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

Part 1:

Architecture and Basic Multilingual Plane

TECHNICAL CORRIGENDUM 2

Technologies de l'information — Jeu universel de caractères codés à plusieurs octets —

Partie 1: Architecture et table multilingue

RECTIFICATIF TECHNIQUE 2

Technical Corrigendum 2 to International Standard ISO/IEC 10646-1:1993 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology,* Subcommittee SC 2, *Coded character sets.*

_____

4 Definitions

Replace 4.2 (block) with:

4.2 block: A contiguous range of code positions to which a set of characters that share common characteristics, such as script, are allocated. A block cannot overlap another block. One or more of the code positions within a block may have no character allocated to it.

Renumber 4.11 to 4.15 as 4.12 to 4.16.

Renumber 4.16 to 4.37 as 4.18 to 4.39 (see Amendment 1).

After 4.10 (code table) insert new definition as follows:

4.11 collection: A set of coded characters which is numbered and named and which consists of those coded characters whose code positions lie within one or more identified ranges.

NOTE - If any of the identified ranges include code positions to which no character is allocated, the repertoire of the collection will change if an additional character is assigned to any of those positions at a future amendment of this part of ISO/IEC 10646. However it is intended that the collection number and name will remain unchanged in future editions of this part of ISO/IEC 10646.

After 4.16 (default state) insert new definition as follows:

4.17 fixed collection: A collection in which every code position within the identified range(s) has a character allocated to it, and which is intended to remain unchanged in future editions of this part of ISO/IEC 10646.

19 Block names

Replace entire clause with:

Named blocks of contiguous code positions are specified within a plane for the purpose of allocation of characters sharing some common characteristic, such as script. The blocks specified within the BMP are listed in A.2 of Annex A, and are illustrated in Figures 3 and 4 (see Amendment 5).

Annex A

Replace Annex A with revised text as on the following pages.

Annex A
(normative)

Collections of graphic characters for subsets

A.1 Collections of coded graphic characters

The collections listed below are ordered by collection number. An * in the "positions" column indicates that the collection is a fixed collection.

See Note 2 for an alphabetically-ordered index of the principal terms used in the names of these collections.

> NOTE 1 - Use of implementation levels 1 and 2 restricts the repertoire of some character collections (see 23.4). Collections which include combining characters are 7, 10, 13 to 26, 35, 49 , 50, 63, 65 and 72.

The following collections are from the Basic Multilingual Plane.

Collection number and name	Positions
1 BASIC LATIN	0020 - 007E *
2 LATIN-1 SUPPLEMENT	00A0 - 00FF *
3 LATIN EXTENDED-A	0100 - 017F *
4 LATIN EXTENDED-B	0180 - 024F
5 IPA EXTENSIONS	0250 - 02AF
6 SPACING MODIFIER LETTERS	02B0 - 02FF
7 COMBINING DIACRITICAL MARKS	0300 - 036F
8 BASIC GREEK	0370 - 03CF
9 GREEK SYMBOLS AND COPTIC	03D0 - 03FF
10 CYRILLIC	0400 - 04FF
11 ARMENIAN	0530 - 058F
12 BASIC HEBREW	05D0 - 05EA*
13 HEBREW EXTENDED	0590 - 05CF 05EB - 05FF
14 BASIC ARABIC	0600 - 065F
15 ARABIC EXTENDED	0660 - 06FF
16 DEVANAGARI	0900 - 097F 200C, 200D
17 BENGALI	0980 - 09FF 200C, 200D
18 GURMUKHI	0A00 - 0A7F 200C, 200D
19 GUJARATI	0A80 - 0AFF 200C, 200D

