



万水计算机编程技术与应用系列

XML

基础与应用

(原书第三版)

XML A Primer (3rd Edition)

[美] Simon St.Laurent 著

云舟工作室 译



中国水利水电出版社
www.waterpub.com.cn



万水计算机编程技术与应用系列

XML 基础与应用

[美] Simon St.Laurent 著

云舟工作室 译

中国水利水电出版社

内 容 提 要

本书是一本全面介绍 XML 的书籍。全书分为四部分, 第一部分(第 1 章至第 4 章)介绍了 XML 标记语言的演进历史以及基本概念和语法; 第二部分(第 5 章至第 10 章)介绍了 XML 的相关文档, 包括 XML 处理过程、文档类型的定义以及大纲和 RELAX 等; 第三部分(第 11 章至第 17 章)相关支持规范, 介绍了浏览器的处理以及 CSS, XPath, XSL, XPointer, 超文本、查询等方面的内容; 第四部分(第 18 章至第 24 章)是 XML 的实践部分, 着眼于 XML 的规划、商业应用、文档管理以及数据库驱动等, 并且介绍了 XML 的创建和管理方面的许多问题。

该书是一本入门书, 适合对 XML 感兴趣的程序员和其他读者。

Original English language edition Copyright © 2001 by Hungry Minds, Inc. All rights reserved including the right of reproduction in whole or in part in any form. This translation published by arrangement with Hungry Minds, Inc.

北京市版权局著作权合同登记号: 图字 01-2001-4708 号

图书在版编目(CIP)数据

XML 基础与应用 / (美) 劳伦特 (Laurent,S.S.) 著; 云舟工作室译. —北京: 中国水利水电出版社, 2002

(万水计算机编程技术与应用系列)

书名原文: XML: A Primer

ISBN 7-5084-1003-3

I .X… II.①劳…②云… III.可扩展语言, XML—程序设计 IV.TP312

中国版本图书馆 CIP 数据核字 (2002) 第 010878 号

书 名	XML 基础与应用
作 者	[美] Simon St.Laurent 著 云舟工作室 译
出版、发行	中国水利水电出版社(北京市三里河路 6 号 100044) 网址: www.waterpub.com.cn E-mail: mchannel@public3.bta.net.cn (万水) sale@waterpub.com.cn 电话: (010) 68359286 (万水)、63202266 (总机)、68331835 (发行部)
经 售	全国各地新华书店
排 版	北京万水电子信息有限公司
印 刷	北京天竺颖华印刷厂印刷
规 格	787×1092 毫米 16 开本 20.75 印张 451 千字
版 次	2002 年 3 月第一版 2002 年 3 月北京第一次印刷
印 数	0001—5000 册
定 价	30.00 元

凡购买我社图书, 如有缺页、倒页、脱页的, 本社发行部负责调换

版权所有·侵权必究

译者序

网络技术的飞速发展，改变了人们熟悉的学习、生活和工作方式，拓宽了人们获取知识和信息的途径，也大大改变了人们的思维方式。新思想、新概念、新创造和新技术的不断涌现，令人应接不暇，XML 技术作为一项新技术，从 1996 年开始发展得十分迅速。

XML 可扩展标记语言是 SGML（标准通用标记语言）的一个子集，它旨在为 Web 信息管理人员提供更大的灵活性。它允许 Web 开发者和设计人员创建自己的标记，相比早期的 HTML 文档编码系统，其在组织和信息显示方面具有更加强大的功能和自由度。

本书作者 Simon St.Laurent 是计算机网络，特别是 Internet 方面的专家，具有丰富的 Web 站点开发和网络管理的实践经验，而且有关于 XML 和网页创作方面的多部著作。作者以平实易懂的语言向作者深入浅出地讲述了 XML 的由来、特性、功能和使用等各方面的信息。本书不要求读者具备任何 SGML 的知识，结构严谨、语言流畅、示例详实是本书的一大特点，因此无论初学者还是有经验的开发人员都可以从阅读中受益匪浅。

全书分为四部分，第一部分（第 1 章至第 4 章）介绍了 XML 标记语言的演进历史以及基本概念和语法；第二部分（第 5 章至第 10 章）介绍了 XML 的相关文档，包括 XML 处理过程、文档类型的定义以及大纲和 RELAX 等；第三部分（第 11 章至第 17 章）介绍了浏览器的处理以及 CSS, XPath, XSL, XPointer, 超文本，查询等方面的内容；第四部分（第 18 章至第 24 章）是 XML 的实践部分，着眼于 XML 的规划、商业应用、文档管理以及数据库驱动等，并且介绍了 XML 的创建和管理方面的许多问题。

