

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.

<u>Collection number and name</u>	<u>Positions</u>		
1 BASIC LATIN	0020 - 007E *	23 KANNADA	0C80 - 0CFF 200C, 200D
2 LATIN-1 SUPPLEMENT	00A0 - 00FF *	24 MALAYALAM	0D00 - 0D7F 200C, 200D
3 LATIN EXTENDED-A	0100 - 017F *	25 THAI	0E00 - 0E7F
4 LATIN EXTENDED-B	0180 - 024F	26 LAO	0E80 - 0EFF
5 IPA EXTENSIONS	0250 - 02AF	27 BASIC GEORGIAN	10D0 - 10FF
6 SPACING MODIFIER LETTERS	02B0 - 02FF	28 GEORGIAN EXTENDED	10A0 - 10CF
7 COMBINING DIACRITICAL MARKS	0300 - 036F	29 HANGUL JAMO	1100 - 11FF
8 BASIC GREEK	0370 - 03CF	30 LATIN EXTENDED ADDITIONAL	1E00 - 1EFF
9 GREEK SYMBOLS AND COPTIC	03D0 - 03FF	31 GREEK EXTENDED	1F00 - 1FFF
10 CYRILLIC	0400 - 04FF	32 GENERAL PUNCTUATION	2000 - 206F
11 ARMENIAN	0530 - 058F	33 SUPERSCRIPTS AND SUBSCRIPTS	2070 - 209F
12 BASIC HEBREW	05D0 - 05EA *	34 CURRENCY SYMBOLS	20A0 - 20CF
13 HEBREW EXTENDED	0590 - 05CF 05EB - 05FF	35 COMBINING DIACRITICAL MARKS FOR SYMBOLS	20D0 - 20FF
14 BASIC ARABIC	0600 - 065F	36 LETTERLIKE SYMBOLS	2100 - 214F
15 ARABIC EXTENDED	0660 - 06FF	37 NUMBER FORMS	2150 - 218F
16 DEVANAGARI	0900 - 097F 200C, 200D	38 ARROWS	2190 - 21FF
17 BENGALI	0980 - 09FF 200C, 200D	39 MATHEMATICAL OPERATORS	2200 - 22FF
18 GURMUKHI	0A00 - 0A7F 200C, 200D	40 MISCELLANEOUS TECHNICAL	2300 - 23FF
19 GUJARATI	0A80 - 0AFF 200C, 200D	41 CONTROL PICTURES	2400 - 243F
20 ORIYA	0B00 - 0B7F 200C, 200D	42 OPTICAL CHARACTER RECOGNITION	2440 - 245F
21 TAMIL	0B80 - 0BFF 200C, 200D	43 ENCLOSED ALPHANUMERIC	2460 - 24FF
22 TELUGU	0C00 - 0C7F 200C, 200D	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	85	SYRIAC	0700 - 074F
52	BOPOMOFO	3100 - 312F 31A0 - 31BF	86	THAANA	0780 - 07BF
53	HANGUL COMPATIBILITY JAMO	3130 - 318F	87	BASIC MYANMAR	1000 - 104F 200C, 200D
54	CJK MISCELLANEOUS	3190 - 319F	88	KHMER	1780 - 17FF 200C, 200D
55	ENCLOSED CJK LETTERS AND MONTHS	3200 - 32FF	89	MONGOLIAN	1800 - 18AF
56	CJK COMPATIBILITY	3300 - 33FF	90	EXTENDED MYANMAR	1050 - 109F
57, 58, 59	(These collection numbers shall not be used, see Note 2.)		91	TIBETAN	0F00 - 0FFF
60	CJK UNIFIED IDEOGRAPHS	4E00 - 9FFF	The following collections specify characters used for alternate formats and script-specific formats. See annex F for more information.		
61	PRIVATE USE AREA	E000 - F8FF	200	ZERO-WIDTH BOUNDARY INDICATORS	200B - 200D FEFF
62	CJK COMPATIBILITY IDEOGRAPHS	F900 - FAFF	201	FORMAT SEPARATORS	2028 - 2029
63	ALPHABETIC PRESENTATION FORMS	FB00 - FB4F	202	BI-DIRECTIONAL FORMAT MARKS	200E - 200F
64	ARABIC PRESENTATION FORMS-A	FB50 - FDFF	203	BI-DIRECTIONAL FORMAT EMBEDDINGS	202A - 202E
65	COMBINING HALF MARKS	FE20 - FE2F	204	HANGUL FILL CHARACTERS	3164, FFA0
66	CJK COMPATIBILITY FORMS	FE30 - FE4F	205	CHARACTER SHAPING SELECTORS	206A - 206D
67	SMALL FORM VARIANTS	FE50 - FE6F	206	NUMERIC SHAPE SELECTORS	206E - 206F
68	ARABIC PRESENTATION FORMS-B	FE70 - FEFE	207	IDEOGRAPHIC DESCRIPTION CHARACTERS	2FF0 - 2FFF
69	HALFWIDTH AND FULLWIDTH FORMS	FF00 - FFEF	The following specify collections which are the union of particular collections defined above.		
70	SPECIALS	FFF0 - FFFD	250	GENERAL FORMAT CHARACTERS	Collections 200 - 203
71	HANGUL SYLLABLES	AC00 - D7A3 *	251	SCRIPT-SPECIFIC FORMAT CHARACTERS	Collections 204 - 207
72	BASIC TIBETAN	0F00 - 0FBF	The following specify other collections.		
73	ETHIOPIC	1200 - 137F	270	COMBINING CHARACTERS	characters specified in annex B.1
74	UNIFIED CANADIAN ABORIGINAL SYLLABICS	1400 - 167F	271	COMBINING CHARACTERS B-2	characters specified in annex B.2
75	CHEROKEE	13A0 - 13FF	299	(This collection number shall not be used, see A.3.2.)	
76	YI SYLLABLES	A000 - A48F	300	BMP	0000 - D7FF E000 - FFFD
77	YI RADICALS	A490 - A4CF	301	BMP-AMD.7	see A.3.1 *
78	KANGXI RADICALS	2F00 - 2FDF			
79	CJK RADICALS SUPPLEMENT	2E80 - 2EFF			
80	BRAILLE PATTERNS	2800 - 28FF			
81	CJK UNIFIED IDEOGRAPHS EXTENSION A	3400 - 4DBF FA1F, FA23			
82	OGHAM	1680 - 169F			
83	RUNIC	16A0 - 16FF			
84	SINHALA	0D80 - 0DFF			

302 BMP SECOND EDITION see A.3.3 *

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 1 - Use of implementation levels 1 and 2 restricts the repertoire of some character collections (see 24.4). Collections which include combining characters are 7, 10, 13 to 26, 35, 49, 50, 63, 65, 72, 84, 85, 86, 87, 88, 89, 90, and 91.

NOTE 2 - Collections numbered 57, 58, and 59 were specified in the First Edition of this international Standards but have now been deleted.

