

最新XML

入门与应用

林锦雀 编著
江高举 策划

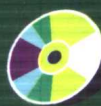
本书特色

本书从标记语言的历史到XML与其相关技术进行一系列的介绍，以使读者先有一个基础的概念。接着以如何建立第一个XML文件，XML语法介绍，文件格式定义，到设置XML样式，让读者在浏览器中便可看到以XML设计出的精美网页。

除了介绍XML本身的语法外，在第7章中还介绍了XSL。它是另一种样式表语言，不但可以对XML文件进行格式的美化，而且还提供了转换的功能，可以将XML文件转换成目前Web中最常见的HTML格式，让XML的应用更加广泛。

如何在XML文件之间做链接，以及如何在XML文件内部进行数据的链接，或是如何链接到其他的资源，是本书第9章和第10章的重点。XLink和XPointer则是对应的标准。这些与XML相关的标准，也都是由W3C所制定。

本书最后的三章，重点在于介绍XML与HTML和ASP的相关整合，以及XML如何应用在电子商务中，最后以简单的范例说明如何整合这些技术，让XML更能发挥它超强的功能！



内附精美光盘

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

最新 XML

入门与应用



中国铁道出版社

2001年·北京

(京)新登字 063 号

北京市版权局著作权合同登记号: 01-2001-3845 号

版 权 声 明

本书中文繁体字版由台湾金禾资讯股份有限公司出版(2001)。本书中文简体字版经台湾金禾资讯股份有限公司授权由中国铁道出版社出版(2001)。任何单位或个人未经出版者书面允许,不得以任何手段复制或抄袭本书内容。

图书在版编目(CIP)数据

最新 XML 入门与应用/林锦雀编著. —北京: 中国铁道出版社, 2001. 10

ISBN 7-113-04393-3

I. 最… II. 林… III. 可扩充语言, XML—程序设计 IV. TP312

中国版本图书馆 CIP 数据核字(2001)第 070442 号

书 名: 最新 XML 入门与应用

作 者: 林锦雀 江高举

出版发行: 中国铁道出版社(100054, 北京市宣武区右安门西街 8 号)

策划编辑: 苏 茜

特邀编辑: 袁秀珍

封面设计: 孙天昭

印 刷: 北京兴顺印刷厂

开 本: 787×1092 1/16 印张: 17.5 字数: 414 千

版 本: 2001 年 11 月第 1 版 2001 年 11 月第 1 次印刷

印 数: 1~5000 册

书 号: ISBN 7-113-04393-3/TP·624

定 价: 31.00 元

版权所有 盗版必究

凡购买铁道版的图书, 如有缺页、倒页、脱页者, 请与本社计算机图书批销部调换。

IS710/01

出版说明

本书结构合理、由浅入深，作者用最简单的字句向读者全面介绍 XML 这一发展迅猛的新技术。每一章在容易混淆、不易理解的地方，配合了插图，使读者在最短的时间内准确地理解书中的内容。随书所附光盘，包含书中所有范例，使读者能即刻实践，所见即所得。

本书由台湾金禾资讯股份有限公司提供版权，经中国铁道出版社计算机图书项目中心审选，陈强、柯研、梁秀玲、关超、史广顺、陈辑超、马超、杨小平、段小明、杨军、裘伟力、陈贤淑及孟丽花等同志参与了本书的整稿及编排工作。

中国铁道出版社

2001 年 10 月

目 录

第 1 章 XML 相关技术简介	1
1-1 标识语言	2
1-1-1 标识语言历史	2
1-1-2 SGML	4
1-1-3 HTML	4
1-2 明日之星 XML	6
1-2-1 XML 优越性	6
1-2-2 XML 文件的特性	7
1-2-3 XML 相关技术	7
DTD	8
CSS	9
XSL	10
XLink	11
XPath 与 XPointer	12
1-2-4 XML 相关软件	12
第 2 章 建立一个 XML 文件	15
2-1 XML 文件	16
2-1-1 XML 文件结构	16
2-1-2 XML 编辑器	17
XML Notepad	17
EXML	20
2-1-3 XML 文件中使用中文	23
2-1-4 浏览器中查看 XML	25
2-2 加入排版样式	26
2-2-1 CSS 样式表	26
2-2-2 XML 文件中套用 CSS	27
2-2-3 来点样式变化	29
修改原有的 CSS 文件	30
建立新的 CSS 文件	30

