

ICS 35.240.30
L 74



中华人民共和国国家标准

GB/T 18793—2002

信息技术 可扩展置标语言(XML)1.0

Information technology—Extensible Markup Language (XML)1.0

2002-07-18 发布

2002-12-01 实施

中华人民共和国
国家质量监督检验检疫总局 发布

目 次

前言	Ⅲ
引言	Ⅳ
1 范围	1
2 引用标准	1
3 定义	2
4 文件	3
5 逻辑结构	9
6 物理结构	16
7 一致性	23
8 记法	24
9 中文处理	25
附录 A(标准的附录) 字符类别	27
附录 B(提示的附录) XML 和 SGML	29
附录 C(提示的附录) 实体引用和字符引用的展开	29
附录 D(提示的附录) 确定性内容模型	30
附录 E(提示的附录) 字符编码的自动检测	31

前 言

本标准非等效采用 W3C(World Wide Web Consortium)RFC-xml-19980210《可扩展置标语言(XML)1.0》。同时参考了截至 2000 年 7 月 W3C 对 RFC-xml-19980210 的勘误(XML 1.0 Specification Errata,见 <http://www.w3.org/XML/xml-19980210-errata>)。在充分吸收、尊重 RFC-xml-19980210 的基础上,以基于已颁布的国家标准的内容替换了部分国际标准的内容,增加了中文处理的内容。

本标准的附录 A 是标准的附录,附录 B、附录 C、附录 D 和附录 E 是提示的附录。

本标准由中华人民共和国信息产业部提出。

本标准由信息产业部电子工业标准化研究所归口。

本标准起草单位:北京信息工程学院、信息产业部电子技术标准化研究所。

本标准主要起草人:李宁、王凌、高健。

引 言

本标准完整地描述了可扩展置标语言(Extensible markup language,XML),它是标准通用置标语言(Standard generic markup language,SGML)的一个子集。其目的在于使得在 Web 上能以现有超文本置标语言(Hypertext markup language,HTML)的使用方式提供、接收和处理通用的 SGML 成为可能。XML 的设计既考虑了实现的方便性,同时也顾及了与 SGML 和 HTML 的互操作性。

XML 描述了一类称为 XML 文件的数据对象,同时也部分地描述了处理这些数据对象的计算机程序的行为。XML 是一个 SGML 的具体应用或受限形式。从构造来看,XML 文件是合乎规范的 SGML 文件。

XML 文件由称为实体的存储单元组成,实体中包含分析的或不分析的数据。分析的数据由字符组成,其中某些字符形成字符数据,还有一些字符形成置标。置标对文件存储总体布局和逻辑结构的描述进行编码。XML 提供了一种机制,对文件的存储布局和逻辑结构加以约束。

一个称为 XML 处理器的软件模块用来读入 XML 文件,存取其中的内容和结构。这里假定 XML 处理器代表另外一个称为应用程序的模块进行工作。本标准描述了 XML 处理器应有的行为:即它必须怎样读入 XML 数据,以及它必须提供给应用程序什么信息。

中华人民共和国国家标准

信息技术 可扩展置标语言(XML)1.0

GB/T 18793—2002

Information technology— Extensible Markup Language(XML)1.0

1 范围

本标准完整地描述了可扩展置标语言(XML)。XML 是 SGML 的一个子集。XML 的设计目的如下:

- a) XML 应在因特网上直接使用;
- b) XML 支持的应用应十分广泛;
- c) XML 应与 SGML 兼容;
- d) 应易于编写处理 XML 文件的程序;
- e) XML 中可选特性的数目应降至最低,最好没有;
- f) XML 文件应是可读的并适度清晰为宜;
- g) 应能快速进行 XML 设计;
- h) XML 的设计应是形式化的和简明的;
- i) XML 文件应易于创建;
- j) XML 置标的简洁性的重要程度最低。

2 引用标准

下列标准中的条文,通过在本标准中引用而成为本标准的条文。本标准出版时,所示版本均为有效。所有标准都会被修订,使用本标准的各方应探讨使用下列标准最新版本的可能性。

GB/T 1988—1998 信息技术 信息交换用七位编码字符集(eqv ISO 646:1991)

GB/T 2311—2000 信息处理 字符代码结构与扩充技术(idt ISO/IEC 2022:1994)

GB 2312—1980 信息交换用汉字编码字符集 基本集

GB/T 2659—2000 世界各国和地区名称代码(eqv ISO 3166-1:1997)

GB/T 4880—1991 语种名称代码(eqv ISO 639:1988)

GB 13000.1—1993 信息技术 通用多八位编码字符集(UCS) 第一部分:体系结构与基本多文种平面(idt ISO/IEC 10646-1:1993)

GB/T 14814—1993 信息技术 文本和办公系统 标准通用置标语言(SGML)(idt ISO 8879:1986)

GB/T 15273.1—1994 信息技术 八位单字节编码图形字符集 第1部分:拉丁字母一(idt ISO 8859-1:1987)

GB/T 15273.2—1995 信息技术 八位单字节编码图形字符集 第2部分:拉丁字母二(idt ISO 8859-2:1987)

GB/T 16965—1997 信息技术 超媒体/ 时基结构化语言(HyTime)(idt ISO/IEC 10744:1992)

GB 18030—2000 信息技术 信息交换用汉字编码字符集 基本集的扩充