NOTE 3 - 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	300 301 302 (299)
Box drawing	44
Bopomofo	52
Braille patterns	80
Canadian Aboriginal	74
Cherokee	75
CJK	49 54 55 56 60 62 66 78 81
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
Ethiopic	73
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 81 207
IPA extensions	5
Jamo	29 53
Kangxi	78
Kannada	23
Katakana	51
Khmer	88
Lao	26
Latin	1 2 3 4 30
Letter	36 55
Malayalam	24
Mathematical operators	39
Mongolian	89
Months	55
Myanmar	87 90
Number	37
Ogham	82
Optical character recognition	42
Oriya	20
Presentation forms	63 64 68
Private use	61 400 500
Punctuation	32 49
Radicals	77 78 79
Runic	83
Shape, shaping	205 206
Sinhala	84
Small form	67
Spacing modifier	6
Specials	70
Subscripts, superscripts	33
Syllables, syllabics	71 74 76
Symbols	9 34 35 36 47 49
Syriac	85
Tamil	21
Technical	40
Telugu	22
Thaana	86
Thai	25
Tibetan	72 91
Yi 76 77	
Zero-width	200

iTech STANDARD IN REVIEW
(standards.iteh.ai)
ISO/IEC 10646-1:2000
<https://standards.iteh.ai/catalog/standards/sist/2211632-59-4266-ae0-8d25aead12e2/iso-iec-10646-1-2000>

A.2 Blocks in the BMP

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

<u>Block name</u>	<u>from</u> <u>to</u>
BASIC LATIN	0020 - 007E
LATIN-1 SUPPLEMENT	00A0 - 00FF
LATIN EXTENDED-A	0100 - 017F
LATIN EXTENDED-B	0180 - 024F
IPA (INTERNATIONAL PHONETIC ALPHABET) EXTENSIONS	0250 - 02AF
SPACING MODIFIER LETTERS	02B0 - 02FF
COMBINING DIACRITICAL MARKS	0300 - 036F
GREEK AND COPTIC	0370 - 03FF
CYRILLIC	0400 - 04FF
ARMENIAN	0530 - 058F
HEBREW	0590 - 05FF
ARABIC	0600 - 06FF
SYRIAC	0700 - 074F
THAANA	0780 - 07BF

DEVANAGARI	0900 - 097F
BENGALI	0980 - 09FF
GURMUKHI	0A00 - 0A7F
GUJARATI	0A80 - 0AFF
ORIYA	0B00 - 0B7F
TAMIL	0B80 - 0BFF
TELVUGU	0C00 - 0C7F
KANNADA	0C80 - 0CFF
MALAYALAM	0D00 - 0D7F
SINHALA	0D80 - 0DFF
THAI	0E00 - 0E7F
LAO	0E80 - 0EFF
TIBETAN	0F00 - 0FFF
MYANMAR	1000 - 109F
GEORGIAN	10A0 - 10FF
HANGUL JAMO	1100 - 11FF
ETHIOPIIC	1200 - 137F
CHEROKEE	13A0 - 13FF
UNIFIED CANADIAN ABORIGINAL SYLLABICS	1400 - 167F
OGHAM	1680 - 169F
RUNIC	16A0 - 16FF
KHMER	1780 - 17FF
MONGOLIAN	1800 - 18AF
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 ENCLOSED ALPHANUMERIC	2440 - 245F
ENCLOSED ALPHANUMERIC	2460 - 24FF
BOX DRAWING	2500 - 257F
BLOCK ELEMENTS	2580 - 259F
GEOMETRIC SHAPES	25A0 - 25FF
MISCELLANEOUS SYMBOLS	2600 - 26FF
DINGBATS	2700 - 27BF
BRILLE PATTERNS	2800 - 28FF
CJK RADICALS SUPPLEMENT	2E80 - 2EFF
KANGXI RADICALS	2F00 - 2FD0
IDEOGRAPHIC DESCRIPTION CHARACTERS	2FF0 - 2FFF
CJK SYMBOLS AND PUNCTUATION	3000 - 303F
HIRAGANA	3040 - 309F
KATAKANA	30A0 - 30FF
BOPOMOFO	3100 - 312F
HANGUL COMPATIBILITY JAMO	3130 - 318F
KANBUN (CJK miscellaneous)	3190 - 319F
BOPOMOFO EXTENDED	31A0 - 31BF
ENCLOSED CJK LETTERS AND MONTHS	3200 - 32FF
CJK COMPATIBILITY	3300 - 33FF
CJK UNIFIED IDEOGRAPHS EXTENSION A	3400 - 4DBF
CJK UNIFIED IDEOGRAPHS	4E00 - 9FFF
YI SYLLABLES	A000 - A48F
YI RADICALS	A490 - A4CF
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

A.3.1 301 BMP-AMD.7

The collection 301 BMP-AMD.7 is specified below as a fixed collection (4.19). It comprises only those coded characters that were in the BMP after amendments up to, but not after, AMD.7 were applied to the First Edition of this International Standard. 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 International Standard.

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
04	A3-CE D0-D6 DA DC DE E0 E2-F3
05	01-0C 0E-4F 51-5C 5E-86 90-C4 C7-C8 CB-CC D0-EB EE-F5 F8-F9
06	31-56 59-5F 61-87 89 91-A1 A3-B9 BB-C4 D0-EA F0-F4
09	0C 1B 1F 21-3A 40-52 60-6D 70-B7 BA-BE C0-CE D0-ED F0-F9
0A	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
0B	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
0C	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
0D	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
0E	02-03 05-0C 0E-10 12-28 2A-39 3E-43 46-48 4A-4D 57 60-61 66-6F
0F	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	00 02-7A
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

A.3.2 299 BMP FIRST EDITION

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 International Standard. This collection is not now in conformity with this International Standard.

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:

<u>rows</u>	<u>positions</u>
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
AC-D7	[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 International Standard since the First Edition.

A.3.3 302 BMP SECOND EDITION

The fixed collection 302 BMP SECOND EDITION comprises only those coded characters that are in the BMP in this Second Edition of ISO/IEC 10646-1. The repertoire of this collection is not subject to change if new characters are added to the BMP by any subsequent amendments.

302 BMP SECOND EDITION is specified by the following ranges of code positions as indicated for each row or contiguous series of rows.

<u>Rows</u>	<u>Positions (cells)</u>
00	20-7E A0-FF
01	00-FF
02	00-33 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 91-A1 A3-B9 BB-C4 D0-EA F0-F4
06	0C 1B 1F 21-3A 40-55 60-6D 70-ED F0-FE
07	00-0D 0F-2C 30-4A 80-BF
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 82-83 85-96 9A-B1 B3-BB BD C0-C6 CA CF-D4 D6 D8-DF F2-F4
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 49-6A 71-8B 90-97 99-BC BE-CC CF
10	00-21 23-27 29-2A 2C-32 36-39 40-59 A0-C5 D0-F6 FB
11	00-59 5F-A2 A8-F9
12	20-26 28-46 48 4A-4D 50-56 58 5A-5D 60-86 88 8A-8D 90-AE B0 B2-B5 B8-BE C0 C2-C5 C8-CE D0-D6 D8-EE F0-FF
13	00-0E 10 12-15 18-1E 20-46 48-5A 61-7C A0-F4
14-15	1401-15FF
16	00-76 80-9C A0-F0
17	80-DC E0-E9
18	00-0E 10-19 20-77 80-A9
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	32	00-1C 20-43 60-7B 7F-B0 C0-CB D0-FE
		33	00-76 7B-DD E0-FE
20	00-46 48-4D 6A-70 74-8E A0-AF D0-E3	34-4D	3400-4DBF
21	00-3A 53-83 90-F3	4E-9F	4E00-9FA5
22	00-F1	A0-A3	A000-A3FF
23	00-7B 7D-9A	A4	00-8C 90-A1 A4-B3 B5-C0 C2-C4 C6
24	00-26 40-4A 60-EA	AC-D7	AC00-D7A3
25	00-95 A0-F7	E0-F8	E000-F8FF
26	00-13 19-71	F9-FA	F900-FA2D
27	01-04 06-09 0C-27 29-4B 4D 4F-52 56 58-5E 61-67 76-94 98-AF B1-BE	FB	00-06 13-17 1D-36 38-3C 3E 40-41 43-44 46-B1 D3-FF
28	00-FF	FC	00-FF
2E	80-99 9B-F3	FD	00-3F 50-8F 92-C7 F0-FB
2F	00-D5 F0-FB	FE	20-23 30-44 49-52 54-66 68-6B 70-72 74 76- FC FF
30	00-3A 3E-3F 41-94 99-9E A1-FE	FF	01-5E 61-BE C2-C7 CA-CF D2-D7 DA-DC E0-E6 E8-EE F9-FD
31	05-2C 31-8E 90-B7		