20 ORIYA	0B00 - 0B7F 200C, 200D
21 TAMIL	0B80 - 0BFF 200C, 200D
22 TELUGU	0C00 - 0C7F 200C, 200D
23 KANNADA	0C80 - 0CFF 200C, 200D
24 MALAYALAM	0D00 - 0D7F 200C, 200D
25 THAI	0E00 - 0E7F
26 LAO	0E80 - 0EFF
27 BASIC GEORGIAN	10D0 - 10FF
28 GEORGIAN EXTENDED	10A0 - 10CF
29 HANGUL JAMO	1100 - 11FF
30 LATIN EXTENDED ADDITIONAL	1E00 - 1EFF
31 GREEK EXTENDED	1F00 - 1FFF
32 GENERAL PUNCTUATION	2000 - 206F
33 SUPERSCRIPTS AND SUBSCRIPTS	2070 - 209F
34 CURRENCY SYMBOLS	20A0 - 20CF
35 COMBINING DIACRITICAL MARKS FOR SYMBOLS	20D0 - 20FF
36 LETTERLIKE SYMBOLS	2100 - 214F
37 NUMBER FORMS	2150 - 218F
38 ARROWS	2190 - 21FF
39 MATHEMATICAL OPERATORS	2200 - 22FF
40 MISCELLANEOUS TECHNICAL	2300 - 23FF
41 CONTROL PICTURES	2400 - 243F
42 OPTICAL CHARACTER RECOGNITION	2440 - 245F
43 ENCLOSED ALPHANUMERICS	2460 - 24FF
44 BOX DRAWING	2500 - 257F*
45 BLOCK ELEMENTS	2580 - 259F
46 GEOMETRIC SHAPES	25A0 - 25FF
47 MISCELLANEOUS SYMBOLS	2600 - 26FF
48 DINGBATS	2700 - 27BF

49	CJK SYMBOLS AND PUNCTUATION	3000 - 303F
50	HIRAGANA	3040 - 309F
51	KATAKANA	30A0 - 30FF
52	BOPOMOFO	3100 - 312F
53	HANGUL COMPATIBILITY JAMO	3130 - 318F
54	CJK MISCELLANEOUS	3190 - 319F
55	ENCLOSED CJK LETTERS AND MONTHS	3200 - 32FF
56	CJK COMPATIBILITY	3300 - 33FF
57	[deleted at AMD.5]	
58	[deleted at AMD.5]	
59	[deleted at AMD.5]	
60	CJK UNIFIED IDEOGRAPHS	4E00 - 9FFF
61	PRIVATE USE AREA	E000 - F8FF
62	CJK COMPATIBILITY IDEOGRAPHS	F900 - FAFF
63	ALPHABETIC PRESENTATION FORMS	FB00 - FB4F
64	ARABIC PRESENTATION FORMS-A	FB50 - FDFF
65	COMBINING HALF MARKS	FE20 - FE2F
66	CJK COMPATIBILITY FORMS	FE30 - FE4F
67	SMALL FORM VARIANTS	FE50 - FE6F
68	ARABIC PRESENTATION FORMS-B	FE70 - FEFE
69	HALFWIDTH AND FULLWIDTH FORMS	FF00 - FFEF
70	SPECIALS	FFF0 - FFFD
71	HANGUL SYLLABLES	AC00 - D7A3*
72	BASIC TIBETAN	0F00 - 0FBF

The following collections specify characters used for alternate formats and script-specific formats. See annex D for more information.

200	ZERO-WIDTH BOUNDARY INDICATORS	200B - 200D FEFF
201	FORMAT SEPARATORS	2028 - 2029
202	BI-DIRECTIONAL FORMAT MARKS	200E - 200F
203	BI-DIRECTIONAL FORMAT EMBEDDINGS	202A - 202EI
204	HANGUL FILL CHARACTERS	3164, FFA0

205	CHARACTER SHAPING SELECTORS	206A - 206D
206	NUMERIC SHAPE SELECTORS	206E - 206F

The following specify collections which are the union of particular collections defined above.

250	GENERAL FORMAT CHARACTERS	Collections 200 - 203
251	SCRIPT-SPECIFIC FORMAT CHARACTERS	Collections 204 - 206

The following specify other collections.