第 3 章 XML 语法介绍	33
3-1 XML 基本语法	34
3-1-1 标记命名原则	34
3-1-2 属性设置	34
3-1-3 加入注释	36
3-1-4 特殊字符的使用	38
实体参照	38
字符参照	39
3-1-5 CDATA 标记	41
3-2 XML 文件构成要素	41
3-2-1 声明	42
3-2-2 元素	43
3-2-3 处理指令	43
3-2-4 满足 Well-Formed 的 XML 文件	44
常见的错误	44
3-3 XML 解析器	44
3-3-1 解析器的功能与种类	45
3-3-2 使用 Internet Explorer 5.0 来解析 XML 文件	46
3-3-3 其他解析器	48
第 4 章 文件格式定义	49
4-1 文件格式定义	50
4-1-1 文件格式定义的用途	50
4-1-2 调用文件格式定义	50
内部文件格式定义	50
外部文件格式定义	52
4-1-3 正确合法的 XML 文件	53
4-2 元素声明与使用	55
4-2-1 元素的种类	55
4-2-2 元素数据类型	56
含子元素	56
#PCDATA	56
ANY	57
EMPTY	57
4-2-3 元素数目设置	57
元素次数	58
选择性元素	58
混合型	59

4-3	元素属性	59
4-3-1	属性声明	59
4-3-2	属性应用	60
	CDATA 与 Enumerated	60
	ID、IDREF 与 IDREFS	61
	NMTOKEN 与 NMTOKENS	62
4-3-3	XML 内定属性	63
	xml:space	63
	xml:lang	65
4-4	DTD 的其他主题	65
4-4-1	DTD 中加入注释	66
4-4-2	实体声明	66
	内部实体	66
	外部实体	68
	参数实体	69
4-4-3	NOTATION 声明	70
	ENTITY 与 ENTITIES 属性类型	72
	NOTATION 属性类型	73
4-4-4	善用 INCLUDE 和 IGNORE	74
第 5 章	设置 XML 样式	77
5-1	CSS 样式表	78
5-1-1	CSS 基本介绍	78
	加入注释	80
5-1-2	元素的选取	80
	图样对应 (Pattern Matching)	80
	优先级 (Priority)	81
5-1-3	CSS 属性值	82
	长度	82
	URL	82
	颜色	83
	关键词	83
5-1-4	DTD 与 CSS 整合应用	83
	调用多个 CSS 样式表	83
	@规则	87
	DTD 与 CSS	88
5-2	网页相关属性	91
5-2-1	区块观念	91
	Width 和 Height 属性	91

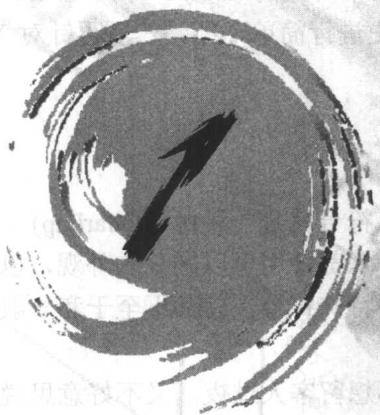
5-2-2	Display 属性.....	91
5-2-3	Margin 与 Padding 属性.....	92
5-2-4	Border 属性	93
5-2-5	Position 属性	95
5-3	文字与字型属性	97
5-3-1	字型属性	97
5-3-2	文字属性	97
5-4	其他视觉属性.....	99
5-4-1	背景属性	99
5-4-2	Float 和 Clear 属性.....	101
5-4-3	Overflow 属性	102
第 6 章	名称空间	105
6-1	名称空间介绍.....	106
6-1-1	名称空间能解决什么问题.....	106
6-1-2	声明名称空间	107
6-2	名称空间的特性.....	109
6-2-1	范围	109
6-2-2	默认值	111
6-3	应用名称空间.....	112
6-3-1	利用名称空间引用 HTML 标记.....	112
6-3-2	名称空间与 XSL	115
第 7 章	另一种样式表语言 XSL	117
7-1	XSL 样式表语言.....	118
7-1-1	XSL 与 CSS 的比较.....	118
7-1-2	XSL 中设置排版样式	118
7-1-3	XSL 转换语言	120
	XSL 格式对象	120
7-2	XSLT	121
7-2-1	2XSLT 的工作原理	121
7-2-2	样板规则	123
7-2-3	对应式	123
7-3	应用 XSL 元素	124
7-3-1	基本 XSL 元素.....	124
	xsl:stylesheet.....	124
	xsl:template 与 xsl:apply-templates	124
	xsl:value-of.....	126

