北京大学信息技术系列教材

BEIJINGDAXUEXINXIJISHUXILIEJIAOCAI

本书附光盘

从HTML到XML

主编　蔡翠平

编著　周宏滔　任吉治

北方交通大学出版社

http://www.press.njtu.edu.cn

清华大学出版社

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

北京大学信息技术系列教材

从HTML到XML

主编　蔡翠平
编著　周宏滔　任吉治

北方交通大学出版社
Northern Jiaotong University Press
清华大学出版社
Tsinghua University Press
北京 · BEIJING

内 容 简 介

本书由浅入深，以循序渐进的方式介绍了 HTML 和 XML。书中通过大量的实例，阐述了 HTML 和 XML 在网页设计中的应用。本书的前半部分主要介绍如何用 HTML 来设计网页，HTML 各种常用标记的基本介绍。而本书的后半部分则侧重于 XML，介绍了 XML 的基本概念，如何用 XML 来设计网页，CSS 和 XSL 的介绍等。

对于熟悉或不熟悉 HTML 的读者或渴望学习 XML 的读者来说，都可通过本书的学习，掌握 HTML 和 XML。

本书语言简炼，实例丰富，配套光盘中提供了书中实例的源代码，以方便读者的学习。可作为大中专院校 HTML、XML 语言程序设计的课程教材，也可作为计算机应用技术人员的参考书。

图书在版编目（CIP）数据

从 HTML 到 XML/周宏滔，任吉治编著 .—北京：北方交通大学出版社，2002.7
北京大学信息技术系列教材
ISBN 7-81082-061-3

Ⅰ.从 ... Ⅱ.①周 ...②任 ... Ⅲ.①超文本标记语言，HTML—程序设计—高等学校—教材②可扩充语言，XML—程序设计—高等学校—教材 Ⅳ.TP312

中国版本图书馆 CIP 数据核字（2002）第 031914 号

丛 书 名：北京大学信息技术系列教材
书　　名：从 HTML 到 XML
主　　编：蔡翠平
编　　著：周宏滔　任吉治
责任编辑：孙秀翠
排版制作：科事洁电脑打印中心
印 刷 者：北京东光印刷厂
装 订 者：三河桃园装订厂
出版发行：北方交通大学出版社　邮编：100044　电话：010-62237564　51686045
　　清华大学出版社　　邮编：100084
经　　销：各地新华书店
开　　本：787×1092　1/16　印张：15　字数：374 千字　附光盘：1 张
版　　次：2002 年 7 月第 1 版　2002 年 7 月第 1 次印刷
书　　号：ISBN 7-81082-061-3 / TP·21
印　　数：5000 册　　定价：26.00 元

北京大学信息技术系列教材

编委会成员名单

北京大学信息技术系列教材

序　言

人类已进入21世纪，科学技术突飞猛进，知识经济初见端倪，特别是信息技术和网络技术的迅速发展和广泛应用，对社会的政治、经济、军事、科技和文化等领域产生越来越深刻的影响，也正在改变着人们的工作、生活、学习和交流方式。信息的获取、处理、交流和应用能力，已经成为人们最重要的能力之一。培养一大批掌握和应用现代信息技术和网络技术的人才，在全球信息化的发展中占据主动地位，不仅是经济和社会发展的需要，也是计算机和信息技术教育者的历史责任。

加入WTO，意味着我国要在同一个网络平台上参与国际竞争，同世界接轨。这对我们既是一个机遇，也是一个挑战。为此我们必须加强全民的信息技术教育，以提高国民的整体素质，抓住国际大环境给我国经济腾飞带来的难得机遇，迎接挑战。

教育部提出，要在全国的中小学中逐步开设信息技术必修课，从小培养获取、分析、处理、发布和应用信息的能力和素养，在条件成熟时，考虑作为普通高校招生考试的科目。国家经贸委也提出，要像抓3年改革和脱困的两大目标那样，把企业管理信息化建设作为新世纪经贸工作的历史性任务抓紧、抓好，推进企业应用计算机管理软件和网络信息技术，用3年左右的时间，在国家重点企业中建立完善的企业管理信息系统。

为了适应这个大的形势，满足各大专院校非计算机专业学生和社会各阶层从事信息技术和急需掌握信息技术人们的需要，我们组织编写了这套《北京大学信息技术系列教材》。目的是让更多的人以最快的速度掌握计算机信息技术，学会运用国际互联网络平台，不断提高自身素质和专业水平，在传统产业改造、升级、实现跨越式发展中更好地展示自己的才能，为祖国的现代化建设服务。

