Annex A
(normative)

Collections of graphic characters for subsets

The following collections are from the Basic Multilingual Plane.

NOTE - Use of implementation level 1 restricts the repertoire of some character collections (see clause 23.6). Collections which include combining characters are 7, 8, 9, 11 to 24, 31, 45, 46 and 51.

Collection number and name	Positions
1 ISO-646 IRV	0020 - 007E
2 LATIN-1 SUPPLEMENT	00A0 - 00FF
3 EXTENDED LATIN-A	0100 - 017F
4 EXTENDED LATIN-B	0180 - 024F
5 IPA EXTENSIONS	0250 - 02AF
6 SPACING MODIFIER LETTERS	02B0 - 02FF
7 COMBINING DIACRITICAL MARKS	0300 - 036F
8 GREEK	0370 - 03FF
9 CYRILLIC	0400 - 04FF
10 ARMENIAN	0530 - 058F
11 HEBREW	0590 - 05FF
12 ARABIC	0600 - 06FF
13 DEVANAGARI	0900 - 097F
14 BENGALI	0980 - 09FF
15 GURMUKHI	0A00 - 0A7F
16 GUJARATI	0A80 - 0AFF
17 ORIYA	0B00 - 0B7F
18 TAMIL	0B80 - 0BFF
19 TELUGU	0C00 - 0C7F
20 KANNADA	0C80 - 0CFF
21 MALAYALAM	0D00 - 0D7F
22 THAI	0E00 - 0E7F
23 LAO	0E80 - 0EFF
24 TIBETAN	1000 - 105F
25 GEORGIAN	10A0 - 10FF
26 ADDITIONAL EXTENDED LATIN	1E00 - 1EFF
27 GREEK EXTENSIONS	1F00 - 1FFF
28 GENERAL PUNCTUATION	2000 - 206F
29 SUPERSCRIPTS AND SUBSCRIPTS	2070 - 209F
30 CURRENCY SYMBOLS	20A0 - 20CF
31 COMBINING DIACRITICAL MARKS FOR SYMBOLS	20D0 - 20FF
32 LETTERLIKE SYMBOLS	2100 - 214F
33 NUMBER FORMS	2150 - 218F
34 ARROWS	2190 - 21FF
35 MATHEMATICAL OPERATORS	2200 - 22FF
36 MISCELLANEOUS TECHNICAL	2300 - 23FF
37 CONTROL PICTURES	2400 - 243F
38 OPTICAL CHARACTER RECOGNITION	2440 - 245F
39 ENCLOSED ALPHANUMERICS	2460 - 24FF
40 BOX DRAWING	2500 - 257F
41 BLOCK ELEMENTS	2580 - 259F
42 GEOMETRIC SHAPES	25A0 - 25FF
43 MISCELLANEOUS DINGBATS	2600 - 26FF
44 DINGBATS	2700 - 27BF
45 CJK SYMBOLS AND PUNCTUATION	3000 - 303F
46 HIRAGANA	3040 - 309F

47 KATAKANA 30A0 - 30FF

48 BOPOMOFO 3100 - 312F

49 HANGUL JAMO 3130 - 318F

50 CJK MISCELLANEOUS 3190 - 319F

51 COMBINING HANGUL JAMO 31A0 - 31FF

52 ENCLOSED CJK LETTERS AND MONTHS 3200 - 32FF

53 CJK COMPATIBILITY WORDS AND HOURS 3300 - 337F

54 CJK COMPATIBILITY ABBREVIATIONS AND DAYS 3380 - 33FF

55 HANGUL 3400 - 3D2F

56 SUPPLEMENTARY HANGUL 3E00 - 458F

57 OLD HANGUL 4600 - 4C8F

58 CJK UNIFIED IDEOGRAPHS 4E00 - 9FFF

59 PRIVATE USE AREA E000 - F7FF

60 CJK COMPATIBILITY IDEOGRAPHS F900 - FAFF

61 ALPHABETIC PRESENTATION FORMS FB00 - FB4F

62 ARABIC PRESENTATION FORMS-A FC50 - FDFF

63 CJK COMPATIBILITY FORMS FE30 - FE4F

64 SMALL FORM VARIANTS FE50 - FE6F

65 ARABIC PRESENTATION FORMS-B FE70 - FEFE

66 HALFWIDTH AND FULLWIDTH FORMS FF00 - FFEF

67 SPECIALS FFF0 - FFFD

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

80 ZERO-WIDTH BOUNDARY INDICATORS 200B - 200D FEFF

81 FORMAT SEPARATORS 2028 - 2029

82 BI-DIRECTIONAL FORMAT MARKS 200E - 200F

83 BI-DIRECTIONAL FORMAT EMBEDDINGS 202A - 202E

84 HANGUL FILL CHARACTERS 3164, FFA0

85 CHARACTER SHAPING SELECTORS 206C - 206D

86 NUMERIC SHAPE SELECTORS 206E - 206F

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

90 GENERAL FORMAT CHARACTERS Collections 80 - 83

91 SCRIPT-SPECIFIC FORMAT CHARACTERS Collections 84 - 86

The following specify other collections.

92 COMBINING CHARACTERS characters specified in ANNEX B

100 BMP 0000 - FFFD

300 PRIVATE USE PLANES G=00, P=E0 - FF

400 PRIVATE USE GROUPS G=60 - 7F

Annex B
(normative)

List of combining characters

The characters in the blocks COMBINING DIACRITICAL MARKS (0300 to 036F), COMBINING DIACRITICAL MARKS FOR SYMBOLS (20D0 to 20FF) and COMBINING HANGUL JAMO (31A0 to 31FF) are combining characters. In addition, the following characters are combining characters.