iTeh STANDARD PREVIEW (standards.iteh.ai)

[ISO/IEC 10646-1:2000](https://standards.iteh.ai/catalog/standards/sist/23a1f63a-fe59-4266-ae0-8d25aead12e2/iso-iec-10646-1-2000)

<https://standards.iteh.ai/catalog/standards/sist/23a1f63a-fe59-4266-ae0-8d25aead12e2/iso-iec-10646-1-2000>

Annex B (normative)

List of combining characters

B.1 List of all combining characters

The characters in the subset collections COMBINING DIACRITICAL MARKS (0300 to 036F), COMBINING DIACRITICAL MARKS FOR SYMBOLS (20D0 to 20FF), and COMBINING HALF MARKS (FE20 to FE2F) are combining characters. In addition, the following characters are combining characters.

0483	COMBINING CYRILLIC TITLO	05B7	HEBREW POINT PATAH
0484	COMBINING CYRILLIC PALATALIZATION	05B8	HEBREW POINT QAMATS
0485	COMBINING CYRILLIC DASIA PNEUMATA	05B9	HEBREW POINT HOLAM
0486	COMBINING CYRILLIC PSILI PNEUMATA	05BB	HEBREW POINT QUBUTS
0488	COMBINING CYRILLIC HUNDRED THOUSANDS SIGN	05BC	HEBREW POINT DAGESH OR MAPIQ
0489	COMBINING CYRILLIC MILLIONS SIGN	05BD	HEBREW POINT METEG
0591	HEBREW ACCENT ETNAHTA	05BF	HEBREW POINT RAFE
0592	HEBREW ACCENT SEGOL	05C1	HEBREW POINT SHIN DOT
0593	HEBREW ACCENT SHALSHELET	05C2	HEBREW POINT SIN DOT
0594	HEBREW ACCENT ZAQEF QATAN	05C4	HEBREW MARK UPPER DOT
0595	HEBREW ACCENT ZAQEF GADOL	064B	ARABIC FATHATAN
0596	HEBREW ACCENT TIPEHA	064C	ARABIC DAMMATAN
0597	HEBREW ACCENT REVIA	064D	ARABIC KASRATAN
0598	HEBREW ACCENT ZARQA	064E	ARABIC FATHA
0599	HEBREW ACCENT PASHTA	064F	ARABIC DAMMA
059A	HEBREW ACCENT YETIV	0650	ARABIC KASRA
059B	HEBREW ACCENT TEVIR	0651	ARABIC SHADDA
059C	HEBREW ACCENT GERESH	0652	ARABIC SUKUN
059D	HEBREW ACCENT GERESH MUQDAM	0653	ARABIC MADDAH ABOVE
059E	HEBREW ACCENT GERSHAYIM	0654	ARABIC HAMZA ABOVE
059F	HEBREW ACCENT QARNEY PARA	0655	ARABIC HAMZA BELOW
05A0	HEBREW ACCENT TELISHA GEDOLA	0670	ARABIC LETTER SUPERSCRIPIT ALEF
05A1	HEBREW ACCENT PAZER	06D7	ARABIC SMALL HIGH LIGATURE QAF WITH LAM WITH ALEF MAKSURA
05A3	HEBREW ACCENT MUNAH	06D8	ARABIC SMALL HIGH MEEM INITIAL FORM
05A4	HEBREW ACCENT MAHAPAKH	06D9	ARABIC SMALL HIGH LAM ALEF
05A5	HEBREW ACCENT MERKHA	06DA	ARABIC SMALL HIGH JEEM
05A6	HEBREW ACCENT MERKHA KEFULA	06DB	ARABIC SMALL HIGH THREE DOTS
05A7	HEBREW ACCENT DARGA	06DC	ARABIC SMALL HIGH SEEN
05A8	HEBREW ACCENT QADMA	06DD	ARABIC END OF AYAH
05A9	HEBREW ACCENT TELISHA QETANA	06DE	ARABIC START OF RUB EL HIZB
05AA	HEBREW ACCENT YERAH BEN YOMO	06DF	ARABIC SMALL HIGH ROUNDED ZERO
05AB	HEBREW ACCENT OLE	06E0	ARABIC SMALL HIGH UPRIGHT RECTANGULAR ZERO
05AC	HEBREW ACCENT ILUY	06E1	ARABIC SMALL HIGH DOTLESS HEAD OF KHAH
05AD	HEBREW ACCENT DEHI	06E2	ARABIC SMALL HIGH MEEM ISOLATED FORM
05AE	HEBREW ACCENT ZINOR	06E3	ARABIC SMALL LOW SEEN
05AF	HEBREW MARK MASORA CIRCLE	06E4	ARABIC SMALL HIGH MADDA
05B0	HEBREW POINT SHEVA	06E7	ARABIC SMALL HIGH YEH
05B1	HEBREW POINT HATAF SEGOL	06E8	ARABIC SMALL HIGH NOON
05B2	HEBREW POINT HATAF PATAH	06EA	ARABIC EMPTY CENTRE LOW STOP
05B3	HEBREW POINT HATAF QAMATS	06EB	ARABIC EMPTY CENTRE HIGH STOP
05B4	HEBREW POINT HIRIQ	06EC	ARABIC ROUNDED HIGH STOP WITH FILLED CENTRE
05B5	HEBREW POINT TSERE	06ED	ARABIC SMALL LOW MEEM
05B6	HEBREW POINT SEGOL	0711	SYRIAC LETTER SUPERSCRIPIT ALAPH
		0730	SYRIAC PTHAHA ABOVE
		0731	SYRIAC PTHAHA BELOW
		0732	SYRIAC PTHAHA DOTTED
		0733	SYRIAC ZQAPHA ABOVE
		0734	SYRIAC ZQAPHA BELOW
		0735	SYRIAC ZQAPHA DOTTED
		0736	SYRIAC RBASA ABOVE
		0737	SYRIAC RBASA BELOW