本系列教材包括《计算机信息技术基础》、《计算机网络应用技术》、《办公自动化软件》、《多媒体应用技术》、《网络程序设计——ASP》、《数据库技术——SQL》、《Visual Basic程序设计》、《Visual FoxPro程序设计》、《C++语言程序设计》、《网页制作技术》、《从HTML到XML》、《计算机局域网实用技术》。随着信息技术的发展和读者的需要，我们还将不断对这一系列教材进行补充或增删，以期形成读者欢迎的动态系列教材。此系列教材可作为大专院校非计算机专业信息技术普及教材，也可供社会各种信息技术培训班选用。

本系列教材具有以下编写特点。

1. 适合不同层次的读者选用

此系列教材从内容上讲，跨度较大，从计算机基础知识一直到动态网站制作，这样可以满足不同领域和不同层次的读者需要，读者可以根据自己的水平像吃自助餐一样自主选用。

2. 选材超前，出版周期短

目前，计算机图书市场火爆，需求旺盛，但是，选一本合适的教材又非易事，其原因之一是读者急需使用的高版本软件对应的书上市甚少，造成这种现象的原因有三：一是信息技术发展速度太快；二是选材没有注意超前量；三是出版周期太长。鉴于以上原因，本系列教材在内容上尽量注意超前量，如每一个软件必须选择当前最高版本，例如：动态网站制作我们选择当前流行的ASP技术和SQL网上数据库以及VB编程技术；在出版上尽量缩短出版周期，此系列教材从策划到出版在8个月内完成。其目的都是为了适应信息技术的飞速发展，满足读者的需要。

3. 实用性强

本系列教材的主要对象是非计算机专业人员，因此，在内容上强调实用，尽量不涉及高深的与软件使用无关的理论问题。比如《多媒体应用技术》，作者着重阐述多媒体信息的获取、处理、传输、保存、制作等实用技术，不涉及多媒体的理论问题。又如《计算机局域网实用技术》，作者重点介绍局域网的构架、服务器的安装、各种网上信息服务的建立以及网络安全管理方面的内容，读者可按照书中所讲的内容自己独立构建局域网。

4. 充分体现案例教学

在本系列丛书中读者会发现，凡是操作型软件都是以一个案例为主线进行阐述，这是本系列书作者多年来在教学第一线经验的总结。案例教学引人入胜，易理解，易掌握，能使读者举一反三，技术掌握扎实。

5. 写作风格通俗易懂

介绍每一个软件开门见山，语言简明扼要，重点突出，难点翔实编写，同一功能决不重复。并每章附有习题，有的例题配有光盘，适合自学。

参加本系列教材编写的作者都是在大学从事信息技术课一线教学的中、青年教师，他们都有极强的敬业精神，本系列教材凝聚了他们多年丰富的教学经验和心血。

本系列教材得到了北京大学教育学院教育技术系各位老师和北京大学信息管理系余锦凤教授的支持和帮助，在此表示诚挚的感谢。

由于本系列教材从策划到出版仅仅用了不到一年的时间，编写者又都担负着繁重的教学任务，在时间紧、任务重的情况下，肯定有不少不尽人意之处，诚挚接受广大读者的批评、指正。

蔡翠平

2002年1月于北京大学

前　言

人类发展的脚步永远不会停止。WWW 的出现，使互联网从此深入人心，也给人类的生活、工作带来了极大的影响。而这里面，HTML 可谓立下了汗马功劳。HTML 使网页变得简单易学、亲切动人。读者甚至只要用记事本写几个字，然后存储成以 htm 为扩展名的格式，就制成了网页。但是，正是由于 HTML 的这种简单、通用性，也限制了其发展。这时 XML 便应运而生。

HTML 的制定是在 1989 年，而 XML 的第一版出现却是在 1998 年。而且，现在 XML 还在不断地发展之中，围绕着 XML，也出现了越来越多的新技术。很多读者可能对 HTML 比较熟悉了，但对 XML 却可能不太熟。其实，在大多数读者的计算机里都有 XML 的影子，那就是 Internet Explorer 4.0 里出现的推播频道。现在，电子商务、手机上网所用到的网页设计语言 WML 都是 XML 的应用。可以说，虽然现在很多网页都还是用 HTML 制作的，可是，XML 的革命已经悄悄地发生了。