270	COMBINING CHARACTERS	characters specified in annex B.1
271	COMBINING CHARACTERS B-2	characters specified in annex B.2
[299	BMP FIRST EDITION]	see A.3 *
300	BMP	0000 - D7FF E000 - FFFD
301	BMP-AMD.7	see A.3 *

The following collections are outside the Basic Multilingual Plane.

400	PRIVATE USE PLANES	G=00, P=0F, 10 & E0 - FF
500	PRIVATE USE GROUPS	G=60 - 7F

NOTE 2 - The principal terms (keywords) used in the collection names shown above are listed below in alphabetical order. The entry for a term shows the collection number of every collection whose name includes the term. These terms do not provide a complete cross-reference to all the collections where characters sharing a particular attribute, such as script name, may be found. Although most of the terms identify an attribute of the characters within the collection, some characters that possess that attribute may be present in other collections whose numbers do not appear in the entry for that term.

Alphabetic	63
Alphanumeric	43
Arabic	14 15 64 68
Armenian	11
Arrows	38
Bengali	17
Bi-directional	202 203
Block elements	45
BMP	300 301 (299)
Box drawing	44
Bopomofo	52
CJK	49 54 55 56 60 62 66
Combining	7 35 65 270 271
Compatibility	53 56 62 66
Control pictures	41
Coptic	9

Currency	34
Cyrillic	10
Devanagari	16
Diacritical marks	7 35
Dingbats	48
Enclosed	43 55
Format	201 202 203 250 251
Fullwidth	69
Geometric shapes	46
Georgian	27 28
Greek	8 9 31
Gujarati	19
Gurmukhi	18
Half (marks, width)	65 69
Hangul	29 53 71 204
Hebrew	12 13
Hiragana	50
Ideographs	60 62
IPA extensions	5
Jamo	29 53
Kannada	23
Katakana	51
Lao	26
Latin	1 2 3 4 30
Letter	36 55
Malayalam	24
Mathematical operators	39
Months	55
Number	37
Optical character recognition	
	42
Oriya	20
Presentation forms	63 64 68
Private use	61 400 500
Punctuation	32 49
Shape, shaping	205 206
Small form	67
Spacing modifier	6
Specials	70
Subscripts, superscripts	33
Syllables	71
Symbols	9 34 35 36 47 49
Tamil	21
Technical	40
Telugu	22
Thai	25
Tibetan	72
Zero-width	200

A.2 Blocks in the BMP

The following blocks are specified in the Basic Multilingual Plane. They are ordered by code position.

Block name	from	to
BASIC LATIN	0020	007E
LATIN-1 SUPPLEMENT	00A0	00FF
LATIN EXTENDED-A	0100	017F
LATIN EXTENDED-B	0180	024F
IPA EXTENSIONS	0250	02AF
SPACING MODIFIER LETTERS	02B0	02FF
COMBINING DIACRITICAL MARKS	0300	036F
BASIC GREEK	0370	03CF
GREEK SYMBOLS AND COPTIC	03D0	03FF

