

Proposal to encode the character “subscript ten”

2008-01-19

Leonid Broukhis, leob@mailcom.com

The following GOST 10859-64 character is proposed:

10

U+208x/9x SUBSCRIPT TEN

- approx. <sub> U+2081 U+2080 lower ten, decimal exponent
- uses one character position in monospace fonts

Example (the proposed character has been encoded as 208F within this proposal for demonstration):

Using 2081+2080 characters, monospace: 1 . 2 3 4 _{1 0} 5 6

Using proposed character, monospace: 1 . 2 3 4 ₁₀ 5 6

Using 2081+2080 characters, proportional: 1.234₁₀56

Using proposed character, proportional: 1.234₁₀56

The **subscript ten** entity is present in the [Revised Report on the Algorithmic Language Algol 60](#) (see <separator>) and has been included as an atomic character in the [State Standard of the USSR](#) (GOST 10859-64) and in the [German standard DIN](#) (ALCOR) to represent the exponent separator (sometimes named “lower ten”) for the scientific notation of real numbers. The standard-conforming hardware capable of printing the character has been produced (e.g. drum printers, console typewriters, punch card stations). Its vertical alignment was higher than the alignment for a subscript while lower than the baseline, but as different fonts sometimes [treat the superscript/subscript characters as numerator/denominator](#) (also see examples of the varying alignment above), this distinction is probably irrelevant. Samples of printouts demonstrating the proposed character are, regrettably, hard to find.

The GOST 10859 encoding was in widespread use among the programming community that used Soviet-manufactured computers (mainly BESM-6 – several hundred manufactured, and Minsk-32 – several thousand manufactured). With the advent of clones of Western computers, the new encodings striving to achieve ASCII-compatibility were adopted as standard. (The low popularity of ALGOL-60 in the U.S. could have prevented the ALGOL character set from being considered as a source of ideas for ASCII extensions.)

As of today, **subscript ten** remains the only character in GOST 10859 (and ALCOR) not present in Unicode.

BESM-6 or [its emulators](#) were reportedly being used in the Russian Federation as late as 2006. This means that the need to transcode GOST-based software and documentation can still arise: legacy numerical algorithms optimized for the non-IEEE floating point representation of BESM-6 cannot be simply recompiled and be expected to work reliably, and some human intervention may be necessary.

The importance of presenting numerical data in a tabulated monospace format, minimizing the probability of a human error, necessitates encoding of the **subscript ten** character as a separate entity. A position within the “Superscripts and Subscripts” range (2070-209F) would be desirable.

It should also be noted that there exists a semantic difference between **number 10 expressed with subscript characters** and the **subscript ten** character: the former denotes a numerical base or index (an arbitrary integer), whereas the latter introduces the exponent of a number in the scientific notation.

Consistently using **subscript ten** instead of 'E' or 'e' in appropriate contexts might simplify automatic processing of documents, avoiding potential ambiguity with hexadecimal numbers.

[NB. This proposal does not suggest replacing **number 10 expressed with subscript characters** as index, numerical base, or vulgar fraction denominator with the new character for stability reasons, therefore such uses of the proposed character, albeit possible, are not discussed.]

Satisfied criteria for encoding: H.3.1 (scientific/engineering notation), H.4 (1) character specific to a particular programming language, H.4 (2) compatibility for text mode tabular display.

Strengthening criteria: H.6 computer application, well-defined usage, occurs with numbers, used in tabular lists, part of a notational system, has well-defined semantics that lend to computer processing, H.8 completes a class of symbols already in the standard.

Weakening criteria H.7: (arguably) composable.

**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.

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:	Addition of the subscript ten character from GOST 10859-64
2. Requester's name:	Leonid Broukhis
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution
4. Submission date:	2008-01-19
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	yes
(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):	no
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	yes
Name of the existing block:	2070-209F Superscripts and Subscripts
2. Number of characters in proposal:	1
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input checked="" type="checkbox"/>
B.2-Specialized (large collection) <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested 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?	yes
a. If YES, are the names in accordance with the character naming guidelines in Annex L of P&P document?	yes
b. Are the character shapes attached in a legible form suitable for review?	yes
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?	Leonid Broukhis (draft)
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	http://mailcom.com/unicode/LowerTen.ttf using FontForge, based on FreeMono font
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	yes
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	no
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)?	no
8. Additional Information:	

¹ Form number: N3102-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)

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.

The 208F encoding is used as an example:

```
208F;SUBSCRIPT TEN;No;0;EN;<sub> 0030 0031;;;10;N;DECIMAL EXPONENT;;;;
```

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	<i>no</i>
If YES explain	
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.)?	<i>yes</i>
If YES, with whom?	<i>BESM-6 emulation enthusiasts</i>
If YES, available relevant documents:	<i>personal communication</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	<i>yes</i>
Reference:	<i>see proposal text</i>
4. The context of use for the proposed characters (type of use; common or rare)	<i>rare</i>
Reference:	<i>see proposal text</i>
5. Are the proposed characters in current use by the user community?	<i>yes</i>
If YES, where? Reference:	<i>Users of the emulation software, see proposal text</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<i>yes</i>
If YES, is a rationale provided?	<i>yes</i>
If YES, reference:	<i>see proposal text</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>n/a</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<i>yes</i>
If YES, is a rationale for its inclusion provided?	<i>yes</i>
If YES, reference:	<i>see proposal text</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	<i>no</i>
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	<i>no</i>
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	<i>no</i>
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:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<i>no</i>
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility character(s)?	<i>no</i>
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	