在我写这本书之前，曾经感到很为难，HTML 已经发展得很成熟了，大多数读者都能看懂甚至书写 HTML 代码，而 XML 却仍旧在不断地发展之中，要将这两种东西放在一本书里介绍的确有些困难，不过，我始终认为“温故而知新”是不会错的，只有巩固了旧知识，找出旧知识与新知识之间的联系，才能更好地学习新的知识，在这本书里，我想给读者的正是这样的一种联系。希望读者能够由浅入深，通过作者将 HTML 和 XML 两者之间的对比，来更好地掌握 XML。

同时，我要感谢北京大学的蔡翠平老师和北方交通大学出版社的孙秀翠编辑对作者和本书所给予的极大帮助。

本书的配套光盘中包括本书所用的所有实例，读者可以将实例进行任意地修改，以促进读者的学习。同时，在本书的附录中，还包括 HTML 的简明参考手册，便于读者查询使用。由于本书对 CSS 部分只进行了简要的介绍，因此在附录中还包括 CSS 的简明参考手册。

本书仓促成章，很多方面难免有疏漏之处，希望各位读者不吝赐教。

周宏滔

2002 年 7 月于北京大学燕园

目　录

第1章 HTML 简 介

本章要点：

- ☑ 网络基础知识
- ☑ HTML的基础知识
- ☑ HTML语言的结构
- ☑ 编辑HTML文档的常用工具

1.1 网络基础知识

1. WWW概述

20世纪40年代以来人们就梦想能拥有一个世界性的信息库。在这个数据库中数据不仅能被全球的人们存取，而且应该能轻松地链接其他地方的信息，以便用户可以方便快捷地获得重要的信息。

随着科学技术的迅猛发展，人们的这个梦想已经变成了现实。目前正在使用的最流行的系统叫“World Wide Web”，中文可译为万维网，缩写为WWW，或W³，W3。它是在全世界范围内所有位于HTTP服务器上相互链接的超文本文档。简而言之，WWW是一个以Internet为基础的计算机网络，它允许用户在一台计算机上通过Internet存取另一台计算机上的信息。从技术角度上说，万维网是Internet上那些支持超文本传输协议HTTP(Hyper Text Transport Protocol)的客户机与服务器的集合，透过它可以存取世界各地的超媒体文件，内容包括文字、图形、声音、动画、资料库以及各式各样的软件。

Web首先是在瑞士的欧洲粒子物理实验室(CERN)研究中心开发出来的。其目的是想给CERN 的物理学家提供一种共享他们的工作和利用集团信息的一种工具。在万维网之前就有许多超文本系统的执行程序。而Berners-Lee在CERN与其他人合作，制定了基于Internet的体系结构，它是开放的、公开的标准规范并附免费的样本执行程序。因为是公开的，所以都可以建立Web客户机或Web服务器；因为是免费的，所以一些开发人员能够选择编制或定义系统的每一部分。也正因为这两个因素促使其他人员都可加入这一项目。WWW的发展使得全世界范围内的人都可协同工作。

理论上说来，万维网包含所有的Web站点、FTP档案库、Telnet公共存取账号、News新闻讨论区等等。所以万维网可以说是当今全世界最大的电子资料世界，已经可以把World Wide Web当成是Internet的同义词了。

位于WWW上的文档称为页面或Web页面，它用HTML语言编写，并使用URL进行标识，URL指明了特定的计算机和路径名，用户通过它可对文件进行访问，并在HTTP协议下将文件在节点间进行传输，直至传输到最终用户。

在WWW世界中，它的资源可以互相链接，全世界目前大概有数万个Web站点，每个Web

站点都可以通过超链接(Hyper Link)与其他Web站点链接，任何人都可以设计自己的主页(Homepage)，放到Web站点，然后在主页上面产生链接，与其他人的主页链接，或是连到其他的Web站点。别人也一样可以连到你的主页，或是你的Web站点，整个信息网就这样编织起来了，形成一个巨大的环球信息网。本节将简要介绍一下万维网的一些概念：浏览器、服务器、协议等。

2. 浏览器与服务器

先撇开互联网络的技术问题，来思考一下谁是网络中最重要的参与者。无疑，他们是“网页提供者”与“上网者”。网页提供者将其制作好的网页放在服务器上，而上网者则是准备好上网的机器，然后通过互联网来浏览网页制作者所提供的网页。因此，把放置网页的部分称为服务器(Server)，而上网浏览的部分则称为浏览器(Browser)。

