

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

Doc Type: Working Group Document

Title: Draft PDAM text for Amendment 1 to ISO/IEC 10646-1:2000

Source: Michael Everson, Contributing Editor

Status: Expert Contribution

Action: For consideration by JTC1/SC2/WG2

Date: 2000-09-25

This document is to be used together with N2285 and N2281.

Note that this is a draft which will be updated according to the resolutions taken at the Athens meeting of WG2.

Information technology — Universal Multiple-Octet Coded Character Set (UCS) —

Part 1:

Architecture and Basic Multilingual Plane

AMENDMENT 1: Mathematical symbols and other characters

Item 1. Restriction of code positions

Page 6, Figure 2

In the legend “Private use planes 0F, 10, E0-FF” delete “E0-FF”, and delete the upward arrows pointing to planes E0 to FF.

Page 8, clause 9

Amend 9.1 to read as follows:

9.1 Planes reserved for future standardization

Planes 11 to D~~FF~~ in Group 00 and all planes in all other groups (i.e. Planes 00 to FF in Groups 01 to 57F) are reserved for future standardization, and thus those code positions shall not be used for any other purpose.

Code positions in these planes do not have a mapping to the UTF-16 form (see Annex C).

NOTE - To ensure continued inter-operability between the UTF-16 form and other coded representations of the UCS, it is intended that no characters will ever be allocated to code positions above 0010 FFFF.

In 9.2 (Planes accessible by UTF-16) delete 2nd paragraph, which reads as follows:

Code positions in Planes 11 to FF of Group 00, or in Planes 00 to FF of other groups, do not have a mapping to the UTF-16 form.

Page 8, clause 10

In 10.2 (Code positions for private use characters) amend the first two paragraphs to read as follows:

~~The code positions of the 32 groups from Group 60 to Group 7F shall be for private use.~~

The code positions of Plane 0F and Plane 10, ~~and of the 32 planes from Plane E0 to Plane FF,~~ of Group 00 shall be for private use.

Item 2. Extension of U+ notation

Page 8, clause 6

Change “four-digit form” and “4-digit form” to “four-to-six-digit form”.

In clause 6.5.b change “It is not defined if the first four digits of the eight-digit form are not all zeros” to “It is not defined if the eight-digit form is greater than U-0010FFFF” and append at the end of the text “Leading zeros are suppressed for values greater than U-0000FFFF.”.

Change “{+}xxxx” in the BNF form to “{+}(xxxx | xxxxx | xxxxxx)”.

Item 3. UCS Sequence Identifiers.

Add a new clause 6.6 (UCS Sequence Identifiers):

“An entity that is represented by a sequence of ‘n’ code positions from the standard, is identified by a UCS Sequence Identifier (USI) having the following form:

<UID1, UID2, UID3, .. UIDn>

where UID1, UID2, etc. represent the unique identifiers of the corresponding characters from the standards, in the same sequence as needed to represent the identified entity. The syntax for UID1, UID2, etc. is specified in clause 6.5. A COMMA character (optionally followed by a SPACE character) separates the UIDs. The UCS Sequence Identifier shall include at least two UIDs and begins with a LESS-THAN SIGN and is terminated by a GREATER-THAN SIGN.”

Item 3. Reserved characters for internal processing uses

Clause 7

Remove the note.

Clause 8

Replace the last paragraph as follows:

“Code positions U-0000FDD0 to U-0000FDEF, U-0000FFFE, and U-0000FFFF are reserved.

NOTE - Code position U-0000FFFE is reserved for “signature” (see annex H). Code positions U-0000FDD0 to U-0000FDEF and U-0000FFFF can be used for internal processing uses requiring numeric values that are guaranteed not to be coded characters, such as in terminating tables, or signaling end-of-text. Furthermore, since U-0000FFFF is the largest BMP value, it may also be used as the final value in binary or sequential searching index withing the context of the BMP.”

Change the annotations for the characters at FDD0–FDEF and FFFE and FFFF to “(This position is permanently reserved)”.

Item 4. Collections for MES

At the end of Annex A add a new clause as follows.

A.4 Other collections within the BMP

The collections specified within this clause are entirely within Plane 00.

NOTE - The acronym MES indicates Multilingual European Subset.

