轻轻松松学电脑

XML 实用培训教程

北京希望电子出版社　总策划
张健飞　编

科学出版社
www.sciencep.com

轻轻松松学电脑

XML 实用培训教程

北京希望电子出版社 总策划
张健飞 编

科学出版社
www.sciencep.com

内 容 简 介

本书用浅显易懂的语言以及丰富的实例向读者展示 XML 技术。

全书共分 10 章：第 1 至第 3 章介绍 XML 的语法以及 DTD 和 Schema 等基础知识；第 4 至第 6 章介绍 DOM 和 SAX 两种 XML 编程接口，其中第 5 章介绍 Xlink 和 Xpointer；第 7 至第 8 章介绍 XML 模式以及显示方式；第 9 章介绍数据库的技术；第 10 章对 XHTML 做了简单的介绍。本书面向有一定网页设计基础和编程经验的读者，并可以作为 XML 爱好者和高等院校相关专业师生的教学参考书。

读者在本书使用过程中遇到的技术问题，请与电子邮件地址：jcxt@sina.com 联系。

需要本书或需要得到技术支持的读者，请与北京中关村 083 信箱（邮编 100080）发行部联系。电话：010-62528991，62524940，62521921,62521724，82610344，82675588（总机） 传真：010-62520573，

E-mail：yanmc@bhp.com.cn

图书在版编目（CIP）数据

XML 实用培训教程 / 张健飞编. —北京：科学出版社，2003

轻轻松松学电脑

ISBN 7-03-012299-2

Ⅰ.X... Ⅱ.张... Ⅲ.可扩充语言，XML—程序设计—技术培训—教材 Ⅳ.TP312

中国版本图书馆 CIP 数据核字（2003）第 088784 号

责任编辑：安源 / 责任校对：杨如林

责任印刷：双青 / 封面设计：梁运丽

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

http://www.sciencep.com

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2003 年 12 月第 一 版 开本：787×1092 1/16

2003 年 12 月第 一次印刷 印张：18 1/10

印数：1—4 000 字数：419 400

定价：27.00 元

前　言

XML 的出现为 Web 技术带来了一次新的革命。在 2001 年 3 月 19 日，W3C 将规范的 XML1.0 提升至正式标准的地位。规范的主要作用在于和 XML Signature 一起确保 XML 文档所表示的信息在不同的 XML 处理器看来都是无差别的，这种无差别性对于电子商务的实现而言无疑是至关重要。随着 XML 被广泛地接受，XML 最初的宗旨被扩大化。XML 不仅被用作用自然语言表达的文档，而且还被用于描述各种各样的信息。数据库的所有者和厂商想要能够将任意的关系型和面向对象型的数据库以 XML 的形式发布到 Web 中，应用开发者想要将 XML 用于所有的信息交换。以上两种人都期望有更好的方法来定义应用中需要的特定 XML 语言，基于以上的原因，对以 XML 编码的信息集合进行查询是非常有用的。XML 要在电子商务或者相关的应用中使用，数字签名是必不可少的，要使得签名更为通行，需要有明显的方法可以将 XML 文档转换为规范一致的形式。综上所述，任何 XML 的新应用都会带来标准化的需求。

对于许多网页设计人员来说，也许已经对 HTML，ASP，PERL，JSP 等相当的熟悉，但 W3C 的 XML 标准为所有的网页设计者打开了一片新的视野。本书从基本入手，为刚刚接触 XML 的设计人员了解 XML 的结构、显示、框架合理性等做了详尽的介绍。

本书内容和介绍上有如下特色：

1. XML 本身是实践性非常强的技术，因此本书将注重书籍内容的实践性，将用更多的代码说话，而不是理论说话。
2. XML 的生命在于应用，而应用的生命在于为应用而产生的方法和观念上的革命性变化。因此本书同时也会更多地和实际应用、实际的产品以及实际的场景结合，重点讨论 XML 技术对于行业特别是一些网络管理系统所带来的好处和如何实现好处等问题。
3. 由于新技术层出不穷，本书也将对新技术新理论给予足够的关注，提出一些理论性很强的观点，给出一些有非常的前瞻性和敏锐的洞察力的建议和意见。