	xsl:for-each.....	127
7-3-2	以 XSL 元素建立新对象.....	129
	xsl:element 与 xsl:attribute.....	129
	xsl:text	131
	xsl:comment	131
7-3-3	条件式的应用	132
	xsl-if.....	132
	xsl:choose.....	134
7-3-4	合并多个 XSL 样式表.....	136
	xsl:import	136
	xsl:include.....	136
7-3-5	输出元素的排序	137
	xsl:sort	138
7-3-6	XSL 与 HTML 整合应用.....	138
第 8 章	连接 XML.....	141
8-1	数据之间的链接.....	142
8-1-1	HTML 中的链接	142
	HTML 链接的缺点	143
8-1-2	XLink	144
	XLink 声明.....	144
8-2	简单链接.....	145
8-2-1	简单链接的属性	145
8-2-2	简单链接范例	146
	建立简单链接 XLink 元素	146
8-3	延伸链接.....	149
8-3-1	延伸链接的属性	149
8-3-2	延伸链接范例	151
	建立延伸链接 XLink 元素	151
	XML 元素中加入延伸链接属性	152
8-3-3	Inline 链接 与 Out-of-line 链接	155
第 9 章	Xpointer.....	157
9-1	XML 文件内的连接设置.....	158
9-1-1	HTML 中的 Pointer	158
9-1-2	为何使用 XPointer.....	161
9-1-3	XPointer 寻址方法.....	161
	简单名称(Bare Name)	163

	子序列(Child Sequence)	163
	完整型(Full)	164
9-2	XPath	164
9-2-1	XPath 简介	164
	相对位置路径	164
	绝对位置路径	165
9-2-2	寻址步骤	165
	轴(Axis)	165
	节点测试(Node Test)	166
	陈述(Predicate)	167
9-2-3	缩写表示法	167
9-3	XPointer 延伸 XPath	168
9-3-1	XPointer 提供额外的函数	168
	string-range 函数	169
	range 相关函数	169
	here 函数	170
	unique 函数	170
9-3-2	XPointer 的错误	170
第 10 章	XML 与 HTML 整合	171
10-1	相关整合技术	172
10-1-1	Data Islands	172
10-1-2	Data Source Object	173
	内部 XML 文件	174
10-2	在 HTML 中显示 XML 文件	175
10-2-1	由 Data Island 读取所有 XML 数据	175
	内部 XML 文件	176
	外部 XML 文件	177
10-2-2	由 Data Island 读取目前的 XML 元素	179
10-2-3	显示 XML 元素的属性	181
10-2-4	复杂元素的显示	184
10-3	善用 DSO 对象	186
10-3-1	显示所有 XML 文件内容	187
10-3-2	一笔一笔显示 XML 元素	189
10-3-3	分页显示 XML 文件	191
第 11 章	XML 文件对象模型	195
11-1	文件对象模型	196

