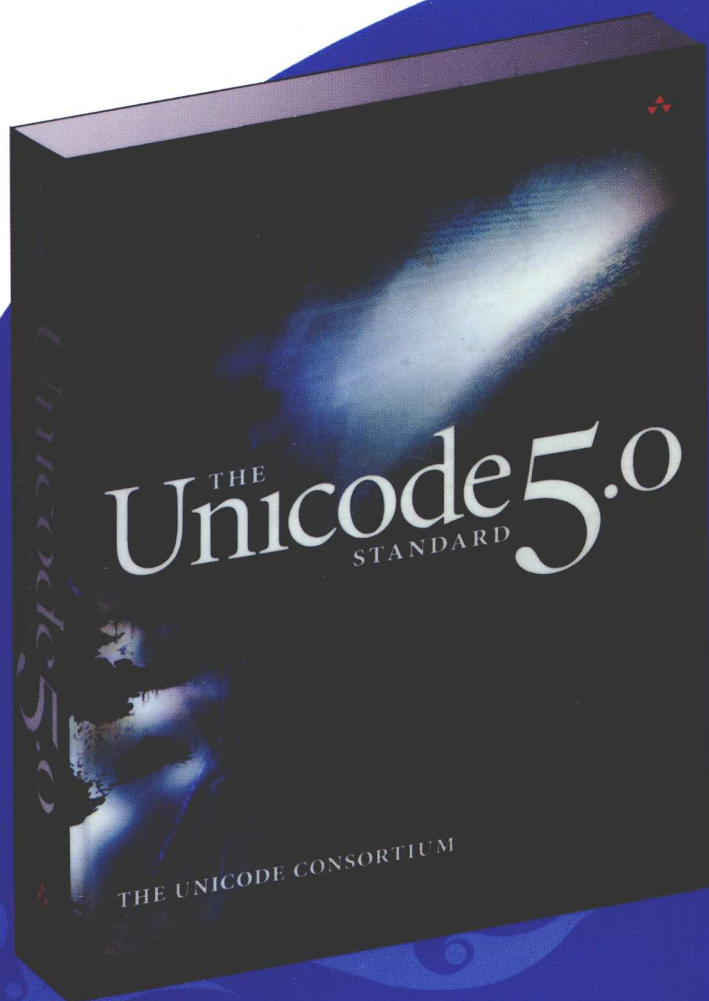


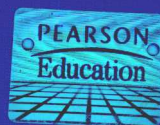


The Unicode Standard, Version 5.0

Unicode 5.0标准



Unicode协会 著
孙伟峰 李德龙 译



清华大学出版社

Unicode 5.0 标准

Unicode 协会 著

孙伟峰 李德龙 译

清华大学出版社

北 京

Authorized translation from the English language edition, entitled The Unicode Standard, Version 5.0, 0321480910 by The Unicode Consortium, published by Pearson Education, Inc, publishing as Addison-Wesley, Copyright © 2007.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

CHINESE SIMPLIFIED language edition published by PEARSON EDUCATION ASIA LTD., and TSINGHUA UNIVERSITY PRESS Copyright © 2010.

北京市版权局著作权合同登记号 图字: 01-2007-2721

本书封面贴有 Pearson Education(培生教育出版集团)防伪标签, 无标签者不得销售。

版权所有, 侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

Unicode.5.0 标准/Unicode 协会 著; 孙伟峰, 李德龙 译.—北京: 清华大学出版社, 2010.1

书名原文: The Unicode Standard, Version 5.0

ISBN 978-7-302-21244-7

I. U… II. ① U… ② 孙… ③ 李… III. 符号—编码—国际标准 IV. TP391.1-65

中国版本图书馆 CIP 数据核字(2009)第 180688 号

责任编辑: 王 军 崔 伟

装帧设计: 孔祥丰

责任校对: 胡雁翎

责任印制: 王秀菊

出版发行: 清华大学出版社

<http://www.tup.com.cn>

社 总 机: 010-62770175

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者: 清华大学印刷厂

装 订 者: 北京市密云县京文制本装订厂

经 销: 全国新华书店

开 本: 185×250 印 张: 86.75 字 数: 2057 千字

附光盘 1 张

版 次: 2010 年 1 月第 1 版 印 次: 2010 年 1 月第 1 次印刷

定 价: 198.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题, 请与清华大学出版社出版部联系调换。联系电话: (010)62770177 转 3103 产品编号: 025905-01

