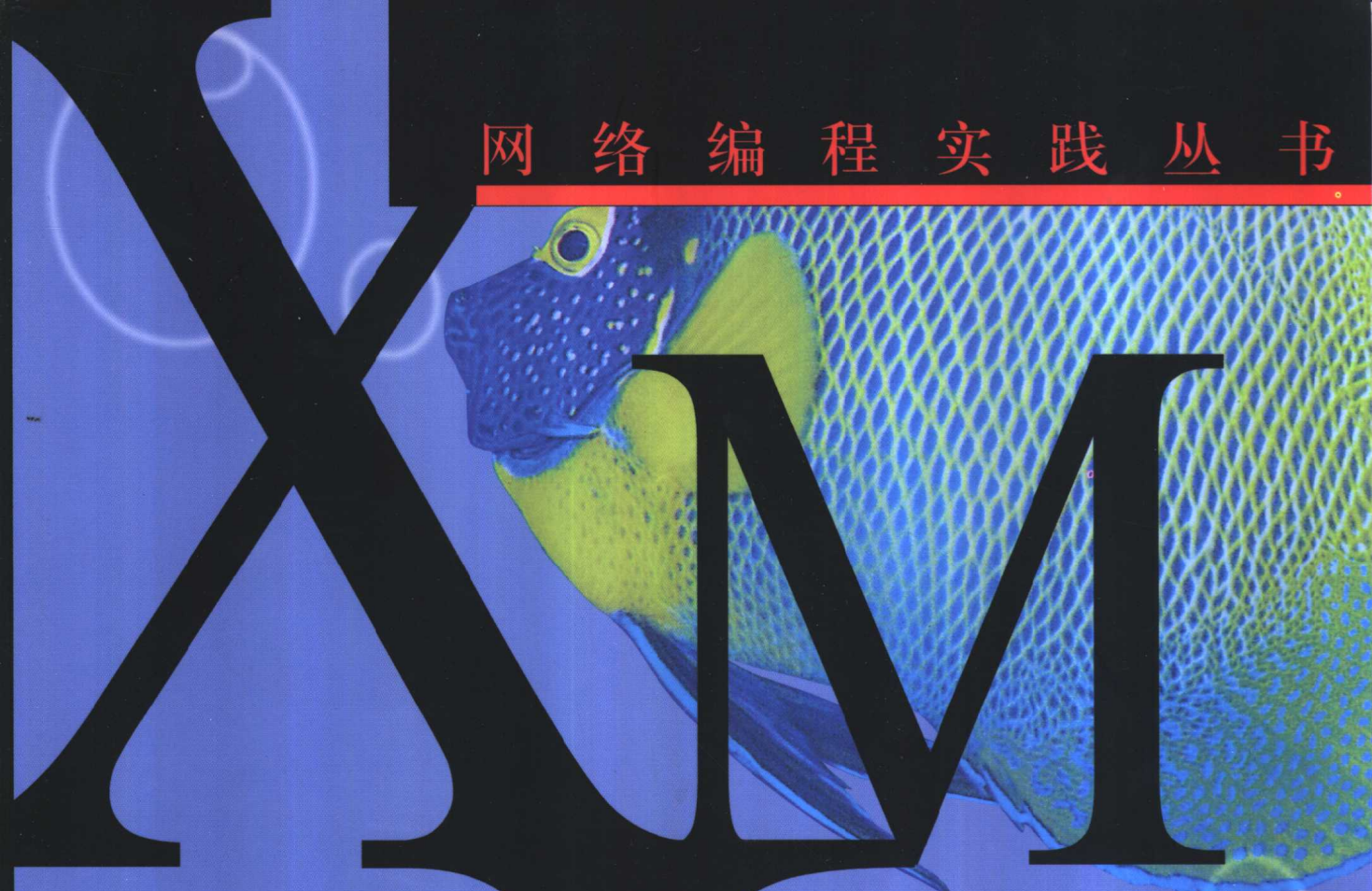


网络编程实践丛书



邵敏 李力鸿
郑震坤 何川 编著
都志辉 主编
刘鹏 陈渝 副主编

XML编程实践

网络上的世界语



清华大学出版社

网络编程实践丛书

XML 编程实践

——网络上的世界语

邵敏 李力鸿

郑震坤 何川 编著

都志辉 主编

刘鹏 陈渝 副主编

清华大学出版社

内 容 简 介

本书从编程实践的角度,对 XML 进行了比较全面的介绍,向读者展示了如何利用这一新工具进行 Web 开发。其内容包括:XML 的个性化显示,解决 XML 中的多语言问题,XML 与 Servlet 和 JSP 的联合编程,XML 与数据库的联系以及利用 SAX 和 DOM 来处理 XML。在本书的最后,给出了一个利用 XML、Java 编程的综合实例。