本书由芮昀负责全书的设计、统稿和修改，并编写第 5 和 8 章；翁宇翔负责编写第 3、4、6、9 和 10 章，并负责实例的测试；张健飞负责编写第 1、2 和 7 章。此外，张东、李晓、范智育、王宏生、李光龙、王瑾、吴浩、李炎、刘伟、刘华刚、朱峰、赵晓燕、李晓、马苍、郝春容、韦勇、成美华、萧峰、李菊、张浩然、李欣、张浩和李想等同志在整理材料方面给予了作者很大的帮助。

由于时间仓促，加之编者的水平有限，缺点和错误在所难免，恳请专家和广大读者不吝赐教，批评指正。

编　者

目　录

第 1 章　XML 基础

本章知识点

- XML 产生以及发展
- XML 相关技术
- XML 的应用
- XML 与电子商务
- XML 编写

本章导读

XML 全称是 Extensible Markup Language（可扩展标识语言），它不像 HTML 那样有固定的形式，所以使得 SGML 标准能在互联网上应用自如。XML 并不是一个独立的、预定义的标识语言，它属于一种元语言，即用来描述其他语言的语言。XML 允许用户自己设计自己的标识，必然与其他许多技术相关，所以不能独立的对待。由于 XML 具有以上的优点，所以目前它在业界中得到广泛的应用。

以上就是本章介绍的主要内容，除此之外，编者还将介绍使用 XML 编辑器来编写代码的规则。

1.1　XML 的历史

1.1.1　XML 产生

XML 是可扩展标记语言的简称，它结构化定义数据标准的格式。它的先驱是 SGML 和 HTML，它们都是非常成功的置标语言，但在某些方面存在着天生的缺陷，XML 正是为了解决它们的不足而诞生的。

XML 的产生来源于最直接而且是最有效的思想：即把强大的 SGML 运用于具有活力和应用前景的 Web，同时把 HTML 加以规范，把数据内容从表现中分离出来。XML 被描述为 SGML 非常简单的子集，它能够使通用的 SGML 像 HTML 一样，在 Web 上进行服务、获取和处理等操作。

1. SGML

60 年代 IBM 公司的技术人员就在研究一种描述文档及其格式的通用标记语言（GML）。1986 年国际标准组织（ISO）最终采纳该语言为标准通用标记语言（SGML）。SGML 提供一套复杂的系统来对文档进行标记化操作，使得文档外观独立于特定的应用软件。SGML 语言很庞大、功能很强、选项很多，适用于需要有严格文档标准的操作。同时 SGML 也可用于创建成千上万的置标语言，为其语法提供异常强大的工具，而且还有极好的扩展性。因此在分类和索引数据中非常有用，80 年代 SGML 较多用于科技文献和政府

办公文件中。

但是 SGML 强大功能的背后是它的复杂性，所以不适合在 Web 中快速简便地发布。不仅如此 SGML 还非常昂贵，目前比较便宜的 SGML 软件之一是 Adobe FrameMaker，其标准版本价格为 850 美元，而 Adobe FrameMaker+SGML 是以 1995 美元售出的。还有一点最为关键，几个主要的浏览器厂商都明确拒绝支持 SGML，这无疑是 SGML 在网上传播遇到的最大障碍。

2. HTML

HTML（超文本标记语言）是一种特定的 SGML 文档类型。由于它的学习和实现十分容易，而且免费提供源代码，所以很早就得到 Web 浏览器厂商的支持。HTML 最初于 1990 年由 CERN 设计，它是非常简单的 SGML 语言。后来 W3C 承担了 HTML 的开发和标准化的责任，从 1993 年的 HTML1.0 标准到现在的 HTML4.0 标准，经历着不断的完善和发展。

3. HTML 的局限

HTML 过于简单的缺点很快体现出来。在 Web 发展的早期，所用到的文档都基于文本格式，只有标题、项目列表和指向其他文档的超链接等。随着应用的发展，有了对多媒体和页面设计方面的需求。这就开始了 HTML 痛苦的成长历程，首先直接的内联图形是不错的，可是用户无法很好地放置图形，因此页面设计就遇到了问题。然后是图形的映射（其中有超链接的图形）也产生了新的问题有待解决。后来出现闪烁的文本、表格、帧和 DHTML。

虽然人们费尽心思给 HTML 增加新的东西，但每次都带来新的不兼容性和对新标准的需求。为什么会这样？原因很简单，HTML 不具有可扩展性。多年来，微软添加了许多只有适用于 IE 的标记，而网景公司添加只用于 Navigator 的标记，然而作为网页创作者却无法添加自己的标记，开发人员呼唤接受和运用已经引入的新标记和元素，