11-1-1	何谓 XML DOM	196
11-1-2	DOM 对象	198
11-1-3	建立 XML DOM 对象	199
11-2	常用的 DOM 对象	201
11-2-1	Node 对象	201
11-2-2	NodeList 对象	203
11-2-3	NamedNodeMap 对象	204
11-2-4	Document 对象	204
11-2-5	Element 对象	206
11-2-6	Attribute 对象和 Text 对象	207
	Text 对象	207
11-3	XML DOM 的应用	208
11-3-1	存取指定 XML 元素	208
11-3-2	处理 XML 元素的属性	210
11-3-3	新增 XML 元素	212
11-3-4	删除 XML 元素	214
11-3-5	一个 XML DOM 范例	215
第 12 章	XML 在电子商务的应用	219
12-1	电子商务	220
12-1-1	何谓电子商务	220
12-1-2	B-to-C 电子商务	221
12-1-3	B-to-B 电子商务	222
12-2	Web 应用程序架构	223
12-2-1	主从式架构	223
12-2-2	三层式架构与 N 层式架构	224
12-3	XML 在各层次所扮演的角色	225
12-3-1	XML 在 Client 层	225
12-3-2	XML 在 Middle 层	226
12-3-3	XML 在 Back-End 层	227
附录 A	HTML 简介	229
A-1	何谓 HTML	230
A-1-1	HTML 文件	230
	HTML 文件的结构	230
A-1-2	HTML 组成的基本组件	231
A-2	背景设置	231
A-3	文字属性的变化	232
A-3-1	设置文字的字型、字体大小及颜色	232

A-3-2	文字的样式	233
A-3-3	空格符	234
	换行	234
A-3-5	预先编排好的文字	235
	编号及项目符号	235
	项目符号	237
A-4	图文并茂的文件	237
A-4-1	指定文文件来源	237
	绝对路径	238
	相对路径	238
A-4-2	指定图形的宽度及高度	238
A-4-3	图片外框	239
A-4-4	显示图形或文字	239
	如何取消显示图形	239
	显示图形或文字	240
A-4-5	对象居中	240
A-5	超级链接	241
A-5-1	文字超级链接	241
A-5-2	图片超级链接	241
A-5-3	页内超级链接	242
A-5-4	超级链接电子邮件帐号	243
A-6	表格	245
A-6-1	列与行	246
A-6-2	列与行的背景颜色	246
A-6-3	表格框线的颜色	246
A-6-4	表格间距	247
A-6-5	合并单元格	248
	合并列	248
	合并行	248
A-7	段落	249
A-8	水平线	250
A-9	何谓 CSS	251
A-9-1	CSS 类型	251
	内嵌 CSS	251
	局部引用 CSS	253
	外部链接 CSS	254
A-9-2	CSS 在超级链接的运用	255
A-9-3	实务范例	256
	CSS 入门范例	256



XML 相关技术简介

本书主要介绍 XML 技术及其应用。全书共分 10 章。第 1 章介绍 XML 的起源、特点、优势、应用及与 HTML 的关系。第 2 章介绍 XML 的语法、文档结构、命名空间、DTD 等。第 3 章介绍 XML 的解析、生成、验证、转换等。第 4 章介绍 XML 的序列化、反序列化、传输、存储等。第 5 章介绍 XML 的查询、更新、删除、插入等。第 6 章介绍 XML 的加密、解密、签名、验证等。第 7 章介绍 XML 的压缩、解压、归档、解压等。第 8 章介绍 XML 的模板、渲染、输出、输入等。第 9 章介绍 XML 的数据库、查询、更新、删除、插入等。第 10 章介绍 XML 的 Web 服务、调用、发布、消费等。

1-1 标识语言

标识语言(Markup Language)或称为标记语言，是目前在互联网上当红的名词，它在不同计算机系统之间的数据交换扮演着重要的角色！我们常听到的 HTML 和本书要详细介绍的 XML，都属于标识语言。本节将先对标识语言的历史进行简单介绍，然后再针对 XML 与其相关技术分别介绍。

1-1-1 标识语言的历史

您可以想像吗？我们平常在文章中所使用的标点符号就是一种标识(Markup)，它保留在文章中，以便告诉读者如何去中断文字。大部分的标识都使用可以区分的外观，以便和原有的文字分隔。例如：我们所使用的标点符号都是使用特殊符号，使其不至于和一般的字母或数字等文字内容混淆。