本书由云舟工作室组织翻译，王伟、魏宏刚、陈维强、曹毅、吴刚、杨作梅等同承担了主要翻译工作。另外，吴峰、郑伟、李训青、许瑞清、唐忠华、钟晓利、曹国权等也参加了翻译、校对和排版工作。我们有幸能参加本书的翻译工作并把这本好书介绍给您，希望能为您掌握 XML 这一极具应用前景的工具尽绵薄之力。由于时间仓促，水平有限，疏漏之处在所难免，敬请批评指正。

译者

2001 年 9 月

前 言

可扩展标记语言 (Extensible Markup Language, XML) 早已被新闻媒体大肆炒作, 通过这本 XML 入门的第三版, 我希望能展示一个真实的 XML, 探索它目前的现状和几种可能的未来发展方向。

XML 已经由以前的规范发展成付诸实施的流行语, 成千上万, 甚至上百万开发者和使用者在不知不觉的情况下正在享受基于 XML 的服务, 商业开发者、系统集成员、网站开发者、文档管理者, 还有广大范围的程序员正在把 XML 作为一个万能胶把各种信息和数不胜数的不同系统连接起来。微软就曾在电视使用“他不知道什么是 XML”来作为醒目的广告语, 而 XML 更是频频出现在各种值得关注的新技术列表清单中。

自从三年前我开始写该入门教程的第一版以来, XML 已经发生了许多激动人心的变化。那时候, 看起来似乎可以用薄薄的一册书就能详尽描述 XML1.0 的所有可能性及其相关的规范, 并提供一个简洁而且全面的介绍。该书第一版是当时的 XML 规范页数的好几倍, 而且尽管说明规范比较难以阅读, 他们所描述的结构却并不十分复杂。而现在的技术规范却大大超出了本书的份量, 而且有越来越多的开发者需要这项技术的简要说明。这些规范的其中之一——《W3C XML 纲要》甚至就提供了多达 92 页的入门教程。

自从 XML1.0 首次问世以来, XML 大家庭已经有了实质性的分化。最初, XML 被称作“为互联网设计的 SGML”而广为人知, 它的诞生地在万维网联盟 (World Wide Web Consortium, W3C), 该联盟的重点在于超文本和对现存的互联网技术进行的集成。这些都使得 XML 看上去像是供网页开发者选择的下一代工具。SGML 家族提供了最初的 XML 家族核心, 并提供了大量经验、工具和对于以文档为中心的信息模型的理解。在这些族群已经并且还将继续为 XML 作出实质性贡献的同时, 大多数真正建构基于 XML 的应用程序的社团是由那些计算机之间相互通讯的开发者构造渠道所组成的。

《XML 基础与应用》保留了首次使用 XML 的用户和开发者的风格, 但在处理方法上又作出了相应的调整。读者会发现有关 XML 核心部分的详细介绍——XML1.0, 名字空间以及 DTD——有必要补充技术方面的信息, 我们将指导读者参考相应的资源。该版本涉及到 XML 的处理风格, 而其详细内容将留给那些直接定位于具体方法、具体内容的各种指南。

绝大多数使用 XML 的开发者努力寻找一套他们习惯的工具, 然后在此基础上工作直到出现分歧。在该书中一开始我着重讲述开发者所需要理解的核心工具, 然后讨论其他一些不断发展并常有争议的工具箱, 使用的示例有助于在创建自己文档的过程中启发思考。

XML 起初本来是较小而且相对简单的技术, 因此没有多少值得我们感谢的预言家和大胆开拓的先驱。它的不断发展有时候相当费劲, 但开发者可以从众多开发者创建的巨大工具组中获得益处, 这些开发者中的许多人有效地分发了他们的工作成果, 或是以开放源码软件的

形式，或是在社区论坛上的帮助言论，使得 XML 除了一套技术规范之外还有一个丰富的工具库和具体实现。我希望这本新版的《XML 基础与应用》将会吸引更多一些的成员加入到这个社区及其它所创建和共享的工具组中。