A4.1 Collection Name: MES-1

Type of Collection: fixed

Number of characters: 335

Rows Positions (cells)

00	20-7E A0-FF
01	00-13 16-2B 2E-4D 50-7E
02	C7 D8-DB DD
20	15 18-19 1C-1D AC
21	22 26 5B-5E 90-93
26	6A

A4.2 Collection Name: MES-2

Type of Collection: fixed

Number of characters: 1052

Rows Positions (cells)

00	20-7E A0-FF
01	00-7F 8F 92 B7 DE-EF FA-FF
02	18-1B 1E-1F 59 7C 92 BB-BD C6-C7 C9 D8-DD EE
03	74-75 7A 7E 84-8A 8C 8E-A1 A3-CE D7 DA-E1
04	00-5F 90-C4 C7-C8 CB-CC D0-EB EE-F5 F8-F9
1E	02-03 0A-0B 1E-1F 40-41 56-57 60-61 6A-6B 80-85 9B F2-F3

1F	00-15 18-1D 20-45 48-4D 50-57 59 5B 5D 5F-7D 80-B4 B6-C4 C6-D3 D6-DB DD-EF F2-F4 F6-FE
20	13-15 17-1E 20-22 26 30 32-33 39-3A 3C 3E 44 4A 7F 82 A3-A4 A7 AC AF
21	05 16 22 26 5B-5E 90-95 A8
22	00 02-03 06 08-09 0F 11-12 19-1A 1E-1F 27-2B 48 59 60-61 64-65 82-83 95 97
23	02 10 20-21 29-2A
25	00 02 0C 10 14 18 1C 24 2C 34 3C 50-6C 80 84 88 8C 90-93 A0 AC B2 BA BC C4 CA-CB D8-D9
26	3A-3C 40 42 60 63 65-66 6A-6B
FB	01-02
FF	FD

A4.3 Collection Name: MES-3A

Type of Collection: not fixed.

This collection consists of the following collections from A.1 above.

Collection number and name

1	BASIC LATIN
2	LATIN-1 SUPPLEMENT
3	LATIN EXTENDED-A
4	LATIN EXTENDED-B
5	IPA EXTENSIONS
6	SPACING MODIFIER LETTERS
7	COMBINING DIACRITICAL MARKS
8	BASIC GREEK
9	GREEK SYMBOLS AND COPTIC
10	CYRILLIC
11	ARMENIAN
27	BASIC GEORGIAN
30	LATIN EXTENDED ADDITIONAL
31	GREEK EXTENDED
32	GENERAL PUNCTUATION
33	SUPERSCRIPTS AND SUBSCRIPTS
34	CURRENCY SYMBOLS
35	COMBINING DIACRITICAL MARKS FOR SYMBOLS
36	LETTERLIKE SYMBOLS
37	NUMBER FORMS
38	ARROWS
39	MATHEMATICAL OPERATORS
40	MISCELLANEOUS TECHNICAL
42	OPTICAL CHARACTER RECOGNITION
44	BOX DRAWING
45	BLOCK ELEMENTS
46	GEOMETRIC SHAPES
47	MISCELLANEOUS SYMBOLS
63	ALPHABETIC PRESENTATION FORMS
65	COMBINING HALF MARKS
70	SPECIALS

A4.4 Collection Name: MES-3B

Type of Collection: fixed
 Number of characters: 2819
 Rows Positions (cells)

00	20-7E A0-FF
01	00-FF
02	00-1F 22-33 50-AD B0-EE
03	00-4E 60-62 74-75 7A 7E 84-8A 8C 8E-A1 A3-CE D0-D7 DA-F3
04	00-86 88-89 8C-C4 C7-C8 CB-CC D0-F5 F8-F9
05	31-56 59-5F 61-87 89-8A
10	D0-F6 FB
1E	00-9B A0-F9
1F	00-15 18-1D 20-45 48-4D 50-57 59 5B 5D 5F-7D 80-B4 B6-C4 C6-D3 D6-DB DD-EF F2-F4 F6-FE
20	00-46 48-4D 6A-70 74-8E A0-AF D0-E3
21	00-3A 53-83 90-F3
22	00-F1
23	00-7B 7D-9A
24	40-4A
25	00-95 A0-F7
26	00-13 19-71
FB	00-06 13-17
FE	20-23
FF	F9-FD

Item 5. Various additional characters

1. Modifications to existing tables.

Insert the following additional character glyphs and names at the indicated positions in the tables given below in Item X, the character names replacing the existing entries which read “(This position shall not be used)”. Note that the table numbers in that item are to be corrected to the existing table numbers in the standard.