真正设计网站还需要更多的东西，诸如 CGI，JavaScript，ASP，PHP 和 Java 等，它们使得 HTML 更加强大。层叠样式表（CSS）和 DHTML 等技术的发展为更完备的定制 Web 设计提供必要的能力，可是这些新增的简单技术反而使正在严重的问题更为突出，要实现网页设计的完全定制化和自动化还只是美妙的幻想。

尽管 HTML 不提供任何扩展性，但父系统 SGML 缺乏完全可扩展。所以在 SGML 中产生一套完全定制的文档集合，并开发控制集合中所有文档的 DTD，虽然非常耗时而且极其复杂，但还是起到了作用。

4. XML

1996 年人们开始致力于描述一个新的标记语言，它既具有 SGML 的强大功能和可扩展性，同时又具有 HTML 的简单性。W3C 决定专门成立一个专家小组来从事此项工作，并由 Sun 公司的 Jon Bosak 担任小组的指挥。

SGML 中所有非核心的、未被使用的和含义模糊的部分将被删除，剩下的就成为短小精干的标记工具 XML。它保留 SGML80%的功能，而其复杂程度降低到原来的 20%。1997 年春，XML 草案已被拟定，W3C 于 1998 年 2 月 10 日批准 XML1.0 版本，一个崭新而大

有前途的语言诞生了。

1998 年在 XML 协调小组等部门的指导下，XML 的设计工作分成 5 个新的工作组：XML Schema 工作组、XML Fragment 工作组、XML Linking 工作组（XLink and Xpointer）、XML 信息集（Information Set）工作组和 XML 语法工作组。这些工作组和 W3C 的 XSL 和 DOM 工作组一起为 XML 技术的发展竭尽所能。

5. XML 产生的必然

回顾计算机科学特别是网络技术的发展，会发现 XML 的诞生正处于历史螺旋式上升的时期。60 年代时计算的中心最初是数据——不同系统、不同格式的非标准数据。70 年代开始用逻辑化的方式描述处理过程。80 年代重点在于面向对象的编程——OO，以对象为单位封装的数据和处理。90 年代出现不同的组件模型体系，如 COM，CORBA 和 EJB。21 世纪初网络的盛行，特别是电子商务的深入发展，网络数据交互和业务集成的需求又使数据重新回到人们视线的焦点上，但这一次的角色已经是用 XML 统一改造的数据。

在 Internet 上实现企业商务流程的集成是全球电子商务的基础，这种集成需要一种被全世界所接受的通讯和交流语言，并且语言要足够强大才能表达商务世界的各种需求，XML 就是拥有这种潜能的语言。

基于文档的信息通信和处理还刚刚起步，但其优越已经初现轮廓，其中包括：可靠的异步处理、通过符合客户需求的确定步骤来处理、通过精确的分离和接口实现商业流程的松耦合。在新一代电子商务中，XML 将扮演重要的角色，虽然 XML 是标准的通用和易于扩展的语言，但还需要建立更多的模块，其中包括。

（1）通信的技术基础。使得可以进行 XML 文档的交换，内容和其他所有典型中间服务的映射，如安全、目录等。进一步重要的处理方面是伸缩性、稳定性和优异的性能。

（2）对 XML 文档灵活控制和处理的组件。这些组件的一部分规定了执行的和处理顺序的通用规则协议，这是标准具有优越性的又一个体现，它们允许用用各自的信息在外部扩展 XML 文档。

（3）基本商务数据文档的定义。

1.1.2　XML 的发展前景

XML 的发展是属于未来的，现在所做的努力只是为将来它能更好的为我们服务。作为计算机的学习者，除非对自己现在所学到的知识感到满足，否则就应该学习 XML，因为这不是一件难事。

网络的未来属于商业，而那时的商业操作已经不在是简单的网页广告，或有限信息的反馈和狭窄的交易空间，取而代之的是多元化的服务，友好的界面，丰富的数据共享。XML 需要强大的新工具用来在文档中显示丰富复杂的数据，今后用户可以在分层的动态变化的数据上，映射用户友好的显示层来实现这一目的。

现在顾客服务会从电话购物或亲自到商场转移到 Web 站点上来，而且将会由于 XML 的强大功能受益更多。由于大多数商业应用软件包括数据处理和转移，如订货单、发货清单、顾客信息等，XML 将改革终端用户在 Internet 上的行为，许多商业应用将能实现。另外通过使用基于 XML 的面向企业内部互联网的词汇库，Web 站点上的信息，无论是存储