本书的使用对象

这本书将有助于任何一个刚刚开始使用 XML 的读者，重点偏向信息处理的文档，但对使用 XML 来交换原始数据感兴趣的读者将会得到同样有用的信息。HTML 的开发者们将会高人一筹，带着对基本标记语言的理解来学习 XML，但我希望任何对网页有一些基本了解的读者都能理解它。该书绝对是一本入门读物，读者不会发现一些 XML 的阐述或是为程序员准备的代码例程，但将会对 XML 标记语言具有一个扎实的基础知识。

本书中使用的一些约定

贯穿全书，XML 和 HTML 内容以 `courier` 字体的形式出现，如果在一般句子中遇到 `courier` 字体，表示对 XML 或 HTML 结构或内容的应用。

本书中使用的图标

本书绝大部分内容采用文本和图片的一般性混合编排，但经常有一些重要信息需要单独列出，如一些有用的旁注、小提示、参考或者是最重要的警告。

注意

“注意”为当前讨论的主题提供额外的细节，但并非所有人都会感兴趣，它提供额外的有用信息，但并非关键性的信息。

提示

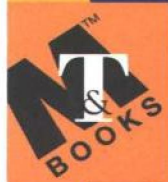
“提示”是那种通常不编入传统文档而通过实践经验学到的信息，它适合于任何场合，而且使用它们能真正地节省时间。

警告

“警告”图标非常重要，你不应该把它当成是出现了问题，而是那些看似平常的操作可能产生可怕的后果。为了避免潜在的问题发生，请务必阅读这些警告。

致谢

Tracey Cranston 亲历了该书三个版本的发行，而当前版本非常感谢她提出的意见和支持。Ann Lush 在该书刚开始启动时就看到了它的潜力，而 Grace Buechlein 则保证了该版正常运作。感谢 Susan Christophersen 和 Eric Newman 为本书发行所做的工作，感谢 S.B.Kleinman 所做的语言校对工作，感谢 Ethan Cerami 在技术检查上的辛苦工作，其余错误纰漏之处，皆出于我。



责任编辑：黄海

XML 基础与应用 (原书第三版)

本系列新书

- ◆ C#语言与程序设计
- ◆ ASP编程技术与实例分析
- ◆ OpenGL图形图像程序设计指南
- ◆ Windows 2000注册表即时详解
- ◆ C# COM+编程指南
- ◆ XML基础与应用

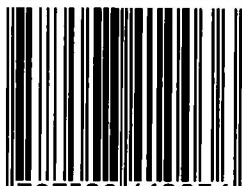
XML地图

在过去的几年中XML已经完成了从规范到实现的转变。现在它正被修改，以适应最新W3C说明和XLink、XPath、XSLT等开发工具。即使以前对XML和SGML毫无经验，这本清晰易懂、例子丰富的入门书也展示了如何把握这门强大的网络语言，来开始创建数据驱动的应用程序。

将XML投入到工作中：

- ◆ 理解最新的W3C XML说明
- ◆ 用手边最好的开发工具加速开发
- ◆ 将XML与CSS, DHTML, XSL及其他技术集成到一起
- ◆ 创建文档类型定义和用户标签的实现
- ◆ 建立在线目录和在线文档结构
- ◆ 为数据库和网页间信息交换开发数据格式

ISBN 7-5084-1003-3



9 787508 410036 >

北京万水电子信息有限公司

Beijing Multi-Channel Electronic Information Co., Ltd.

地址：北京市西直门外榆树馆一巷永康商务写字楼

邮编：100044

电话：(010)6835.9286, 6835.9167

传真：(010)6835.9284

E-mail: mchannel@public3.bta.net.cn

ISBN 7-5084-1003-3/TP · 385

定价：30.00 元

目 录

译者序
前言

第 I 部分 入门

第 1 章 结构化标签信息	1
1.1 文档方面	2
1.1.1 “所见即所得”带来的灾难	2
1.1.2 HTML 的快速成长	3
1.1.3 导航 HTML 碎片	4
1.1.4 追本溯源：结构和 SGML	5
1.1.5 HTML 的根源	6
1.1.6 XML 的兴起：具有 SGML 20% 的复杂性及其 80% 的功能	7
1.2 数据方面	8
1.2.1 二进制格式	8
1.2.2 XML 冗长，但是结构化、标签化、易于交换	8
1.3 谁来控制 XML	9
1.4 不断前进	10
第 2 章 在标记库基础上工作	11
2.1 检阅核心部分：标记文本	11
2.2 能用 XML 做很多事情	15
2.2.1 描述和共享数据及文档格式	15
2.2.2 创建、显示和维护文档	16
2.2.3 创建读写 XML 的应用程序	17
2.2.4 创建、管理和提交数据	18
2.3 深入 XML 丛林	19
第 3 章 XML 核心：元素、属性、文本和名字空间	20
3.1 标准组件：元素结构	20
3.1.1 构建清晰结构：一个 HTML 示例	21
3.1.2 创建元素标记	23