0370	GREEK COMBINING IOTA BELOW
0371	GREEK COMBINING DASIA PNEUMATA
0372	GREEK COMBINING PSILI PNEUMATA
0384	GREEK COMBINING TONOS
0385	GREEK COMBINING DIAERESIS TONOS
0483	CYRILLIC COMBINING TITLO
0484	CYRILLIC COMBINING PALATALIZATION
0485	CYRILLIC COMBINING DASIA PNEUMATA
0486	CYRILLIC COMBINING PSILI PNEUMATA
05B0	HEBREW POINT SHEVA
05B1	HEBREW POINT HATAF SEGOL
05B2	HEBREW POINT HATAF PATAH
05B3	HEBREW POINT HATAF QAMATS
05B4	HEBREW POINT HIRIQ
05B5	HEBREW POINT TSERE
05B6	HEBREW POINT SEGOL
05B7	HEBREW POINT PATAH
05B8	HEBREW POINT QAMATS
05B9	HEBREW POINT HOLAM
05BB	HEBREW POINT QUBUTS
05BC	HEBREW POINT DAGESH
05BD	HEBREW POINT METEG
05BF	HEBREW POINT RAFE
05C1	HEBREW POINT SHIN DOT
05C2	HEBREW POINT SIN DOT
05F5	HEBREW POINT VARIKA
064B	ARABIC FATHATAN
064C	ARABIC DAMMATAN
064D	ARABIC KASRATAN
064E	ARABIC FATHAH
064F	ARABIC DAMMAH
0650	ARABIC KASRAH
0651	ARABIC SHADDAH
0652	ARABIC SUKUN
0670	ARABIC ALEF ABOVE
06D6	SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA
06D7	SMALL HIGH LIGATURE QAF WITH LAM WITH ALEF MAKSURA
06D8	SMALL HIGH MEEM INITIAL FORM
06D9	SMALL HIGH LAM ALEF
06DA	SMALL HIGH JEEM
06DB	SMALL HIGH THREE DOTS
06DC	SMALL HIGH SEEN
06DD	END OF AYAH
06DE	START OF RUB EL HIZB
06DF	SMALL HIGH ROUNDED ZERO
06E0	SMALL HIGH UPRIGHT RECTANGULAR ZERO
06E1	SMALL HIGH DOTLESS HEAD OF KHAH
06E2	SMALL HIGH MEEM ISOLATED FORM
06E3	SMALL LOW SEEN
06E4	SMALL HIGH MADDA
06E7	SMALL HIGH YEH
06E8	SMALL HIGH NOON
06EA	EMPTY CENTER LOW STOP
06EB	EMPTY CENTER HIGH STOP
06EC	ROUNDED HIGH STOP WITH FILLED CENTER
06ED	SMALL LOW MEEM
0901	DEVANAGARI SIGN CANDRABINDU
0902	DEVANAGARI SIGN ANUSVARA
093C	DEVANAGARI SIGN NUKTA
093F	DEVANAGARI VOWEL SIGN I
0941	DEVANAGARI VOWEL SIGN U
0942	DEVANAGARI VOWEL SIGN UU
0943	DEVANAGARI VOWEL SIGN VOCALIC R
0944	DEVANAGARI VOWEL SIGN VOCALIC RR
0945	DEVANAGARI VOWEL SIGN CANDRA E
0946	DEVANAGARI VOWEL SIGN SHORT E
0947	DEVANAGARI VOWEL SIGN E
0948	DEVANAGARI VOWEL SIGN AI
094D	DEVANAGARI SIGN VIRAMA
0951	DEVANAGARI STRESS SIGN UDATTA
0952	DEVANAGARI STRESS SIGN ANUDATTA
0953	DEVANAGARI GRAVE ACCENT
0954	DEVANAGARI ACUTE ACCENT
0962	DEVANAGARI VOWEL SIGN VOCALIC L
0963	DEVANAGARI VOWEL SIGN VOCALIC LL
0981	BENGALI SIGN CANDRABINDU
09BC	BENGALI SIGN NUKTA
09BF	BENGALI VOWEL SIGN I
09C1	BENGALI VOWEL SIGN U
09C2	BENGALI VOWEL SIGN UU
09C3	BENGALI VOWEL SIGN VOCALIC R
09C4	BENGALI VOWEL SIGN VOCALIC RR
09C7	BENGALI VOWEL SIGN E
09C8	BENGALI VOWEL SIGN AI
09CB	BENGALI VOWEL SIGN O
09CC	BENGALI VOWEL SIGN AU
09CD	BENGALI SIGN VIRAMA
09E2	BENGALI VOWEL SIGN VOCALIC L
09E3	BENGALI VOWEL SIGN VOCALIC LL
0A02	GURMUKHI SIGN BINDI
0A3C	GURMUKHI SIGN NUKTA
0A3F	GURMUKHI VOWEL SIGN I
0A41	GURMUKHI VOWEL SIGN U
0A42	GURMUKHI VOWEL SIGN UU