在文档中还是在数据库中都可以被标识，这些词汇也能够对那些需要在顾客和供应商之间交换的中小型机企业提供帮助。

XML 作为表示结构化数据的一个工业标准，为组织、软件开发者、Web 站点和终端使用者提供许多有利的条件。这些条件使更多的纵向市场数据格式建立起来，并被应用于关键市场，诸如高级的数据库搜索、网上银行、医疗、法律事业、电子商务和其他领域，这使得机会更进一步的扩大。当站点更多的进行分发数据，而不仅仅是提供数据浏览时，特别的机会就产生了。另外一个有待开发的市场是建立 Web 站点的工具，包括用来从数据库信息和使用者界面中产生 XML 数据的工具，其中开发用来描述从数据库中产生 XML 的可视化工具是个很好的想法。

1.2 XML 相关技术

XML 并不是在真空中操作的，如果将 XML 用于多种数据格式的操作，就需要与相关的技术进行结合，其中包括 HTML，CSS，XSL，URL 与 URI，XLL 和 Unicode 字符集等。

1.2.1 超文本标记语言 Hypertext Markup Language

Mozilla 5.0 和 Internet Explorer 5.0 是首先对 XML 提供支持的浏览器，若让大多数用户升级到这 2 种浏览器的版本上可能还要花两年的时间，所以在今后还需要将 XML 内容转化为经典的 HTML 格式。

使用者应该了解如何将一个页面与另一个页面进行链接，了解如何在文档中包括图像，如何使文本变成粗体等操作。由于 HTML 是 XML 中最普通的输出格式，所以对 HTML 了解得越多，也就越容易了解如何创建所需的效果。另一方面，如果已经熟悉利用表格或用单像素的 GIF 来安排页面上的对象，或是开始借助于画出草图而不是借助于内容来创建 Web 站点，那么也就必须要忘记某些坏的习惯。正如前面所讨论的，XML 将文档的内容与文档的外观相分离，首先开发内容，然后再用样式单将格式附加其上，这既改善文档内容也改善文档外观。除此之外，还允许作者和设计者更加互相独立地工作。但对于设计 Web 站点来说，确实需要有不同的思路，如果涉及多人的话，或许要利用不同的项目管理技术。

1.2.2 级联样式单 Cascading Style Sheets

由于 XML 允许在文档中包括任意的标记，所以对于浏览器来说，没有办法事先知道如何显示每个元素。当文档送给用户时还要向用户发送样式单，通过样式单告诉浏览器如何格式化每个元素。

样式单可以使用 CSS 级联样式单。它原来是为 HTML 服务的，定义字号、字族、字重、段落缩进、段落对齐和其他样式等格式化的属性，这些属性可以施加到个别元素上。如 CSS 允许 HTML 文档来指定所有 H1 元素应该被格式化为 32 磅、中间对齐的 Helvetica 字体的粗体。单独的样式可以施加到大多数 HTML 标记上，它能够覆盖浏览器的缺省设置。多个样式单可施加到一个文档上，而多个样式也可用于单个元素上，因为样式是根据特定的一套规则级联起来的。

向 XML 施加 CSS 规则也很容易，只要改变施加规则于其上的标记名称即可。Mozilla 5.0 直接支持 CSS 样式单与 XML 的结合，虽然此浏览器时常发生崩溃。

1.2.3　可扩展的样式语言 Extensible Style Language

XSL 可扩展的样式语言是更先进的专门用于 XML 文档的样式单语言，它本身就是结构完整的 XML 文档。

XSL 文档包括一系列的适用于特定 XML 元素样式的规则，它读取 XML 文档并将其读入的内容与样式单中的模式相比较，当在 XML 文档中识别出 XSL 样式单中的模式时，对应的规则将输出某些文本的组合。与 CSS 不同，输出的文本比较任意，也不局限于输入文本加上格式化信息。CSS 只能改变特定元素的格式，以元素为基础，而 XSL 可以重新排列元素并对元素进行重排序。XSL 可以隐藏一些元素而显示另外一些元素。可以选择应用样式的标记，但选择不仅是这一种，因为也基于标记的内容和特性，或基于标记在文档中相对于其他元素的位置，甚至基于各种其他的准则。

