

Universal Multiple-Octet Coded Character Set  
 International Organization for Standardization  
 Organisation Internationale de Normalisation  
 Международная организация по стандартизации

**Doc Type: Working Group Document**

**Title: Proposal to encode two dashes required by the Chicago Manual of Style in the UCS**

**Source: Karl Pentzlin**

**Status: Individual Contribution**

**Action: For consideration by JTC1/SC2/WG2 and UTC**

**Date: 2010-01-29** (Rev. 1: A paragraph regarding the line breaking behavior was added.)

**Supplemental Punctuation: Dashes**

——	U+2Ex1	TWO-EM DASH = omission dash → 2014 em dash
————	U+2Ex2	THREE-EM DASH

**Properties:**

2Ex1; TWO-EM DASH; Pd; 0; ON; ; ; ; ; N; ; ; ; ;  
 2Ex2; THREE-EM DASH; Pd; 0; ON; ; ; ; ; N; ; ; ; ;

**Background information**

The Chicago Manual of Style (15th ed. [2003], ISBN 0-226-10403-6) explains different 2-em and 3-em dashes for different punctuation requirements, as is shown in the following excerpt of p.265 (sections 6.95/6.96):

2-EM AND 3-EM DASHES

6.95 *2-em dash.* A 2-em dash represents a missing word or part of a word, either omitted to disguise a name (or occasionally an expletive) or else missing or illegible in quoted or reprinted material. When a whole word is missing, space appears on both sides of the dash. When only part of a word is missing, no space appears between the dash and the existing part (or parts) of the word; when the dash represents the end of a word, a normal word space follows it. See also 7.66, 11.66–67.

“The region gives its —— to the language spoken there.”  
 Admiral N—— and Lady R—— were among the guests.  
 David H——h [Hirsch?] voted aye.

(Although a 2-em dash sometimes represents material to be supplied, it should not be confused with a blank line to be filled in; a blank in a form should appear thus \_\_\_\_.)

6.96 *3-em dash.* In a bibliography, a 3-em dash followed by a period represents the same author or editor named in the preceding entry (see 16.84–89, 16.103–5).

———. *The Last Dinosaur Book*. Chicago: University of Chicago Press, 1998.

In absence of the characters now proposed, they usually are substituted by sequences of 2 resp. 3 em dashes. However, the character U+2014 EM DASH is not necessarily left-joining and right-joining (as Box Drawing characters, or characters explicitly marked such like e.g. U+0332).

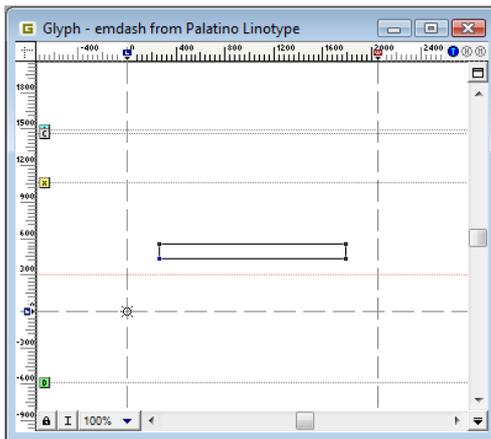
Thus, this trick may work with some fonts like Microsoft's Times New Roman: — ——

but it does not work with common fonts like Palatino Linotype: — — — — —

or DejaVu Serif: — — — —— (each: em dash + space + 2x em dash + space + 3x em dash).

This is due to the fact that dashes may intentionally have side bearings, when the font designer considers this adequate for good typography for the primary use of the em dash—like this.

E.g., this screenshot from FontLab shows the em dash in Palatino Linotype:



Thus, sequences of single em dashes cannot be regarded as equivalent to the two dashes proposed here.

Moreover, the two dashes proposed here are not variants of already encoded characters. The most similar character is U+2015 HORIZONTAL BAR. In fact, in most fonts this character is only slightly longer than the em dash (if not of the same length), while the proposed dashes have a specified length quite larger in relation to the em dash (implied by their name).

E.g., em dash + horizontal bar in Microsoft's Times New Roman: — — ; in DejaVu Serif: — —

Also, the alternative name for U+2015 HORIZONTAL BAR: "≠ quotation dash", suggests that the horizontal bar was introduced as a variant of the em dash, which according to the Chicago Manual of Style is also the appropriate character for starting quotations (see p.264, section 6.93). This is in accordance with its usual length similar to the em dash.

It also is to be noted that the two dashes proposed here are no recent inventions.

They are already mentioned in the 4th edition of the "Chicago Manual of Style" from 1906, see: [http://www.archive.org/stream/manualofstylebei00univrich/manualofstylebei00univrich\\_djvu.txt](http://www.archive.org/stream/manualofstylebei00univrich/manualofstylebei00univrich_djvu.txt)  
Also, they existed in lead type, as it was mentioned on the Unicode mailing list: <http://www.unicode.org/mail-arch/unicode-ml/y2010-m01/0226.html>

## Line breaking behavior

The two-em dash, in its function as an omission dash, usually behaves like the letter sequence which it represents, i.e. it acts like a non-breaking hyphen (see the example "H—h" above). Thus, in LineBreak.TXT, it should have the attribute "GL".

The three-em dash shows nothing special regarding line breaking, thus, in LineBreak.TXT, it should have the attribute BA (like the other dashes except the em dash).

**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

**A. Administrative**

1. Title:	<i>Proposal to encode two dashes required by the Chicago Manual of Style in the UCS</i>
2. Requester's name:	<i>Karl Pentzlin</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>2010-01-27</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/>
(or) More information will be provided later:	<input type="checkbox"/>

**B. Technical – General**

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input type="checkbox"/> No	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/> Yes	
Name of the existing block:	<i>Supplemental Punctuation</i>	
2. Number of characters in proposal:	<i>2</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?	<input checked="" type="checkbox"/> Yes	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/> Yes	
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/> Yes	
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?	<i>The author (font released into the Public Domain)</i>	
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	<i><a href="http://www.europatastatur.de/material/longdash.ttf">http://www.europatastatur.de/material/longdash.ttf</a></i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> Yes	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> Yes	
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)?	<input type="checkbox"/> No	

**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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

<sup>1</sup> Form number: N3152-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)

**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before? If YES explain		No
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:		No
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>All typesetters and readers in the USA and several other areas</i>	Yes:
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>see above</i>	common
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>see above</i>	Yes
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>To keep them up with other Supplemental Punctuation</i>	Yes Yes
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		Yes
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:		No
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:		No
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:		No
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:		No
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)		No
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:		No