本书用大量的实例,生动地阐述了 XML 编程要点,使读者可以动态地掌握编程方法,而非静态地学习编程规则;本书着重介绍应用编程经常会用到的一些技术,但并不面面俱到,其目的是让读者尽可能简洁地掌握编程的要旨。随书所附光盘包括了本书所有的程序和一些相关的 XML 规范。

本书可作为广大的网络编程人员以及网络爱好者开发网络程序的参考资料。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

书 名: XML 编程实践——网络上的世界语

作 者: 都志辉 主编

出 版 者: 清华大学出版社(北京清华大学学研大厦,邮编 100084)

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

责任编辑: 刘 彤

印 刷 者: 北京四季青印刷厂

发 行 者: 新华书店总店北京发行所

开 本: 787×1092 1/16 **印张:** 26.5 **字数:** 595 千字

版 次: 2002 年 12 月第 1 版 2002 年 12 月第 1 次印刷

书 号: ISBN 7-900643-85-0

印 数: 0001~4000

定 价: 48.00 元(含盘)

丛书编委会

主 编 都志辉

副主编 刘 鹏 陈 渝

编 委 何 川 方 兴 郑震坤
李力鸿 邵 敏 顾 雷
房 冰 齐 畅 毛佳音
刘 琳



丛书序言

XML

编程实践

网络上的世界语

20 世纪 70 年代的 ARPANET, 还只是少数科学家的网络实验环境, 而从 ARPANET 发展起来的 Internet, 在短短几十年里, 已经迅速成为科研机构、政府部门、公司企业乃至整个社会高效、正常运转必不可少的基础性设施和平台。Internet 的覆盖范围在不断扩大, 用户数量在迅速增加, 功能和性能在不断完善。Internet 为我们提供了一个充满机遇和挑战、将梦幻与真实完美结合的、无比广阔的舞台和空间。熟悉并掌握 Internet 技术, 不仅可以为我们的生活增添绚丽的色彩, 而且可以使我们在竞争日益激烈的时代, 增强自身的竞争力, 更好地把握成功, 充满自信地迎接挑战!

为了充分地把握和利用网络, 网络编程者较普通的网络用户, 需要更多的网络专门知识, 从而可以更自由、更全面地利用网络, 实现各种复杂和强大的网络功能, 同时通过网络以简单和方便的形式提供这些功能, 这是一般的普通网络用户很难做到的。

网络热也引起了网络图书的出版热潮, 广大的网络编程者在这些图书面前应如何选择呢? 本套丛书不是对网络的技术规范进行详细描述, 其侧重点不是网络技术的理论, 而是网络技术的实践和应用。这套网络编程丛书告诉读者的是当前先进和前沿的网络编程技术有哪些, 使用这些技术可以用来做什么, 在进行网络编程时会遇到什么样的问题, 应该如何解决这些问题等内容。

本套丛书具有如下特点:

1. 内容实用。重点介绍的是网络编程中常见的、有一定难度的实际问题的解决方法和思路, 而不是对网络技术本身进行全面、详尽的介绍。因此本套丛书适合于对网络编程技术有一定了解, 但又缺乏大量的实际经验, 希望能够进一步提高实际编程技术的读者。

2. 语言通俗。由于本套丛书介绍的网络编程技术对于普通的初学

III

者来说是有一定难度的,因此用深奥难懂的理论和专业表达方式就十分不利于读者的学习和理解。为此,本套丛书在叙述过程中,力求语言简洁、通俗,将有难度的地方、容易误解和出错的地方专门标识出来,根据作者在实际编程和学习过程中积累的丰富经验,进行有针对性的介绍和讲解。

3. 经验丰富。本书的作者虽然都很年轻,但是他们已有长期的编程经验和丰富的程序设计技术积累。在某种程度上说,程序设计是充满朝气的年青一代的天下。本书所有作者年龄都在 32 岁以下,这正是程序设计人员的黄金年龄,他们长期积累的丰富编程经验在本套丛书中得到了充分体现,这些难得的财富使本套丛书更具价值。

网络技术的发展日新月异,本套丛书提供的是最先进的网络编程技术和经验,希望通过本套丛书的学习,读者能少走弯路,在较短的时间内,掌握网络这一强大工具,成为网络技术的弄潮儿,在网络的世界里冲浪!