CSS 的优越性在于具有广泛的浏览器支持，但 XSL 更为灵活和强大，而且 XSL 样式单的 XML 文档可以很容易地转换成 CSS 样式单的 HTML 文档。

1.2.4　URL 和 URI

互联网中 XML 文档被统一资源定位符 URL 所引用，如地址http://finance.sina.com.cn/nz/dhwto/index.shtml可以找到经济学家和企业家对话关于 WTO 的信息。虽然 URL 得到人们的广泛理解，但 XML 规范使用的是更为通用的统一资源标识符 URI，它定位资源是更为通用的架构，更为注重资源而不太注重位置。理论上说，URI 可找出镜像文档的最为近的副本，或是找出已经从一个站点移动到另一站点的文档，实际上 URI 仍然处于进一步的研究之中。

1.2.5　XLink 和 XPointer

只要将 XML 粘贴到 Internet 上，用户当然希望能够对此文档寻址并且可以将众多文档链接起来。标准 HTML 链接标记可用在 XML 文档中，而且 HTML 文档也可与 XML 文档加以链接，如下面的 HTML 代码，它将链接指向前文中提到的以 XML 形式出现的链接副本。

```
<a href="http://finance.sina.com.cn/nz/dhwto/index.shtml">
著名经济学家和企业家对话WTO
</a>
```

XML 利用 XLink 来与文档相链接，用 XPointer 来确定文档个别部分的位置。XLink 使任意元素成为链接，而不只是 A 元素。进一步说，链接可以是双向的、多向的或是指向多个镜像的站点，当然要选择站点中最近的一个，XLink 利用普通的 URL 来标识它链接的站点。

XPointer 能使链接不仅指向特定的位置，而且还可指向特定文档的特定部分。它可以引用文档中特定的元素，如像第一个、第二个或是第十七个等特定的元素。XPointer 提供非常强大的文档间连接功能，而文档不必有包括附加标记的目的文档，所以其中的个别部分才可以被链接。

进一步说，与 HTML 的 anchor（锚）不同，XPointer 不只是引用文档中的一点，同时

可以指向一个范围或是一个区域，因而 XPointer 可以用来选择文档的特定部分。

1.2.6 Unicode 字符集

Web 到目前为止其上主要文本部分仍是英文，XML 是改变这种状况的开始，它对双字节的 Unicode 字符集及其紧凑的表示提供完全的支持，这一字符集几乎可以支持地球上的每一种常用的字符。遗憾的是，光有 XML 还是不够的，为了阅读一种文字，还需要 3 个条件：

（1）该种文字的字符集；

（2）该字符集的字体；

（3）操作系统和应用软件能够理解这种字符集。

如果想要以这种文字写作，还需要文字的输入法。当然 XML 定义了字符引用，可使用户利用纯 ASCII 字符将未列在本地字符集中的字符加以编码。这对于偶尔引用希腊或是中文字符也足够了，当然不能指望用这种办法以其他语言来写一部小说。

1.3 XML 的应用

WWW 无疑是最近两年 Internet 上最具生命力的一种应用，由于它操作简单且功能强大，不仅能够传输文本数据，而且可以进行声音、图像、多媒体等数据的传输，因此越来越多受到用户的喜爱。随着 Web 文件的复杂，内容提供商已经开始感受到普通 HTML 已经无法提供的用于大规模商业出版所需要的扩展性、结构和数据检查功能。由于 Java 语言的发展，越来越多的客户端应用要用到 Java applet，由于 Java applet 能够往 Web 客户端嵌入强大的数据控制能力，这使得当前 HTML 在传输文件数据方面的不足更加明显。

为了满足商业 Web 出版和解决 Web 技术在新的分布式文件处理领域应用的需求，W3C 开发一种可扩展的标记语言，这就是 XML，用于那些目前 HTML 无法满足要求的应用。本文介绍 XML 技术的发展，并且讨论由 XML 产生的新的基于 Java 的 Web 应用。

1.3.1 HTML 和 SGML

1. HTML 和 SGML

Web 上的绝大部分文件是以 HTML 形式存储和传输的，它是一种最简单的 Web 页面标记语言，非常适合于标记超文本、多媒体和显示较小较简单的文件。HTML 是在 SGML 基础上发展来的，SGML 是一个用于定义和使用 Web 文件格式的国际标准，即 ISO 8879 标准。

