

Appendix B

Submitting New Characters

The Unicode Consortium accepts proposals for inclusion of new characters or 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 documentation justifying the proposal. The identification of the sponsor(s) must be included, along with a postal or electronic mail address, and/or a phone number. General guidelines for the preparation of a proposal appear next, and more detailed information is available electronically through the consortium's World Wide Web server.

The Unicode Standard definition of character is stated in the Glossary. Before preparing a proposal, sponsors should note in particular the distinction between the terms *character* and *glyph* as therein defined. 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 "non-characters," mainly for cataloging and research purposes; however, they are generally not acceptable for character proposals.

B.1 Proposals

Each proposal received will be evaluated by members of the Unicode Technical Committee, 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..

Send proposals to:

The Unicode Consortium
P.O. Box 700519
San Jose, CA 95170-0519
USA

Internet: unicode-inc@unicode.org

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

1. 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 equiva-

lents for the proposed character and, if so, why each was rejected.

2. 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.
3. If the character is part of a dead language or obsolete/rare script, cite the most important modern sources of information on the script. Names (including academic affiliation) of researchers in the relevant field are welcomed.
4. Proposals to include entire scripts (Egyptian hieroglyphics, for example) should also 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.
5. The Unicode Consortium works closely with ISO 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 form N-1116-F, *Proposal Summary Form to Accompany Submissions for Additional Repertoire of ISO/IEC 10646*, has been designed for this purpose, and its use is encouraged for all proposals. The form is available via anonymous FTP (at the time of writing, it is available as <ftp://dkuug.dk/JTC1/SC2/WG2/n1252.doc>), and 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, PostScript, or 96x96 bit-map) 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

In the meantime, 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.

A few living scripts are very likely to be included in the near future, including Ethiopian, Mongolian, Sinhala, Burmese, Khmer, and Cree/Inuktitut. The Unicode Consortium has also published three technical reports that contain proposals or draft proposals for encoding a number of additional scripts. These reports and the scripts they cover are as follows:

Unicode Technical Report #1, Draft Proposals, October 1992.

Burmese, Khmer, and Ethiopian

Unicode Technical Report #2, Preliminary Draft Proposals, October 1992.

Mongolian, Sinhala, and Tibetan (superseded by the Tibetan block in the Unicode Standard, Version 2.0).

Unicode Technical Report #3, Exploratory Proposals, October 1992.

Early Aramaic, Balti, Batak, Buginese, Cherokee, Etruscan, Glagolitic, Kirat (Limbu), Linear B, Maldivian (Dihevi), Manipuri (Meithei), Meroitic, Ogham, Pahlavi/Avestan, Old Persian Cuneiform, Phoenician, Róng (Lepcha), Northern Runes, Epigraphic South Arabian, Syriac, Tagalog and Mangyan (Buhid), Tai Lu, Tai Mau (Tai Nua), Tifinagh (Numidian), and Ugaritic Cuneiform.

Additional characters that have been proposed to ISO JTC1/SC2/WG2 include the following:

Aboriginal Canadian Scripts (Syllabics), Additional CJK Typesetting symbols, Arabic Script Extensions for Uighur, Kazakh, and Kirghiz, Braille, Cham, Chũ Nôm, Irish Gaelic Precomposed Latin Letters, Egyptian Hieroglyphics, Unified CJK Ideograph extensions, Xishuang Banna Dai, and Yi.