0A47 GURMUKHI VOWEL SIGN EE
0A48 GURMUKHI VOWEL SIGN AI
0A4B GURMUKHI VOWEL SIGN OO
0A4C GURMUKHI VOWEL SIGN AU
0A70 GURMUKHI TIPPI
0A71 GURMUKHI ADDAK
0A81 GUJARATI SIGN CANDRABINDU
0A82 GUJARATI SIGN ANUSVARA
0ABC GUJARATI SIGN NUKTA
0ABF GUJARATI VOWEL SIGN I
0AC1 GUJARATI VOWEL SIGN U
0AC2 GUJARATI VOWEL SIGN UU
0AC3 GUJARATI VOWEL SIGN VOCALIC R
0AC4 GUJARATI VOWEL SIGN VOCALIC RR
0AC5 GUJARATI VOWEL SIGN CANDRA E
0AC7 GUJARATI VOWEL SIGN E
0AC8 GUJARATI VOWEL SIGN AI
0ACD GUJARATI SIGN VIRAMA
0B01 ORIYA SIGN CANDRABINDU
0B3C ORIYA SIGN NUKTA
0B3F ORIYA VOWEL SIGN I
0B41 ORIYA VOWEL SIGN U
0B42 ORIYA VOWEL SIGN UU
0B43 ORIYA VOWEL SIGN VOCALIC R
0B47 ORIYA VOWEL SIGN E
0B48 ORIYA VOWEL SIGN AI
0B4B ORIYA VOWEL SIGN O
0B4C ORIYA VOWEL SIGN AU
0B4D ORIYA SIGN VIRAMA
0BC0 TAMIL VOWEL SIGN II
0BC1 TAMIL VOWEL SIGN U
0BC2 TAMIL VOWEL SIGN UU
0BC6 TAMIL VOWEL SIGN E
0BC7 TAMIL VOWEL SIGN EE
0BC8 TAMIL VOWEL SIGN AI
0BCA TAMIL VOWEL SIGN O
0BCB TAMIL VOWEL SIGN OO
0BCC TAMIL VOWEL SIGN AU
0BCD TAMIL SIGN VIRAMA
0C3E TELUGU VOWEL SIGN AA
0C3F TELUGU VOWEL SIGN I
0C40 TELUGU VOWEL SIGN II
0C46 TELUGU VOWEL SIGN E
0C47 TELUGU VOWEL SIGN EE
0C48 TELUGU VOWEL SIGN AI
0C4A TELUGU VOWEL SIGN O
0C4B TELUGU VOWEL SIGN OO
0C4C TELUGU VOWEL SIGN AU
0C4D TELUGU SIGN VIRAMA
0C55 TELUGU LENGTH MARK
0C56 TELUGU AI LENGTH MARK
0CBF KANNADA VOWEL SIGN I
0CC0 KANNADA VOWEL SIGN II
0CC6 KANNADA VOWEL SIGN E
0CC7 KANNADA VOWEL SIGN EE
0CC8 KANNADA VOWEL SIGN AI
0CCA KANNADA VOWEL SIGN O
0CCB KANNADA VOWEL SIGN OO
0CCC KANNADA VOWEL SIGN AU
0CCD KANNADA SIGN VIRAMA
0D41 MALAYALAM VOWEL SIGN U
0D42 MALAYALAM VOWEL SIGN UU
0D43 MALAYALAM VOWEL SIGN VOCALIC R
0D46 MALAYALAM VOWEL SIGN E
0D47 MALAYALAM VOWEL SIGN EE
0D48 MALAYALAM VOWEL SIGN AI
0D4A MALAYALAM VOWEL SIGN O
0D4B MALAYALAM VOWEL SIGN OO
0D4C MALAYALAM VOWEL SIGN AU
0D4D MALAYALAM SIGN VIRAMA
0E31 THAI VOWEL SIGN MAI HAN-AKAT
0E34 THAI VOWEL SIGN SARA I
0E35 THAI VOWEL SIGN SARA II
0E36 THAI VOWEL SIGN SARA UE
0E37 THAI VOWEL SIGN SARA UEE
0E38 THAI VOWEL SIGN SARA U
0E39 THAI VOWEL SIGN SARA UU
0E3A THAI VOWEL SIGN PHINTHU
0E47 THAI VOWEL SIGN MAI TAI KHU
0E48 THAI TONE MAI EK
0E49 THAI TONE MAI THO
0E4A THAI TONE MAI TRI
0E4B THAI TONE MAI CHATTAWA
0E4C THAI THANTHAKHAT
0E4D THAI NIKKHAHIT
0E4E THAI YAMAKKAN
0E70 THAI PHONETIC ORDER VOWEL SIGN SARA E
0E71 THAI PHONETIC ORDER VOWEL SIGN SARA AE
0E72 THAI PHONETIC ORDER VOWEL SIGN SARA O
0E73 THAI PHONETIC ORDER VOWEL SIGN SARA MAI MUAN
0E74 THAI PHONETIC ORDER VOWEL SIGN SARA MAI MALAI
0EB1 LAO VOWEL SIGN MAI KAN
0EB4 LAO VOWEL SIGN I
0EB5 LAO VOWEL SIGN II
0EB6 LAO VOWEL SIGN Y
0EB7 LAO VOWEL SIGN YY
0EB8 LAO VOWEL SIGN U
0EB9 LAO VOWEL SIGN UU
0EBB LAO VOWEL SIGN MAI KON
0EBC LAO SEMIVOWEL SIGN LO
0EC8 LAO TONE MAI EK
0EC9 LAO TONE MAI THO
0ECA LAO TONE MAI TI
0ECB LAO TONE MAI CATAWA
0ECC LAO CANCELLATION MARK
0ECD LAO NIGGAHITA
0EF0 LAO PHONETIC ORDER VOWEL SIGN E
0EF1 LAO PHONETIC ORDER VOWEL SIGN EI
0EF2 LAO PHONETIC ORDER VOWEL SIGN O
0EF3 LAO PHONETIC ORDER VOWEL SIGN AY
0EF4 LAO PHONETIC ORDER VOWEL SIGN AI
1026 TIBETAN VOWEL SIGN I
1027 TIBETAN VOWEL SIGN SHORT I
1028 TIBETAN VOWEL SIGN U
1029 TIBETAN VOWEL SIGN E
102A TIBETAN VOWEL SIGN O
102E TIBETAN ANUSVARA
1030 TIBETAN UNDER RING
1036 TIBETAN CANDRABINDU
1037 TIBETAN CANDRABINDU WITH ORNAMENT
103B TIBETAN HONORIFIC UNDER RING
103D TIBETAN VOWEL SIGN AI
103E TIBETAN VOWEL SIGN AU
104B TIBETAN VIRAMA
104C TIBETAN LENITION MARK
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
(informative)

Mirrored characters in bi-directional context

In the context of right-to-left (bi-directional) text, the following characters have semantic meaning. To preserve the meaning in right-to-left text, the character symbol should be rendered as the mirror image of the associated graphical symbol from the left-to-right context. These characters included mathematical symbols and paired characters such as the SQUARE BRACKETS. For example in a right-to-left text segment, the GREATER-THAN SIGN (rendered as ">" in left-to-right text) would be rendered as the "<" graphic symbol.