SGML 允许一个文件来描述它们自己的语法，即允许文件自己确定用在文件中的标记集合和这些标记所代表的结构上的联系。标准的 HTML 规范是 SGML 规范的一个严格定义的子集合，它规定固定数据的标签集合，不允许用户定义自己的扩展标签，这样用户在开发 Web 页面文件时不必考虑语言规范，因此可以节省开发时间和精力。但是同时也导致标准 HTML 语言在几个重要方面如可扩展性、结构和有效性等的严重不足，表现如下。

（1）可扩展性 HTML 并不允许用户根据在 Web 上表达一些特殊数据的需要，去定义专用的标签或属性。

（2）结构 HTML 并不支持表达数据库结构或面向对象的分级结构，所需要的深层结构的规范。

（3）有效性 HTML 并不允许利用应用来检查数据的结构上的有效性。

虽然 HTML 是在 SGML 基础上发展而来的，但它在上述几个方面的做法却与 SGML 相反，一个标准的 SGML 应用应该可以支持任意复杂的 SGML 语言规范，并且具有标准 HTML 中所没有的可扩展性、结构和有效性检查功能。SGML 的出现使人们有可能定义自己专用的文件格式来处理庞大而又复杂的 Web 文件，并管理具有大量信息的数据库。然而全部的 SGML 规范包含许多一般的 Web 应用，并不需要可选的特性，实际上正是这些可选的特性使 SGML 过于复杂而无法得到普及。

2. 与 XML 的关系

XML 是一个专门为 Web 应用设计，且是简化的 SGML 子集。XML 保留 SGML 可扩展性、结构和有效性等方面的优点，保留 SGML 中绝大部分的实用功能又使得用户更容易学习、和使用。XML 与 HTML 的不同主要体现在 3 个方面。

（1）信息提供商能够根据自己的需要随意定义新的标签和属性。

（2）文件结构能够具有任意深度的结构层次。

（3）任意一个 XML 文件都能够包含一个可选的描述自身的语法，以供需要进行结构的有效性检查的使用。

XML 在设计之初就要求具有最强大的表达功能、最大限度的适合教学、最大限度的易于实现，因此它一经产生就得到了用户的普遍欢迎。XML 语言并不后向兼容现有的 HTML 文件，但是遵守 W3C HTML 3.2 规范的文件能够很容易转换成符合 XML 格式的文件，这样用户就不必担心原有的 HTML 文件无法在 XML 环境中使用，最大限度的保持了用户在 HTML 方面的原有投资，而且许多厂商专门推出一些专门的 XML 转换工具。

1.3.2　基于 XML 的 Web 应用

基于 XML 的 Web 应用可以被分为 4 类：

（1）需要 Web 客户端在两个或更多异质数据库之间进行通信的应用；

（2）试图将大部分处理负载，从 Web 服务器转到 Web 客户端的应用；

（3）需要 Web 客户端将同样的数据以不同的浏览形式提供给不同的用户的应用；

（4）需要智能 Web 代理，根据个人用户的需要裁减信息内容的应用。

若没有 XML 想要满足上述应用，就需要使用专门编写的 script 代码嵌入到 HTML 文件中，并同专门的浏览器插件或 Java applet 一起提供给用户。XML 在设计时是基于这样一个理念的：数据应该属于它的创建者，内容提供商应该具有最好的数据结构以便使他们不被束缚到某种特定的 script 语言中，只有这样才能为不同的写作和分发工具实现自由竞争提供一个标准的、厂商中立的页面标记语言，下面就分别来介绍这 4 种应用。

1. 数据库交换

典型第一类 XML 应用就是数据库交换技术，应用实例就是美国家庭健康医疗机构的信息跟踪系统，它收集全国病人各种医疗的记录，以满足联邦医疗机构和健康维护组织的

需要。

一般需要将病人的资料提交给家庭健康医疗机构的信息系统，这些是基于纸文件的信息材料，包括病人的病史、个人医生、医院、药店和保险机构的账单等。然后将病人的资料输入到数据库，这也是工作量最大的且需要手工操作的环节。

Web 技术的产生，使那些盼望通过电子手段减少工作人员输入负担的医疗信息机构看到了希望。不幸的是，现有的 Web 应用仍然还不能满足这种应用的需要。医院最开始为信息机构提供类似的解决方案：登录到医院的 Web 站点，成为一个授权用户，使用 Web 浏览器访问病人的医疗记录并从浏览器中打印出记录。

读者可能觉得方案太过于原始，但它的确是较早采用医疗信息系统的最初做法。方案更完善的做法是：操作员从 Web 浏览器上读取病人的数据，并且将数据直接输入到另一个浏览器窗口中的医疗信息机构的在线表格中，而不必先将数据打印出来。