3.2	用属性补充元素	25
3.3	名字空间：标识元素和属性	28
3.4	文本内容	30
3.5	使用语法分析器和浏览器测试文档	32
第 4 章	增加更多的部分	36
4.1	<?xml?>：一个特别的声明	36
4.1.1	版本号	36
4.1.2	编码声明	36
4.1.3	独立的声明	38
4.2	CDATA 部分	39
4.3	注释	40
4.4	处理指令	40
4.5	预定义属性	41
4.5.1	xml:space	41
4.5.2	xml:lang	43

第 II 部分 描述文档

第 5 章	XML 处理过程的类型	45
5.1	语法分析器和程序	45
5.2	校验和非校验语法分析器	46
5.3	名字空间和大纲检测语法分析器	48
5.4	语法分析器选择	49
第 6 章	文档类型定义	51
6.1	入门简介	51
6.2	文档类型声明	59
6.3	元素类型声明	61
6.4	属性列表声明	66
6.5	注释	70
6.6	表示法声明	70
6.7	实体	71
6.7.1	通用实体	71
6.7.2	未解析实体	73
6.7.3	参数实体	73
6.8	DTD 中的记号部分：IGNORE 和 INCLUDE	76

第 7 章 大纲：下一代产品？	78
7.1 DTD 的局限	78
7.1.1 数据类型	78
7.1.2 名字空间处理	79
7.1.3 扩展性	79
7.1.4 管理	80
7.2 大纲派系	81
7.3 大纲构建工具	83
7.3.1 大纲的手工创建	83
7.3.2 XML 著作权威	83
7.3.3 XML Spy	84
第 8 章 W3C XML 大纲：数据类型	85
8.1 基本数据类型	85
8.2 其他内建类型	87
8.3 数据类型扩展	88
第 9 章 W3C XML 大纲入门：结构	90
9.1 使用对象方法描述 XML	90
9.2 理解结构	90
9.2.1 名字空间和 W3C XML 大纲	91
9.2.2 基本元素结构	91
9.3 浏览高级特性	96
9.3.1 结构：扩展、限制和复用	96
9.3.2 开放内容模型	97
9.3.3 替换组	97
9.3.4 用于文档记录的注释及其扩展	98
9.3.5 路径测试	98
第 10 章 RELAX 入门	100
10.1 大纲与 RELAX	100
10.2 使用 RELAX 术语	101
10.2.1 创建模块	101
10.2.2 在模块之内定义元素	101
10.2.3 在模块内定义属性	103
10.2.4 使用注释进行文档管理	104
10.2.5 用于复杂内容模型的栅栏规则	105
10.2.6 作用、标签和上下文相关性	106

10.2.7 重新构建木材产品清单	106
10.3 在 RELAX 核心上工作	108

第 III 部分 相关支持规范

第 11 章 处理手段：浏览器、DOM、SAX 及其他	109
11.1 浏览器中的 HTML 和 XML 语法分析器	109
11.2 浏览器之外：应用程序中的语法分析器	112
11.2.1 建构和操作树状结构：DOM	113
11.2.2 检测事件：SAX	114
11.2.3 新的竞争者	115
第 12 章 基于结构的格式化：CSS	116
12.1 将样式表与文档联系起来	116
12.2 应用带有结构化格式的样式表	117
12.2.1 规范的发展	117
12.2.2 将 XML 和 CSS 集成起来	118
12.2.3 创建样式表	125
第 13 章 XPath：标识文档部分	129
13.1 描述位置	129
13.2 从位置测试创建路径	131
13.2.1 创建绝对 XPath	132
13.2.2 创建相对 XPath	133
第 14 章 XSL	135
14.1 理解树状结构（转换）	136
14.2 格式化对象	143
14.3 使用 Schematron 来创建规则	145
第 15 章 使用 XPointer 标识文档片断	146
15.1 URL 和 URI	146
15.2 XPath 扩展	147
15.2.1 转义字符	147
15.2.2 两个非 XPath 处理手段	148
15.2.3 基于 XPath 的 XPointer	149
15.3 XPointer 的承诺	150
第 16 章 超文本 XML：XLink 和 XML Base	152
16.1 XLink 名字空间	152