出版说明

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 章是汉字偏旁一笔画索引,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。

我们以这种方式出版国外的经典著作,也是希望能够更加便于读者学习和参考,如果大家有更好的建议,不妨赐教至读者邮箱 wkservice@vip.163.com,我们不胜感激。

清华大学出版社

2009 年 12 月

序

在过去的十年中，没有过多的宣扬，Unicode 已经完全改变了软件和通信的基础。无论何时您在计算机上进行读写操作，都要使用 Unicode。无论何时您在 Google、Yahoo、MSN、Wikipedia 或其他 Web 站点上搜索信息，也要用到 Unicode。Unicode 5.0 已成为一个重要的里程碑，世界上任何地方的人们通过它都可以在计算机上使用自己的语言。

我们开始使用 Unicode 时只有一个简单的目标：统一几百种不同的字符编码方法，替换成一个全球通用的标准。我们现在使用的编码字符既不完整又不一致：两种编码方式可能对不同的字符使用相同的内码，而对相同的字符使用不同的内码；没有编码可以表示一些小语种。无论何时文本数据在不同的程序或平台间转换，都存在被破坏的风险。过去程序被硬编码以支持特定的编码，而开发国际通用版本的代价很大，测试和资金都是不小的问题。最终，产品投放到国外市场也是成本昂贵且滞后的。这种状况使得公司及其客户都不满意。对发展中国家尤其影响很大，支持小型市场也不是切实可行的。诸如数学这样的技术领域也会处于不利地位；这迫使人们使用特殊的字体来表示任意字符，但当这些字体无效时，内容就会变得混乱。

Unicode 从根本上改变了这一状况。如今，对于所有的文本，程序仅仅需要使用唯一的表示法就可以实现对世界上所有语言的编码。使用所有从代码中分离出来的可译材料，以唯一的表示法很容易对程序进行构建。这种表示法也为多种语言的快速普及提供了基础。因此，一个程序的多语言版本可以以很低的投入迅速得以开发，即使是 Microsoft Office 和 OpenOffice 这样复杂的程序。

字符的分配仅仅是 Unicode 标准及其相关规范中很小的一部分。他们通过扩展性的描述和大量的数据来说明如何使用字符：怎样构词和断行，怎样在不同的语言中排序文本；怎样设置数字、日期、时间及其他不同语言元素的格式；怎样显示那些书写形式是从右到左的语言(如阿拉伯文和希伯来文)，以及那些书写形式是分离的、合成的、重新排序的语言(例如南亚语言)；怎样安全地处理世界上字母表中许多外形相近的字符。没有 Unicode 标准及其辅助说明中的属性、算法和其他规范，不同工具间的互通是不可能的。

随着网络的快速发展，文字显示的唯一性对于全球通信来说变得至关重要。于是，HTML 和 XML 的文本内容都按照 Unicode 进行定义——每个程序处理 XML 必须使用 Unicode 标准。搜索引擎使用 Unicode 标准也有许多的好处。即使用旧的字符编码制作的网页，搜索引擎索引网页唯一最有效的方法也是把这些字符转换成 Unicode 标准，这样网页上所有文本就都用相同的程序代码存储、搜索和匹配。由于所有的搜索引擎都将网页转换成了 Unicode 标准，最可靠的搜索网页的方法就是把所有字符都用 Unicode 标准表示出来。

Unicode 5.0 版本取代和废除了前几个版本的所有标准。本书虽然篇幅有限，价格不高，但却提供了大量的新资料 and 许多新修订的资料。Unicode 当然也有瑕疵。本书将通过其发展历程帮助您认识 Unicode 标准。更重要的是，还将帮助您理解将来可能发生改变的一些特性，使得您能适当地优化实现过程。您也可以在 Unicode 的站点上找到大量的其他信息。如果您对参与 Unicode 未来开发的讨论感兴趣，并且想更进一步了解 Unicode 标准正在进行的工作，可以在 Unicode 的站点找到加入 Unicode 协会的相关信息。

您正在学习的内容是全世界的专家多年经验积累的结果。我相信您将会发现本书对您是非常有益的。

Unicode 协会主席
Mark Davis 博士

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。

详细的变化信息。从附录 D 中可以找到关于 Unicode 标准以前版本变化的详细信息, 包括字符计数、稳定性保证、Unicode 字符数据库更新和 Unicode 标准附件等。