两种方案唯一的不同点在于第二种方案节省了打印输出的纸张，但它并不能从根本上解决问题。较为实际一些的方案如下：登录到医院的 Web 站点上，成为一个授权用户，通过 Web 方式访问病人的医疗记录，医疗信息系统为每一个病人的记录放在一个文件夹中。用户将存放记录的文件夹从 Web 应用中拖到内部的数据库应用中，即将其放到数据库中。然而这个过程用现在标准的 HTML 语言是无法完成，原因主要有 3 个方面。

（1）HTML 标签集合太有限了，因而无法表现或区别混合在病人医疗数据文件中的大量数据库的字段。HTML 不能够表现这些复杂的文件所具有的各种各样的结构。

（2）HTML 缺乏任何能够在接受试图将自己输入到目标数据库中的应用之前对其进行数据结构有效性检查的机制。

（3）HTML 将文档内容和文档结构合在一起限制了与应用程序格式（如 EDI 或者电子数据交换）的内容传递。

一种无缝实现医疗记录的共享交换技术的可行的办法，要求所有的医院和健康医疗机构使用一种由政府制定统一的标准格式系统，然而这要求几乎是不实际的。

允许异质系统之间交换数据的另一种办法是采用一种业界统一的交换格式，它为所有的输出系统制定统一的输出格式，并且为所有的输入系统制定统一的输入格式。这实际上就是 SGML 最初的设计目标，XML 也继承了这个理念。许多行业都已经在使用统一格式的语言来进行数据交换。一般 W3C 制定的主要标准文件格式就是文件类型定义（Document Type Definition，DTD），它确定标记语言的标签集和语法。然后 DTD 可以用任何的编辑工具以业界标准的语言进行标记的文件一起发送，接收端的任何标准的应用都可以对其进行接收和处理。

XML 的目标是提供一个系统独立、厂商独立的解决方案，它的出现将可以解决上述医疗信息系统的问题，其中最关键的就是 XML 能够针对特定的应用来定义自己的标签。有趣的是 XML 1.0 规范发布的同一天，SGML 也声明 HL7（一个由医疗 IS 厂商组成的标准组织）将开发一个用来解决上面例子中的问题的医疗标记语言。

其他的基于 XML 的第一类应用的例子还有：司法出版，政府药品批准过程，联合 CAD/CAM 努力，跨系统的联合日历管理等。

2. 利用 Java 的分布式处理

XML 应用的典型代表是半导体工业设计的数据传输系统，每个半导体制造商都维护着包括产生所有 IC 庞大的技术信息。为了交换这些数据，几年前由 Intel，National Semiconductor，Philips，Texas Instrument 和 Hitachi 等公司成立一个工业论坛，来设计一个专门用于半导体业的 SGML 规范，论坛 1995 年完成，目前都在使用这个规范来进行数据的传输和交换。

可能会有人认为随着 HTML 变得越来越普及，Pinnacles 的成员会重新考虑它们的作法，但实际上 HTML 的限制已经使他们确信它们最初的思路是对的。Pinnacles Group 最早的想法使用专门的 SGML 标记语言产生的数据不仅可以将半导体数据单显示成可读的文件，而且可以促进半导体芯片的设计效率。现在人们都认识到这个方法的好处，它具有分布式 Java applet 概念，将来的版本将允许工程师访问制造商的 Web 站点，不仅下载关于某种特定的集成电路的可读数据，而且下载一个允许他们调整这些电路各种组合的 Java applet。

这样的半导体应用很好的体现 XML 的优点。

（1）它需要专门的标记。由于半导体工具中有许多专门复杂的数据，为了表达它们就必须要有专门的标记标签，而这些专门的标记是目前固定 HTML 标签集合所无法提供的。

（2）它需要数据表达是平台和厂商中立的，这样来自各种数据源的数据才能在分布式应用中使用。Interent 本身就是一个异质的网络，组成网络的操作系统和所使的数据资源都可能各不相同，只有成为平台独立和厂商独立的应用，XML 才能用于各种异质数据库之间的通信。

（3）XML 最终结果就是使得原有的必须在服务器上完成的大量计算过程，变成是由用户自己的 Web 客户端同服务器的简单交互操作。

虽然有效性同时也很重要，但它并不总是在这类应用中起到关键的作用，这与在数据进入数据库前必须检查其结构完整性的应用是不同的。为了使处理过程尽可能富有效率，XML 在设计时就允许有效性在它并不需要的地方是可选的而不是必须的。