丛书编委会



前言

XML

编程实践

网络上的世界语

XML 被称为是下一代的 Web 语言,它是一种类似于 HTML 的语言。但是却不同于传统的 HTML。

XML 是被设计用来描述数据的语言。在 XML 中标志(在 HTML 中就是类似与<a>,<input>之类的东西)是没有预先定义的,必须要自定义需要的标志。XML 是能够自解释(self describing)的语言,它使用 DTD(document type definition,文档类型定义)来显示这些数据。XML 是关于数据的语言,它是可扩展的,同时 XML 允许用户自定义标志和文档结构。将来网站开发者会使用 XML 来描述网站所需数据和网站的结构,而用 HTML 来格式化和显示这些数据。XML 会成为网站开发的一个重要特色,也许它将挑起所有数据的处理和传输任务。

根据笔者经验,只要对 HTML 有一些初步的了解,哪怕一开始对 XML 一无所知,只要通过一些初步的学习,就可以很快地掌握 XML 的一些基本功能。有鉴于此,本书的编写也是按照从易到难、从简单到复杂、从低级到高级的顺序进行。书中的第 1 章~第 2 章简单介绍了一些 XML 的基本知识;第 3 章~第 7 章对 XML 与 Java、Servlet 等语言的联合编写,与数据库的联系以及 DOM、SAX 在 XML 中的应用进行了详细的讲解;第 8 章~第 9 章给出了两个综合实例,使读者从感性上直接看到 XML 的应用。

从 XML 进入中国至今,关于 XML 的书籍层出不穷。但是大部分都仅仅限于对 XML 基本知识和理论的讲解。而本书的最大特色,就是由一批有 XML 实践编程经验的人来向读者展示他们的思想和方法。对读者来说,这可能更具实用价值。本书的作者在 XML 编程方面有丰富的经验,并努力把自己的亲身感受付诸笔端。

本书的第 1 章、第 2 章和第 4 章由邵敏完成,第 3 章和第 9 章由郑震坤完成,第 5 章和第 6 章由李力鸿完成,第 7 章和第 8 章由何川完成。

希望读者在阅读此书时,将发现的错误或建议及时告诉笔者([mailto: shaomin@tirc.cs.tsinghua.edu.cn](mailto:shaomin@tirc.cs.tsinghua.edu.cn)),以便在再版的时候进一步提高本书的质量,我们将非常感谢!

作 者

2001 年 10 月 14 日于清华大学



目 录

XML

编程实践

网络上的世界语

第 1 章 在 XML 中实现个性化显示	1
1.1 用 CSS 控制 XML 数据展示风格	1
1.1.1 CSS 项目信息罗列	1
1.1.2 用 CSS 给文本上妆	4
1.1.3 用 CSS 为文本整容	8
1.1.4 用 CSS 包装文本外围格式	11
1.2 用 XSL 使 XML 数据展示出不同的风格	18
1.2.1 为何选用 XSL	18
1.2.2 用 XSL 将 XML 文件转化成 HTML 文件	20
1.2.3 使用 XSL 筛选 XML 文件	23
1.2.4 运用 XSL 中的各种函数为显示提供更大的 灵活性	28
1.2.5 用 XSL 控制 XML 文件的显示	41
1.3 利用 XPath 对 XML 元素进行定位	45
1.3.1 XPath 简介	45
1.3.2 书写定位路径表达式	45
1.3.3 书写复杂表达式	47
1.4 用 XSLT 进行 XML 文档的转换	49
1.4.1 XSLT 概述	49
1.4.2 XSLT 工作原理	50
1.4.3 一个简单的 XML 转换	51
1.4.4 改变 XML 文件元素的顺序	54
1.4.5 复制源文件及转换 XML 文件编码	59
1.4.6 在 XML 文件中引入条件判断	62

