

Electronic Edition

This file is part of the electronic edition of *The Unicode Standard, Version 5.0*, provided for online access, content searching, and accessibility. It may not be printed. Bookmarks linking to specific chapters or sections of the whole Unicode Standard are available at

<http://www.unicode.org/versions/Unicode5.0.0/bookmarks.html>

Purchasing the Book

For convenient access to the full text of the standard as a useful reference book, we recommend purchasing the printed version. The book is available from the Unicode Consortium, the publisher, and booksellers. Purchase of the standard in book format contributes to the ongoing work of the Unicode Consortium. Details about the book publication and ordering information may be found at

<http://www.unicode.org/book/aboutbook.html>

Joining Unicode

You or your organization may benefit by joining the Unicode Consortium: for more information, see [Joining the Unicode Consortium](http://www.unicode.org/consortium/join.html) at

<http://www.unicode.org/consortium/join.html>

This PDF file is an excerpt from *The Unicode Standard, Version 5.0*, issued by the Unicode Consortium and published by Addison-Wesley. The material has been modified slightly for this electronic edition, however, the PDF files have not been modified to reflect the corrections found on the Updates and Errata page (<http://www.unicode.org/errata/>). For information on more recent versions of the standard, see <http://www.unicode.org/versions/enumeratedversions.html>.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed with initial capital letters or in all capitals.

The Unicode® Consortium is a registered trademark, and Unicode™ is a trademark of Unicode, Inc. The Unicode logo is a trademark of Unicode, Inc., and may be registered in some jurisdictions.

The authors and publisher have taken care in the preparation of this book, but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

The *Unicode Character Database* and other files are provided as-is by Unicode®, Inc. No claims are made as to fitness for any particular purpose. No warranties of any kind are expressed or implied. The recipient agrees to determine applicability of information provided. *Dai Kan-Wa Jiten*, used as the source of reference Kanji codes, was written by Tetsuji Morohashi and published by Taishukan Shoten.

Cover and CD-ROM label design: Steve Mehallo, www.mehallo.com

The publisher offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales, which may include electronic versions and/or custom covers and content particular to your business, training goals, marketing focus, and branding interests. For more information, please contact U.S. Corporate and Government Sales, (800) 382-3419, corpsales@pearsoned.com. For sales outside the United States please contact International Sales, international@pearsoned.com

Visit us on the Web: www.awprofessional.com

Library of Congress Cataloging-in-Publication Data

The Unicode Standard / the Unicode Consortium ; edited by Julie D. Allen ... [et al.]. — Version 5.0.
p. cm.

Includes bibliographical references and index.

ISBN 0-321-48091-0 (hardcover : alk. paper)

1. Unicode (Computer character set) I. Allen, Julie D.

II. Unicode Consortium.

QA268.U545 2007

005.7'22—dc22

2006023526

Copyright © 1991–2007 Unicode, Inc.

All rights reserved. Printed in the United States of America. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. For information regarding permissions, write to Pearson Education, Inc., Rights and Contracts Department, 75 Arlington Street, Suite 300, Boston, MA 02116. Fax: (617) 848-7047

ISBN 0-321-48091-0

Text printed in the United States on recycled paper at Courier in Westford, Massachusetts.

First printing, October 2006

Praise for *The Unicode Standard, Version 5.0*

“The world is a global village, trade crosses language barriers, and yet every one of us likes to feel comfortable within their own mother tongue. Unicode enabled us to give the local sense to every one of our users, while connecting the world of trade—which is the reason we will support Unicode in all of our products.”

—**Shai Agassi**, Member
SAP Executive Board

“The W3C was founded to develop common protocols to lead the evolution of the World Wide Web. The path W3C follows to making text on the Web truly global is Unicode. Unicode is fundamental to the work of the W3C; it is a component of W3C Specifications, from the early days of HTML, to the growing XML Family of specifications and beyond.”

—**Sir Tim Berners-Lee**, KBE
Web Inventor and Director of the World Wide Web Consortium (W3C)

“The IETF has made the Unicode-compatible UTF-8 format of ISO 10646 the basis for its preferred default character encoding for internationalization of Internet application protocols, so I am delighted to see the official release of Unicode 5.0.”

—**Brian E. Carpenter**, Chair, Internet Engineering Task Force
Distinguished Engineer, Internet Standards & Technology, IBM

“Google’s objective is to organize the world’s information and to make it accessible. Unicode plays a central role in this effort because it is the principal means by which content in every language can be represented in a form that can be processed by software. As Unicode extends its coverage of the world’s languages, it helps Google accomplish its mission.”

—**Vint Cerf**, Chief Internet Evangelist
Google, Inc.

“Unicode Standard Version 5.0 is a great milestone for the Unicode Standard, which has been critical to computing since it was first published in 1991. With extended script and character support, this new version will help us bridge the digital divide by enabling more people to access computing in the language they use every day. The comprehensive set of mathematics symbols simplifies support for technical documents in business software. For more than a decade, Unicode has been a foundation for many Microsoft products and technologies: Unicode Standard Version 5.0 will help us deliver important new benefits to users.”

—**Bill Gates**, Chairman
Microsoft Corporation

“Unicode transformed characters from being a random collection of bits to things of meaning. Without Unicode, Java wouldn’t be Java, and the Internet would have a harder time connecting the people of the world.”

—**James Gosling**, Inventor of Java
Sun Microsystems, Inc.