Unicode 标准 5.0 版本和 ISO/IEC 10646:2003 相一致, 并且增加了修正方案 1 和 2, 同时修正方案 3 中增加了 4 种字符以支持信德语。

本书的组织结构

本书和 Unicode 字符数据库定义了 Unicode 标准 5.0 版本。本书提供了总则、一致性需求、程序员指导方针、字符编码表和名称, 以及 Unicode 标准附件。

概念、体系结构、一致性和指导方针。Unicode 标准 5.0 版本的前 5 章介绍了 Unicode 标准, 提供了实现一致性的基本信息, 并且描述了基本的文本处理、合成标记和编码方式。第 5 章则主要回答了许多实现 Unicode 时出现的问题。

第 1 章介绍了 Unicode 标准的基本概念、设计基础, 并且全面讨论了文本处理的基本需求。

第 2 章介绍 Unicode 标准的基本原理, 并且介绍了一些特定的主题, 例如文本处理、整个字符属性和合成标记的使用等。

第 3 章是关于一致性的描述。该章还为三种处理过程提供了标准化的算法, 这三种处理过程分别为: 合成标记的规范排序、连接 jamo 的韩语音节编码和默认大小写算法。

第 4 章详细描述了字符属性, 包括字符的标准化和信息化。Unicode 字符数据库中提供了字符属性的其他信息。

第 5 章讨论了实现的问题, 包括压缩、处理未知和不支持的字符策略, 以及转换到其他标准的策略。

字符块描述。第 6~16 章是关于字符块描述的内容, 这些章节提供了关于每种文字或者符号组的基本信息, 并且还讨论了特殊的字符及相关的设计信息。其中一些信息是生成文

字和其他字符集的一致性实现所必需的。

编码表。第 17 章提供了编码表和字符名称列表。编码表包括标准的字符编码分配，名称列表包括标准信息、有用的参考文献和注释等。

汉字偏旁-笔画索引。第 18 章为 CJK 表意文字的 IICore 子集提供了汉字偏旁-笔画索引。这个索引的目的是便于查找常见的表意文字在 Unicode 标准中的编码。

附录。附录详细介绍了 Unicode 标准的历史背景和它与 ISO/IEC 10646 的关系。

附录 A 说明了标准中用到的一些符号的具体规定。

附录 B 提供了 Unicode 技术报告的摘要，罗列了重要的 Unicode 资源。

附录 C 详细介绍了 Unicode 标准和 ISO/IEC 10646 的关系。

附录 D 罗列了自 4.0 版本以来 Unicode 标准的版本发展历程。

附录 E 描述了在 Unicode 标准中的统一汉字字符集的历史。

附录 F 生成了 Unicode 协会关于字符编码稳定性的策略。

术语表、参考文献和索引。在附录之后是术语表、参考文献和 Unicode 字符名称索引。

Unicode 标准附件

Unicode 标准附件在本书的最后。这些附件构成了 Unicode 标准的一部分。Unicode 标准版本的一致性包括附件的一致性。

Unicode 标准附件#9 描述了诸如阿拉伯语和希伯来语这样混合方向文本的字符定位说明。

Unicode 标准附件#11 提供了 Unicode 字符属性的说明，当和东亚字符进行互操作时是很有用的。

Unicode 标准附件#14 提供了 Unicode 字符断行属性的说明。

Unicode 标准附件#15 描述了四种 Unicode 文本的标准化格式。

Unicode 标准附件#24 给所有的 Unicode 字符代码点分配了文字名称。

Unicode 标准附件#29 描述了确定文本元素之间默认边界的策略：字形串、字和句子。

Unicode 标准附件#31 描述了在 Unicode 标识符定义和句法结构中推荐使用的默认规范。

Unicode 标准附件#34 定义了 Unicode 名称字符序列概念和一系列规则，这些规则强制性地可将可用名字应用到字符序列里。

Unicode 标准附件#41 包括其他 Unicode 标准附件使用的参考文献。

任何一个 UAX 的 5.0.0 版本在 CD-ROM 中都能找到。所有版本包括最新的 Unicode 标准附件版本，都能在 Unicode 的 Web 站点上找到：<http://www.unicode.org/reports/>。

Unicode 字符数据库

Unicode 字符数据库(Unicode Character Database, UCD)是数据文件的集合，它包括字符编码、字符名称和字符属性数据(4.1 节有更加详细的介绍)。所有版本包括最新的 Unicode 字符数据库版本，在 Unicode 的 Web 站点上都有介绍：<http://www.unicode.org/ucd/>。