CYRILLIC	0400 - 04FF
ARMENIAN	0530 - 058F
HEBREW	0590 - 05FF
BASIC ARABIC	0600 - 065F
ARABIC EXTENDED	0660 - 06FF
DEVANAGARI	0900 - 097F
BENGALI	0980 - 09FF
GURMUKHI	0A00 - 0A7F
GUJARATI	0A80 - 0AFF
ORIYA	0B00 - 0B7F
TAMIL	0B80 - 0BFF
TELUGU	0C00 - 0C7F
KANNADA	0C80 - 0CFF
MALAYALAM	0D00 - 0D7F
THAI	0E00 - 0E7F
LAO	0E80 - 0EFF
BASIC TIBETAN	0F00 - 0FBF
GEORGIAN EXTENDED	10A0 - 10CF
BASIC GEORGIAN	10D0 - 10FF
HANGUL JAMO	1100 - 11FF
LATIN EXTENDED ADDITIONAL	1E00 - 1EFF
GREEK EXTENDED	1F00 - 1FFF
GENERAL PUNCTUATION	2000 - 206F
SUPERSCRIPTS AND SUBSCRIPTS	2070 - 209F
CURRENCY SYMBOLS	20A0 - 20CF
COMBINING DIACRITICAL MARKS FOR SYMBOLS	
	20D0 - 20FF
LETTERLIKE SYMBOLS	2100 - 214F
NUMBER FORMS	2150 - 218F
ARROWS	2190 - 21FF
MATHEMATICAL OPERATORS	2200 - 22FF
MISCELLANEOUS TECHNICAL	2300 - 23FF
CONTROL PICTURES	2400 - 243F
OPTICAL CHARACTER RECOGNITION	2440 - 245F
ENCLOSED ALPHANUMERICS	2460 - 24FF
BOX DRAWING	2500 - 257F
BLOCK ELEMENTS	2580 - 259F
GEOMETRIC SHAPES	25A0 - 25FF
MISCELLANEOUS SYMBOLS	2600 - 26FF
DINGBATS	2700 - 27BF
CJK SYMBOLS AND PUNCTUATION	3000 - 303F
HIRAGANA	3040 - 309F
KATAKANA	30A0 - 30FF
BOPOMOFO	3100 - 312F
HANGUL COMPATIBILITY JAMO	3130 - 318F
CJK MISCELLANEOUS	3190 - 319F
ENCLOSED CJK LETTERS AND MONTHS	
	3200 - 32FF
CJK COMPATIBILITY	3300 - 33FF
CJK UNIFIED IDEOGRAPHS	4E00 - 9FFF
HANGUL SYLLABLES	AC00 - D7A3
PRIVATE USE AREA	E000 - F8FF
CJK COMPATIBILITY IDEOGRAPHS	F900 - FAFF
ALPHABETIC PRESENTATION FORMS	FB00 - FB4F
ARABIC PRESENTATION FORMS-A	FB50 - FDFF
COMBINING HALF MARKS	FE20 - FE2F
CJK COMPATIBILITY FORMS	FE30 - FE4F
SMALL FORM VARIANTS	FE50 - FE6F
ARABIC PRESENTATION FORMS-B	FE70 - FEFE
HALFWIDTH AND FULLWIDTH FORMS	FF00 - FFEF
SPECIALS	FFF0 - FFFD

A.3 Fixed collections of the whole BMP

The collection 301 BMP-AMD.7 is specified below as a fixed collection (4.17). It comprises only those coded characters that were in the BMP after amendments up to, but not after, Amendment 7 were applied to this part of ISO/IEC 10646. Accordingly the repertoire of this collection is not subject to change if new characters are added to the BMP by any subsequent amendments.

> NOTE - The repertoire of the collection 300 BMP is subject to change if new characters are added to the BMP by an amendment to this part of ISO/IEC 10646.

301 BMP-AMD.7 is specified by the following ranges of code positions as indicated for each row or contiguous series of rows.