0738	SYRIAC DOTTED ZLAMA HORIZONTAL	09C4	BENGALI VOWEL SIGN VOCALIC RR
0739	SYRIAC DOTTED ZLAMA ANGULAR	09C7	BENGALI VOWEL SIGN E
073A	SYRIAC HBASA ABOVE	09C8	BENGALI VOWEL SIGN AI
073B	SYRIAC HBASA BELOW	09CB	BENGALI VOWEL SIGN O
073C	SYRIAC HBASA-ESASA DOTTED	09CC	BENGALI VOWEL SIGN AU
073D	SYRIAC ESASA ABOVE	09CD	BENGALI SIGN VIRAMA
073E	SYRIAC ESASA BELOW	09D7	BENGALI AU LENGTH MARK
073F	SYRIAC RWAHA	09E2	BENGALI VOWEL SIGN VOCALIC L
0740	SYRIAC FEMININE DOT	09E3	BENGALI VOWEL SIGN VOCALIC LL
0741	SYRIAC QUSHSHAYA	0A02	GURMUKHI SIGN BINDI
0742	SYRIAC RUKKAKHA	0A3C	GURMUKHI SIGN NUKTA
0743	SYRIAC TWO VERTICAL DOTS ABOVE	0A3E	GURMUKHI VOWEL SIGN AA
0744	SYRIAC TWO VERTICAL DOTS BELOW	0A3F	GURMUKHI VOWEL SIGN I
0745	SYRIAC THREE DOTS ABOVE	0A40	GURMUKHI VOWEL SIGN II
0746	SYRIAC THREE DOTS BELOW	0A41	GURMUKHI VOWEL SIGN U
0747	SYRIAC OBLIQUE LINE ABOVE	0A42	GURMUKHI VOWEL SIGN UU
0748	SYRIAC OBLIQUE LINE BELOW	0A47	GURMUKHI VOWEL SIGN EE
0749	SYRIAC MUSIC	0A48	GURMUKHI VOWEL SIGN AI
074A	SYRIAC BARREKH	0A4B	GURMUKHI VOWEL SIGN OO
07A6	THAANA ABAFILI	0A4C	GURMUKHI VOWEL SIGN AU
07A7	THAANA ABAAFILI	0A4D	GURMUKHI SIGN VIRAMA
07A8	THAANA IBIFILI	0A70	GURMUKHI TIPPI
07A9	THAANA EEBEFILI	0A71	GURMUKHI ADDAK
07AA	THAANA UBUFILI	0A81	GUJARATI SIGN CANDRABINDU
07AB	THAANA OBOOFILI	0A82	GUJARATI SIGN ANUSVARA
07AC	THAANA EBEFILI	0A83	GUJARATI SIGN VISARGA
07AD	THAANA EYBEFILI	0ABC	GUJARATI SIGN NUKTA
07AE	THAANA OBOFILI	0ABE	GUJARATI VOWEL SIGN AA
07AF	THAANA OABOAFILI	0ABF	GUJARATI VOWEL SIGN I
07B0	THAANA SUKUN	0AC0	GUJARATI VOWEL SIGN II
0901	DEVANAGARI SIGN CANDRABINDU	0AC1	GUJARATI VOWEL SIGN U
0902	DEVANAGARI SIGN ANUSVARA	0AC2	GUJARATI VOWEL SIGN UU
0903	DEVANAGARI SIGN VISARGA	0AC3	GUJARATI VOWEL SIGN VOCALIC R
093C	DEVANAGARI SIGN NUKTA	0AC4	GUJARATI VOWEL SIGN VOCALIC RR
093E	DEVANAGARI VOWEL SIGN AA	0AC5	GUJARATI VOWEL SIGN CANDRA E
093F	DEVANAGARI VOWEL SIGN I	0AC7	GUJARATI VOWEL SIGN E
0940	DEVANAGARI VOWEL SIGN II	0AC8	GUJARATI VOWEL SIGN AI
0941	DEVANAGARI VOWEL SIGN U	0AC9	GUJARATI VOWEL SIGN CANDRA O
0942	DEVANAGARI VOWEL SIGN UU	0ACB	GUJARATI VOWEL SIGN O
0943	DEVANAGARI VOWEL SIGN VOCALIC R	0ACC	GUJARATI VOWEL SIGN AU
0944	DEVANAGARI VOWEL SIGN VOCALIC RR	0ACD	GUJARATI SIGN VIRAMA
0945	DEVANAGARI VOWEL SIGN CANDRA E	0B01	ORIYA SIGN CANDRABINDU
0946	DEVANAGARI VOWEL SIGN SHORT E	0B02	ORIYA SIGN ANUSVARA
0947	DEVANAGARI VOWEL SIGN E	0B03	ORIYA SIGN VISARGA
0948	DEVANAGARI VOWEL SIGN AI	0B3C	ORIYA SIGN NUKTA
0949	DEVANAGARI VOWEL SIGN CANDRA O	0B3E	ORIYA VOWEL SIGN AA
094A	DEVANAGARI VOWEL SIGN SHORT O	0B3F	ORIYA VOWEL SIGN I
094B	DEVANAGARI VOWEL SIGN O	0B40	ORIYA VOWEL SIGN II
094C	DEVANAGARI VOWEL SIGN AU	0B41	ORIYA VOWEL SIGN U
094D	DEVANAGARI SIGN VIRAMA	0B42	ORIYA VOWEL SIGN UU
0951	DEVANAGARI STRESS SIGN UDATTA	0B43	ORIYA VOWEL SIGN VOCALIC R
0952	DEVANAGARI STRESS SIGN ANUDATTA	0B47	ORIYA VOWEL SIGN E
0953	DEVANAGARI GRAVE ACCENT	0B48	ORIYA VOWEL SIGN AI
0954	DEVANAGARI ACUTE ACCENT	0B4B	ORIYA VOWEL SIGN O
0962	DEVANAGARI VOWEL SIGN VOCALIC L	0B4C	ORIYA VOWEL SIGN AU
0963	DEVANAGARI VOWEL SIGN VOCALIC LL	0B4D	ORIYA SIGN VIRAMA
0981	BENGALI SIGN CANDRABINDU	0B56	ORIYA AI LENGTH MARK
0982	BENGALI SIGN ANUSVARA	0B57	ORIYA AU LENGTH MARK
0983	BENGALI SIGN VISARGA	0B82	TAMIL SIGN ANUSVARA
09BC	BENGALI SIGN NUKTA	0B83	TAMIL SIGN VISARGA
09BE	BENGALI VOWEL SIGN AA	0BBE	TAMIL VOWEL SIGN AA
09BF	BENGALI VOWEL SIGN I	0BBF	TAMIL VOWEL SIGN I
09C0	BENGALI VOWEL SIGN II	0BC0	TAMIL VOWEL SIGN II
09C1	BENGALI VOWEL SIGN U	0BC1	TAMIL VOWEL SIGN U
09C2	BENGALI VOWEL SIGN UU	0BC2	TAMIL VOWEL SIGN UU
09C3	BENGALI VOWEL SIGN VOCALIC R	0BC6	TAMIL VOWEL SIGN E