相信大家都读过这样一个故事：主人在下雨天不想留客人过夜，又不好意思说出来，于是写了一个字条，希望客人能知趣自行走人：

下雨，天留客，天留我不留。

客人想留下来，便修改了原文的标点符号，使语意和主人完全不同。

下雨天，留客天，留我不？留。

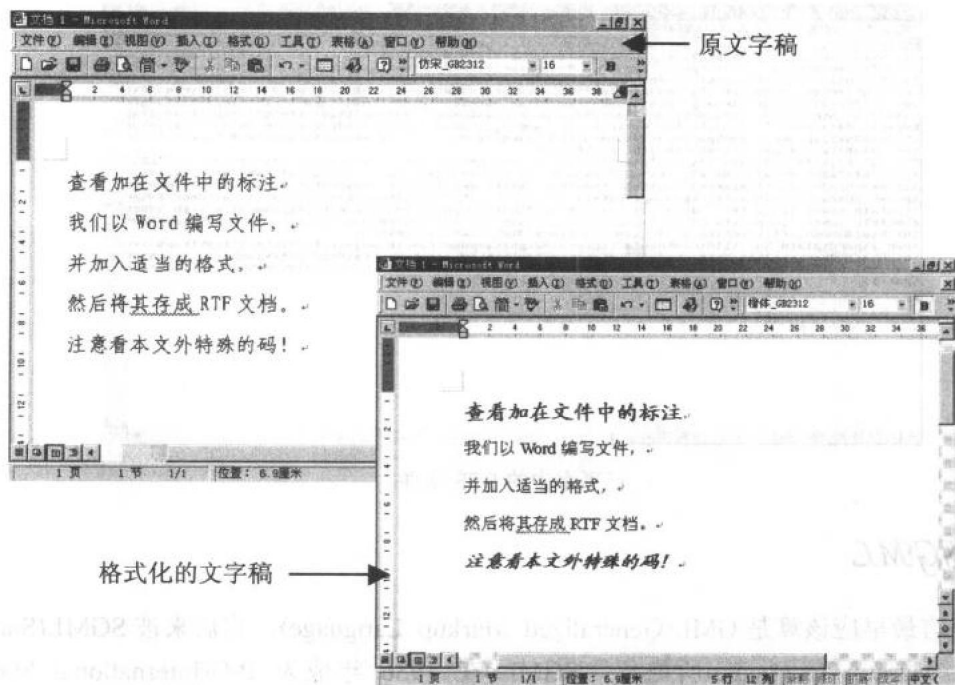
即使在印刷技术已经发明的时代，当手写文件要打印之前，作者必须在原稿中加入一些注释，以便告诉印刷者做适当的格式设置和版面设计，这些不属于本文且与格式或版面编排有关的注释，就称为标识(Markup)。若这些标识的集合，遵照一定的文法和语法，就可以被称为一种语言(Language)。

说明

Explain

大部分的标识语言都以<>为标记，来区分原文和所加入的标识。

随着计算机的普及，文件的处理可以利用计算机来完成，这些电子文件一般的文书处理可以借助适当的文字处理软件轻松完成。在输入完文稿的内容后，为了文件的美观起见，可以利用文字处理软件所提供的格式指令，对文件作适当的格式美化和版面编排。这些额外的信息就是标识，它们被存储成特殊的代码并且和原文件一起被保存，当文件打印时，打印机就根据这些额外的信息打印出用户所要的文件。

