

<<Unicode 5.0标准>>

图书基本信息

书名：<<Unicode 5.0标准>>

13位ISBN编号：9787302212447

10位ISBN编号：7302212449

出版时间：2010-1

出版时间：清华大学出版社

作者：Unicode协会

页数：1358

字数：2057000

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<Unicode 5.0标准>>

前言

本书和Unicode字符数据库是Unicode字符编码标准5.0版本的权威来源。

5.0版本和以前的版本有很大不同。

它对支持Unicode的需求表述得更清楚，并且它提供了更清晰的编程指导以满足新技术和新兴市场快速增长的需要，同时满足用户对安全、健壮软件的需求。

购买本书的意义 Unicode标准5.0版本更小更方便，并且包含更多的文本内容。

最重要的是，该书包括所有的Unicode标准附件(Unicode Standard Annexes)，提供了像文本标准化、双向算法和标识符分解等重要进程的说明。

5.0版本包含世界范围内多年来积累的经验知识并且对此进行了加强：本书吸纳了15年来用户的反馈，向用户提供了关于Unicode问题的详细解释，并且更容易入门——因为本书改进了许多图表，并对内容进行了修订。

五分之四的图是新的。

三分之二的定义是新的。

二分之一的Unicode标准附件是新的。

三分之一的一致性子句是新的。

四分之一的表格是新的。

另外，本书反映了计算机书写系统的新发展。

它极大地改进了对印度文翻译的描述，以满足市场的需求——印度政府支持基于Unicode的编程，本书解释了如何构建它们。

5.0版本还支持最近发布的字符核心CJK子集，IICore，在东亚市场上它对于翻译和互操作是非常重要的。

简言之，本书能够使开发者为世界范围内的软件用户快速实现最新的先进技术，同时把握高速增长的市场需求。

从版本3.0，4.0到5.0的改变是非常重要的——这是一本所有的Unicode开发者都需要拥有的书。

升级到5.0版本的必要性 Unicode标准5.0版本对3.0，4.0版本作了很大改变。

业界已经注意到并且很快迁移到5.0版本——Windows Vista就是在5.0版本下运行的；ICU，Google和Yahoo也已经计划升级到5.0版本。

Internet和W3C协议都建立在Unicode之上，并且继续保持适应最新的版本。

国际标准ISO/IEC10646也是和5.0版本同步的。

最新版本的Unicode标准是Unicode安全机制、Unicode校勘算法、通用区域数据仓库(Common Locale Data Repository)提供的区域数据的基础，并支持正则表达式中的Unicode。

改进的Unicode编码模型使得程序员能够更清楚地知道如何在UTF-8中支持Unicode文本的表示和其他编码方式。

字符属性已经被系统化并更好地帮助开发者进行文本处理。

标准还为Casefolding和标识符的稳定性制定了规则，对于互操作性和形式语言使用的后向兼容，以及在其他上下文环境中标识符精确使用和匹配都是非常重要的。

为了满足现代信息技术的需求，5.0版本提供了稳定、实用的字符处理模型。

Unicode提供了如下功能： 与中国GB18030和HKSCS标准的前后兼容性。

对最新建立的核心CJK字符子集、IICore的说明。

改进了大小写转换(casing)和双向行为以满足业界的需求。

改进了印度文的翻译指导。

更好地处理合成字符、Unicode字符串、变量选择、断行和分段。

如果程序员想要和业界发展同步，利用稳定的安全性，和最新校勘及区域数据定义一致，最重要的是拓展市场，那么要尽快将标准升级到5.0版本。

<<Unicode 5.0标准>>

内容概要

在计算机发展的初期,由于各个计算机系统都是相对独立诞生的,因此曾出现过多达数百种编码系统,但每种编码系统都不能包含全部的字符,并且系统相互之间并不兼容,给信息交换带来了麻烦和安全隐患。

于是,很多国家都制定了计算机信息交换用的字符编码集,如美国的扩展ASCII码、中国的GB2312-80、日本的JIS等,作为相应国家信息处理的基础,起着统一编码的重要作用。

但随着Web的广泛应用,尤其是近年来支持多语言(如中、英、日)的应用已成为大势所趋。这些国家标准字符集编码也出现了范围重叠、相互间的信息交换比较困难的情况,造成软件各个本地化版本的维护成本较高,基于不同环境的程序编写越来越复杂。