16.2	简单链接	152
16.2.1	HTML 中的链接	153
16.2.2	XML 中的简单链接	153
16.3	XHTML 和 XLink 之间的冲突	155
16.4	更多复杂的链接	156
16.5	Roles 和 Arcroles	158
16.6	链接库	159
16.7	从规范到实践	160
16.8	Xinclude: 添加内容, 而不只是嵌入	161
16.9	XML Base: 处理相对 URI	162
第 17 章	查询 XML	164
17.1	管理数据的所有类型	164
17.2	路径、指针和查询	165
17.3	正在建设中	167
17.4	展望未来: Quilt	167

第 IV 部分 建立自己的标签

第 18 章	现在计划, 为将来做准备	169
18.1	XML 的使用对象	170
18.2	注意结构	171
18.2.1	文档结构	171
18.2.2	数据结构	174
18.3	书写计划	178
第 19 章	用 XML 重建网页和纸张文档	180
19.1	从 HTML 到 XML	180
19.2	创作本书	189
19.3	第一关: 一种看起来具有老式样式的 DTD	191
19.4	章节 DTD 的 CSS 样式	198
19.5	第二关: 走向更精简的 DTD	204
19.6	创建包装文档	208
第 20 章	XML 在商业上的应用	210
20.1	XML 面向的对象	210
20.2	展示货物或服务	211
20.2.1	用 DTD 描述文档结构	212

20.2.2	用 W3C XML 大纲描述文档结构.....	216
20.2.3	用 RELAX 模块描述文档结构.....	220
20.3	收取订单——商务处理.....	222
20.3.1	为订单设计 DTD.....	223
20.3.2	走向订单的大纲.....	228
20.4	直接连接：信息的交换.....	232
第 21 章	XML 在文档管理方面的应用.....	234
21.1	XML 的继承性：SGML 和文档管理.....	234
21.2	XML 文档管理的未来.....	235
21.3	通向无纸办公的一小步.....	236
21.4	创建历史：用于公司记录的一个 DTD.....	246
21.5	发布文件的方法.....	252
21.5.1	用 ICE 来组织文档.....	252
21.5.2	利用 RSS 来生成页头标题.....	252
21.5.3	信道定义格式的应用.....	255
第 22 章	XML 在数据驱动程序中的应用.....	257
22.1	用于交换的数据.....	257
22.2	用于控制的数据.....	258
22.2.1	一个房间的照明.....	258
22.2.2	定义一个使用 DTD 进行开关照明的文件接口.....	259
22.2.3	用 XML 大纲重新生成控制结构.....	265
22.3	控制工具和工具标识语言.....	270
22.4	目标文件.....	272
22.4.1	粒媒标识语言（BeanML）.....	272
22.4.2	Quick 和 QARE.....	273
22.4.3	XML-RPC.....	274
22.4.4	SOAP.....	275
22.5	元数据：使用 XML 来描述资源.....	275
22.5.1	XSA（XMLSoftware Autoupdate）.....	276
22.5.2	资源描述框架（RDF）.....	277
22.5.3	Dublin Core.....	277
22.6	展望.....	279
第 23 章	存储、管理和处理 XML.....	280
23.1	创建 XML.....	280
23.1.1	编辑 XML.....	280

23.1.2 生成 XML	283
23.2 使用中介	284
23.2.1 储存库-XML 存储	284
23.2.2 中介	286
23.3 用户交流（浏览器与其他应用）	287
23.3.1 浏览器：剖析和展望	287
23.3.2 入口和其他网关	291
23.4 XML 与其他技术结合	291
23.5 创建常用的 XML 应用程序	293
23.6 选择 XML 编程工具	294
第 24 章 不断扩展的 XML 网络	296
24.1 XML 和传统网络	296
24.1.1 升级 HTML：用于 XML 语法	297
24.1.2 在传统网页上使用 XML	299
24.2 XML 和一种新型网页的创建	303
24.3 网页、网页、更多的网页	305
术语表	306