上面讲的半导体应用不仅代表基于 Java 的应用，实际上人们都希望能够在客户端上对任何使用标准化数据的应用进行控制，能够代表这类应用的最典型的例子如下所示。

（1）设计应用。当设计者无法用其他办法通过服务器，集中处理考虑的各种任务如：电子、工程、建筑、菜单计划等。

（2）日程表应用。当顾客用其他方法无法通过服务器来享受各种服务如：飞机、火车、公共汽车、地铁、就餐、电影、戏剧和音乐会等，这些应用正是分布式基于 Java 的应用可以大展身手的地方。

（3）商业应用。这种应用允许消费者通过提供不同的购物原则进行选择如：不动产、汽车、器械等。

（4）各种教育性应用。如那些经常被称之为 online help（在线帮助）的应用。

（5）各种消费者支持应用。可以从最简单的割草机维护，一直到最复杂的计算机的技术支持等各种应用。

其中最后一类应用的代表是 Solution Exchange Standard，它是由 60 多个硬件、软件和通信公司组成的联合会公布的一种 SGML 标记语言，其目的是促进技术支持信息在厂商、系统集成商和企业帮助台之间的传输和交换。在 Solution Exchange Standard 的声明中有这样的话："标准应该设计成灵活的，它独立于任何平台、厂商和应用，因此它能够用于交换信息而不管它们来自或到达什么样的系统。除此之外，标准应该设计成具有很长的生命期。SGML 提供很大的可扩展性，因此标准应该能够很容易容纳变化迅速的支持环境。"

随着 Web 技术的发展，Java 将会和 XML 技术相得益彰，这样的应用也会变得越来越重要。

3. 用户选择浏览方式

第三类 XML 应用就是同样的数据，可以以不同的浏览方式出现在浏览器中，而数据并不需要再次从 Web 服务器上下载。

这类应用的早期的例子是 Web 上的动态表格。使用建立在面向对象数据库基础上的 Web 服务器，现在有可能将一个内容表格做成一个巨大的数据库，可以通过单击鼠标来打开这些数据库的"内容提要"，此时就可以看到整个数据库内容结构的更详细的信息。一种具体的方案就是下载整个结构化的"内容提要"到客户端，而不是仅仅由服务器产生的"内容提要"的浏览模式。这样用户就能够在客户端根据自己的需要和爱好更快的对"内容提要"进行扩展、缩小、移动等操作，不仅加快操作的速率，而且通过减少与服务器的交互操作减轻了 Internet 网络的负载。

Sun 公司的一个工作小组将这种方案作为一个基于 Java 的 HTML 帮助浏览器的一部分，但是 HTML 的限制对工作组人员提出更高的要求。在应用中，"内容提要"以手工方式使用非标准的扩展标签构造（由于标准 HTML 中缺乏结构化，因此不可能从文件中直接产生"内容提要"），然后"内容提要"条目由一个 HTML 页面中的评论进行包装，以在 Web 浏览器中隐藏非标准的标记。随 HTML 文件一起下载的 Java applet，解释隐藏的标记并且提供基于客户端的"内容提要"的浏览模式。

在实践中，用 HTML 设计的这种方案虽然工作得很好，但它对设计人员提高了要求。在一个 XML 环境中，"内容提要"的手工创建和隐藏非标准的标记扩展都不是必需的。相反，标准的 XML 编辑器可以将"内容提要"下载到可以自动的创建和显示"内容提要"的浏览器中，这个浏览器使用一个下载的 Java applet 或一个标准的 JavaHelp 类库，来完成结构化内容的创建和显示工作。

XML 允许创建和定义新的标记和结构数据的能力大大扩展了数据在客户端的显示控制方式，这表现在如下方面。

（1）一个提供有 Solaris 操作系统的 Sparc 和 X86 两种版本的技术手册，根据用户的爱好在客户端可以显示成专门针对 Sparc 制作的，也可以显示成专门为 X86 制作的，只需用鼠标单击一下"爱好"开关即可。

（2）一个以多种语言提供的安装说明书，可以根据用户的需求来选择一种语言显示在用户的浏览器上。

（3）一个包含许多注释的文件，可以选择在浏览器上只显示文本、也可转换到只显示注释的模式或这两者都显示的，所有这一切只需要通过一个菜单选择就可以完成。