1.4.7 综合实例:综合运用多种 XSLT 元素	67
第 2 章 解决 XML 中的多语言问题	73
2.1 如何解决 servlet 动态产生的 XML 文档中的中文问题	73
2.1.1 产生中文乱码的主要原因	73
2.1.2 主要解决方法	78
2.2 实现其他字符集的显示	79
第 3 章 用 Servlet 和 JSP 动态生成 XML 文档	83
3.1 用 Servlet 动态生成 XML 文档	84
3.1.1 使用 Servlet 服务器小程序	84
3.1.2 Servlet 技术和 XML	86
3.1.3 配置使用 Servlet	88
3.1.4 用 Java Servlet 创建简单的 XML 文档	91
3.1.5 用 Servlet 传递参数	94
3.1.6 用 Servlet 处理表单	97
3.1.7 直接用 Java 程序解析 XML 文档	101
3.1.8 关于 DOM 和 SAX	108
3.2 用 JSP 动态生成 XML	110
3.2.1 JSP 概念	110
3.2.2 JSP 与 XML 结合的优势	111
3.2.3 直接将 XML 在 JSP 中输出	112
3.2.4 使用 JavaBean 直接生成 XML	115
第 4 章 实现 XML 与数据库数据的交互	119
4.1 XML 在数据库中的广泛应用	119
4.1.1 XML 在数据库中应用的优点	119
4.1.2 XML 数据源的分类	120
4.1.3 XML 中数据读取的主要技术	120
4.2 现有数据库系统对 XML 的支持	121
4.2.1 主要数据库系统对 XML 的支持	121
4.2.2 SQL2K 对 XML 的支持	123
4.3 如何将 XML 数据导入导出数据库	134
4.3.1 数据库中 XML 的存储	135
4.3.2 通过查询结果生成 XML	135
4.3.3 把 XML 写到表中	142
4.3.4 通过 XML 实现异构数据库之间的数据交换	144

4.3.5	异构数据库基本概念	144
4.3.6	异构数据库传统的访问方法	145
4.3.7	利用 XML 访问异构数据库	147
4.3.8	利用 XML 在异构数据库之间进行数据交换实例	149
第 5 章	在 XML 中实现对数据合法性的判断	163
5.1	概述	163
5.1.1	数据合法性检验	163
5.1.2	规定数据合法性的方法	164
5.1.3	检验数据合法性的工具	165
5.2	使用 DTD 对 XML 数据合法性判断	165
5.2.1	文档类型定义(DTD)	165
5.2.2	使 DTD 与 XML 文档相关联	166
5.2.3	用 DTD 指定元素间的包含关系元素的出现顺序	167
5.2.4	用 DTD 控制元素个数	170
5.2.5	元素声明中的选择	172
5.2.6	元素的嵌套	172
5.2.7	标识 DTD 中的元素	174
5.2.8	元素属性的声明	177
5.2.9	设定元素属性的默认值	178
5.2.10	DTD 综合实例	178
5.3	使用 XDR 对 XML 数据合法性判断	181
5.3.1	Schema 和 XDR	181
5.3.2	初识 XDR 结构	182
5.3.3	在 XDR 中用<datatype>来指定数据类型(datatype)	184
5.3.4	在 XDR 中声明属性	185
5.3.5	在 XDR 中声明元素	186
5.3.6	在 XDR 中声明元素的子元素和属性	190
5.3.7	在 XDR 中对子元素进行分组	192
5.3.8	在 XDR 中加入文档描述	192
5.3.9	XDR 综合实例	194
5.4	使用 XSD 对 XML 数据合法性判断	197
5.4.1	XSD 概念	197
5.4.2	第一个简单的 XSD	197
5.4.3	声明一个元素类型	199
5.4.4	给元素和属性分组	201
5.4.5	扩展自己的数据类型	204