Unicode有效解决了这些问题,它几乎包括所有的(超过十万个)语言字符。基于Unicode编码标准的信息和程序能够在不同平台和环境下高效地进行交换和运行,极大简化了应用的开发过程。

为此,Windows早在90年代就已经将Unicode作为自身的编码标准,XML和Python、PERL、Mac OS和Linux等编程语言也都提供了对Unicode编码的支持。

目前,国内关于Unicode的书籍非常少,为了满足读者的迫切需求,清华大学出版社引进了这本由致力于Unicode编码系统研究、推广和应用的国际权威组织——The Unicode Consortium(Unicode协会)组织编写的经典图书。

为了保持原书的系统性和规范性,尽可能为读者提供标准的第一手资料,我们对该书的中文版译本做了一些调整,主要体现在以下几个方面: 第1~6章是完整的中文译本。

前5章主要是关于Unicode标准的基本概念、原理、一致性、字符属性和实现方法等内容,第6章主要是世界范围内使用的书写系统和标点的相关介绍。

第7~16章采用英文影印的方式,最大化地保留了原书的内容,但为了方便读者查找相关内容,我们加了中文导读信息。

这部分主要是关于各种语言文字及符号的基本介绍,如拉丁字符、中东字符、南亚字符、东亚字符、新增的现代字符、各种符号、特殊区域和格式字符等基本信息。

第17章是完整的Unicode编码表,这部分内容按照不同字符块的分区,以表格的形式体现出来,以便于读者查阅。

第18章是汉字偏旁-笔画索引,因这部分内容是我们日常使用最广泛的信息,故完全采用中文译本。

附录A~F、术语表、参考文献、Unicode名称索引以及Unicode标准附件全部采用英文影印方式,以便于读者查阅最原始的资料、文献和技术标准。

<<Unicode 5.0标准>>

书籍目录

第1章 绪论	1.1 覆盖范围	1.1.1 Unicode标准覆盖面	1.1.2 新字符	1.2 设计目标	1.3
文本处理	第2章 总结构	2.1 前后结构关系	2.1.1 基本文本处理过程	2.1.2 文本要素、	
字符和文本处理过程	2.1.3 文本处理和编码	2.2 Unicode设计原则	2.2.1 通用性	2.2.2 逻辑顺序	
2.2.2 有效性	2.2.3 字符, 而非字形	2.2.4 语义	2.2.5 纯文本	2.2.6 逻辑顺序	
2.2.7 一致性	2.2.8 动态合成	2.2.9 稳定性	2.2.10 可转换性	2.3 兼容性字符	
2.3.1 兼容性变量	2.3.2 兼容性可分解字符	2.3.3 映射兼容性字符	2.4 代码点和字符		
2.5 编码格式	2.5.1 UTF-32	2.5.2 UTF-16	2.5.3 UTF-8	2.5.4 UTF-32、UTF-16	
和UTF-8优点的比较	2.6 编码方案	2.7 Unicode字符串	2.8 Unicode分配	2.8.1 平面	
2.8.2 分配区域和字符代码块	2.8.3 代码点的分配	2.9 字符分配	2.9.1 Plane 0 (BMP)		
2.9.2 Plane 1	2.9.3 Plane 2	2.9.4 其他平面	2.10 书写方向	2.11 合成字符	
2.11.1 基本字符序列和读音符号	2.11.2 多重合成字符	2.11.3 连字式多重基本字符			
2.11.4 隔离中的非间隔标记	2.11.5 “字符”和字形串	2.12 等价序列和规范化	2.13 特殊		
字符和非字符	2.13.1 特殊非字符代码点	2.13.2 字节顺序标记 (BOM)	2.13.3 布局和		
格式控制字符	2.13.4 替换字符	2.13.5 控制码	2.14 Unicode标准的一致性	2.14.1 一	
致性实现的特征	2.14.2 不可接受的行为	2.14.3 可接受的行为	2.14.4 支持的子集	第3	
章 一致性	第4章 字符属性	第5章 实现原则	第6章 书写系统与标点符号	第7章 欧洲字符	第8章
中东字符	第9章 南亚字符-I	第10章 南亚字符-	第11章 东南亚字符	第12章 东亚字符	第13章
附加的现代字符	第14章 古代字符	第15章 符号	第16章 特殊区域和格式字符	第17章 编码表	
第18章 汉字偏旁-笔画索引	参考文献	名家书评			