随书赠送的光盘中包含 5.0.0 版本的所有 Unicode 字符数据库文件。

下面站点中有全部的 Unicode 标准版本信息：<http://www.unicode.org/versions/>。

Unicode 技术标准和 Unicode 技术报告

Unicode 技术标准和 Unicode 技术报告是不同的，并不作为 Unicode 标准的一部分。

所有版本的 Unicode 技术标准和 Unicode 技术报告能够在下面的 Web 站点上得到：
<http://www.unicode.org/reports/>。

随书赠送的光盘中也都包含每个最新版本的文档(参见附录 B)。

光盘中的资料

光盘中还包括额外的信息，例如示例代码，这些信息可以在下面的 FTP 网站中找到：
<ftp.unicode.org>。HTTP 网址为：<http://www.unicode.org/Public/>。

对于光盘中的详细内容，请参见 `readme.txt` 文档。

更新资料和勘误表

Unicode 标准中包括 Unicode 字符数据库和 Unicode 标准附件中的错误报告，它是通过在线报告形式进行更正的，具体网址为：<http://www.unicode.org/reporting.html>。

Unicode 的 Web 站点上也能找到已知的勘误表：<http://www.unicode.org/errata/>。

今后的版本都会陆续将当前修订的勘误表包含进去。

致 谢

Unicode 标准 5.0 版本是许多人多年不断奉献的结果。感谢以下作者，他们为本书的设计、创作和评审做了很大贡献。

Julie D.Allen 本书的主编。作为高级编辑和项目负责人，她重写了许多文字并且规划了完成该书的时间进度。Julie 同时也负责术语表的更新，以及与发行商、艺术家和其他人员的合作。

Joe Becker 创作了最初的 Unicode 简介，并且还担任本书的编辑。

Richard Cook 修改和升级了统一汉字数据库及其文档，同时他还是 Unicode 协会的代表。

Mark Davis 对 5.0 版本的发展做出了重要贡献，领导了 Unicode 标准设计的多个方面。他在修订和改进上做了很多努力：一致性的综述、封装行为、标识符的稳定性、文本的界限、双向书写行为、执行指导方针、标准化和 Unicode 字符数据库属性的附加等等。Mark 是 Unicode 标准附件的三位作者之一，同时在制定 Unicode 安全机制方面也有突出的贡献。

Michael Everson 将少数历史文字增加到 5.0 版本中，是文字描述的主要贡献者。这些文字包括巴厘文、埃及古文、古斯拉夫文、新泰文、古波斯文、腓尼基文和苏美尔-阿卡德楔形文字等。他提供了这些语言在标准中的字体和编码表、字符名称以及注释等。

Asmus Fretag 对内容组织和章节特色做出了突出贡献，并且将他的研究重点放到符号上。他负责标点、符号、特殊领域和形式字符的升级，还对欧洲字母、双向书写、断行等方面有突出贡献。他设计了许多附加图表，提出了许多改进建议。Asmus 在定义 Unicode 字符属性模型上做了很大努力，是三位 Unicode 标准附件的作者之一，并且很乐意将附件加入该书中。他还创建了自定义格式化软件、negotiate 字体和编码表。

John H.Jenkins 作为 Unicode 协会的代表，他致力于统一汉字数据库的扩展，将古汉语基本笔画索引扩展到标准的表意的内容，并且准备了 IICore 汉语子集的基本笔画索引。John 还负责组织古汉语列表，为康熙及 CJK 的字提供字体。

Mike Ksar 作为 JTC1/SC2/WG2 和 SC2 联络的会议召集人，努力使 5.0 版本和 ISO/IEC 10646: 2003 修正版本 1 和 2 同步。他负责中东文字和附录 C，并检查了所有新增的文字。

Rick McGowan 负责协调编码新文字的工作，对许多新文字的编辑工作都有贡献。他修改并且制作了书中的 100 多张图，而且还负责了光盘的制作等。

Lisa Moore 作为 Unicode 技术委员会的主席，审校了整个 5.0 版本的内容。她编辑了迦娄士梯文，重写了附录 D 和许多字体格式，对本书的编辑也做了突出贡献。

