

Appendix B

Submitting New Characters

The Unicode Consortium accepts proposals for inclusion of new characters and scripts in the Unicode Standard. All proposals must be in writing, must include at least one picture of each proposed character (normally from a printed source), and must include significant documentation justifying the proposal. Those considering submitting a proposal should first determine whether a particular script or character has already been proposed. The identification of the sponsor(s) must be included, along with postal address and an electronic mail address or phone number. General guidelines for the preparation of a proposal follow. Because this information may be updated periodically, please consult the Unicode Consortium's Web site (<http://www.unicode.org>) for the most current guidelines. The Web site also provides information on proposed new scripts and characters, which may help to determine whether a script or character has already been proposed.

Before preparing a proposal, sponsors should note in particular the distinction between the terms *character* and *glyph* as defined in this standard. Because of this distinction, graphics such as ligatures, conjunct consonants, minor variant written forms, or abbreviations of longer forms are generally not acceptable as Unicode characters. The Unicode Consortium is interested in obtaining information on known glyphs, minor variants, precomposed characters (including ligatures, conjunct consonants, and accented characters) and other such "noncharacters," mainly for cataloging and research purposes; however, they are generally not acceptable for character proposals.

Experience has shown that it is often helpful to discuss preliminary proposals before submitting a detailed proposal. One open forum for such discussion is the Unicode mailing list. Please see the Unicode Web site for instructions on how to subscribe to the mailing list. Sponsors are urged to send a message of inquiry or a preliminary proposal there before formal submission. Many problems and questions can be dealt with there, minimizing the severity of later revisions.

Each proposal received will be evaluated initially by technical officers of the Unicode Consortium and the result of this evaluation will be communicated to the sponsor(s) of the proposal. All proposals, whether successful or not, will be retained by the Unicode Consortium as a matter of record.

B.1 Proposal Guidelines

The sponsor(s) proposing the addition of a new character to the Unicode Standard should follow these guidelines:

- Before proceeding, determine that each proposed addition is a character according to the definition given in the Unicode Standard and that the

proposed addition does not already exist in the standard. Ensure that documentation supporting the proposal states whether any Unicode characters were examined as possible equivalents for the proposed character and, if so, why each was rejected.

- If the character is a Han (ideographic) character, strong evidence for its uniqueness, indispensability, and value to users must be submitted, including printed contexts where it is used. If it could be considered a variant of a character already included in the Unicode Standard, the proposed character should be shown in a context that demonstrates why it must be distinguished as a separate character and not considered a variant of the character already included.
- Determine and list the proposed (or recommended) character properties for each character being proposed, especially when proposing entire scripts for encoding. (See *Chapter 4, Character Properties*.)
- Proposals to include entire scripts (Egyptian hieroglyphics, for example) must cite modern, definitive sources of information regarding such scripts. Sponsorship by the relevant academic bodies (such as the International Association of Egyptologists) may be helpful in determining the proper scope for encoding of characters in such cases. (Before submitting full script proposals, sponsors should determine that a proposal does not already exist for that script.)
- If a proposed character is part of a dead language or obsolete/rare script that is already encoded, cite the most important modern sources of information on the script and the proposed additions. Names, including academic affiliations, of researchers in the relevant field are welcomed.
- New symbol characters are evaluated on the basis of how they are used in plain text and whether they have well-defined semantics. The proposal should demonstrate their usage in computer applications and should clarify whether they are part of a notation system, whether they are searchable or indexable, and whether they are used in tabular lists as a shorthand for specified characteristics. For further information on the criteria for encoding symbols, see the Unicode Web site.
- The Unicode Consortium works closely with the relevant committee responsible for ISO/IEC 10646, namely JTC1/SC2/WG2, in proposing additions as well as monitoring the status of proposals by various national bodies. Therefore, proposals may eventually be formulated as ISO/IEC documents and significant detailed information will be required.
- The standardized form “ISO/IEC JTC 1/SC 2/WG 2 Proposal Summary Form to Accompany Submissions for Additions to the Repertoire of ISO/IEC 10646” has been designed for the purpose of obtaining detailed information for ISO purposes and for the Unicode Technical Committee. Use of this form is required for all proposals. It is available via the Unicode Web site.