第 I 部分 入门

- 第 1 章 结构化标签信息
- 第 2 章 在标记库基础上工作
- 第 3 章 XML 核心：元素、属性、文本和名字空间
- 第 4 章 增加更多的部分

第 1 章 结构化标签信息

可扩展的标记语言 XML 承诺要让网页的基本结构超越 HTML，并用一种功能更加强大、扩展性更高的体系结构取而代之。XML 为开发者提供了一个途径来创建他们自己的格式，这些格式用于文档和数据时组织严密而且标签清晰。XML 将会把网页从基于格式的结构中解放出来，而使得网页中的数据和文档一样重要。随着 XML 对越来越多的公众敞开门户，那些主导网页浏览器的开发者们正在逐步丧失对网页术语发展和实现的控制。XML 同时还向应用开发者们许诺，不管他们是否在网页上进行开发，使用 XML 都将获得一种极其方便的格式来存储各种不同的信息。

XML 使得开发人员可以创建他们自己的标记语言，同时可以使用那些他们自己用来协同工作的专业术语作为标记，并且更加简单也更加容易做到这一点。在创建过程中开发人员可以使用 HTML，但不是仅仅局限于使用 HTML。XML 使得开发者可以创建基于逻辑内容而不是基于格式的标记结构，这种智能标记将使得人们和计算机检索信息变得非常容易，他们可以在整个文档中检索专门内容的信息，而不仅仅是在某个页面的全文中进行检索。

XML 最初起源于文档处理，但它现在更多的应用在于数据交换。“文档和数据的分离”，即对于不同的需求产生的持续争论是 XML 用户群的一贯特征，但 XML 更多的是为两方面都提供了大量的支持，还有可能一劳永逸地消除上述分歧。把 XML 和风格页技术一起使用，将使得网页作者能够制作富有吸引力而且易于管理维护的页面。与此同时，它让各种开发者对于他们的信息的控制具备了一种更新层次掌握能力，并且这种控制拥有多种多样的灵活性。

1.1 文档方面

XML 起初完全是处理文档的，它使得我们能够描绘文档时不仅仅局限于固有的格式，同时对集成各种数据结构的能力提供支持。尽管当前 XML 应用于文档创建和维护的现状还不是那么令人满意，但至少为这种可能性提供了光明的前景。XML 使得人们可以不受“所见即所得”方法和 HTML 提供的“一个尺寸适合所有”等术语的限制，可随意创建自己的文档。

1.1.1 “所见即所得”带来的灾难

我曾经用过的第一个文字处理器是一个非常简单的文本编辑器，当时看到通过在屏幕四处移动光标就能使只显示 40 列字符的屏幕将 80 列字符的页面显示出来，我觉得非常了不起，但大多数时候它仅仅有助于做家庭作业和其他一些类似的枯燥烦人的文档，然后用我心爱的点阵打印机把它们打印出来。

在计算机上一边编程一边咒骂它，工作了几年后，我终于放弃了，买了一台电动打字机。它让我能够做一些有趣的事情，如不需要输入奇异的换码字符就可以给文本加上下划线。尽管没有输入黑体文本的好方法，但我不用再担心因为一个换码字符的打字错误而浪费大量的纸张。这个打字机以传统的白纸黑字的方式给了我“所见即所得”的感性认识。

我坚持使用这台打字机好几年，直到发现了 Macintosh 计算机。当 Mac 计算机第一次出现时，我对它并没有什么好感，因为当时到手上的每一本杂志封面上都有这样一台我从未拥有过的昂贵的计算机，而它甚至连一个像样的编程软件包都没有。但四年后当我再次遇到 Mac 计算机时，我被惊呆了，用它写论文真是有趣极了，因为我可以锁定所有的格式信息，可以将页面分成多栏，甚至可以偶尔使用 72 点字型，尽管在我的 ImageWriter 打印机上输出的页面看上去不是很好，但比我原来的点阵式计算机文本强多了。在上交的论文中，我使用了大字标题，适当的斜体参考书目，将页面分成了几栏，甚至还使用了一两张图片。写作再也不是简单的组织语句，而完全可以创建子标题、表格和脚注，还可以使用各种格式，使得一篇即使很短的论文也可以包含多种结构，从而看上去非常漂亮。通过使用样式，可以一次应用一组格式化工具，给这组格式化工具命名后，以后可以重复调用，这看上去跟魔术一样。

