

Proposal to add the Open Four character to Unicode

Aaron Franke

2018-10-09 (October 9th)

1. Introduction

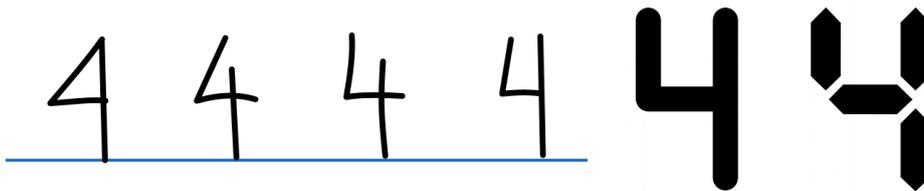
The international symbol 4, or “Digit Four”, has been in Unicode since its inception, as it was present in ASCII. This symbol represents the number four in most languages and is often written with three straight lines, each intersecting the others, with the right vertical line extending to the bottom to differentiate it from a right triangle. However, the number four is not always written this way. It is common to see the symbol written using three straight lines with an open top, with two vertical lines not touching. Visually, the closest existing character is ౪ (Telugu).

This four with an open top, which I will be referring to as “Open Four”, was originally created out of a necessity to allow the displaying of the number four on a seven-segment display. Since then, it has become a popular alternative way to hand-write the number four, especially in the United States. Some fonts define the Unicode “Digit Four” to be in this style instead of the traditional version. The ability to explicitly represent this Open Four shape would be a very useful addition to Unicode.

Left Image: Four variations of hand-writing the number four.

Middle Image: The reference image for the Open Four character.

Right Image: The seven-segment display version of the number four.



2 Suitability for Inclusion

There are already several dozen possible explicit representations of the number four in Unicode according to FileFormat.info.^{[A][B]} Several of these exist because of explicit mathematical notation purposes, and the Open Four is justified for similar reasons. Having the Open Four character in Unicode would allow: (1) hand-written documents to be transcribed and viewed with the same symbols as written, (2) documentation of equipment using Open Fours to have the correct symbol encoded as text, and (3) a way for anyone to explicitly prefer the Open Four shape over the standard Digit Four.

I have provided a reference image depicting the shape, called “openfour.png”, and a few other images depicting the open-top four. The reference image was created by me, and I am placing it in the public domain with no restrictions on its use. The “Seven-segment_4.png” and “Handwriting_variations_numeral_4.png” files are part of Wikimedia Commons, in the public domain, and are considered ineligible for copyright.

A. <http://www.fileformat.info/info/unicode/category/Nd/list.htm>

B. <http://www.fileformat.info/info/unicode/category/No/list.htm>

3 Evidence of Use

An Internet search for “open top four” gives some examples of its use. You can also find many examples if you search for “handwritten four” and “digital four”. Some of the results for “number font”, “four”, “4”, etc, also depict the Open Four.

Aside from Internet searches, you can also find examples of its use on many handwritten documents, at least in the United States.

4 Character Properties

Suggested character properties for the proposed Open Four symbol:

Property	Suggested Value
Code point	
Name	Open Four
General Category	Nl
Canonical Combining Class	0
Bidirectional Class	EN
Decomposition Type/Mapping	
Numeric Type	Digit
Numeric Value	4
Bidi Mirrored	N
Unicode 1 Name	
ISO Comment	
Simple Uppercase Mapping	
Simple Lowercase Mapping	
Simple Titlecase Mapping	

4.1 Ordering and Usage

In the context of both numeric and lexicographical sorting, the Open Four character should ideally be treated as equal to the 4 or “Digit Four” Unicode character. The Open Four should be sorted such that 3 is below it and 5 is above it. The sorting order between the Open Four and the existing “Digit Four” Unicode character is not important. This character is uncased. There is no special line-breaking behavior required. The open four should resolve to the number four when evaluated as a numeric value. The Open Four can have its lines thickened or thinned, and be used anywhere, as long as an ordinary user can recognize the symbol correctly.

5 Anticipated Objections

I can think of three possible objections to the inclusion of the open four character in Unicode.

The first objection that I can think of is that some people may not consider it necessary. Some may consider that the shape of the number four should be dependent on the font used and not by the character used. However, I believe the existence of several mathematical characters for the sole purpose of defining a notation style using a character, and the fact that the Open Four's shape differs significantly from the existing Digit Four character, justify it being included.

The second objection is that some may argue that other alternative writing styles for numbers should be included as separate characters. For example, the striked-through seven (~~7~~), the one with/without an underline and/or a stroke at the top, or the flat-topped three (looks similar to 3). I would not be against including these as characters as well. However, I believe these would be less significant additions than the Open Four. The flat-topped three is a minor difference, and the striked-through seven can be achieved with a seven and a strikethrough. Other people may wish to propose other characters, but the argument of this proposal is purely about the Open Four.

The third objection is that this proposal may be dismissed due to either containing errors or being unprofessional. This is my first ever Unicode character proposal and I apologize in advance for any mistakes. I created this proposal by following the example of the IEC power symbols proposal linked on the proposals page of www.unicode.org. I hope that the Unicode team can see this proposal for what it tries to achieve. If there are any questions or concerns about this proposal, I can be reached by E-mail at "arnfranke@yahoo.com".

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹
Please fill all the sections A, B and C below.

A. Administrative

1. Title:	<i>Proposal to add the Open Four character to Unicode</i>
2. Requester's name:	<i>Aaron Franke</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>2018-10-09</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>Yes</i>
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<i>Yes</i>
Name of the existing block:	<i>Numeric Digit category</i>
2. Number of characters in proposal:	<i>1</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input checked="" type="checkbox"/> B.1-Specialized (small collection) <input type="checkbox"/> B.2-Specialized (large collection) <input type="checkbox"/>	
C-Major extinct <input type="checkbox"/> D-Attested extinct <input type="checkbox"/> E-Minor extinct <input type="checkbox"/>	
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/> G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<i>Yes</i>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	
b. Are the character shapes attached in a legible form suitable for review?	<i>Yes</i>
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>Yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>There are references in general, and other uses</i>
7. Special encoding issues:	
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?	<i>Yes</i>

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>No</i> <i>If it has been submitted before, I am not aware of it</i>
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	<i>No</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>Yes</i> <i>Proposal document</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>common</i> <i>Proposal document</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>Yes</i> <i>Hand-written documents and seven-segment displays</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>Yes</i> <i>Yes</i> <i>Proposal document</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>Yes</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>Yes</i> <i>Yes</i> <i>Proposal document, its images, and their sources</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	<i>No</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	<i>No</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>No</i> <i>I am unsure what Ideographic compatibility characters are</i>