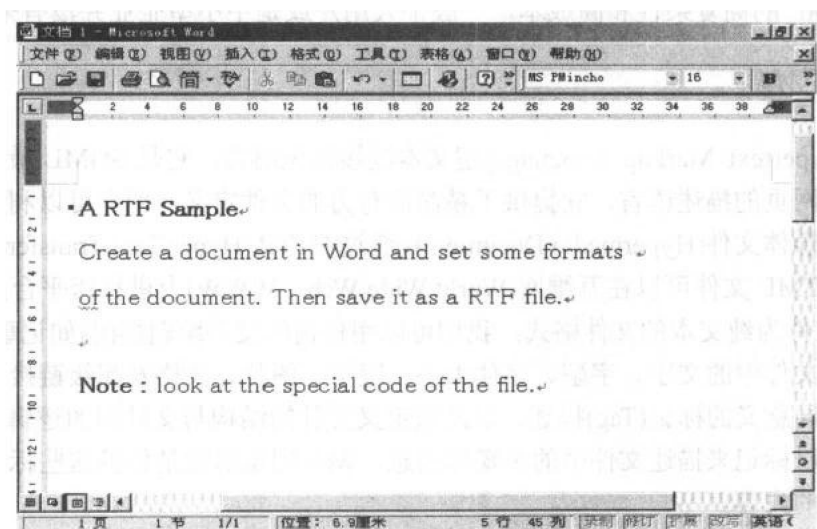


由于 Word 具有打印预览的功能，所以在 Word 中所看到的就是以后打印所看到的结果。为了让读者区分原文和编码的不同，以下将先在 Word 中建立一个英文的范例，然后将其存为 RTF 文件，并在 Windows 的记事本(只能显示文本文件)中打开。

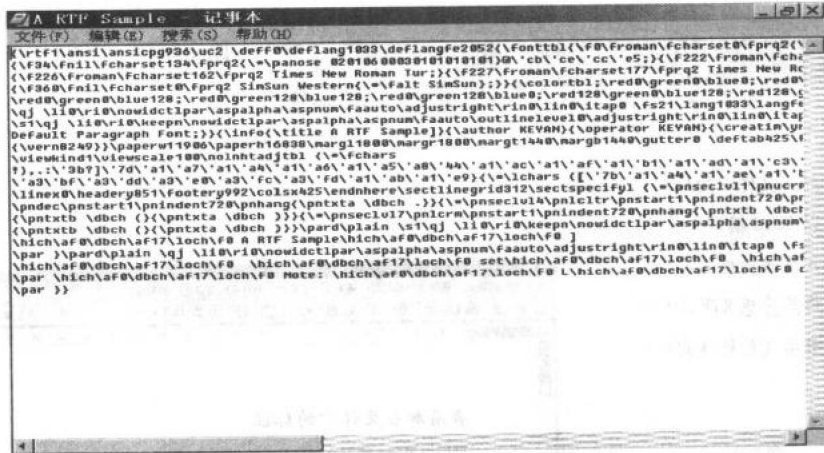
说明

Explain

因为中文存为 RTF 文件时，无法显示原有的中文，所以以英文稿作范例。



原英文范例



记事本中的 RTF 文件

1-1-2 SGML

标识语言最早应该算是 GML (Generalized Markup Language)，它后来被 SGML (Standard Generalized Markup Language) 所取代，SGML 在 1986 年成为 ISO (International Standard Organization) 标准。SGML 允许用户根据与语意有关的标记 (Tag) 和文件属性，去定义一个文法来标识文件，也就是说 SGML 是一个 meta-language，它可以用来定义其他的语言。

在信息爆炸的时代，各式各样的文件充斥在人们的生活中，实在无法用一个单纯的文件结构满足所有用户的需要。例如各式各样的科技文件：生物的、化学的、数学的、物理的、医学的、音乐的以及各式的 Web 网页等。SGML 并不是提供一个固定的标记集合来满足所有用户的需要；相反的是 SGML 提供一个语言，它让作者们去描述他们的文件结构和对文件进行适当的标识。

SGML 本身是功能很强且复杂性高的标识语言，它目前被广泛应用于各国政府单位、出版界和大型制造厂，这些 SGML 文件最后都可以被转换成适当的格式，以便出版或是打印输出。由于 SGML 的高复杂性和高成本，一般个人用户或是中小型企业并没有采用它！

1-1-3 HTML

HTML (Hypertext Markup Language) 超文本链接标识语言，它是 SGML 最有名的应用程序，也是建构网页的描述语言，它提供了精简而有力的文件定义，用户可以利用它来设计出多彩多姿的超媒体文件 (Hypermedia Document)，通过 HTTP (Hyper Text Transfer Protocol) 通讯协议，使得 HTML 文件可以在万维网 (World Wide Web, WWW) 上进行跨平台的文件交换。