0BC7	TAMIL VOWEL SIGN EE	0DD2	SINHALA VOWEL SIGN KETTI IS-PILLA
0BC8	TAMIL VOWEL SIGN AI	0DD3	SINHALA VOWEL SIGN DIGA IS-PILLA
0BCA	TAMIL VOWEL SIGN O	0DD4	SINHALA VOWEL SIGN KETTI PAA-PILLA
0BCB	TAMIL VOWEL SIGN OO	0DD6	SINHALA VOWEL SIGN DIGA PAA-PILLA
0BCC	TAMIL VOWEL SIGN AU	0DD8	SINHALA VOWEL SIGN GAETTA-PILLA
0BCD	TAMIL SIGN VIRAMA	0DD9	SINHALA VOWEL SIGN KOMBUVA
0BD7	TAMIL AU LENGTH MARK	0DDA	SINHALA VOWEL SIGN DIGA KOMBUVA
0C01	TELUGU SIGN CANDRABINDU	0DDB	SINHALA VOWEL SIGN KOMBU DEKA
0C02	TELUGU SIGN ANUSVARA	0DDC	SINHALA VOWEL SIGN KOMBUVA HAA AELA-PILLA
0C03	TELUGU SIGN VISARGA	0DDD	SINHALA VOWEL SIGN KOMBUVA HAA DIGA AELA-PILLA
0C3E	TELUGU VOWEL SIGN AA	0DDE	SINHALA VOWEL SIGN KOMBUVA HAA GAYANUKITTA
0C3F	TELUGU VOWEL SIGN I	0DDF	SINHALA VOWEL SIGN GAYANUKITTA
0C40	TELUGU VOWEL SIGN II	0DF2	SINHALA VOWEL SIGN DIGA GAETTA-PILLA
0C41	TELUGU VOWEL SIGN U	0DF3	SINHALA VOWEL SIGN DIGA GAYANUKITTA
0C42	TELUGU VOWEL SIGN UU	0E31	THAI CHARACTER MAI HAN-AKAT
0C43	TELUGU VOWEL SIGN VOCALIC R	0E34	THAI CHARACTER SARA I
0C44	TELUGU VOWEL SIGN VOCALIC RR	0E35	THAI CHARACTER SARA II
0C46	TELUGU VOWEL SIGN E	0E36	THAI CHARACTER SARA UE
0C47	TELUGU VOWEL SIGN EE	0E37	THAI CHARACTER SARA UEE
0C48	TELUGU VOWEL SIGN AI	0E38	THAI CHARACTER SARA U
0C4A	TELUGU VOWEL SIGN O	0E39	THAI CHARACTER SARA UU
0C4B	TELUGU VOWEL SIGN OO	0E3A	THAI CHARACTER PHINTHU
0C4C	TELUGU VOWEL SIGN AU	0E47	THAI CHARACTER MAITAIKHU
0C4D	TELUGU SIGN VIRAMA	0E48	THAI CHARACTER MAI EK
0C55	TELUGU LENGTH MARK	0E49	THAI CHARACTER MAI THO
0C56	TELUGU AI LENGTH MARK	0E4A	THAI CHARACTER MAI TRI
0C82	KANNADA SIGN ANUSVARA	0E4B	THAI CHARACTER MAI CHATTAWA
0C83	KANNADA SIGN VISARGA	0E4C	THAI CHARACTER THANTHAKHAT
0CBE	KANNADA VOWEL SIGN AA	0E4D	THAI CHARACTER NIKHAHIT
0CBF	KANNADA VOWEL SIGN I	0E4E	THAI CHARACTER YAMAKKAN
0CC0	KANNADA VOWEL SIGN II	0EB1	LAO VOWEL SIGN MAI KAN
0CC1	KANNADA VOWEL SIGN U	0EB4	LAO VOWEL SIGN I
0CC2	KANNADA VOWEL SIGN UU	0EB5	LAO VOWEL SIGN II
0CC3	KANNADA VOWEL SIGN VOCALIC R	0EB6	LAO VOWEL SIGN Y
0CC4	KANNADA VOWEL SIGN VOCALIC RR	0EB7	LAO VOWEL SIGN YY
0CC6	KANNADA VOWEL SIGN E	0EB8	LAO VOWEL SIGN U
0CC7	KANNADA VOWEL SIGN EE	0EB9	LAO VOWEL SIGN UU
0CC8	KANNADA VOWEL SIGN AI	0EBB	LAO VOWEL SIGN MAI KON
0CCA	KANNADA VOWEL SIGN O	0EBC	LAO SEMIVOWEL SIGN LO
0CCB	KANNADA VOWEL SIGN OO	0EC8	LAO TONE MAI EK
0CCC	KANNADA VOWEL SIGN AU	0EC9	LAO TONE MAI THO
0CCD	KANNADA SIGN VIRAMA	0ECA	LAO TONE MAI TI
0CD5	KANNADA LENGTH MARK	0ECB	LAO TONE MAI CATAWA
0CD6	KANNADA AI LENGTH MARK	0ECC	LAO CANCELLATION MARK
0D02	MALAYALAM SIGN ANUSVARA	0ECD	LAO NIGGAHITA
0D03	MALAYALAM SIGN VISARGA	0F18	TIBETAN ASTROLOGICAL SIGN -KHYUD PA
0D3E	MALAYALAM VOWEL SIGN AA	0F19	TIBETAN ASTROLOGICAL SIGN SDONG TSHUGS
0D3F	MALAYALAM VOWEL SIGN I	0F35	TIBETAN MARK NGAS BZUNG NYI ZLA
0D40	MALAYALAM VOWEL SIGN II	0F37	TIBETAN MARK NGAS BZUNG SGOR RTAGS
0D41	MALAYALAM VOWEL SIGN U	0F39	TIBETAN MARK TSA -PHRU
0D42	MALAYALAM VOWEL SIGN UU	0F3E	TIBETAN SIGN YAR TSHES
0D43	MALAYALAM VOWEL SIGN VOCALIC R	0F3F	TIBETAN SIGN MAR TSHES
0D46	MALAYALAM VOWEL SIGN E	0F71	TIBETAN VOWEL SIGN AA
0D47	MALAYALAM VOWEL SIGN EE	0F72	TIBETAN VOWEL SIGN I
0D48	MALAYALAM VOWEL SIGN AI	0F73	TIBETAN VOWEL SIGN II
0D4A	MALAYALAM VOWEL SIGN O	0F74	TIBETAN VOWEL SIGN U
0D4B	MALAYALAM VOWEL SIGN OO	0F75	TIBETAN VOWEL SIGN UU
0D4C	MALAYALAM VOWEL SIGN AU	0F76	TIBETAN VOWEL SIGN VOCALIC R
0D4D	MALAYALAM SIGN VIRAMA	0F77	TIBETAN VOWEL SIGN VOCALIC RR
0D57	MALAYALAM AU LENGTH MARK	0F78	TIBETAN VOWEL SIGN VOCALIC L
0D82	SINHALA SIGN ANUSVARAYA	0F79	TIBETAN VOWEL SIGN VOCALIC LL
0D83	SINHALA SIGN VISARGAYA	0F7A	TIBETAN VOWEL SIGN E
0DCA	SINHALA SIGN AL-LAKUNA	0F7B	TIBETAN VOWEL SIGN EE
0DCF	SINHALA VOWEL SIGN AELA-PILLA		
0DD0	SINHALA VOWEL SIGN KETTI AEDA-PILLA		
0DD1	SINHALA VOWEL SIGN DIGA AEDA-PILLA		