0028	LEFT PARENTHESIS
0029	RIGHT PARENTHESIS
003C	LESS-THAN SIGN
003E	GREATER-THAN SIGN
005B	LEFT SQUARE BRACKET
005D	RIGHT SQUARE BRACKET
007B	LEFT CURLY BRACKET
007D	RIGHT CURLY BRACKET
103C	TIBETAN LEFT BRACE
102F	TIBETAN RIGHT BRACE
207D	SUPERSCRIPT LEFT PARENTHESIS
207E	SUPERSCRIPT RIGHT PARENTHESIS
208D	SUBSCRIPT LEFT PARENTHESIS
208E	SUBSCRIPT RIGHT PARENTHESIS
2201	COMPLEMENT
2202	PARTIAL DIFFERENTIAL
2203	THERE EXISTS
2204	THERE DOES NOT EXIST
2208	ELEMENT OF
2209	NOT AN ELEMENT OF
220A	SMALL ELEMENT OF
220B	CONTAINS AS MEMBER
220C	DOES NOT CONTAIN AS MEMBER
220D	SMALL CONTAINS AS MEMBER
2211	N-ARY SUMMATION
2215	DIVISION SLASH
2216	SET MINUS
221A	SQUARE ROOT
221B	CUBE ROOT
221C	FOURTH ROOT
221D	PROPORTIONAL TO
221F	RIGHT ANGLE
2220	ANGLE
2221	MEASURED ANGLE
2222	SPHERICAL ANGLE
2224	DOES NOT DIVIDE
2226	NOT PARALLEL TO
222B	INTEGRAL
222C	DOUBLE INTEGRAL
222D	TRIPLE INTEGRAL
222E	CONTOUR INTEGRAL
222F	SURFACE INTEGRAL
2230	VOLUME INTEGRAL
2231	CLOCKWISE INTEGRAL
2332	CLOCKWISE CONTOUR INTEGRAL
2333	ANTICLOCKWISE CONTOUR INTEGRAL
2239	EXCESS
223B	HOMOTHETIC
223C	TILDE OPERATOR
223D	REVERSED TILDE
223E	INVERTED LAZY S
223F	SINE WAVE
2240	WREATH PRODUCT
2241	NOT TILDE
2242	MINUS TILDE
2243	ASYMPTOTICALLY EQUAL TO
2244	NOT ASYMPTOTICALLY EQUAL TO
2245	APPROXIMATELY EQUAL TO
2246	APPROXIMATELY BUT NOT ACTUALLY EQUAL TO
2247	NEITHER APPROXIMATELY NOR ACTUALLY EQUAL TO
2248	ALMOST EQUAL TO
2249	NOT ALMOST EQUAL TO
224A	ALMOST OR EQUAL TO
224B	TRIPLE TILDE
224C	ALL EQUAL TO
2252	APPROXIMATELY EQUAL TO OR THE IMAGE OF
2253	IMAGE OF OR APPROXIMATELY EQUAL TO
2254	COLON EQUAL
2255	EQUAL COLON
225F	QUESTIONED EQUAL TO
2260	NOT EQUAL TO
2262	NOT IDENTICAL TO
2264	LESS-THAN OR EQUAL TO
2265	GREATER-THAN OR EQUAL TO
2266	LESS-THAN OVER EQUAL TO
2267	GREATER-THAN OVER EQUAL TO
2268	LESS-THAN BUT NOT EQUAL TO
2269	GREATER-THAN BUT NOT EQUAL TO
226A	MUCH LESS-THAN
226B	MUCH GREATER-THAN
226E	NOT LESS-THAN
226F	NOT GREATER-THAN
2270	NEITHER LESS-THAN NOR EQUAL TO
2271	NEITHER GREATER-THAN NOR EQUAL TO
2272	LESS-THAN OR EQUIVALENT TO
2273	GREATER-THAN OR EQUIVALENT TO
2274	NEITHER LESS-THAN NOR EQUIVALENT TO
2275	NEITHER GREATER-THAN NOR EQUIVALENT TO
2276	LESS-THAN OR GREATER-THAN
2277	GREATER-THAN OR LESS-THAN
2278	NEITHER LESS-THAN NOR GREATER-THAN
2279	NEITHER GREATER-THAN NOR LESS-THAN
227A	PRECEDES
227B	SUCCEEDS
227C	PRECEDES OR EQUAL TO
227D	SUCCEEDS OR EQUAL TO
227E	PRECEDES OR EQUIVALENT TO
227F	SUCCEEDS OR EQUIVALENT TO
2280	DOES NOT PRECEDE

2281 DOES NOT SUCCEED
2282 SUBSET OF
2283 SUPERSET OF
2284 NOT A SUBSET OF
2285 NOT A SUPERSET OF
2286 SUBSET OF OR EQUAL TO
2287 SUPERSET OF OR EQUAL TO
2288 NEITHER A SUBSET OF NOR EQUAL TO
2289 NEITHER A SUPERSET OF NOR EQUAL TO
228A SUBSET OF OR NOT EQUAL TO
228B SUPERSET OF OR NOT EQUAL TO
228C MULTISET
228F SQUARE IMAGE OF
2290 SQUARE ORIGINAL OF
2291 SQUARE IMAGE OF OR EQUAL TO
2292 SQUARE ORIGINAL OF OR EQUAL TO
2298 CIRCLED DIVISION SLASH
22A2 RIGHT TACK
22A3 LEFT TACK
22A6 ASSERTION
22A7 MODELS
22A8 TRUE
22A9 FORCES
22AA TRIPLE VERTICAL BAR TURNSTILE
22AB DOUBLE VERTICAL BAR DOUBLE RIGHT TURNSTILE
22AC DOES NOT PROVE
22AD NOT TRUE
22AE DOES NOT FORCE
22AF NEGATED DOUBLE VERTICAL BAR DOUBLE RIGHT TURNSTILE
22B0 PRECEDES UNDER RELATION
22B1 SUCCEEDS UNDER RELATION
22B2 NORMAL SUBGROUP OF
22B3 CONTAINS AS NORMAL SUBGROUP
22B4 NORMAL SUBGROUP OF OR EQUAL TO
22B5 CONTAINS AS NORMAL SUBGROUP OR EQUAL TO
22B6 ORIGINAL OF
22B7 IMAGE OF
22B8 MULTIMAP
22BE RIGHT ANGLE WITH ARC
22BF RIGHT TRIANGLE
22C9 LEFT NORMAL FACTOR SEMIDIRECT PRODUCT
22CA RIGHT NORMAL FACTOR SEMIDIRECT PRODUCT
22CB LEFT SEMIDIRECT PRODUCT
22CC RIGHT SEMIDIRECT PRODUCT
22CD REVERSE TILDE EQUALS
22D0 DOUBLE SUBSET
22D1 DOUBLE SUPERSET
22D6 LESS-THAN WITH DOT
22D7 GREATER-THAN WITH DOT
22D8 VERY MUCH LESS-THAN
22D9 VERY MUCH GREATER-THAN
22DA LESS-THAN EQUAL TO OR GREATER-THAN
22DB GREATER-THAN EQUAL TO OR LESS-THAN
22DC EQUAL TO OR LESS-THAN
22DD EQUAL TO OR GREATER-THAN
22DE EQUAL TO OR PRECEDES
22DF EQUAL TO OR SUCCEEDS
22E0 DOES NOT PRECEDE OR EQUAL
22E1 DOES NOT SUCCEED OR EQUAL
22E2 NOT SQUARE IMAGE OF OR EQUAL TO
22E3 NOT SQUARE ORIGINAL OF OR EQUAL TO
22E4 SQUARE IMAGE OF OR NOT EQUAL TO
22E5 SQUARE ORIGINAL OF OR NOT EQUAL TO
22E6 LESS-THAN BUT NOT EQUIVALENT TO
22E7 GREATER-THAN BUT NOT EQUIVALENT TO
22E8 PRECEDES BUT NOT EQUIVALENT TO
22E9 SUCCEEDS BUT NOT EQUIVALENT TO
22EA NOT NORMAL SUBGROUP OF
22EB DOES NOT CONTAIN AS NORMAL SUBGROUP
22EC NOT NORMAL SUBGROUP OF OR EQUAL TO
22ED DOES NOT CONTAIN AS NORMAL SUBGROUP OR EQUAL
22F0 UP RIGHT DIAGONAL ELLIPSIS
22F1 DOWN RIGHT DIAGONAL ELLIPSIS
2308 LEFT CEILING
2309 RIGHT CEILING
230A LEFT FLOOR
230B RIGHT FLOOR
2320 TOP HALF INTEGRAL
2321 BOTTOM HALF INTEGRAL
2329 LEFT-POINTING ANGLE BRACKET
232A RIGHT-POINTING ANGLE BRACKET
3008 LEFT ANGLE BRACKET
3009 RIGHT ANGLE BRACKET
300A LEFT DOUBLE ANGLE BRACKET
300B RIGHT DOUBLE ANGLE BRACKET
300C LEFT CORNER BRACKET
300D RIGHT CORNER BRACKET
300E LEFT WHITE CORNER BRACKET
300F RIGHT WHITE CORNER BRACKET
3010 LEFT BLACK LENTICULAR BRACKET
3011 RIGHT BLACK LENTICULAR BRACKET
3014 LEFT TORTOISE SHELL BRACKET
3015 RIGHT TORTOISE SHELL BRACKET
3016 LEFT WHITE LENTICULAR BRACKET
3017 RIGHT WHITE LENTICULAR BRACKET
3018 LEFT WHITE TORTOISE SHELL BRACKET
3019 RIGHT WHITE TORTOISE SHELL BRACKET
301A LEFT WHITE SQUARE BRACKET
301B RIGHT WHITE SQUARE BRACKET