十年后我仍旧使用大字标题和子标题为我的文档安排格式，幸运的是，我不再关心脚注了。但这时出现了一个新问题：很难重复使用过去的文档。当我在撰写某学年的论文时这无关紧要——写完上交以后就再也不用想它了，但我现在整天跟若干年前由远隔几千里外的人所写的大量信息打交道时，我至少需要关心如何将这文件转化成相同的文字处理格式。我发现自己经常花费数小时所作的并不是编辑这些材料，而是将它们进行重新格式化，这当然不是因为我喜欢这样做。滥用制表符和空格（使用打字机养成的习惯）成长起来的一代人创建的文档无法剪切然后粘贴到其他文档中，因为所有东西都断开了。各行的断点、换行符彻底错误，文本全部堆积到左边或右边，表格已经不成形状，甚至像行间隔之类简单的东西也

出现了问题，以前为创建所见即所得外观的文档带来过极大方便的那些神奇的格式化手段，现在却带来了这么多乱七八糟的问题。

另外还有一些小的问题。那些年我认为自己是在创建大字标题和子标题，而实际上却不是这样。我所创建的不过是格式上像标题的文本，我们或者可以将这些样式称为“大字标题”，但对于计算机来说除了字体大小之外“大字标题”什么也不是。“所见即所得”同样也改变了人们，那些可能连五年级美术班都没有毕业的人也可以在页面上使用三十种以上的字体。当这些字体渐渐失去新意以后，他们中的许多人开始使用较为保守的方法来排布格式，其意图仍旧是要使他们的文档看上去符合他们所想要的样子。设计人员已经习惯于指定位置时精确到千分之一英寸——仿佛人人都可以分辨出这么小的尺寸。

在“所见即所得”出现之前，文档无疑是够美观的，但它们也有其他一些优点，然而这些优点往往不被人们所注意。为了让计算机能够有效地管理大规模文档库，也有人曾经酝酿过一些创建文档管理和文档标记系统的行动。纯文本或许比较单调呆板，但与常见的文字处理程序或桌面出版程序的输出文件相比，纯文本要容易管理得多，没有格式需要移除，没有不可翻译的字符，从一个应用程序转移到另一个应用程序时也不会出现那些稀奇古怪的东西。这些文档管理工具在“所见即所得”的早期还是很新颖的，直到后来用户逐渐习惯了使用基于纸张也就是“所见即所得”的媒介之后，这些工具才慢慢变得不再那么昂贵而且逐渐适用起来。目前市场上大多数程序的最终目标仍旧是文档的打印输出结果。

1.1.2 HTML 的快速成长

当国际互联网在 1994 年第一次受到广泛关注的时候，一些业余爱好者和专业的设计人员开始着手创建他们能够创建的最激动人心的页面，但多数人很快就打了退堂鼓，因为习惯了完全“所见即所得”环境之后，他们实在由于 HTML 缺乏格式化的工具而倍感失望。更要命的是，各种浏览器之间莫名其妙的差异使得开发者很难预料一个页面在浏览器中会是什么样子，而公司用户需要能够像控制纸张文档一样控制他们的电子文档。有一段时间，这些种种限制几乎扼杀了 HTML 的发展。

就像 HTML 以前的许多网络技术一样，HTML 是因为它本身非常有趣而被那些出于兴趣爱好的热心者所使用的。HTML 非常简单，只需要一两天就能学会，但它却为人们提供了一种全新的阅读感受和写作体验。这些早期爱好者们所产生的冲击力和新闻媒体的相关报道，给 HTML 后来成为万众瞩目的事物提供了巨大的潜能。

要让 Web 发展成为一块经济上很有生命力的市场，HTML 还必须改变自身来满足用户的要求。设计人员和他们的雇主希望能够创建出看上去完全符合他们要求的网页，而且希望能够达到像一般桌面出版系统所提供的那种控制水平。经常爆发的浏览器大战（这时受害的往往是网页开发人员）使 HTML 变得越来越强大，尽管它们还不能相互兼容。国际互联网联盟（World Wide Web Consortium，一个管理互联网上标准的机构）已经或多或少地兼顾了各方，出台了 HTML 4.0。尽管准确地说，HTML 对普通用户而言还仍然不是一个简单的页面排版