Eric Muller 检查了书中的所有章节，对本书内容的清晰和一致性做了很多改进。他对古汉语数据的验证做出了贡献，并且提供了重要的 PDF 技术。

Markus Scherer 检查了 5.0 版本的章节结构和一致性, 并且为标准中的指导方针做出重要更新, 还对所有权作了验证。

Michel Suignard 作为 10646 的项目编辑, 是 Unicode 和 ISO/IEC 10646 两种标准的负责人。他负责编辑 ISO/IEC 10646:2003、修正案 1 和 2, 并且提供 Unicode 版本 4.1 和 5.0 的无缝接合。他将 IRG 资源添加到古汉语里, 并且还是 Unicode 安全机制的主要制定者。

Ken Whistler 是 Unicode 5.0 版本的重要编辑者。他付出了很多努力使本书的内容更精悍, 但是还包括了扩展的文字和所有 Unicode 标准附件。他负责所有的出版和校对工作。Ken 升级了 Unicode 字符数据库, 增加了所有的新字符和它们的属性。他还维护了字符名称列表, 并提供了更多的注释。Ken 还负责了结构和一致性的重写, 使合成类、合成标记的应用与子句及定义的一致性关联。

字体对该书的出版非常重要。在 Michael Everson 的支持下, Asmus Freytag 组织了字体收集工作, 进一步改进 John Jenkins 收集的 Unicode 2.0 版本的字体。除了以上对本书做出贡献的个人、组织和公司, 在字体方面给予支持的还有以下人员: Patrick Andries (Tifinagh), Cora Chang (Braille), Oliver Corff (Yi), Anton Dumbadze and Irakli Garibashvili (Georgian), Andrew Glass (Kharoshthi), Yannis Haralambous (Greek, Syriac, and Thai), George Kiraz (Syriac), Svante Lagman (Runic), Raymond Mercier (Greek zero), Stephen Morey and Michael Everson (Tai Le), Paul Nelson and Sarmad Hussain (Syriac and Sindhi/Urdu numbers), David Perry and James Kass (Greek musical symbols), Peter Martin (Phonetic Additions), Hector Santos (Philippine scripts), Yayasan Bali Simbar (Balinese), Ngakham Southichack (Lao), Michael Stone (Armenian), Steve Tinney and Michael Everson (Cuneiform), Dirk VanDamme (Coptic), Al Webster (Cherokee), Andrew West (Phags-pa), and K. Yarang, J. R. Pandhak, Y. Lawoti, and Y. P. Yakwa (Limbu)。

Michael Everson 为加拿大音节、许多历史文字(包括伽娄士梯文字、B 类线形文字、欧甘碑文和古波斯文等)、符号和拉丁文、希腊文和古代斯拉夫字符提供了字体。John M. Fiscella 设计了符号和许多字母的字体。平壤信息中心的 Yang Song Jin 提供了 CJK 兼容性符号。Thomas Milo 设计了阿拉伯字体。SIL 提供了 Jonathan Kew、Peter Martin 和 Victor Gaultney 字体的设计。CJK 扩展 A 和 B 的字体是由北京中易(郑码)信息技术有限公司提供的。A 扩展是由技术主管 Zhenglong 和 Hua Weicang 设计的, Asmus Freytag 创造了许多特殊字符或者符号的表意文字。

对字体的注释是由以下人员完成的: Heidi Jenkins, Dr. Virach Sornlertlamvanich(Thai), Dr. Sarmad Hussain(Urdu), Roozbeh Pournader(Farsi), Barbara Beeton and Patrick Ion (Mathematical Symbols)及其他人。许多个人和组织提供了 5.0 版本中使用的字体。

增加的新图使得本书内容更加生动。Heidi Jenkins, Dr. Virach Sornlertlamvanich(Thai), Dr. Sarmad Hussain(Urdu), Roozbeh Pournader (Farsi), Barbara Beeton and Patrick Ion (Mathematical Symbols)为希腊编辑符号画了大量的图片。Parisian Schola Cantorum 和 Hymns of Faith 是阿拉伯音乐章节的资料来源。图 10-5 是根据 Glass (2000)改编的。

Steve Mehallo 设计了本书的封面。他还更新了章节的艺术性并且为 5.0 版本做了许多额外的艺术工作。Kamal Mansour 进行了图形设计并且保持本书的封面设计与内容的一致

性。单面版增加了封面设计、光盘设计和章节分割艺术工作的成本。

如果没有 Unicode 公司办公室工作人员和 Mike Kernaghan(Unicode 办公室主任)的帮助,本书是不可能完成的。感谢 Magda Danish,他曾无数次帮助我们:包括协助编辑 Unicode 标准附件,验证汉语数据库的拼音,对技术参考文献的增补和更正,感谢 Sarasvati 设计了邮件表,特别感谢微软公司为我们提供了办公地点。