Annex D
(informative)

Alternate Format Characters

There is a special class of characters called Alternate Format Characters which are included for compatibility with some industry practices. These characters do not have printable graphic symbols, and are thus represented in the character code tables by dotted boxes.

The function of most of these characters is to indicate the correct presentation of a sequence of characters. For any text processing other than presentation (such as sorting and searching), the alternate format characters can be ignored by filtering them out. The alternate format characters are not intended to be used in conjunction with bi-directional control functions from ISO/IEC 6429.

There are collections of graphic characters for selected subsets which consists of Alternate Format Characters (see Annex A).

D.1 General format characters

D.1.1 Zero-width boundary indicators

The following characters are used to indicate whether or not the adjacent characters should be separated by a word boundary.

ZERO WIDTH SPACE (200B): This character behaves like a SPACE in that it indicates a word boundary, except that it has no width. For example, this character could be used to indicate word boundaries in Thai, which does not use visible spaces to separate words.

ZERO WIDTH NO-BREAK SPACE (FEFF): This character behaves like a NO-BREAK SPACE in that it indicates the absence of word boundaries, except that it has no width. For example, this character could be inserted after the fourth character in the text "base+delta" to indicate that there is to be no word break between the "e" and the "+".

NOTE - For additional usages of this character for "signature", see annex E.

The following characters are used to indicate whether or not the adjacent characters should be joined together in rendering (cursive joiners).

ZERO WIDTH NON-JOINER (200C): This character indicates that the adjacent characters should not be joined together in a ligature or cursive connection, even when normal rendering would join the characters. For examples:

1. ZERO WIDTH NON-JOINER between LATIN SMALL LETTER F and LATIN SMALL LETTER I indicates that they should not join to form a ligature "fi" in rendering.
2. ZERO WIDTH NON-JOINER between ARABIC LETTER NOON and ARABIC LETTER MEEM indicates that the characters should not be the normal cursive connection form being rendered.
3. In the sequence DEVANAGARI KA followed by DEVANAGARI SIGN VIRAMA followed by DEVANAGARI KA, ZERO WIDTH NON-JOINER between the second and third characters indicates that the characters should not form the conjunct "kka".

ZERO WIDTH JOINER (200D): This character indicates that the adjacent characters should be represented with joining forms, even if the normal rendering would not join the characters. However, the joiner should not be used to cause a spacing character to be rendered as a combining mark. For examples:

1. In the sequence SPACE followed by ARABIC LETTER BAA, ZERO WIDTH JOINER can be inserted between the two characters to display the initial form of the ARABIC LETTER BAA.
2. ZERO WIDTH JOINER inserted between LATIN SMALL LETTER S and LATIN SMALL LETTER T can be used to indicate that the "st" ligature be rendered if possible, even if

the ligature would not normally be rendered.

D.1.2 Format separators

The following characters are used to indicate formatting boundaries between lines or paragraphs.

LINE SEPARATOR (2028): This character indicates where a new line should start; although the text should continue to the next line, it does not start a new paragraph; e.g. no inter-paragraph indentation might be applied.

PARAGRAPH SEPARATOR (2029): This character indicates where a new paragraph should start; e.g. the text should continue on the next line and inter-paragraph line spacing or paragraph indentation might be applied.

D.1.3 Bi-directional text formatting

The following characters are used in formatting bi-directional text. If the specification of a subset includes these characters, then text containing right-to-left characters are to be rendered with an implicit bi-directional algorithm.

An implicit algorithm uses the directional character properties to determine the correct display order of characters on a horizontal line of text.

The following characters are format characters that act exactly like right-to-left or left-to-right characters in terms of affecting ordering (Bi-directional format marks). They have no visible graphic symbols, and they do not have any other semantic effect.

Their use can be more convenient than the explicit embeddings or overrides, since their scope is much more local.

LEFT-TO-RIGHT MARK (200E): In bi-directional formatting, this character acts like a left-to-right character (such as LATIN SMALL LETTER A).

RIGHT-TO-LEFT MARK (200F): In bi-directional formatting, this character acts like a right-to-left character (such as ARABIC LETTER NOON).

The following format characters indicate that a piece of text is to be treated as embedded, and is to have a particular ordering attached to it (Bi-directional format embeddings). For example, an English quotation in the middle of an Arabic sentence can be marked as being an embedded left-to-right string. These format characters nest in blocks, with the embedding and override characters initiating (pushing) a block, and the pop character terminating (popping) a block.

The function of the embedding and override characters are very similar; the main difference is that the embedding characters specify the implicit direction of the text, while the override characters specify the explicit direction of the text. When text has an explicit direction, the normal directional character properties are ignored, and all of the text is assumed to have the ordering direction determined by the override character.

LEFT-TO-RIGHT EMBEDDING (202A): This character is used to indicate the start of a left-to-right implicit embedding.

RIGHT-TO-LEFT EMBEDDING (202B): This character is used to indicate the start of a right-to-left implicit embedding.

LEFT-TO-RIGHT OVERRIDE (202D): This character is used to indicate the start of a left-to-right explicit embedding.

RIGHT-TO-LEFT OVERRIDE (202E): This character is used to indicate the start of a right-to-left explicit embedding.

POP DIRECTIONAL FORMATTING (202C): This character is used to indicate the termination of an implicit or explicit directional embedding initiated by the above characters.

D.2 Script-specific format characters

D.2.1 Hangul fill characters

The following format characters have a special usage for Hangul characters.

HANGUL FILL (3164): This character represents the fill value used with the standard spacing Jamos.

HALFWIDTH HANGUL FILL (FFA0): As with the other halfwidth characters, this character is included for compatibility with certain systems that provide halfwidth forms of characters.