浏览器是一个需要某些东西的程序，而服务器则是提供某些东西的程序。也就是说，浏览器是请求资源的，而服务器是提供资源的。浏览器是一种客户端应用程序，它允许用户查看位于WWW、另一网络以及用户计算机上的HTML文档；允许用户沿着文档中的超链接进行浏览或传输文件。服务器是对浏览器的请求做出反应的计算机或程序。这与通常所说的“客户”与“服务”的概念之间有一定的相似性。这就像商家与消费者之间的关系，商家提供资源，提供服务；消费者享受(或接受)资源，享受(或接受)服务。消费者可以去不同的商家，商家也给许多不同的消费者提供服务。一个浏览器可以向许多不同的服务器请求，一个服务器也可以向多个不同的浏览器提供服务。通常情况下，一个浏览器启动与某个服务器的对话，也就是请求服务。服务器通常是等待浏览器请求的一个自动程序。浏览器通常是作为某个用户请求或类似于用户的某个程序提出的请求而运行的。浏览器的作用就是把从服务器传回的超媒体信息展现在用户面前，它知道如何去解释和显示在WWW上找到的Hypertext(用HTML语言编写)，HTML语言本身包含了各种格式化超文本的方法，从而允许浏览器根据它格式化每一种文本类型，以获得WWW页面(Web Page或HomePage)设计者当初设计时的屏幕显示效果。此外，大多数浏览器都可以自动调用其他应用程序(Helper Applications Program)，以显示特殊类型的文档，如audio或者video的文件格式。协议是浏览器请求服务器和服务器如何应答请求的各种方法的定义，它就好像消费者与商家之间的交易规则，定义了如何请求服务与如何提供服务。WWW中最常用的协议就是HTTP协议，它负责管理超文本转换协议，HTTP是一个客户机/服务器协议，通过HTML接收和传送请求，它是支持WWW上信息交换的Internet标准，是定义Web服务器如何响应文件请求的Internet协议。

浏览器(Browers)是为了使用Web，需要一个Web客户端程序，这一程序能够解释并显示超文本文件，它知道如何找到并显示由链接指向的文件。第一个WWW浏览器，是一种文本行式浏览器，它是由CERN中的小组完成的。Mosaic，是由伊利诺斯大学的NCSA开发的浏览器，由马克·安德森和艾瑞克·比拉开发的。它的出现对Web的增长起到了巨大的推动作用。Mosaic第一个版本是基于UNIX系统下的X Windows上开发的。随着Internet用户群体的增加，NCSA扩展了Mosaic的开发成果，开发出了基于Microsoft Windows和Macintosh版本的浏览器。NCSA还允许为商业软件开发商发放Mosaic源代码。结果，从Mosaic演变并出现了许多商业WWW浏览器，包括微软的Internet Explorer。

常用的浏览器主要包括Netscape Navigator、Internet Explorer(也就是通常所说的IE)等。

在Web中，浏览器的任务是：

- 帮助制作一个请求(通常在单击某个链接点时启动)。
- 将请求发送给某个服务器。
- 呈交HTML文档和传递各种文件给相应的“观察器”(Viewer)，把请求所得的结果进行报告。

一个观察器是一个可被WWW客户机调用而呈现特定类型文件的程序。当一个声音文件被某个浏览器查阅并下载时，它只能用某些程序(例如Windows下的“媒体播放器”)来“观察”。

网络浏览器为网络服务器提供了一个图片似地、以文本为基础的终端界面。这种终端方法在用户与网络服务器间提供了一个界面，网络浏览器负责把网络服务器送来的HTML在浏览器内转换成图形用户界面。

通常浏览器不仅可以向Web服务器发出请求，还可以向其他服务器(例如FTP、邮件服务器)发出请求。

网络服务器有一些责任，即所有的中心必须围绕把HTML传送到请求的客户浏览器上。一个Web服务器的任务是：

- 接受请求。
- 请求的合法性检查，包括安全性屏蔽。
- 针对请求获取并制作数据，包括Java脚本和程序、CGI脚本和程序、为文件设置适当的MIME类型来对数据进行前期处理和后期处理。
- 把信息发送给提出请求的浏览器。

3. 统一资源定位符(URL)