“In the Directorate-General for Translation of the European Commission, the databases for internal document management and the interfaces of software applications and hardware equipment—including keyboards—have been built around Unicode, allowing representation of alphabets of all languages. Therefore, introducing languages using Cyrillic characters, or any other character set recognised by Unicode, is no problem.”

—**Tytti Granqvist**, Coordinator for External Communication
Directorate-General for Translation of the European Commission

“Because the character primitive in Java is Unicode, the global market readiness of internationalized Java applications depends on the features and coverage of the scripts that Unicode provides. As a member of the Java community, I greatly welcome Unicode 5.0. Developers will find it easy to implement—the standard is far more thoroughly explained than ever before. Computer users in global markets will also appreciate its larger coverage of scripts.”

—**Kazuhiro Kazama**, Senior Research Engineer
Nippon Telegraph and Telephone Corporation, Japan

“The development of Unicode has underscored the Internet’s truly global character. The recorded history of every nation and culture can travel in its natural form across Cyberspace for the use of anyone, anywhere. Through the power of Unicode, a worldwide audience is finally able to share in the breadth of human creativity.”

—**Brendan Kehoe**
Zen and the Art of the Internet

“Hard copy versions of the Unicode Standard have been among the most crucial and most heavily used reference books in my personal library for years. Unicode allows me to celebrate the fact that computer science is a vast worldwide collaboration. And Unicode is perhaps the best tool I know to help bring understanding between people of different cultures.”

—**Donald E. Knuth**, Professor Emeritus of The Art of Computer Programming
Stanford University

“Our innate desire to communicate defines us, both as individuals and as a species. Operating systems such as Solaris 10 use Unicode to enable humans to communicate across the Internet, and to bridge the digital divide.”

—**Tim Marsland**, Software CTO, VP/Fellow
Sun Microsystems, Inc.

“Unicode, as an enabler to support multiple languages and locales across multiple platforms without re-engineering, is a solid foundation for e-business in a global economy. IBM’s implementation of Unicode support across our product lines echoes our overall commitment to the importance of open standards in the evolving global marketplace.”

—Steve Mills, General Manager Solutions and Strategy
IBM Software Group

“XML software tools are well internationalized, thanks to XML’s adoption of Unicode. The addition of JIS X 0213 characters to Unicode 5.0 provides the characters required by the Japanese e-government.”

—Makoto Murata, Research Specialist
Tokyo Research Lab, IBM Japan, Ltd.

“Unicode marks the most significant advance in writing systems since the Phoenicians.”

—James J. O’Donnell, Provost,
Georgetown University

“I applaud the efforts of the Unicode community, ensuring computers worldwide work seamlessly in everyone’s language.”

—Larry Page, co-founder
Google, Inc.

“Unicode and its companion ISO/IEC 10646 overcome the limitations and confusion of all earlier character coding standards. They enable every nation and community to write its own language with computers. They ensure a firm foundation for reliable and efficient interchange of text worldwide.”

—Hugh McGregor Ross
First editor of ISO/IEC 10646

“Apple has been supporting Unicode since the beginning. We’re thrilled to see the growing adoption of the Unicode and welcome Unicode 5.0 as a new milestone in the definition of the standard.”

—Bertrand Serlet, Senior Vice President of Software Engineering
Apple, Inc.

“Unicode is arguably the most widely adopted software standard in the world, reaching into any program, application, or system that displays text. Though starting from a high point, Unicode 5.0 manages to increase quality yet again, which will continue to expand adoption and support integration.”

—Richard Mark Soley, Ph.D., Chairman and CEO
Object Management Group

“If you are a programmer working in 2006 and you don’t know the basics of characters, character sets, encodings, and Unicode, and I catch you, I’m going to punish you by making you peel onions for six months in a submarine.”

—Joel Spolsky
Joel on Software

“The Java™ programming language was designed to be a portable, platform-independent programming language for the World Wide Web—not the ASCII Web or the ISO-Latin-1 Web, but for the entire world. It was for this reason that Java’s designers broke out of the 8-bit strait-jacket and based the character and string datatypes firmly on Unicode, the character encoding standard designed to support all the world’s languages for truly international communication and commerce. Version 5.0 of the Unicode Standard is the most comprehensive and thoroughly documented version yet.”

—Guy L. Steele Jr., Sun Fellow
Sun Microsystems, Inc.

“Justsystems was the first vendor in Japan to implement the Unicode architecture in word processing software. Thus for over a decade Japanese computer users have enjoyed the benefits of Unicode, especially the CJK Unified Ideographs. We are now aggressively launching our new technology ‘xfy’ on XML and Java—thanks to Unicode, it is already internationalized! Justsystems appreciates the Unicode philosophy and architecture of universality, and welcomes the publication of Version 5.0.”

—Kazunori Ukigawa, President and CEO
Justsystems Corp.

“Modern programs must handle Unicode—Python has excellent support for Unicode, and will keep getting better.”

—Guido van Rossum
Inventor of Python

“Unicode is marvellous. It makes it possible for phoneticians throughout the world to use all manner of phonetic symbols in their work and display them on computer screens in the certainty that they will not now be garbled or turned into wingdings (as once used to happen all too often). All alphabetic phonetic symbols officially recognized by the International Phonetic Association are now included in the Unicode Standard.”

—John Wells, President
International Phonetic Association

For additional acclaim, see <http://www.unicode.org/press/quotations.html>.