[EDITOR’S NOTE: In the tables below, the following code positions indicate the new characters added to those tables:

0363-036F, 03D8-03D9, 03F4-03F6, 048A-048B, 04C5-04C6, 04C9-04CA, 04CD-04CE, 07B1, 17DD, 2047, 204E-2051, 2057, 205F-2062, 2071, 20B0-20B1, 20E4-20E8, 213D-214B, 21F5-21FF, 22F2-22FF, 237C, 239B-23CE, 24EB-24FE, 2596-259F, 25F8-25FF, 2616-2617, 2672-2679, 303B-303D, 3095-3096, 309F, 30A0, 30FF, 3251-325F, 32B1-32BF, A4A2-A4A3, A4B4, A4C1, A4C5, FA30-FA68, FE45-FE46

The following code positions indicate new tables to be added:

0500-052F, 1700-171F, 1720-173F, 1740-175F, 1760-177F, 2900-297F, 2980-29FF,

2A00-2A7F, 2A80-2AFF, 31F0-31FF, FE00-FE0F.]

B. New tables.

After page 279

Insert the following additional character glyphs and names at the indicated positions in the tables given below in Item 4, the new character names replacing the existing entries at the specified code locations which read “(This position shall not be used)”. Note that the table numbers in Item 4 are to be corrected to the new table numbers specified below.

Table 144 - Row 03: CYRILLIC SUPPLEMENTARY

Table 145 - Row 17: TAGALOG

Table 146 - Row 17: HANUNOO

Table 147 - Row 17: BUHID

Table 148 - Row 17: TAGBANWA

Table 149 - Row 29: SUPPLEMENTAL ARROWS

Table 150 - Row 29: MISCELLANEOUS MATHEMATICAL SYMBOLS

Table 151 - Row 2A: SUPPLEMENTAL MATHEMATICAL OPERATORS

Table 152 - Row 2A: SUPPLEMENTAL MATHEMATICAL OPERATORS

Table 153 - Row 31: KATAKANA PHONETIC EXTENSIONS

TABLE 154 - Row FE: VARIATION SELECTORS

3. Pages 904 ff, Annex G Alphabetically sorted list of character names

Insert each of the character name entries from Item X below at the appropriate position, ordered alphabetically by the character name, in the list of character names in Annex G.

Page 880, Annex A.1

In the list of collection numbers and names, after 91 TIBETAN

insert new entries as follows:

92 CYRILLIC SUPPLEMENTARY

93 TAGALOG

94 HANUNOO

95 BUHID

96 TAGBANWA

97 SUPPLEMENTAL ARROWS

98 MISCELLANEOUS MATHEMATICAL SYMBOLS

99 SUPPLEMENTAL MATHEMATICAL OPERATORS

100 KATAKANA PHONETIC EXTENSIONS

101 VARIATION SELECTORS

Item 6. Sources for characters**Page XXX. Annex X.**

Add to Annex XXXXX the following references, interfiled in alphabetical order:

{TO BE COMPLETED}

JIS X 0213:2000

Komi dictionary

Philippine - Santos

Math stuff - forthcoming Mathematica documentation

Terminal graphics - too many sources?

Combining letters - too many sources?

2228, 2273 2235 2263 2264 2265 2240 2266 1933
2224 2256 2278

Item 7. Addition to Annex P.

Add an asterisk to the KOPPAs.

The name of this character distinguishes it from 03xx GREEK LETTER KOPPA, which is most commonly used with its numeric value, such as in the dating of legal documentation. GREEK LETTER ARCHAIC KOPPA is primarily used alphabetically to represent the letter used in early Greek inscriptions, etc.

The name of this character distinguishes it from 03xx GREEK SMALL LETTER KOPPA, which is most commonly used with its numeric value, such as in the dating of legal documentation. GREEK SMALL LETTER ARCHAIC KOPPA is primarily used alphabetically to represent the letter used in early Greek inscriptions, etc.

Item 7. Code tables and name lists

See below.

[These will be based on N2281 with updates as agreed in Athens.]