World Wide Web是一个信息资源的网络。Web依靠3种结构来使这些资源为各类用户做好准备：

- 单纯的命名方案——提供在Web上进入资源的统一的方法和路径(Uniform Resource Locator，URL)。
- 协议——允许在Web上交换已命名的资源(HTTP、FTP)。
- 超文本——供在资源之间易于引导(HTML)。

在Web上的任何可用资源——HTML文档、图像、视频、程序等等——都有一个地址可被统一资源定位符解码。“资源”(Resource)是指在网络上所能获取的文字、图像、声音、动画等资料的统称。这些资料实际上都是以各种不同格式的“文件”(File)类型存在，分散于各地的电脑主机中；而“定位符”(Locator)的目的就是要指出这些资料的所在处。

统一资源定位符是互联网上资源的地址。Web浏览器通过使用统一资源定位符(URL)来对互联网上的资源进行定位。在单机系统中，定位一个文件需要路径和文件名，对于遍布全球的Internet，显然还需要知道文件存放在哪个网络的哪台主机中才行。URL与单机系统不同的是：在单机系统中，所有的文件都由统一的操作系统管理，因而不必给出访问该文件的方法；而在Internet上，各个网络，各台主机的操作系统都不一样，因此必须指定访问该文件的方法。一个URL包括了以上所有的信息。URL包含了3个部分：在Web上传输资源所使用的协议名称、数据所在的主机名称、资源本身的路径。它的构成为：

protocol:// machine-name[:port] / directory / filename

其中，protocol是访问该资源所采用的协议，即访问该资源的方法，下面显示了几种协议，它可以是：

- HTTP：超文本传输协议，该资源是HTML文件。最常用的是HTTP，主要用来连接远端的WWW服务器。例如： http://www.pku.edu.cn，即可连接到北京大学的WWW服务器。
- FTP：文件传输协议，用FTP协议访问该资源，主要是用来取得FTP服务器上的文件资源。例如：ftp://ftp.pku.edu.cn，即可显示北京大学的FTP服务器内容。
- Gopher：Gopher协议，该资源是Gopher文件，表明该资源是网络新闻。

machine-name是存放该资源主机的IP地址，通常以字符形式出现，如www.pku.edu.cn。

port，端口号，是服务器在该主机所使用的端口号。一般情况下端口号不需要指定。只有当服务器所使用的端口号不是默认的端口号时才指定。

directory和filename是该资源的路径和文件名。

一个典型的URL为http://www.pku.edu.cn，它表示北京大学主页。再考虑这样一个网页，它是北京大学的学校概况，其URL为：http://www.pku.edu.cn/about/about.htm。这个URL可以这样阅读:使用HTTP协议并且通过www.pku.edu.cn的机器传输，而文件是/about/about.htm。

URL还可能携带更多的信息，我们再来看一个URL地址，它的形式如下所示：http://andyzu:dream@www.dreamworks.com:81/network/intro.htm1#faq，其中，“http://”表明这个地址中的数据是HTTP协议的，需用WWW方式访问；“www.dreamworks.com.cn”是这个地址，其中最后的“cn”是国家标识，二级域名“com”代表公司的含义，“andyzu:dream”表明这个地址需要用户名和密码登录才能访问，而完全公开的网址无需使用用户名和密码；“:81”说明这个服务不在WWW服务的缺少端口80，而是将服务改在81端口上。“network”是子目录名，“intro.htm”是文件名。WWW服务器一般会指定一个默认文件名，当访问一个目录并且没有明确给出文件名的时候，服务器会自动帮助用户调出默认文件，一般是index.htm、default.htm等。WWW服务器上的起始html文件(文件具体存放的路径及文件名)取决于该WWW服务器的配置情况。当一个HTML文本比较长的时候，还可以使用“锚”，如上面的“#faq”，这样可以快速地定位到这个文本的任何一部分。对于使用CGI程序或ASP程序的URL，在文件名后面可以跟随任意形式的参数，如：“http://www.pku.edu.cn/try.asp?user=andyzu&city=beijing”等形式。

与单机系统绝对路径、相对路径的概念类似，统一资源定位符也有绝对URL和相对URL之分。上文所述的是绝对URL。相对URL是相对于用户最近访问的URL。比如一个用户正在观看一个URL为http://www.pku.edu.cn/index.html的文件，如果想看同一个目录下的另一个文件intro.html，可以直接使用intro.html，这时intro.html就是一个相对URL，它的绝对URL为http://www.pku.edu.cn/intro.html。