0F7C	TIBETAN VOWEL SIGN O	1039	MYANMAR SIGN VIRAMA
0F7D	TIBETAN VOWEL SIGN OO	1056	MYANMAR VOWEL SIGN VOCALIC R
0F7E	TIBETAN SIGN RJES SU NGA RO	1057	MYANMAR VOWEL SIGN VOCALIC RR
0F7F	TIBETAN SIGN RNAM BCAD	1058	MYANMAR VOWEL SIGN VOCALIC L
0F80	TIBETAN VOWEL SIGN REVERSED I	1059	MYANMAR VOWEL SIGN VOCALIC LL
0F81	TIBETAN VOWEL SIGN REVERSED II	17B4	KHMER VOWEL INHERENT AQ
0F82	TIBETAN SIGN NYI ZLA NAA DA	17B5	KHMER VOWEL INHERENT AA
0F83	TIBETAN SIGN SNA LDAN	17B6	KHMER VOWEL SIGN AA
0F84	TIBETAN MARK HALANTA	17B7	KHMER VOWEL SIGN I
0F86	TIBETAN MARK LCI RTAGS	17B8	KHMER VOWEL SIGN II
0F87	TIBETAN MARK YANG RTAGS	17B9	KHMER VOWEL SIGN Y
0F90	TIBETAN SUBJOINED LETTER KA	17BA	KHMER VOWEL SIGN YY
0F91	TIBETAN SUBJOINED LETTER KHA	17BB	KHMER VOWEL SIGN U
0F92	TIBETAN SUBJOINED LETTER GA	17BC	KHMER VOWEL SIGN UU
0F93	TIBETAN SUBJOINED LETTER GHA	17BD	KHMER VOWEL SIGN UA
0F94	TIBETAN SUBJOINED LETTER NGA	17BE	KHMER VOWEL SIGN OE
0F95	TIBETAN SUBJOINED LETTER CA	17BF	KHMER VOWEL SIGN YA
0F96	TIBETAN SUBJOINED LETTER CHA	17C0	KHMER VOWEL SIGN IE
0F97	TIBETAN SUBJOINED LETTER JA	17C1	KHMER VOWEL SIGN E
0F99	TIBETAN SUBJOINED LETTER NYA	17C2	KHMER VOWEL SIGN AE
0F9A	TIBETAN SUBJOINED LETTER TTA	17C3	KHMER VOWEL SIGN AI
0F9B	TIBETAN SUBJOINED LETTER TTHA	17C4	KHMER VOWEL SIGN OO
0F9C	TIBETAN SUBJOINED LETTER DDA	17C5	KHMER VOWEL SIGN AU
0F9D	TIBETAN SUBJOINED LETTER DDHA	17C6	KHMER SIGN NIKAHIT
0F9E	TIBETAN SUBJOINED LETTER NNA	17C7	KHMER SIGN REAHMUK
0F9F	TIBETAN SUBJOINED LETTER TA	17C8	KHMER SIGN YUUKALEAPINTU
0FA0	TIBETAN SUBJOINED LETTER THA	17C9	KHMER SIGN MUUSIKATOAN
0FA1	TIBETAN SUBJOINED LETTER DA	17CA	KHMER SIGN TRIISAP
0FA2	TIBETAN SUBJOINED LETTER DHA	17CB	KHMER SIGN BANTOC
0FA3	TIBETAN SUBJOINED LETTER NA	17CC	KHMER SIGN ROBAT
0FA4	TIBETAN SUBJOINED LETTER PA	17CD	KHMER SIGN TOANDAKHIAT
0FA5	TIBETAN SUBJOINED LETTER PHA	17CE	KHMER SIGN KAKABAT
0FA6	TIBETAN SUBJOINED LETTER BA	17CF	KHMER SIGN AHSDA
0FA7	TIBETAN SUBJOINED LETTER BHA	17D0	KHMER SIGN SAMYOK SANNYA
0FA8	TIBETAN SUBJOINED LETTER MA	17D1	KHMER SIGN VIRIAM
0FA9	TIBETAN SUBJOINED LETTER TSA	17D2	KHMER SIGN COENG
0FAA	TIBETAN SUBJOINED LETTER TSHA	17D3	KHMER SIGN BATHAMASAT
0FAB	TIBETAN SUBJOINED LETTER DZA	18A9	MONGOLIAN LETTER AG DAGALGA
0FAC	TIBETAN SUBJOINED LETTER DZHA	302A	IDEOGRAPHIC LEVEL TONE MARK
0FAD	TIBETAN SUBJOINED LETTER WA	302B	IDEOGRAPHIC RISING TONE MARK
0FAE	TIBETAN SUBJOINED LETTER ZHA	302C	IDEOGRAPHIC DEPARTING TONE MARK
0FAF	TIBETAN SUBJOINED LETTER ZA	302D	IDEOGRAPHIC ENTERING TONE MARK
0FB0	TIBETAN SUBJOINED LETTER -A	302E	HANGUL SINGLE DOT TONE MARK
0FB1	TIBETAN SUBJOINED LETTER YA	302F	HANGUL DOUBLE DOT TONE MARK
0FB2	TIBETAN SUBJOINED LETTER RA	3099	COMBINING KATAKANA-HIRAGANA VOICED SOUND MARK
0FB3	TIBETAN SUBJOINED LETTER LA	309A	COMBINING KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK
0FB4	TIBETAN SUBJOINED LETTER SHA	FB1E	HEBREW POINT JUDEO-SPANISH VARIKA
0FB5	TIBETAN SUBJOINED LETTER SSA		
0FB6	TIBETAN SUBJOINED LETTER SA		
0FB7	TIBETAN SUBJOINED LETTER HA		
0FB8	TIBETAN SUBJOINED LETTER A		
0FB9	TIBETAN SUBJOINED LETTER KSSA		
0FBA	TIBETAN SUBJOINED LETTER FIXED-FORM WA		
0FBB	TIBETAN SUBJOINED LETTER FIXED-FORM YA		
0FBC	TIBETAN SUBJOINED LETTER FIXED-FORM RA		
0FC6	TIBETAN SYMBOL PADMA GDAN		
102C	MYANMAR VOWEL SIGN AA		
102D	MYANMAR VOWEL SIGN I		
102E	MYANMAR VOWEL SIGN II		
102F	MYANMAR VOWEL SIGN U		
1030	MYANMAR VOWEL SIGN UU		
1031	MYANMAR VOWEL SIGN E		
1032	MYANMAR VOWEL SIGN AI		
1036	MYANMAR SIGN ANUSVARA		
1037	MYANMAR SIGN DOT BELOW		
1038	MYANMAR SIGN VISARGA		

B.2 List of combining and other characters not allowed in implementation level 2

The characters in the subset collections COMBINING DIACRITICAL MARKS (0300 to 036F), COMBINING DIACRITICAL MARKS FOR SYMBOLS (20D0 to 20FF), HANGUL JAMO (1100 to 11FF) and COMBINING HALF MARKS (FE20 to FE2F) are not allowed in implementation level 2. In addition, the following individual characters are also not allowed.

NOTE - This list is a subset of the list in clause B.1 except for HANGUL JAMO (see 25.1).