文本、编码表和数据都是经过专家们认真修订的。编委会感谢以下专家的贡献和对特殊文字的反馈: Barbara Beeton (mathematical symbols), Peter Constable (New Tai Lue and phonetic extensions), Roozbeh Pournader (Arabic), Lorna Priest (Cyrillic), Andrew West (Phags-pa, Mongolian and Yi)和信息技术国际论坛的成员,以及 Kent Karlsson 等人。Thomas Bishop 检查了 CJK 表意数据并且提供了笔画数据来存储基本笔画索引。我们还要感谢 Patrick Andries, François Yergeau, Alain LaBonté, Jacques André 等人为 4.1 版本中提供法语转换成字符图表注释所做的工作。

许多人为印度文转化做出了突出贡献,他们是: Stefan Baums (Devanagari), Gihan Dias (Sinhala), Naga Ganesan (Malayalam), ManojJain (improvement of the overall Indic text), Gautam Sengupta (Bengali), Sukhjinder Sidhu (Gurmukhi), K. G. Sulochana (Malayalam)和 Om Vikas (improvement of the overall Indic text)。增加了新字符,改进了文字,增加了注释,系统化的编程也建立起来了。

对 Unicode 标准 5.0 版本来说,改进和校勘许多字符的属性和算法是非常重要的。国际 Unicode 组织(ICU)在这里发挥了巨大作用,包括以下人员: Min Cui, Mark Davis, John Emmons, Doug Felt, Deborah Goldsmith, Andy Heninger, Qian Jing, Yan Xuan Liang, Alan Liu, Steven Loomis, Eric Mader, George Rhoten, Markus Scherer, Bei Shu, William Sullivan, Raghuram Viswanadha 和 Vladimir Weinstein。另外, Kent Karlsson 很认真地校勘了字符属性。

Unicode 标准和 ISO/IEC10646 的同步体现了在世界范围内的多年努力。许多大学和研究所为少数的历史文字编码提供了很好的建议。伯克利的加利福尼亚大学的 Deborah Anderson 为文字编码提供了许多很好的建议。基金提供者主要有: National Endowment for the Humanities, the N'Ko Institute of America and Mamady Doumbouya, the Society of Biblical Literature, Association Manden, and UNESCO (Communication & Information Sector, InitiativeB@bel)。Unicode 协会非常感激其他大学和研究机构: Thesaurus Linguae Graecae Project, University of California, Irvine, the Initiative for Cuneiform Encoding (ICE), Johns Hopkins University 和 the International Association for Coptic Studies。

伴随着 Unicode 5.0 版本的出现,Unicode 标准对所有的主流现代文字和大量的历史文字及少数民族文字进行了编码。我们感谢以下专家,因为他们提供了特殊的知识,促进了该标准的完成。

- 阿拉伯文: Jonathan Kew, Michael Everson 和 Roozbeh Pournader。
- 巴厘文: Michael Everson, Made Suatjana, Ida Bagus Adi Sudewa, I Nyoman Suarka, Donny Harimurti, Tudy Harimurti 和 Nyoman Sugiarta。Unicode 得到了 UNESCO, National Endowment for the Humanities 和 the Yayasan Bali Galang 的大力支持。
- 布吉文: Michael Everson。