4. 超文本

超文本文件是与其他数据有关联(links)的数据。包含与其他文档链接的文档；选择链接时自动显示第二个文档。超文本文件的一个简单例子是大百科全书。假设用户正在读“*树*”这个条目，在文章的最末有一个参考这样写到，“相关信息参见‘植物’”。这最末一行就是一

个关联，从“树”到“植物”这个条目。当然这只是一个简单例子，全球网是基于一个远远复杂得多的超文本文件。特别是在文件的任何地方都可能有关联，不仅是在末尾。举一个想像的例子。假设正在用全球网阅读“树”的超文本文件，每次提到一种新的“树”就有一个关联，每个关联都以某种方式标识起来以显得突出。例如，有关联的字可能被加亮或者加了下划线，或者标注了一个数字。如果跟着关联走，将会转到那种类型的“树”的条目去。在主条目里还有其他题目的关联，如“雨林”或“木”。这些关联又引出完整的条目。用户或许还能找到与技术术语的关联，如“落叶科”和“针叶树”。如果沿着这些关联就会找到其定义。在全球网的语言里，一个超文本文件的文件就是由一些数据组成，而这些数据又可能与其他文件相关联。

5．WWW上常用的文件格式

用户在网上查询信息时，经常会遇到不同的文件格式。这些文件中，有一些格式可以直接使用浏览器进行处理，还有一些需要另外的软件来进行处理。

（1）.html或.htm

它是Internet中超文本文档格式文件，在浏览器中可直接浏览。

（2）.au

它是Internet中常见的声音文件格式。

（3）.jpeg或.jpg

它是Internet中常见的图像文件格式，可在浏览器中直接浏览。

（4）.gif

它是Internet中常见的图像文件格式，可在浏览器中直接浏览。

（5）.mov

它是Internet中常见的影视文件格式，可用Quicktime或Realplay等软件播放这类软件。

（6）.gz或.z

它是UNIX中一种常见压缩文件格式，可使用Winzip对这类软件进行解压缩。

（7）.txt

它是一种常见的文本文件，可在浏览器中直接进行浏览。

1.2 HTML基础

World Wide Web(简称WWW)是由无数的商业、教育和娱乐资料充斥着的日益庞大的信息空间网。这些可以通过Internet访问的以超媒体文档形式构成的，资料可以放置在世界上任何地方，但是，无论这些资料怎样产生，绝大多数Web文档都是由超文本标记语言(HTML)编写的。在WWW的世界中，HTML(超文本标记语言，Hyper Text Markup Language)扮演着一个不可或缺的角色。HTML是构成网页最“基础”的要素。

HTML是在1989年由Timothy Berners-Lee所制定出来的。随后，由NCSA推出Mosaic浏览器，将HTML语言推广并得到越来越多用户的欢迎。

HTML是一种描述性语言，是由W3C组织(World Wide Web Consortium)推出的。它是一个按SGML(标准通用标记语言，Standard Generalized Markup Language)定义的语言，和其他

标识语言一样，采用作标记定义文本的特殊格式。标记是对文本的一段进行语义标记。在浏览器解释HTML时，标记本身并不呈现在浏览器上，浏览器按标记的意义来显示被标识的部分。HTML用标记来定义文档中的文本及图像等元素，来指示网络浏览器如何显示这些元素，以及如何响应用户的行为(如通过按键或单击鼠标来激活一个链接)。用HTML可以创建那些能在互联网上传输的信息页——即通常所称的主页(Homepage)或网页。SGML可以通用在各类应用领域的文件上，它允许使用者根据数据结构与形态的需求，制定出实用的文件格式定义(Document Type Definition，简称为DTD)，所以不同领域中的文件，其包含的数据项目与形态可能差异性很大，但只要分别定义出各自的DTD，就可以被各类文件建立时所引用与遵循，以保证同一类的文件都有相似的文件结构。网页本身也就是一种文件，也是互联网中的一种应用领域，所以有关网页的文件结构，当然也可以采用SGML来定义网页适用的DTD。这就是HTML的由来，也是SGML的一种应用，主要是采用SGML的规范来制定网页适用的DTD，而这种特别为网页量身定做而产生的标记语言就称为HTML(Hyper Text Markup Language)，中文可译为超文本标记语言。在本书的后面讲到XML时，将专门讲述DTD的定义与撰写方式，读者就可以体会到上面这段话的含义了。