5.4.6	XSD 综合实例	208
第 6 章	用 MS XML Parser 对 XML 进行解析	213
6.1	MS XML Parser 简介	213
6.1.1	MS XML Parser 概念	213
6.1.2	MS XML Parser 工作原理	213
6.1.3	使用 MS XML Parser	214
6.2	用 Visual Basic 6.0 结合 MS XML Parser 进行 DOM 编程	215
6.2.1	XML DOM 初步	215
6.2.2	几个常用接口简介	219
6.2.3	编程实例 1:加载/保存 XML 文档	220
6.2.4	编程实例 2:遍历 XML 文档	223
6.2.5	编程实例 3:修改 XML 文档的数据	225
6.3	用 Visual C++ 6.0 结合 MS Parser 进行 DOM 编程	236
6.3.1	用 Visual C++ 6.0 构建 XML 应用程序	236
6.3.2	分析 XML 基本应用程序	239
6.3.3	编程实例 1:结合 MFC 统计书目的信息	242
6.3.4	编程实例 2:转换 XML 文件结构(1)	254
6.4	用 Visual Basic 6.0 结合 MS XML Parser 进行 SAX 编程	258
6.4.1	SAX 初步	258
6.4.2	SAX 解析事件	260
6.4.3	编程实例 1:用 Visual Basic 构建一个简单的 SAX 程序	261
6.4.4	编程实例 2:转换 XML 文件结构(2)	267
第 7 章	使用 XML 的 Java 解析器	273
7.1	使用 IBM XML4J Apache Xerces	273
7.2	使用 Java 的 DOM 接口处理 XML	274
7.2.1	生成一棵 DOM 树	274
7.2.2	遍历一棵 DOM 树	281
7.2.3	修改 DOM 树中的内容	290
7.2.4	利用 DOM 进行筛选功能	293
7.2.5	利用 DOM 编写特定 XML 的浏览器	298
7.3	使用 Java 的 SAX 接口处理 XML	304
7.3.1	使用 SAX 对 XML 文档进行统计	304
7.3.2	使用 SAX 遍历 XML 文档的内容	310
第 8 章	综合实例:利用 XML 实现计算机监视工具	315
8.1	总体介绍	315
8.1.1	功能分析	315

8.1.2	技术要点·····	316
8.2	总体设计·····	316
8.2.1	整体结构·····	316
8.2.2	详细设计·····	317
8.2.3	传输数据格式·····	318
8.2.4	开发工具、运行环境·····	320
8.3	总控监视程序·····	320
8.3.1	Java Swing 简介·····	320
8.3.2	模拟监控程序·····	322
8.3.3	实际的监控程序·····	332
8.4	Linux 平台下的探测程序·····	342
8.4.1	进行基本信息的采集·····	342
8.4.2	进行状态信息的采集·····	345
8.5	其他信息的采集和 Windows 平台下的探测程序·····	347
第 9 章	综合实例：实现基于 XML 的 Web 论文管理系统·····	349
9.1	总体介绍·····	349
9.1.1	功能分析·····	349
9.1.2	整体结构·····	349
9.1.3	主要技术·····	350
9.1.4	系统特点·····	351
9.1.5	开发工具与环境·····	351
9.2	数据库的设计连接和中文问题的解决·····	352
9.2.1	数据库的设计·····	352
9.2.2	为 Java Servlet 连接数据库·····	355
9.2.3	Web 表单中的汉字输入处理·····	363
9.3	各个页面的显示与处理·····	365
9.3.1	总体构成·····	365
9.3.2	系统主页·····	366
9.3.3	论文提交·····	373
9.3.4	论文数据的显示、修改与删除·····	390
9.3.5	论文的检索·····	392
9.3.6	论文的各项统计资料·····	399
附录 A	CSS 中一些基本属性及其表示·····	403
附录 B	网上资源·····	407

在 XML 中实现个性化显示

XML 是 eXtensible Markup Language 的缩写,意为可扩展的标记语言。它为 Web 页面上的结构化的文档和数据定义了一套通用的格式,它为将结构化的数据写入文本文件提供了一种方法。XML 是一套定义语义标记的规则,这些标记将文档分成许多部件并对这些部件加以标识。

XML 很像超文本标记语言(Hypertext Markup Language,HTML),它也使用标记。但它们的不同点是:HTML 为每个标记和属性规定了固定的意义,它只能描述一定数目的元素,HTML 中没有规定的标记和属性将无法被处理;XML 中的标记仅用于分割数据,而对每个标记的解释完全依赖于分析 XML 文档的应用程序。在 HTML 中标记 `<p>` 表示段落,而在 XML 文件中的标记 `<p>` 很可能就不是段落的意思,它可能表示价格(price),参数(parameter),人(person)……具体的含义要视上下文而定。在某种意义上,可以将 HTML 视为 XML 的一种不规范的应用。

由于 XML 是一种元标记语言,用户可以定义自己需要的标记。因此各种不同领域的工作者可以开发与自己的特定领域有关的标记语言,这就使得该领域中的人们可以方便地交换数据和信息,而不用担心接收端的人是否有特定的软件来读取数据。像化学领域里的 CML 和数学领域里的 MathML 就是自定义标记语言的成功范例。