Rows	Positions (cells)
00	20-7E A0-FF
01	00-F5 FA-FF
02	00-17 50-A8 B0-DE E0-E9
03	00-45 60-61 74-75 7A 7E 84-8A 8C 8E-A1 A3-CE D0-D6 DA DC DE E0 E2-F3
04	01-0C 0E-4F 51-5C 5E-86 90-C4 C7-C8 CB-CC DA-EB EE-F5 F8-F9
05	31-56 59-5F 61-87 89 91-A1 A3-B9 BB-C4 D0-EA F0-F4
06	0C 1B 1F 21-3A 40-52 60-6D 70-B7 BA-BE C0-CE D0-ED F0-F9
09	01-03 05-39 3C-4D 50-54 58-70 81-83 85-8C 8F-90 93-A8 AA B0 B2 B6-B9 BC BE-C4 C7-C8 CB-CD D7 DC-DD DF-E3 E6-FA
0A	02 05-0A 0F-10 13-28 2A-30 32-33 35-36 38-39 3C 3E-42 47-48 4B-4D 59-5C 5E 66-74 81-83 85-8B 8D 8F-91 93-A8 AA-B0 B2-B3 B5-B9 BC-C5 C7-C9 CB-CD D0 E0 E6-EF
0B	01-03 05-0C 0F-10 13-28 2A-30 32-33 36-39 3C-43 47-48 4B-4D 56-57 5C-5D 5F-61 66-70 82-83 85-8A 8E-90 92-25 99-9A 9C 9E-9F A3-A4 A8-AA AE-B5 B7-B9 BE-C2 C6-C8 CA-CD D7 E7-F2
0C	01-03 05-0C 0E-10 12-28 2A-33 35-39 3E-44 46-48 4A-4D 55-56 60-61 66-6F 82-83 85-8C 8E-90 92-A8 AA-B3 B5-B9 BE-C4 C6-C8 CA-CD D5-D6 DE E0-E1 E6-EF
0D	02-03 05-0C 0E-10 12-28 2A-39 3E-43 46-48 4A-4D 57 60-61 66-6F
0E	01-3A 3F-5B 81-82 84 87-88 8A 8D 94-97 99-9F A1-A3 A5 A7 AA-AB AD-B9 BB-BD C0-C4 C6 C8-CD D0-D9 DC-DD
0F	00-47 49-69 71-8B 90-95 97 99-AD B1-B7 B9
10	A0-C5 D0-F6 FB
11	00-59 5F-A2 A8-F9
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	00-2E 30-46 6A-70 74-8E A0-AB D0-E1
21	00-38 53-82 90-EA
22	00-F1
23	30 32-7A
24	00-24 40-4A 60-EA
25	00-95 A0-EF
26	00-13 1A-6F
27	01-04 06-09 0C-27 29-4B 4D 4F-52 56 58-5E 61-67 76-94 98-AF B1-BE
30	00-37 3F 41-94 99-9E A1-FE
31	05-2C 31-8E 90-9F
32	00-1C 20-43 60-7B 7F-B0 C0-CB D0-FE
33	00-76 7B-DD E0-FE
4E-9F	4E00-9FA5
AC-D7	AC00-D7A3
E0-F8	E000-F8FF
F9-FA	F900-FA2D
FB	00-06 13-17 1E-36 38-3C 3E 40-41 43-44 46-B1 D3-FF
FC	00-FF
FD	00-3F 50-8F 92-C7 F0-FB
FE	20-23 30-44 49-52 54-66 68-6B 70-72 74 76-FC FF
FF	01-5E 61-BE C2-C7 CA-CF D2-D7 DA-DC E0-E6 E8-EE FD

The collection number and collection name:

> 299 BMP FIRST EDITION

have been reserved to identify the fixed collection comprising all of the coded characters that were in the BMP in the First Edition of this part of ISO/IEC 10646. This collection is not now in conformity with this part of ISO/IEC 10646.

> NOTE - The specification of collection 299 BMP FIRST EDITION consisted of the specification of collection 301 BMP-AMD.7 except for the replacement of the corresponding entries in the list above with the entries shown below:
>
> | rows | positions |
> |---|---|
> | 05 | 31-56 59-5F 61-87 89 B0-B9 BB-C3 D0-EA F0-F4 |
> | 0F | [no positions] |
> | 1E | 00-9A A0-F9 |
> | 20 | 00-2E 30-46 6A-70 74-8E A0-AA D0-E1 |
> | 4E-9F | [no positions] |
>
> and by including an additional entry:
> | 34-4D | 3400-4DFF |
>
> for the code position ranges of three collections (57, 58, 59) of coded characters which have been deleted from this part of ISO/IEC 10646 since the First Edition.