- CJK 表意文字、符号和标记: China National Information Technology Standardization Technical Committee, Christopher Cullen, Deborah Goldsmith, John Jenkins, Eric Muller, Michel Suignard 和 Andrew West。
- 哥普特文: Michael Everson, Gerald Browne, Stephen Emmel, 和 the International Association for Coptic Studies。
- 西里尔文: Lorna Priest。
- 埃塞俄比亚文补充: Daniel Yacob。
- 格鲁吉亚文补充: Michael Everson, Georgian State Department of Information Technology, David Tarkhan-Mouravi (Chair)和 Jost Gippert。
- 格拉哥里文: Michael Everson 和 Ralph Cleminson。
- 希腊文: Maria Pantelia, Deborah Anderson, Nick Nicholas 和 Richard Peevers。
- 印度文: 印度政府, 信息技术部, Om Vikas 和 Manoj Jain, INFITT, Michael Kaplan 和 Peter Constable。
- 卡纳达文和天成体: Michael Everson。
- 迦婆士梯文: Andrew Glass, Stefan Baums 和 Richard Salomon。
- 拉丁文: Peter Constable, Mark Davis, Michael Everson, Chris Harvey, Jonathan Kew 和 Lorna Priest。
- 蒙古文: Andrew West。
- 新傣文: Peter Constable, Mark Davis, Michael Everson, Chris Harvey, Jonathan Kew 和 Lorna Priest。
- N’Ko: Michael Everson, Mamadi Doumbouya, Mamadi Baba Diané 和 Karamo Kaba Jammeh. Unicode gratefully acknowledges support from UNESCO Initiative B@bel, N’Ko Institute of America and Mamady Doumbouya 和 Association Manden。
- 古波斯文: Michael Everson。
- 八思巴字: China National Information Technology Standardization Technical Committee, Mongolian Agency for Standardization and Metrology, Andrew West 和 Chris Fynn。
- 腓尼基文: Michael Everson。
- 语音扩展: Peter Constable, Mark Davis, Michael Everson 和 Lorna Priest。
- 苏美尔-阿卡德楔形文: Michael Everson, Karljürgen Feuerherm, Steve Tinney, Madeleine Fitzgerald 和 Cale Johnson。。
- Syloti Nagri 文: Peter Constable, James Lloyd-Williams 和 Sue Lloyd-Williams, Shamsul Islam Chowdhury, Asaddar Ali, Mohammed Sadique 和 Matiar Rahman。
- 符号: Asmus Freytag, Barbara Beeton, Michael Everson, Murray Sargen 和 Andreas Stoetzner。
- 提非纳文: Patrick Andries, François Yergeau 和 Alain LaBonté。
- 汉字: George Bell, Joy Zhao Rouxer 和 Steve Mann。

Unicode 技术委员会(UTC) 制定了 Unicode 标准的技术内容, 该委员会在 Unicode 5.0

版本的发展过程中由 Lisa Moore 任主管, Cathy Wissink 和 Eric Muller 任副主席, Rick McGowan 任秘书长。Unicode 技术委员会的成员包括会员、个人、Unicode 官员和专家。如果没有过去三年来所有成员的创造性工作和深入思考, Unicode 5.0 版本是不可能制定出来的: Mujahid Agha (Pakistan, NLA), Joan Aliprand (formerly of RLG), Deborah Anderson (UC Berkeley, Department of Linguistics), Rick Andrews (VeriSign), Takeshi Asano (Sun), Andreas Bäß (Denic), Christopher Chapman (Monotype Imaging), Daniel Chen (IBM), Steve Cohen (Basis), Peter Constable (Microsoft), Richard Cook (UC Berkeley, Department of Linguistics), Mark Davis (Google, formerly of IBM), Sabine Dolderer (Denic), Attash Durrani (Pakistan, NLA), Peter Edberg (Apple), Asmus Freytag (for Basis Technology), Deborah Goldsmith (Apple), Geoffry Greve (Monotype Imaging), Hideki Hiura (Justsystem, formerly of Sun), Takahiro Imai (formerly of Peoplesoft), Manoj Kumar Jain (India, MIT), John Jenkins (Apple), Bob Jung (Google), Susan Kline (HP), Tatsuo Kobayashi (Justsystem), Gary Krall (VeriSign), Hirobumi Kurosu (formerly of People-Soft), Swaran Lata (India, MIT), Ken Lunde (Adobe), Ian Macleod (Sybase), Kamal Mansour (Monotype Imaging), Benson Margulies (Basis), Shinobu Matsuzuka (Sun), Matthias Mittelstein (SAP), Lisa Moore (IBM), Nobuyoshi Mori (SAP), Eric Muller (Adobe), Mihai Nita (Adobe), Sandra Martin O'Donnell (formerly of Compaq, HP), Dave Opstad (Monotype Imaging), Pierre Ouédraogo (Agence Intergouvernementale de la Francophonie), Shripad Patki (Sun), Addison Phillips (Yahoo!), Toby Phipps (formerly of Peoplesoft), Gabriel Plumlee (formerly of Peoplesoft), Erik van der Poel (Google), Wendy Rannenberg (Compaq, then HP), Mike Van Riper (VeriSign), Lynn Ruggles (HP), Marcos Sanz (Denic), Murray Sargent (Microsoft), Bernhard Schilling (SAP), Karen Smith-Yoshimura (RLG), Michel Suignard (Microsoft), Ienup Sung (Sun), Tex Texin (Yahoo!), V. S. Umamaheswaran (IBM), Om Vikas (India, MIT), Ken Whistler (Sybase), Cathy Wissink (Microsoft), Jianping Yang (Oracle), Michael Yau (Oracle) 和 Weiran Zhang (Oracle)。