HTML 文件为纯文本的文件格式，我们可以用任何的文书编辑器 (例如记事本、PE2 等) 来编辑。至于文件中的文字、字型、字体大小、段落、图片、表格及超级链接，甚至是文件名称都是以不同意义的标记 (Tag) 描述，以此来定义文件的结构与文件间的逻辑关联。简而言之，HTML 是以标记来描述文件中的多媒体信息，Web 浏览器就是依据这些标记来处理文件以及显示文件的内容。

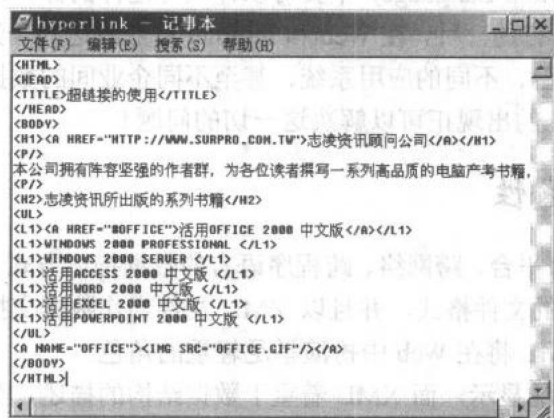
HTML 是目前互联网中用来做数据交换的主要文件格式，只要文件格式为 HTML，几乎

所有的 Web 浏览器均可以正常地打开显示! HTML 的出现也大大改变了人们的阅读习惯,在网页中提供了超级链接(Hyper Link),读者可以依照自己的喜好,点击具有超级链接的文字或图形,系统就会自动切换到对应的网页或数据。

说明

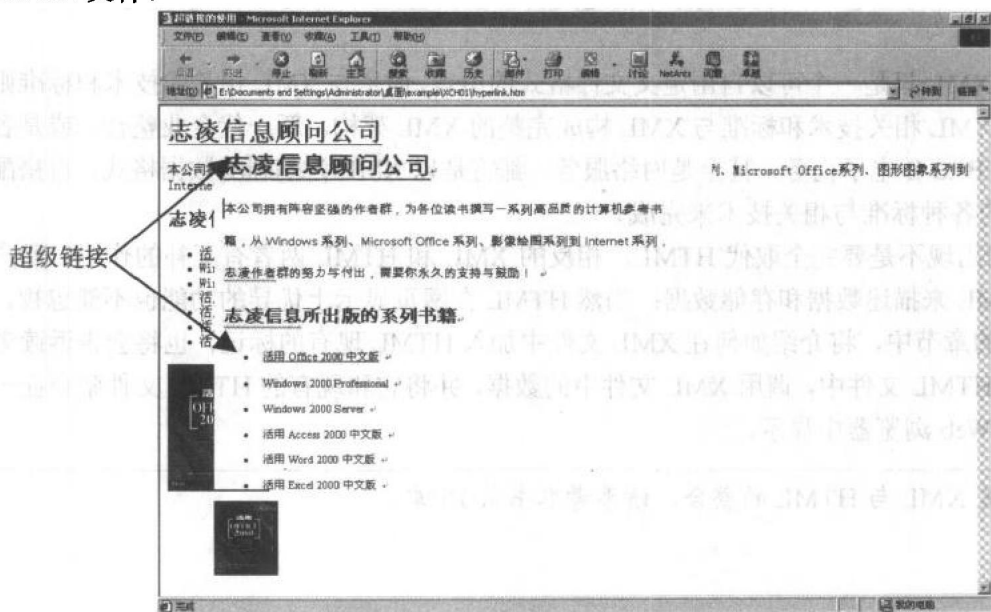
Explain

有关 HTML 的基本介绍,请参考本书附录 A。



HTML 范例

HTML 的专长是将文件数据显示在浏览器中,它提供了很多的标记都和文件的显示有关,这些标记都是系统定义的,用户无法自定义所要的标记。虽然如此,但是就一般的使用,HTML 所提供的标记还是够用的,而且目前市面上很多网页制作软件,均支持 HTML 的文件格式,甚至提供图形化的操作接口以使用户轻松完成所需的网页,最后将这些文件存储成对应的 HTML 文件。



IE 中的 HTML 范例