下面将详细介绍 XML 文件如何以不同的样式在浏览器中显示出来。

1.1 用 CSS 控制 XML 数据展示风格

1.1.1 CSS 项目信息罗列

级联样式表(Cascading Style Sheets,CSS)是在 1996 年作为把有关样式信息如字体和边框加到 HTML 文件中的标准方法而提出来的。CSS 样式表中的每个规则都给出下一个规则所使用的元素的名称,以及此规则要应用哪些元素的样式。从应用上来看,CSS 样式表实际上就是一些项目显示规则信息的罗列。

程序清单 1-1 和程序清单 1-2 分别给出了一个 XML 文件和一个 CSS 文件。该 XML

文件显示了清华大学足球队的信息,用后面的 CSS 文件控制它的显示。

程序清单 1-1 足球队的 XML 文件(1-1.xml)

【\source\chapter1\1-1.xml】

```
<? xml version="1.0" encoding="gb2312" standalone="no"? >
<? xml-stylesheet type="text/css" href="test1.css"? >
<FootballTeams>
  <TeamTotal>2</TeamTotal>
  <Team>
    <TeamName>酒井队</TeamName>
    <Department>计算机系</Department>
    <Captain>
      <CaptainName>张赫</CaptainName>
      <CaptainClass>计 72 班</CaptainClass>
    </Captain>
  </Team>
  <Team>
    <TeamName>精仪队</TeamName>
    <Department>精密仪器系</Department>
    <Captain>
      <CaptainName>赖耿</CaptainName>
      <CaptainClass>精 81 班</CaptainClass>
    </Captain>
  </Team>
</FootballTeams>
```

程序清单 1-2 用于 1-1.xml 的 CSS 样式表(1-2.css)

【\source\chapter1\1-2.css】

```
FootballTeams {display:block}

TeamTotal{}
Team{display:block}

TeamName{}
Department{}
Captain{}

CaptainName{}
CaptainClass{}
```