D.2.2 Character shaping selectors

The following characters are used in conjunction

with Arabic presentation forms. During the presentation process, certain characters may be joined together in cursive connection or ligatures. The following characters indicate that the character shape determination process (CSD) used to achieve this presentation effect is to be either activated or inhibited. Unlike the directional embedding characters, the following characters do not nest.

INHIBIT ARABIC FORM SHAPING (206C): Between this character and the following ACTIVATE ARABIC FORM SHAPING format character (if any) the CSD process is to be inhibited. The stored Arabic presentation forms will be presented without shape modification. This is the default state.

ACTIVATE ARABIC FORM SHAPING (206D): Between this character and the following INHIBIT ARABIC FORM SHAPING format character (if any), the stored Arabic presentation forms should be presented with shape modification by means of the CSD process.

NOTE - These characters have no effect on characters that are not presentation forms: in particular, Arabic nominal characters as from 0600 to 06FF are always subject to character shaping, and are unaffected by these formatting characters.

D.2.3 Numeric shape selectors

The following characters allow the selection of the shapes in which the digits from 0030 to 0039 are to be rendered. Unlike the directional embedding characters, these do not nest.

NATIONAL DIGIT SHAPES (206E): Between this character and the following NOMINAL DIGIT SHAPES format character (if any), digits from 0030 to 0039 are rendered with the appropriate national digit shapes as specified by means of appropriate agreements. For example, they could be displayed with shapes such as the ARABIC-INDIC digits from 0660 to 0669.

NOMINAL DIGIT SHAPES (206F): Between this character and the following NATIONAL DIGIT SHAPES format character (if any), the digits from 0030 to 0039 will be rendered with the shapes as those shown in the code tables for those digits. This is the default state.

Annex E
(informative)

The use of "signatures" to identify UCS

This annex describes a convention for the identification of features of the UCS, by the use of "signatures" within data streams of coded characters. The convention makes use of the character ZERO WIDTH NO-BREAK SPACE, and is applied by a certain class of applications.

When this convention is used, a signature at the beginning of a stream of coded characters indicates that the characters following are encoded in the UCS-2 or UCS-4 coded representation, and indicates the ordering of the octets within the coded representation of each character (see 6.3). It is typical of the class of applications mentioned above, that some make use of the signatures when receiving data, while others do not. The signatures are therefore designed in a way that makes it easy to ignore them.

In this convention, the ZERO WIDTH NO-BREAK SPACE character has the following significance when it is present at the beginning of a stream of coded characters:

UCS-2 signature: FEFF
UCS-4 signature: 0000 FEFF

An application receiving data may either use these signatures to identify the coded representation form, or may ignore them and treat FEFF as the ZERO WIDTH NO-BREAK SPACE character.

If an application which uses one of these signatures recognises its coded representation in reverse sequence (e.g. hexadecimal FFFE), the application can identify that the coded representations of the following characters use the opposite octet sequence to the sequence expected, and may take the necessary action to recognise the characters correctly.

NOTE - The hexadecimal value FFFE does not correspond to any coded character within this International Standard.

Annex F
(informative)

UCS transformation format (UTF)

The following method transforms the coded representation of graphic characters in this coded character set into a form that does not use octet values specified in ISO 2022 as coded representations of C0, SPACE, DEL, or C1 characters, and can thus be used for transmitting text data through communication systems that are sensitive to these octet values.

F.1 Outline of the algorithm

The algorithm can be summarized as follows:

1. A UCS character from 0000 0000 to 0000 009F is mapped to the corresponding octet from 00 to 9F.
2. A UCS character from 0000 00A0 to 0000 00FF is mapped to a sequence of two octets, with the first octet being A0, and the second octet in the range A0 to FF.
3. A UCS character from 0000 0100 to 0000 4015 is mapped to a sequence of two octets, with the first octet in the range from A1 to F5, and the second octet having 190 values in the range 21 to 7E or the range A0 to FF.
4. A UCS character from 0000 4016 to 0003 8E2D is mapped to a sequence of three octets, with the first octet in the range from F6 to FB, and the other octets in the range 21 to 7E or the range A0 to FF.
5. A UCS character at 0003 8E2E or larger is mapped to a sequence of five octets, with the first octet in the range from FC to FF, and the other octets in the range 21 to 7E or the range A0 to FF.

Notice that four-octet sequences are not used, since this maximizes the number of characters that can use the three-octet form.

F.2 Notation

1) All numbers are in hexadecimal notation.
2) Octet boundaries in the transformed text are indicated with semicolons.
3) The symbol "%" indicates the modulo operation, e.g.:

   x % y = x modulo y

   The symbol "/" indicates the integer division operation, e.g.:

   7 / 3 = 2

   Superscripting indicates the power-of operation, e.g.:

   $2^3 = 8$

   Precedence is "^{3}" > "/" > "%", e.g.:

   $x / y^2 \% w = ((x / (y^2)) \% w)$
4) T(z) is defined for z = 00..FF such that:

   z = 00 .. 5D: T(z) = z + 21
   z = 5E .. BD: T(z) = z + 42
   z = BE .. DE: T(z) = z - BE
   z = DF .. FF: T(z) = z - 60

   e.g. T(00) = 21, T(5D) = 7E,
   T(5E) = A0, T(BD) = FF,
   T(BE) = 00, T(DE) = 20,
   T(DF) = 7F, T(FF) = 9F
5) U(z) is the inverse of T(z): that is, U(T(z)) = z, and T(U(z)) = z:

   z = 00 .. 20: U(z) = z + BE
   z = 21 .. 7E: U(z) = z - 21
   z = 7F .. 9F: U(z) = z + 60
   z = A0 .. FF: U(z) = z - 42

   e.g. U(00) = BE, U(20) = DE,
   U(21) = 00, U(7E) = 5D,
   U(7F) = DF, U(9F) = FF,
   U(A0) = 5E, U(FF) = BD
6) The algorithm in this annex has been presented in a descriptive format. The implementation may differ for efficiency. For example, the T and U functions can be implemented with a small table look-up.

F.3 From UCS to UTF format

Condition/UCS	UTF octets
x = 0000 0000 .. 0000 009F:	x;
x = 0000 00A0 .. 0000 00FF:	A0; x;
x = 0000 0100 .. 0000 4015: (y = x - 0000 0100)	A1 + y/BE; T(y%BE);
x = 0000 4016 .. 0003 8E2D: (y = x - 0000 4016)	F6 + y/BE^2; T(y/BE%BE); T(y%BE);
x = 0003 8E2E .. 7FFF FFFF: (y = x - 0003 8E2E)	FC + y/BE^4; T(y/BE^3%BE); T(y/BE^2%BE); T(y/BE%BE); T(y%BE);

F.4 From UTF to UCS format

Condition/UTF	UCS
x = 00 .. 9F;	x
x = A0; y;	y
x = A1 .. F5; y;	(x-A1) × BE + U(y) +100
x = F6 .. FB; y; z;	(x-F6) × BE^2 + U(y) × BE + U(z) + 4016
x = FC .. FF; y; z; v; w;	(x-FC) × BE^4 + U(y) × BE^3 + U(z) × BE^2 + U(v) × BE + U(w) + 38E2E

Annex G
(informative)

Recommendation for combined receiving/originating devices with internal storage

This annex is applicable to a widely-used class of devices that can store received CC-data elements for subsequent retransmission.