0483	COMBINING CYRILLIC TITLO	05A6	HEBREW ACCENT MERKHA KEFULA
0484	COMBINING CYRILLIC PALATALIZATION	05A7	HEBREW ACCENT DARGA
0485	COMBINING CYRILLIC DASIA PNEUMATA	05A8	HEBREW ACCENT QADMA
0486	COMBINING CYRILLIC PSILI PNEUMATA	05A9	HEBREW ACCENT TELISHA QETANA
0591	HEBREW ACCENT ETNAHTA	05AA	HEBREW ACCENT YERAH BEN YOMO
0592	HEBREW ACCENT SEGOL	05AB	HEBREW ACCENT OLE
0593	HEBREW ACCENT SHALSHELET	05AC	HEBREW ACCENT ILUY
0594	HEBREW ACCENT ZAQEF QATAN	05AD	HEBREW ACCENT DEHI
0595	HEBREW ACCENT ZAQEF GADOL	05AE	HEBREW ACCENT ZINOR
0596	HEBREW ACCENT TIPEHA	05AF	HEBREW MARK MASORA CIRCLE
0597	HEBREW ACCENT REVIA	05C4	HEBREW MARK UPPER DOT
0598	HEBREW ACCENT ZARQA	093C	DEVANAGARI SIGN NUKTA
0599	HEBREW ACCENT PASHTA	0953	DEVANAGARI GRAVE ACCENT
059A	HEBREW ACCENT YETIV	0954	DEVANAGARI ACUTE ACCENT
059B	HEBREW ACCENT TEVIR	09BC	BENGALI SIGN NUKTA
059C	HEBREW ACCENT GERESH	09D7	BENGALI AU LENGTH MARK
059D	HEBREW ACCENT GERESH MUQDAM	0A3C	GURMUKHI SIGN NUKTA
059E	HEBREW ACCENT GERSHAYIM	0A70	GURMUKHI TIPPI
059F	HEBREW ACCENT QARNEY PARA	0A71	GURMUKHI ADDAK
05A0	HEBREW ACCENT TELISHA GEDOLA	0ABC	GUJARATI SIGN NUKTA
05A1	HEBREW ACCENT PAZER	0B3C	ORIYA SIGN NUKTA
05A3	HEBREW ACCENT MUNAH	0B56	ORIYA AI LENGTH MARK
05A4	HEBREW ACCENT MAHAPAKH	0B57	ORIYA AU LENGTH MARK
05A5	HEBREW ACCENT MERKHA	0BD7	TAMIL AU LENGTH MARK
		0C55	TELUGU LENGTH MARK
		0C56	TELUGU AI LENGTH MARK
		0CD5	KANNADA LENGTH MARK
		0CD6	KANNADA AI LENGTH MARK
		0D57	MALAYALAM AU LENGTH MARK
		0F39	TIBETAN MARK TSA -PHRU
		302A	IDEOGRAPHIC LEVEL TONE MARK
		302B	IDEOGRAPHIC RISING TONE MARK
		302C	IDEOGRAPHIC DEPARTING TONE MARK
		302D	IDEOGRAPHIC ENTERING TONE MARK
		302E	HANGUL SINGLE DOT TONE MARK
		302F	HANGUL DOUBLE DOT TONE MARK
		3099	COMBINING KATAKANA-HIRAGANA VOICED SOUND MARK
		309A	COMBINING KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK

Annex C (normative)

Transformation format for 16 planes of Group 00 (UTF-16)

UTF-16 provides a coded representation of over a million graphic characters of UCS-4 in a form that is compatible with the two-octet BMP form of UCS-2 (13.1). This permits the coexistence of those characters from UCS-4 within coded character data that is in accordance with UCS-2.

In UTF-16 each graphic character from the UCS-2 repertoire retains its UCS-2 coded representation. In addition, the coded representation of any character from a single contiguous block of 16 Planes in Group 00 (1,048,576 code positions) consists of a pair of RC-elements (4.33), where each such RC-element corresponds to a cell in a single contiguous block of 8 Rows in the BMP (2,048 code positions). These code positions are reserved for the use of this coded representation form, and shall not be allocated for any other purpose.

C.1 Specification of UTF-16

The specification of UTF-16 is as follows:

1. The high-half zone shall be the 4 rows D8 to DB of the BMP, i.e., the 1,024 cells in the S-zone whose code positions are from D800 through DBFF.
2. The low-half zone shall be the 4 rows DC to DF of the BMP, i.e., the 1,024 cells in the S-zone whose code positions are from DC00 through DFFF.
3. All cells in the high-half zone and the low-half zone shall be permanently reserved for the use of the UTF-16 coded representation form.
4. In UTF-16, any UCS character from the BMP shall be represented by its UCS-2 coded representation as specified by the body of this international standard.
5. In UTF-16, any UCS character whose UCS-4 coded representation is in the range 0001 0000 to 0010 FFFF shall be represented by a sequence of two RC-elements from the S-zone, of which the first is an RC-element from the high-half zone, and the second is an RC-element from the low-half zone.

The mapping between UCS-4 and UTF-16 for these characters shall be as shown in C.3; the reverse mapping is shown in C.4.

NOTE - The Unicode Standard, Version 3.0, defines the following forms of UTF-16.

- UTF-16: the ordering of octets (6.3) is not defined and signatures (Annex H) may appear;
- UTF-16BE: in the ordering of octets the more significant octet precedes the less significant octet, as specified in 6.2, and no signatures appear;
- UTF-16LE: in the ordering of octets the less significant octet precedes the more significant octet and no signatures appear.

C.2 Notation

1. All numbers are in hexadecimal notation.
2. Double-octet boundaries in the notations for UTF-16 are indicated with semicolons.
3. The symbol “%” indicates the modulo operation, e.g.: $x \% y = x$ modulo y .
4. The symbol “/” indicates the integer division operation, e.g.: $7 / 3 = 2$.
5. Precedence is -
integer-division > modulo-operation >
integer-multiplication > integer-addition.

C.3 Mapping from UCS-4 form to UTF-16 form

UCS-4 (4-octet)	UTF-16, 2-octet elements
x = 0000 0000 .. 0000 FFFF (see Note 1)	x % 0001 0000;
x = 0001 0000 .. 0010 FFFF	y; z;
where	$y = ((x - 0001\ 0000) / 400) + D800$ $z = ((x - 0001\ 0000) \% 400) + DC00$
x = 0011 0000 .. 7FFF FFFF	(no mapping (is defined)

NOTE - Code positions from 0000 D800 to 0000 DFFF are reserved for the UTF-16 form and do not occur in UCS-4. The values 0000 FFFE and 0000 FFFF also do not occur (see clause 8). The mapping of these code positions in UTF-16 is undefined.

Example:

The UCS-4 sequence [0000 0048] [0000 0069]
[0001 0000] [0000 0021] [0000 0021]

represents “Hi<0001 0000>!!”.

It is mapped to UTF-16 as:

[0048] [0069] [D800] [DC00] [0021] [0021]

If interpreted as UCS-2 this sequence will be

“Hi<RC-element from high-half zone>
<RC-element from low-half zone>!!”

C.4 Mapping from UTF-16 form to UCS-4 form

UTF-16, 2-octet elements UCS-4 (4-octet)

x = 0000; ... D7FF; x
x = E000; ... FFFF; x

pair (x, y) such that

x = D800; ... DBFF; ((x - D800) * 400
y = DC00; ... DFFF; + (y - DC00))
 + 0001 0000

Example:

The UTF-16 sequence

[0048] [0069] [D800] [DC00] [0021] [0021]

is mapped to UCS-4 as

[0000 0048] [0000 0069] [0001 0000]
[0000 0021] [0000 0021]

and represents “Hi<0001 0000>!!”