在浏览器中显示该 XML 文件,效果如图 1-1 所示。

此时页面的显示比较简单,仅仅是 XML 文件中信息的罗列。一般用户在看到页面时,对上面所包含的信息内容中所需要强调的内容没有深刻的印象。在实际应用的时候,

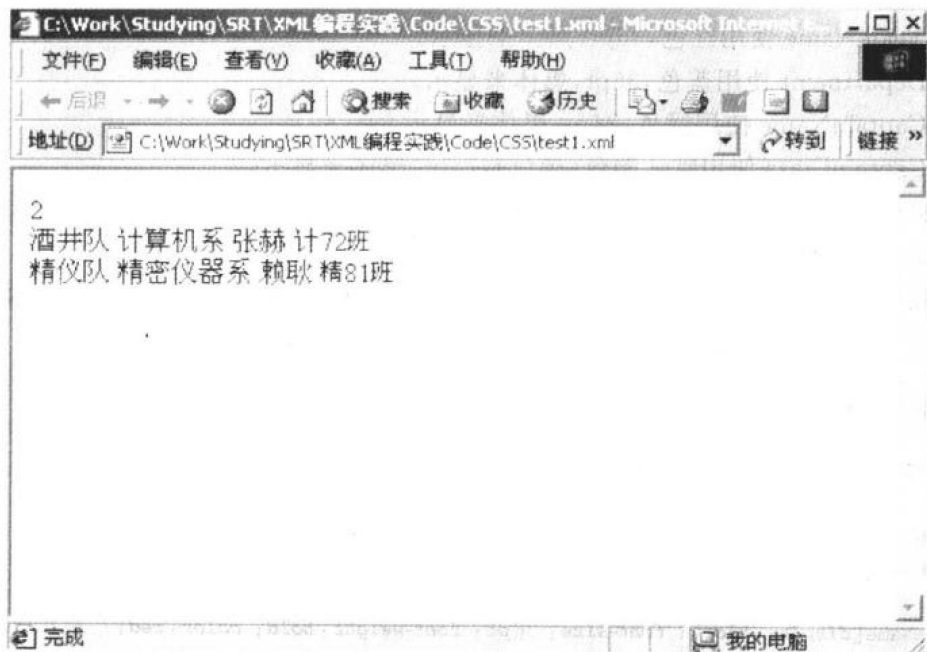


图 1-1 Internet Explorer 5.0 显示的足球队的情况

我们需要对某些信息使用各种方式进行强调,比如加上颜色、改变字体以及加上背景等等。此时就需要对 CSS 样式表做一些改动。

例如试图实现如图 1-2 所示的效果。

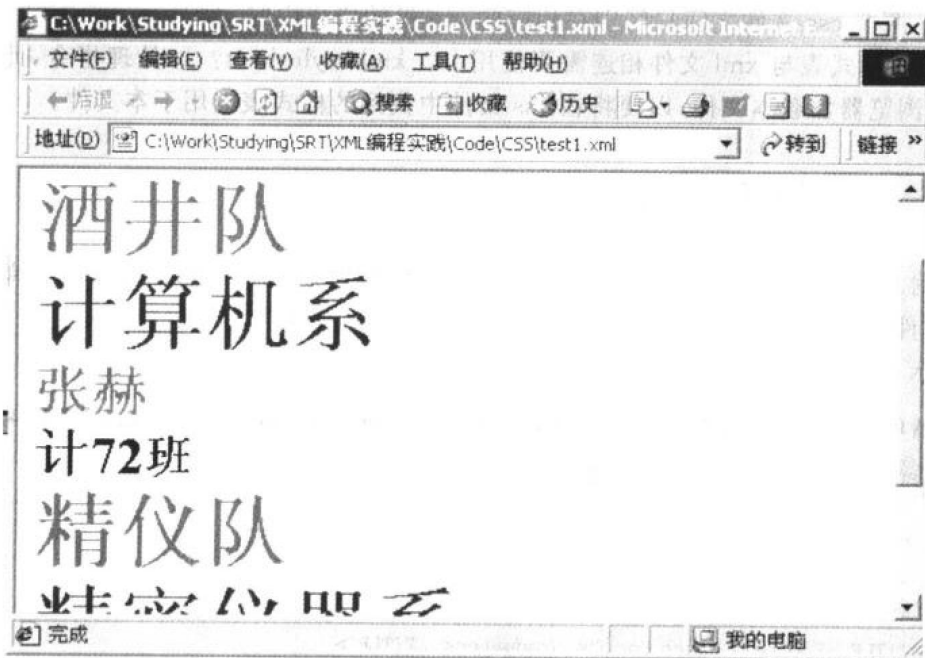


图 1-2 Internet Explorer 5.0 显示的另一种样式表的足球队的情况

仔细观察一下样式的要求,可以发现它们具有如下特征:

- TeamName 使用红色,36pt,黑体来显示。
- Department 使用蓝色,36pt,黑体来显示。
- CaptainName 使用绿色,24pt,黑体来显示。
- CaptainClass 使用默认颜色(黑色),24pt,黑体来显示。

要实现以上的显示效果,需要对 XML 相应 Tag 里面的样式内容进行一些修改,如程序清单 1-3(注意黑体部分,与相应的要求对应)。

按照上面对样式要求分析的结果,依次写出 CSS 样式表如程序清单 1-3 所示。

程序清单 1-3 用于 1-1.xml 的 CSS 样式表(1-3.css)

【\source\chapter1\1-3.css】

```
FootballTeams {display:block}

TeamTotal{display:block; font-size: 36pt; font-weight:bold}
Team{display:block}

TeamName{display:block; font-size: 36pt; font-weight:bold; color:red}
Department{display:block; font-size: 36pt; font-weight:bold; color:blue}
Captain{}

CaptainName{display:block; font-size: 24pt; font-weight:bold; color:green}
CaptainClass{display:block; font-size: 24pt; font-weight:bold;}
```

从以上两个实例可以看出,针对同一份 xml 文件可以使用不同的样式表控制其显示的方式。两个 css 文件之间只区别在相应的 tag 里的样式内容。

把 css 样式表与 xml 文件相连需要使用<? xml-stylesheet? >处理指令,此指令通知 Web 浏览器加载本文件,以便将在 css 文件中找到的样式表应用于本文件。

1.1.2 用 CSS 给文本上妆

在附录中,给出了在 CSS 中可能用到的一些属性值,下面将利用程序清单 1-4 的 XML 文件,依次在本节、1.1.3 节和 1.1.4 节通过实例介绍如何通过 CSS 控制 XML 文件的显示。

程序清单 1-4 七字班毕联足球赛的 XML 文件(为了充分表现出各个属性,使用英文)(1-4.xml)

【\source\chapter1\1-4.xml】

```
<? xml version="1.0"? >
<? xml-stylesheet type="text/css" href="plan.css"? >
<PLAN>
  <TITLE>Football Match For The Graduates</TITLE>

  <TURN>
```