This recommendation is intended to ensure that loss of information is minimised between the receipt of a CC-data-element and its retransmission.

A device of this class includes a receiving device component and an originating device component as in clause 2.3, and can also store received CC-data-elements for retransmission, with or without modification by the actions of the user on the corresponding characters represented within it.

It is therefore recommended that the originating device ensure that any such characters that may be outside the identified subset of the receiving device component be capable of being retransmitted without change to their coded representations, unless modified by the user. Such a device is a "receiving device with full retransmission capability".

If the originating device component can transmit only the coded representations of the characters of the subset adopted by the receiving device component, such a type of device is a "receiving device with subset retransmission capability".

Annex H
(informative)

Notations of octet value representations

Representation of octet values in this International Standard is different from other character coding standards such as ISO 2022, ISO/IEC 6429 and ISO 8859. This section clarifies the relationship between the two notations.

- In this International Standard, the notation used to express an octet value is z, where z is a hexadecimal number in the range 00 to FF.

  For example, the character ESCAPE (ESC) of ISO 2022 is represented by 1B.

- In other character coding standards, the notation used to express an octet value is x/y, where x and y are two numbers in the range 00 to 15. The correspondence between the notations of the form x/y and the octet value is as follows.

  x is the number represented by bit 8, bit 7, bit 6 and bit 5 where these bits are given the weight 8, 4, 2 and 1 respectively;

  y is the number represented by bit 4, bit 3, bit 2 and bit 1 where these bits are given the weight 8, 4, 2 and 1 respectively.

  For example, the character ESC of ISO 2022 is represented by 01/11.

Thus ISO 2022 (and other character coding standards) octet value notation can be converted to ISO/IEC 10646 octet value notation by converting the value of x and y to hexadecimal notation. For example; 04/15 is equivalent with 4F.

Annex J
(informative)

Character naming guidelines

Guidelines for generating and presenting unique names of characters in ISO/IEC JTC1/SC2 standards is listed in this annex for reference. These guidelines are used in information technology coded character set standards such as ISO 646, ISO 6937, ISO 8859, ISO/IEC 10367 as well as in this International Standard.

These Guidelines specify rules for generating and presenting unique names of characters. Rules 1 to 3 are implemented without exceptions. However it must be accepted that in some cases (e.g., historical or traditional usage, unforeseen special cases, difficulties inherent to the nature of the character considered), exceptions to some of the other rules will have to be tolerated. Nonetheless, these rules are applied wherever possible.

Rule 1

By convention, only Latin capital letters A to Z, space, and hyphen shall be used for writing the names of characters.

> NOTE - Names of ideographic characters may also include digits 0 to 9 provided that a digit is not the first character in a word.

Rule 2

The names of control functions shall be coupled with an acronym consisting of Latin capital letters A to Z and, where required, digits. Once the name has been specified for the first time, the acronym may be used in the remainder of the text where required for simplification and clarity of the text.

Exceptionally, acronyms may be used for graphic characters where usage already exists and clarity requires it, in particular in code tables.

Examples:

Name: LOCKING-SHIFT TWO RIGHT
Acronym: LS2R

Name: SOFT-HYPHEN
Acronym: SHY

> NOTE - In ISO/IEC 6429, also the names of the modes have been presented in the same way as control functions.

Rule 3

In some cases, the names of a character can be followed by an additional explanatory statement not part of the name. These statements shall be in parentheses and not in capital Latin letters except the initials of the word where required. See examples in rule 12.

Rule 4

The names of a character shall wherever possible denote its customary meaning, for examples PLUS SIGN. Where this is not possible, names should describe shapes, not usage; for example: UPWARDS ARROW.

Rule 5

Only one name will be given to each character.

Rule 6

The names shall be constructed from an appropriate set of the applicable terms of the following grid and ordered in the sequence of this grid. Exceptions are specified in Rule 11.

1	Script
2	Case
3	Type
4	Language
5	Attribute
6	Designation
7	Mark(s)
8	Qualifier

Examples of such terms:

Script	Latin, Cyrillic, Arabic
Case	capital, small
Type	letter, ligature, digit
Language	Ukrainian
Attribute	final, sharp, subscript, vulgar
Designation	customary name, name of letter

Mark	acute, ogonek, ring above, diaeresis
Qualifier	sign, symbol

Examples of names:

LATIN CAPITAL LETTER A WITH ACUTE
1 2 3 6 7

DIGIT FIVE
3 6

LEFT CURLY BRACKET
5 5 6

NOTE - A ligature is a graphic symbol in which two or more other graphic symbols are imaged as single graphic symbol.

Rule 7

The letters of the Latin script shall be represented within their name by their basic graphic symbols (A, B, C, ...). The letters of all other scripts shall be represented by their transcription in the language of the first published International Standard.

Examples:

K	LATIN CAPITAL LETTER K
Ю	CYRILLIC CAPITAL LETTER YU

Rule 8

In principle when a character of a given script is used in more than one language, no language name is specified. Exceptions are tolerated where an ambiguity would otherwise result.

Examples:

И	CYRILLIC CAPITAL LETTER I
I	CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I

Rule 9

Letters that are elements of more than one script are considered different even if their shape is the same, they shall have different names.

Examples:

A	LATIN CAPITAL LETTER A
Α	GREEK CAPITAL LETTER ALPHA
А	CYRILLIC CAPITAL LETTER A

Rule 10

A character of one script used in isolation in another script, for example as a graphic symbol in relation with physical units of dimension, is considered as a character different from the character of its native script.

Example:

µ	MICRO SIGN

Rule 11

A number of characters have a traditional name consisting of one or two words. It is not intended to change this usage.

Examples:

'	APOSTROPHE
:	COLON
@	COMMERCIAL AT
_	LOW LINE
~	TILDE

Rule 12

In some cases, characters of a given script, often punctuation marks, are used in another script for a different usage. In these cases the customary name reflecting the most general use will be given to the character. The customary name may be followed in the list of characters of a particular standard by the name in parentheses which this character has in the script specified by this particular standard.

Examples:

HORIZONTAL BAR (Parenthetiki pavla)
DIAERESIS (Dialytika)

Rule 13

The above rules shall not apply to ideographic characters. These characters will be identified by alpha-numeric identifiers specified for each ideographic character (see clause 25).