B.2 Requirements of Proposal Form and Process

The proposal summary form requires the following further information (paraphrased):

- The repertoire, including proposed character names
- The name and contact information for a company or individual who would agree to provide a computerized font (True Type or PostScript) for publication of the standard

- References to dictionaries and descriptive texts establishing authoritative information
- Names and addresses of appropriate contacts within national body or user organizations
- The context within which the proposed characters are used (for example, current, historical, and so on)
- Especially for sporadic additions, what similarities or relationships the proposed characters bear to existing characters already encoded in the standard

The international standardization of entire scripts requires a significant effort on the sponsor's part. It frequently takes years to move from an initial draft to final standardization, particularly because of the requirements to synchronize proposals with the work done in the ISO committee responsible for the development of ISO/IEC 10646. Sponsors, particularly of entire scripts, should be prepared to become involved at various times throughout the process—perhaps revising their proposals more than once; collecting further detailed information; organizing online discussions or meetings to dispel controversy; or answering questions posed by committees or national bodies. Without such involvement, any proposal of more than a few characters is unlikely to be successful in the long run.

Interim Solutions

There are ways for programmers and scholarly organizations to make use of Unicode character encoding, even if the script they want to use or transmit is not yet (or may never be) part of the Unicode Standard. Individual groups that make use of rare scripts or special characters can reach a private agreement about interchange and set aside part of the Private Use Area to encode their private set of characters. Individuals with interests in rare scripts or materials relating to them may sometimes be contacted through an electronic mailing list which the Consortium maintains. For information about these mailing lists, please contact the Consortium.

Sending Proposals

Completed proposals may be sent to:

The Unicode Consortium
P.O. Box 391476
Mountain View, CA 94039-1476
USA

For inquiries only: E-mail: info@unicode.org

All proposals are required to be in “hard copy” (paper) form unless prior arrangements have been made for receipt of electronic copy and/or Web pages (machine-readable HTML documents plus GIF or JPEG images).

This PDF file is an excerpt from *The Unicode Standard, Version 3.0*, issued by the Unicode Consortium and published by Addison-Wesley. The material has been modified slightly for this online edition, however the PDF files have not been modified to reflect the corrections found on the Updates and Errata page (see <http://www.unicode.org/unicode/uni2errata/UnicodeErrata.html>). More recent versions of the Unicode standard exist (see <http://www.unicode.org/unicode/standard/versions/>).

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 Addison-Wesley was aware of a trademark claim, the designations have been printed in initial capital letters. However, not all words in initial capital letters are trademark designations.

The authors and publisher have taken care in 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.

ISBN 0-201-61633-5

Copyright © 1991-2000 by Unicode, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher or Unicode, Inc.

This book is set in Minion, designed by Rob Slimbach at Adobe Systems, Inc. It was typeset using FrameMaker 5.5 running under Windows NT. ASMUS, Inc. created custom software for chart layout. The Han radical-stroke index was typeset by Apple Computer, Inc. The following companies and organizations supplied fonts:

Apple Computer, Inc.
Atelier Fluxus Virus
Beijing Zhong Yi (Zheng Code) Electronics Company
DecoType, Inc.
IBM Corporation
Monotype Typography, Inc.
Microsoft Corporation
Peking University Founder Group Corporation
Production First Software

Additional fonts were supplied by individuals as listed in the *Acknowledgments*.

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.

All other company and product names are trademarks or registered trademarks of the company or manufacturer, respectively.

The publisher offers discounts on this book when ordered in quantity for special sales. For more information please contact:

Corporate, Government, and Special Sales
Addison Wesley Longman, Inc.
One Jacob Way
Reading, Massachusetts 01867

Visit A-W on the Web: <http://www.awl.com/cseng/>

First printing, January 2000.