说明：

① **SGML英文全名为Standard Generalized Markup Language，中文可以为标准通用标记语言。SGML最主要的目的是在提供一种描述电子文件的规范，也就是提供一种对文件进行结构化的法则，而当文件采用这种法则来进行结构化的处理后，该SGML文件就可以被广泛地传递与使用，且有关该文件的制作、存取、应用都可以通过计算机来做最有效的处理。**

② **W3C英文全名是World Wide Web Consortium，中文可以译为全球信息网协会，该组织主要负责有关全球信息网应用各类标准的制定，像HTTP、HTML、XML等等都是该协会制定出来的。且该协会主要是由对全球信息网有兴趣的公司所组成的，成员包括Microsoft、HP、IBM、AT&T、Netscape等，所以一般通过该组织推荐的标准都能受到各家厂商的支持。**

HTML语言使用描述性标记符(称为标记Tag)来指明文本的不同内容。标记是区分文本各个组成部分的分界符，用来把HTML文本划分成不同的逻辑部分(或结构)，如段落、标题和表格等。标记描述了文本的结构，它向浏览器提供该文本的格式化信息，以传送文本的外观特征。用HTML语言编写的页面是普通的文本文件(ASCII文件)，不含任何与操作系统和计算机硬件相关的信息，所以HTML文件可以被任何文本编辑器读取。

HTML是一种语言，但并不算是“程序”语言。HTML所定义的范畴仅局限于如何表现文字、图片，以及如何建立文件之间的链接。而程序则是经过规划的一连串命令(或称为“语句”)，而这样的命令可用来驱动操作系统或应用程序执行某些操作。

HTML文本包含两种信息：页面本身的文本和表示页面元素、结构、格式、及其他超文本链接的HTML标记。HTML标记规定了HTML文本的逻辑结构，并且控制其显示格式，也就是说，设计者可以用标记定义HTML文本的逻辑结构，但是文本的实际显示则由浏览器来负责解释。可以使用HTML标记来设置链接、标题、段落、列表和字符加亮区域等等。

大部分HTML标记是以“<标记名>相应内容</标记名>”形式出现的，标记的名字用尖括号括起来。HTML标记一般有起始标记与结束标记两种，分别放在它起作用的文本两边。

起始标记与结束标记极相似，只是结束标记在“<”号后面多了一个斜杠“/”。后面将会看到，某些HTML元素只有起始标记而没有相应的结束标记，例如换行标记
，由于换行不包括相应的内容，所以只使用一个标记就可以了。还有一些元素的结束标记是可以省略的，如列表项结束标记“</LI>”、段落标记“</P>”和表格行结束标记“</TR>”等等。标记名不区分大小写。

起始标记中可以包含属性(Attribute)，其位置是从标记名之后空一格的地方开始，在结束符“>”之前结束，如 <FONT 属性>文字</FONT>。属性向客户端提供了关于页面元素内容以及如何处理的附加信息，用户可以在这个区域中对文本的一些具体属性如大小、颜色、字体等信息进行设置。

1.3 HTML语言的基本构成

本节先以实际的例子来示范HTML语言的基本结构，下面的实例1-1就是一份HTML文件，该HTML文件经过浏览器解读以后，即可显示出如图1-1的网页内容。

例1-1：

```
<HTML>
<HEAD><TITLE>个人资料</TITLE></HEAD>
<BODY>
  <H3>姓名：张华</H3>
  <H3>生日：1978年10月4日</H3>
  <H3>职业：国家公务员</H3>
  <H3>电子邮箱：lily@263.net</H3>
</BODY>
</HTML>
```

显示结果：

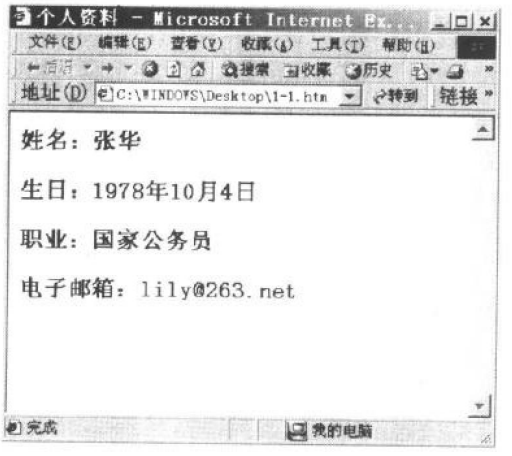

图1-1 例1-1的显示结果