Annex K
(informative)

Sources of characters

Several sources and contributions were used for constructing this coded character set. In particular, all characters of the following national and international standards are included in this International Standard.

ISO 233:1984 Documentation — Transliteration of Arabic characters into Latin characters.

ISO 646:1991 Information processing — ISO 7-bit coded character set for information interchange.

ISO 2047:1975 Information processing — Graphical representations for the control characters of the 7-bit coded character set.

ISO 2022:1986 Information processing — ISO 7-bit and 8-bit coded character sets — Coded extension techniques.

ISO 2033:1983 Information processing — Coding of machine-readable characters (OCR and MICR).

ISO 5426:1981 Extension of the Latin alphabet coded character set for bibliographic information interchange.

ISO 5427:1984 Extension of the Cyrillic alphabet coded character set for bibliographic information interchange.

ISO 5428:1984 Greek alphabet coded character set for bibliographic information interchange.

ISO 6438:1983 Documentation — African coded character set for bibliographic information interchange.

ISO DIS 6861.2:1990 Information and documentation — Glagolitic character set for bibliographic information interchange.

ISO DIS 6862.2:1990 Documentation — Mathematical coded character set for bibliographic information interchange.

ISO DIS 6937:1991 Information processing — Coded character sets for text communication.

ISO 8859 Information processing — 8-bit single-byte coded graphic character sets. Contents:
-Part 1. Latin alphabet No. 1 (1987).
-Part 2. Latin alphabet No. 2 (1987).
-Part 3. Latin alphabet No. 3 (1988).
-Part 4. Latin alphabet No. 4 (1988).
-Part 5. Latin/Cyrillic alphabet (1988)
-Part 6. Latin/Arabic alphabet (1987)
-Part 7. Latin/Greek alphabet (1987)
-Part 8. Latin/Hebrew alphabet (1988)
-Part 9. Latin alphabet No. 5 (1989).

ISO 8879:1986 Information processing — Text and office systems — Standard Generalized Markup Language (SGML).

ISO DP 8957:1987 Hebrew alphabet character sets for bibliographic interchange. (ISO/TC46/SC4 N 205).

ISO 9036:1983 Information processing — Arabic 7-bit coded character set for information interchange.

ISO/IEC 10367:1991 Information processing — Repertoire of standardized coded graphic character sets for use in 8-bit codes.

ISO international register of character sets to be used with escape sequences. (registration procedure ISO 2375:1985) .

ANSI X3.4-1986 American National Standards Institute. Coded character set — 7-bit American national standard code for information interchange.

ANSI X3.32-1973 American National Standards Institute. American national standard graphic representation of the control characters of American national standard code for information interchange.

ANSI Y10.20-1988 American National Standards Institute. Mathematic signs and symbols for use in physical sciences and technology.

ANSI Y14.5M-1982 American National Standard. Engineering drawings and related document practices, dimensioning and tolerances.

ANSI Z39.47-1985 American National Standards Institute. Extended Latin alphabet coded character set for bibliographic use.

ANSI Z39.64-1989 American National Standards Institute. East Asian character code for bibliographic use.

ASMO 449-1982 Arab Organization for Standardization and Methodology. Data processing — 7-bit coded character set for information interchange.

GB 2312-1980 Code of Chinese Graphic Character Set for Information Interchange: Jishu Biaozhun Chubanshe (Technical Standards Publishing).

LTD 37(1610)-1988 Indian standard code for information interchange.

JIS X 0201-1976 Japanese Standards Association. Jouhou koukan you fugou (Code for Information Interchange).

JIS X 0208-1990 Japanese Standards Association. Jouhou koukan you kanji fugoukei (Code of the Japanese Graphic Character Set for Information Interchange).

JIS X 0212-1990 Japanese Standards Association. Jouhou koukan you kanji fugou-hojo kanji (Code of the supplementary Japanese graphic character set for information interchange).

KS C 5601-1987 Korea Industrial Standards Association. Jeongho gyohwanyong buho (Hangul mit Hanja) (Code for Information Interchange (Hangul and Hanja)).

KS C 5657-1991 Korean Industrial Standards Association. Jeongho gyohwanyong buho hwakjang saten (Code of the supplementary Korean graphic character set for information interchange).

TIS 620-2529:1986 Thai Industrial Standard for Thai Character Code for Computer.

Esling, John. Computer coding of the IPA: supplementary report. Journal of the International Phonetic Association, 20:1 (1990), p. 22-26.

International Phonetic Association. The IPA 1989 Kiel Convention Workgroup 9 report: Computer Coding of IPA Symbols and Computer Representation of Individual Languages. Journal of the International Phonetic Association, 19:2 (1989), p. 81-82.

International Phonetic Association. The International Phonetic Alphabet (revised to 1989).

Knuth, Donald E. The TeXbook. — 19th. printing, rev. — Reading, MA : Addison-Wesley, 1990.

Pullum, Geoffrey K. Phonetic symbol guide. Geoffrey K. Pullum and William A. Ladusaw. — Chicago : University of Chicago Press, 1986.

Pullum, Geoffrey K. Remarks on the 1989 revision of the International Phonetic Alphabet. Journal of the International Phonetic Association, 20:1 (1990), p. 33-40.

Selby, Samuel M. Standard mathematical tables. — 16th ed. — Cleveland, OH : Chemical Rubber Co., 1968. Shepherd, Walter.

Shepherd, Walter. Shepherd's glossary of graphic signs and symbols. Compiled and classified for ready reference. — New York : Dover Publications, [1971].

Shinmura, Izuru. Kojien (Shinmura Izuru hen). — Dai 4-han. — Tokyo : Iwanami Shoten, Heisei 3 [1991].

The Unicode Consortium. The Unicode Standard. Worldwide Character Encoding Version 1.0, Volume One. — Reading, MA : Addison-Wesley, 1991.

Annex L
(informative)

Scripts under consideration for future editions of this International Standard

In order to make sure that this International Standard is useful for people using their native scripts, characters included in this International Standard were selected with input and feedback from national standards organisations and/or qualified experts.

Some scripts and symbols were not included in this edition because sufficient input and feedback have not been provided during the preparation and review stages.

It is intended that character code positions for these scripts and symbols will be allocated when sufficient input and review is provided. Such scripts and symbols includes:

- Burmese
- Cree and Inuktitut
- Ethiopian
- Hieroglyphics
- Khmer
- Maldivian
- Mongolian
- Runic
- Sinhalese
- Syriac
- Yi

This list is not exhaustive. Other scripts and symbols as well as additional characters for the included scripts are expected to be included in future editions of this International Standard.