C.5 Identification of UTF-16

When the escape sequences from ISO/IEC 2022 are used, the identification of UTF-16 and an implementation level (see clause 14) shall be by a designation sequence chosen from the following list:

ESC 02/05 02/15 04/10
UTF-16 with implementation level 1

ESC 02/05 02/15 04/11
UTF-16 with implementation level 2

ESC 02/05 02/15 04/12
UTF-16 with implementation level 3

If such an escape sequence appears within a CC-data-element conforming to ISO/IEC 2022, it shall consist only of the sequences of bit combinations as shown above.

If such an escape sequence appears within a CC-data-element conforming to ISO/IEC 10646, it shall be padded in accordance with clause 15.

When the escape sequences from ISO 2022 are used, the identification of a return, or transfer, from UTF-16 to the coding system of ISO 2022 shall be

as specified in 16.5 for a return or transfer from UCS.

C.6 Unpaired RC-elements: Interpretation by receiving devices

According to C.1 an unpaired RC-element (4.33) is not in conformance with the requirements of UTF-16.

If a receiving device that has adopted the UTF-16 form receives an unpaired RC-element because of error conditions either:

- in an originating device, or
- in the interchange between an originating and the receiving device, or
- in the receiving device itself,

then it shall interpret that unpaired RC-element in the same way that it interprets a character that is outside the adopted subset that has been identified for the device (see 2.3c).

NOTE - Since a high-half RC-element followed by a low-half RC-element is a sequence that is in accordance with UTF-16, the only possible type of syntactically malformed sequence is an unpaired RC-element.

Example:

A receiving/originating device which only handles the Basic Latin repertoire, and uses boxes (shown here as à) to display characters outside that repertoire, would display:

“The Greek letter Σ is the capital form of letter σ.”

as:

The Greek letter à is the capital form of letter à.”

Accordingly a similar device that can also interpret a UTF-16 data stream should also display an unpaired RC-element as a box.

C.7 Receiving devices, advisory notes

When a receiving device interprets a CC-data-element that is in accordance with UTF-16 the following advisory notes apply.

1. UTF-16 is designed to be compatible with the UCS-2 two-octet BMP Form (13.1). The high-half and low-half zones are assigned to separate ranges of code positions, to which characters can never be assigned. Thus the function of every RC-element (two-octet unit) within a UTF-16 data stream is always immediately identifiable from its value, without regard to context.

For example, the valid UTF-16 sequence [0048] [0069] [D800] [DC00] [0021] [0021] may also be interpreted, by a receiving device, that has adopted only UCS-2, as the coded representation of

“Hi<unrecognized><unrecognized>!!”

This form of compatibility is possible because RC-elements from the S-zone are interpreted according to UTF-16 by receiving devices that have adopted UTF-16, and as unrecognized characters by receiving devices that have only adopted UCS-2. Consequently an originating device may transmit UTF-16 data even if the receiving device can only interpret that data as UCS-2 characters.

2. Designers of devices may choose to use UTF-16 as an internal representation for processing or other purposes. There are two primary issues for such devices:

- Does the device interpret (i.e., process according to the assigned semantics) some subset of the pairs (high-half + low-half) of RC-elements, e.g., render the pair as the intended single character?
- Does the device guarantee the integrity of every pair (high-half + low-half) of RC-elements, e.g., never separate such pairs in operations such as string truncation, insertion, or other modifications of the coded character sequence?

The decisions on these issues give rise to four possible combinations of capability in a device:

(U) UCS-2 implementations:

- Interpret no pairs.
- Do not guarantee integrity of pairs.

(W) Weak UTF-16 implementations:

- Interpret a non-null subset of pairs.
- Do not guarantee integrity of pairs.

- (A) Aware UTF-16 implementations:
 - Interpret no pairs.
 - Guarantee integrity of pairs.
- (S) Strong UTF-16 implementations:
 - Interpret a non-null subset of pairs.
 - Guarantee integrity of pairs.

Example:

The following sentence could be displayed in four different ways, assuming that both the weak and strong implementations have Etruscan fonts but no hieroglyphic fonts:

“The Greek letter Σ corresponds to <hieroglyphic-High> <hieroglyphic-Low> and to <Etruscan-High> <Etruscan-Low>.”

where <xxx-High> and <xxx-Low> represent RC-elements, from the High-half and Low-half zones respectively, corresponding to a character from the block indicated by xxx. These four ways are shown below.

U: “The Greek letter Σ corresponds to àà and to àà.”

W: “The Greek letter Σ corresponds to àà and to S.”

A: “The Greek letter Σ corresponds to à and to à.”

S: “The Greek letter Σ corresponds to à and to S.”

where S here indicates the letter ES in the Etruscan font.

Annex D (normative)

UCS Transformation Format 8 (UTF-8)

UTF-8 is an alternative coded representation form for all of the characters of the UCS. It can be used to transmit text data through communication systems which assume that individual octets in the range 00 to 7F have a definition according to ISO/IEC 4873, including a C0 set of control functions according to the 8-bit structure of ISO/IEC 2022. UTF-8 also avoids the use of octet values in this range which have special significance during the parsing of file-name character strings in widely-used file-handling systems.

The number of octets in the UTF-8 coded representation of the characters of the UCS ranges from one to six; the value of the first octet indicates the number of octets in that coded representation.

D.1 Features of UTF-8

- UCS characters from the BASIC LATIN collection are represented in UTF-8 in accordance with ISO/IEC 4873, i.e. single octets with values ranging from 20 to 7E.
- Control functions in positions 0000 0000, to 0000 001F, and the DELETE character in position 0000 007F, are represented without the padding octets specified in clause 15, i.e. as single octets with values ranging from 00 to 1F, and 7F respectively in accordance with ISO/IEC 4873 and with the 8-bit structure of ISO/IEC 2022.
- Octet values 00 to 7F do not otherwise occur in the UTF-8 coded representation of any character. This provides compatibility with existing file-handling systems and communications sub-systems which parse CC-data-elements for these octet values.
- The first octet in the UTF-8 coded representation of any character can be directly identified when a CC-data-element is examined, one octet at a time, starting from an arbitrary location. It indicates the number of continuing octets (if any) in the multi-octet sequence that constitutes the coded representation of that character.

D.2 Specification of UTF-8

In the UTF-8 coded representation form each character from this International Standard shall have a coded representation that comprises a sequence of octets of length 1, 2, 3, 4, 5, or 6 octets.

For all sequences of one octet the most significant bit shall be a ZERO bit.

For all sequences of more than one octet, the number of ONE bits in the first octet, starting from the most significant bit position, shall indicate the number of octets in the sequence. The next most significant bit shall be a ZERO bit.

NOTE 1 - For example, the first octet of a 2-octet sequence has bits 110 in the most significant positions, and the first octet of a 6-octet sequence has bits 1111110 in the most significant positions.

All of the octets, other than the first in a sequence, are known as continuing octets. The two most significant bits of a continuing octet shall be a ONE bit followed by a ZERO bit.

The remaining bit positions in the octets of the sequence shall be "free bit positions" that are used to distinguish between the characters of this International Standard. These free bit positions shall be used, in order of increasing significance, for the bits of the UCS-4 coded representation of the character, starting from its least significant bit. Some of the high-order ZERO bits of the UCS-4 representation shall be omitted, as specified below.

Table D.1 below shows the format of the octets of a coded character according to UTF-8. Each free bit position available for distinguishing between the characters is indicated by an x. Each entry in the column "Maximum UCS-4 value" indicates the upper end of the range of coded representations from UCS-4 that may be represented in a UTF-8 sequence having the length indicated in the "Octet usage" column.