已经送给谁一份文件（“CC:”域）
- 谁发出该文件（“From:”域）
- 文件书写日期（“Date:”域）
- 主题（“Re:”域）
- 文件正文

如果你想创建备忘录的标记语言，其结构无疑应包括这六个元素。使这种方法进一步标准化，你可能会给每个元素赋予标注，而不管该文件是在哪里或由谁创建的。例如，文件欲提交给他那个人前面总有To:记号，主题总是使用Re:记号。

备忘录文档类型的这些规则为计算机处理文档建立了一个基本词汇表。这个规则集叫做文档类型定义或DTD（document type definition）。本书后面将详细讨论DTD，第8章“创建DTD”专门介绍创建你自己的XML DTD的过程。现在，只要知道DTD代表给定文档类型中可能出现的一组元素就够了。

既然有了一个概要描述文档可能包含的可能元素的标准文档类型和文档类型定义，你就可以以任何方式处理符合该定义的文档。例如，如果要求计算机查找发给Jeo Cline的所有备忘录，就可以命令查询系统搜索每个文档的To:域来查找匹配的实例。因为每个文档都有这个To:域，不存在仅仅因为备忘录格式不同就找不到文档的危险。

标准化GML

在其后的八年中，Goldfarb继续研究Generalized Markup Language，并充任一个国际开发人员组织的技术领导。他们终于建立了所谓Standard Generalized Markup Language（标准广义标记语言，SGML），而且国际标准化组织（ISO）于1986年最终批准采用SGML。

说明：ISO并不是国际标准化组织（International Organization for Standardization）的字母缩写（该缩写应为IOS），而是来源于希腊字isos，意思是平等。对国际标准团体来说，这是一个恰当的名称，你觉得呢？

ISO是一个包括130多个成员国的国际标准联合会，每个成员国一名代表。该组织成立于1947年，其宗旨是促进科学和学术合作，推动全球商贸和信息交流。

SGML目前由其部分创始者管理，但它现在属于一个称为JTC1/SC34（International Organization for Standardization/International ElectroTechnical Committee, Subcommittee 34）的新小组。除了关于SGML的工作以外，该小组还提出相关工作的标准，如处理超链表示的HyTime和DSSSL（Document Style Semantics and Specification Language，文档样式语义和规范语言），后者标准化样式表的创建并作为W3C的工作基准（关于W3C，见下述附栏）。

管理Web：万维网联合会（World Wide Web Consortium, W3C）

万维网联合会（World Wide Web Consortium, W3C）成立于1994年底，其主要任务是指导HTML的构造和发展。W3C发表了一份行动报告，该报告总结了标准发布程序中涉及的因素，确定其目标并展望未来的方向。可在<http://www.w3.org/MarkUp/Activity>找到该文件。

今天，350多家公司合作制定规范，其范围几乎囊括Web的所有方面，从HTML和XML到描述资源的方法，包括RDF（Resource Description Framework，资源描述框架）、P3P（Platform for Privacy Preferences，保密选项平台）的保密实践指南以及数字签名。

该协会挂靠在三家学术机构：美国的麻省理工学院计算机科学实验室（MIT/LCS）、欧洲的国家信息和自动化研究所（INRIA）和亚洲的庆应大学湘南藤泽学院。W3C由成员组织资助，它将主要行业成员聚集在一起，制定每个人都能使用的Web规范。

SGML

SGML最好通过它不是什么来描述：

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

SGML提供描述文档和创建新的一致性衡量准则所必要的公共框架。几乎所有处理文档的语言都可能至少部分追溯到SGML。确实，如果不实行这种描述标准，我们今天所看到的Web本来是不存在的。