其他成员和专家也对 Unicode 技术委员会的工作做出了贡献, 他们是: Mati Allouche, Harald Alvestrand, Barbara Beeton, Peter Constable, James Do, Martin Dürst, Behdad Esfahbod, Patrik Fältström, John M. Fiscella, Paul Hoffman, Michael Kaplan, Jonathan Kew, Kamal Mansour, Thomas Milo, Paul Nelson, Nick Nicholas, Roozbeh Pournader, Jonathan Rosenne 和 Zhang Zhoucai。

Unicode 技术委员会和国际信息技术标准委员会(ISO/IEC JTC1/SC2/WG2)的技术委员会 L2 有密切的工作往来。我们感谢 Cathy Wissink、Eric Muller 以及 Lisa Moore 副主席的协作, Rick McGowan 维护了基于网页的 L2 文档, 这对 Unicode 5.0 版本的发展是至关重要的。

Unicode 协会和国际标准组织保持互惠的双向合作。我们感谢 ISO/IEC JTC1/SC2/WG2 的成员及 Ideographic Rapporteur Group 成员的支持和努力, 实现了两个标准文档的同步。感谢 WG2 的会议召集人 Mike Ksar, IRG 大会起草人 Lu Qin 博士和前一届大会起草人 Zhang Zhoucai, WG2 的编辑 Michel Suignard, Asmus Freytag, Michael Everson 和 Ken Whistler, 感谢 Asmus Freytag 在 WG2 大会上所做的陈述, 感谢 ISO/IEC JTC1/SC2 成员的校验, 尤其是 Alain LaBonté 和 Ken Whistler, 他们为校验做了定义。

在 Unicode 5.0 版本的改进过程中，特别感谢国际工作组 W3C 的紧密合作。感谢它的成员，特别是 Martin Dürst, Richard Ishida, Addison Phillips, Felix Sasaki, Tex Texin, Misha Wolf 和 François Yergeau，他们及时对该标准做了评价并提出了宝贵的建议。

IETF 在 Unicode 的采用上做出了很大的贡献。感谢它与 Unicode 协会的合作，尤其要感谢 Harald Alvestrand, Patrik Fältström 和 Paul Hoffman，他们帮助我们改进了标准的稳定性。

公司成员的支持对 Unicode 5.0 版本也是很重要的。Adobe 系统公司发放了支持少数民族字体的许可证，Adobe® Framemaker® 7.2 和 Adobe® Creative Suite 2 Premium 制作了该书的文本和图表。另外，感谢 Apple Computer, Inc., Microsoft Corporation 和 Monotype Imaging. PdfLibGmbH donated PDFlib+PDI 6.0.3 for Microsoft Windows, PDFlib TET (Text Extraction Toolkit) 2.1.0 for Microsoft Windows, plus custom-built software for post-production processing of PDFs for the code charts 和 Unicode 标准附件提供了设备和资源。

没有以前版本相关人员的努力付出也就没有 Unicode 5.0 版本，他们是：Glenn Adams, Joan Aliprand, Avery Bishop, Lori Brownell, Lee Collins, Andy Daniels, Burwell Davis, Bill English, Edwin Hart, Masami Hasegawa, Lloyd Honomichl, Liao Huan-Mei, Eric Mader, Dave Opstad, Hugh McGregor Ross, Isai Scheinberg, Ed Smura, Alan Tucker, Bill Tuthill 和 J. G. Van Stee。

感谢以上提到的所有人，任何错误和疏漏都是 Unicode 协会的责任。

Unicode 协会成员

在 Unicode 5.0 版本改进过程中，下列公司和政府都是 Unicode 的成员。一些成员在这一过程中改变了他们的级别，所以会被列出两次：

正式成员

Adobe Systems, Inc.
Agence Intergouvernementale de la
Francophonie
Apple Computer, Inc.
Basis Technology Corporation
Denic e.G.
Google, Inc.
Hewlett-Packard Company
IBM Corporation
India, Ministry of Information Technology
Justsystem Corporation
Microsoft Corporation
Monotype Imaging
Oracle Corporation
Pakistan, National Language Authority