

Doc Type: Working Group Document
 Title: Revised proposal to encode the decimal exponent symbol
 Source: Leonid Broukhis, leob@mailcom.com
 Status: Individual Contribution
 Action: For consideration by JTC1/SC2/WG2 and UTC
 Date: 2008-01-22

The following GOST 10859-64 character is proposed:

10 U+23E8 DECIMAL EXPONENT SYMBOL

- ALGOL-60 token for scientific notation literals

Underscores show the baseline position.

Example (the proposed character has been encoded as 23E8 within this proposal for demonstration), as used by the ALGOL-BESM6 system, emulated by [an online emulator](#):

СИСТЕМА АЛГОЛ-БЭСМ6. ВАРИАНТ 01-05-79.

СЧЕТ БЕЗ КОНТРОЛЯ

1. BEGIN OUTPUT ('E', 355.0/113.0) END

.3141592920₁₀+01

The **decimal exponent symbol** is present in the [Revised Report on the Algorithmic Language Algol 60](#) (<separator> ::= , | . |₁₀ | : | ; | := | - | step | until | while | comment) and has been included as an atomic character in the [State Standard of the USSR](#) (GOST 10859-64, position 020) 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 slightly lower than the baseline, approximately as low as the descenders of Cyrillic letters Л and Ш. Samples of printouts demonstrating the proposed character are, regrettably, hard to find.

Here is a hand-retouched sample of a diagonal printer test from *Lyashenko V. F. "Programming for digital computing machines M-20, BESM-3M, BESM-4, M-220" (1967)*

```

100+-/ , . 10↑ ( )x = : [ ] * 6 ? ≠ < > : АБВГДЕЖЗИ
6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < > : АБВГДЕЖЗ
5 6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < > : АБВГДЕЖ
4 5 6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < > : АБВГДЕ
3 4 5 6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < > : АБВГД
2 3 4 5 6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < > : АБВГ
1 2 3 4 5 6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < > : АБВ
0 1 2 3 4 5 6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < > : АБ
Z 0 1 2 3 4 5 6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < > : А
WZ 0 1 2 3 4 5 6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < > :
VWZ 0 1 2 3 4 5 6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < >
UVWZ 0 1 2 3 4 5 6 7 8 9 + - / , . 10↑ ( )x = : [ ] * 6 ? ≠ < >

```

The GOST 10859-64 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, **decimal exponent symbol** remains the only character in GOST 10859-64 (and ALCOR) not present in Unicode.

BESM-6 or [its emulators](#) were reportedly being used in the Russian Federation as late as 2006, and rumors of a machine operating at a navy training base near St. Petersburg still persist (pers. comm.). This means that the need to transcode GOST-based software and documentation can still arise: legacy numerical algorithms (some of which may be of interest, e.g. for the [automatic landing of the Buran shuttle](#) – Russian language page lamenting the availability of the flight data only in printed form and on legacy tapes) 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.

It should be noted that there exists a semantic difference between **number 10 expressed with subscript characters** and the **decimal exponent symbol**: the former denotes a numerical base or an index (an arbitrary integer), whereas the latter introduces the exponent of a number in the scientific notation, being a fixed token. An example of an attempted decomposition of the proposed symbol causing a change in semantics:

“ $a_{10} - b$ ” means, as a meta-syntactic expression, “ $a \times 10^{-b}$ ”, whereas “ $a_{10} - b$ ” means “*element(a, 10)-b*”.

Consequently, the **decimal exponent symbol** should have no numerical value and should belong to the same category and class as the **plus sign** U+002B and the **minus sign** U+2212 that can also appear in the scientific notation between the mantissa and the exponent, namely “Symbol, Math” and “European Number Separator”.

The importance of adequately representing numerical data encoded in GOST 10859-64, as

well as the ALGOL-60 and ALGOL-68 programs, minimizing the probability of a human error, necessitates encoding of the **decimal exponent symbol** as a separate non-decomposable entity. A position within the “Miscellaneous Technical” range, where the APL-specific characters already reside (tentatively U+23E8) would be desirable.

Furthermore, consistently using **decimal exponent symbol** instead of 'E' or 'e' in appropriate contexts might simplify automatic processing of documents, avoiding potential ambiguity with hexadecimal numbers.

Satisfied criteria for encoding: H.3.1 (scientific/engineering notation), H.4 (1) character specific to programming languages (ALGOL-60, ALGOL-68), 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 (formatted output for monospaced drum printers), part of a notational system (scientific number notation), has well-defined semantics that lend to computer processing, H.8 completes a class of symbols already in the standard.

**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 decimal exponent symbol from GOST 10859-64
2. Requester's name:	Leonid Broukhis
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution
4. Submission date:	2008-01-22
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:	2300-23FF Miscellaneous Technical	
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"/>
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?	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/DecimalExponent.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?	emulated	
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:

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.

23E8;DECIMAL EXPONENT SYMBOL;Sm;0;ES;;;N;ALGOL LOWER TEN TOKEN;;;;

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)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>no</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>yes</i> <i>BESM-6 emulation enthusiasts</i> <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? Reference:	<i>yes</i> <i>see proposal text</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>rare</i> <i>see proposal text</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>yes</i> <i>Users of ALGOL and 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? If YES, is a rationale provided? If YES, reference:	<i>yes</i> <i>yes</i> <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? 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 an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>yes</i> <i>yes</i> <i>provided in proposal</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 character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	<i>no</i>