<<Unicode 5.0标准>>

章节摘录

字母和音节。

字母的概念多用于上下文中。

计算机语言标准通常对标识符进行字符化表示，例如字母、音节、表意文字和数字，但并没有精确指出什么字母、音节、表意文字或者数字，也没有明确定义按照字符编码标准，还是本地规范。

Unicode标准意味着该标准包含了许多书写系统，每种书写系统的区别并不是自身言明的，可能是由于以前设计这些书写系统的主要目的是为西欧语言和日语服务的。

特别是，尽管Unicode标准包含各种字母表和音节表，以及书写系统，这些书写系统是基于前两者产生的。

因此，在字母和音节之间没有明显的属性区别。

字母文字。

字母文字属性是字母表和 / 或音节表的一种重要属性，不论是合成字母还是非合成字母。

这类字符都是复杂字符，它们是合成字符序列的规范等价字符。

该合成字符序列包括：由字母表中的基本字符加上一个或者多个合成字符、双字母组成的合成字符、与上下文相关的字母字符变异体、多字母构成的合成字符、与上下文相关的合成字符变异体、修饰字母、与单个字母兼容等价的类似字母的字符和杂类字母元素组成。

注意，U+00AAFEMININE ORDINAL INDICATOR和U+00BAMSSCULINE ORDINAL INDICATOR都是包括一个拉丁字母的缩写形式，应当认为是字母而不是非字母字符。

表意属性。

表意属性是Unicode字符数据库中定义的信息性属性。

例如，表意属性用于表示断行操作。

具有表意属性的字符包括统一CJK表意字符，CJK兼容表意字符和其他部分的字符。

例如，U+3007表意数字零和U+3006表意封闭符号。

关于汉语表意字符的更多信息，参见12.1节。

更多表意字符和意音文字（logosyllabary）的相关信息可参见6.1节。

<<Unicode 5.0标准>>

媒体关注与评论

多年来，Unicode标准的复制版本已经成为我的个人藏书中最重要和使用频率最高的一本书。

——高德纳，美国著名计算机科学家，图灵奖获得者，被誉为现代计算机科学的鼻祖 十几年来，Unicode已经成为许多微软产品和技术的基础。

Unicode标准5.0版本将给客户提供更多新的功能。

——比尔·盖茨，微软公司创始人，前任董事长兼首席执行官 W3C使网络上的文本真正全球化，而W3C遵从的正是Unicode标准。

——蒂姆·伯纳斯-李，Web发明者和万维网联盟（W3C）主席 没有Unicode，Java就不能称之为Java，互联网也不能很好地连接全世界的人们。

——詹姆斯·戈士林，Java之父，SUN公司全球副总裁

<<Unicode 5.0标准>>

编辑推荐

以上权威人士和其他软件领域的专家都意识到，Unicode已经成为一个支撑与日俱增的全球化市场必不可少的工具。

作为一个表示全世界所有字母的标准的综合系统，Unicode是现代编程技术（Windows、XML、Python、PERL、Mac OS和Linux）和当今操作中每种搜索引擎及浏览器的基础。

Unicode最新版的官方参考手册已经在最新的Unicode标准修订版文档中有了几百页的更新。其中包括对文本、图表、定义和一致性语句的修订，并为常见问题提供了清晰和正确的答案。

《Unicode 5.0标准》首次包含了Unicode标准附件，附件中说明了一些诸如文本标准化和标识符分解等重要的处理过程。

这些改进非常重要，所以Unicode标准5.0版本是微软公司最新的Windows 7操作系统的基础，同时该版本也是Google、Yahoo和ICU升级方案的一部分。

Unicode标准5.0版本与以前版本的不同之处：
Unicode安全机制的稳固基础 Unicode校勘
算法和通用区域数据库（CLDP）的属性数据 改进的UTF-8 Unicode编码模式 严密稳定的大小
写重叠和标识符间更好的互用性及后台兼容性——能够利用新的方法优化代码 更可靠的文本处理
系统框架——涵盖合成字符、Unicode字符串、断行和分段等

<<Unicode 5.0标准>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>