INTERNATIONAL STANDARD

ISO/IEC 10646-1

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

Part 1:
Architecture and Basic Multilingual Plane

AMENDMENT 1: Mathematical symbols and other characters

Technologies de l'information — Jeu universel de caractères codés sur plusieurs octets (JUC) —

Partie 1: Architecture et plan multilingue de base

AMENDEMENT 1: Symboles mathématiques et autres caractères

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

iTeh STANDARD PREVIEW

(standards.iteh.ai)

ISO/IEC 10646-1:2000/Amd 1:2002
https://standards.iteh.ai/catalog/standards/sist/6868c3d4-2888-42fb-9efd-fb0132b341f5/iso-iec-
10646-1-2000-amd-1-2002

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this Amendment may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Amendment 1 to International Standard ISO/IEC 10646-1:2000 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 2, *Coded character sets*.

This corrected version of ISO/IEC 10646-1:2000/Amd.1:2002 includes the electronic attachments that were omitted in the original version.

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

Part 1:
Architecture and Basic Multilingual Plane

AMENDMENT 1: Mathematical symbols and other characters

Page v, Foreword

Replace:

— *Part 2: Secondary Multilingual Plane for scripts and symbols, Supplementary Plane for CJK Ideographs, Special Purpose Plane*

by:

— *Part 2: Supplementary Planes*

Remove the text "Additional parts will specify other planes".

Page 2, clause 3

Append the following references:

ISO/IEC 10646-2:2001, Information Technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 2: Supplementary Planes.

Unicode Standard Annex, UAX#9, The Unicode Bidirectional Algorithm, Version 3.1.0, 2001-03-23.

Page 6, figure 2

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

Page 7, clause 6

Change "four-digit form" *and* "4-digit form" *to* "four-to-six-digit form" *throughout.*

In clause 6.5, replace title, first paragraph and add note as follows:

6.5 Short identifiers for code positions (UIDs)

ISO/IEC 10646 defines short identifiers for each code position, including code positions that are reserved. A short identifier for any code position is distinct from a short identifier for any other code position. If a character is allocated at a code position, a short identifier for that code position can be used to refer to the character allocated at that code position.

> NOTE 1 – For instance, U+DC00 identifies a code position that is permanently reserved for UTF-16, and U+FFFF identifies a code position that is permanently reserved. U+0025 identifies a code position to which a character is allocated; U+0025 also identifies that character (named PERCENT SIGN).

Rename the existing note to NOTE 2, and increment all other note numbers of the sub-clause by one (NOTE 3 and 4).

In clause 6.5.b change "It is not defined if the first four digits of the eight-digit form are not all zeros" *to* "It is not defined if the eight-digit form is greater than 0010FFFF" *and append at the end of the text* "Leading zeros beyond four digits are suppressed.".

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

Add a clause 6.5f:

f. For the 8 digit forms, the characters SPACE or NO-BREAK SPACE may optionally be inserted before the four last digits.

Page 8, clause 6

After clause 6.5, add a new clause 6.6 (UCS Sequence Identifiers) and a note:

"ISO/IEC 10646 defines an identifier for any sequence of code positions taken from the standard. Such an identifier is known as a UCS Sequence Identifier (USI). For a sequence of n code positions it has the following form:

      <UID1, UID2, ..., UIDn>

where UID1, UID2, etc. represent the short identifiers of the corresponding code positions, in the same order as those code positions appear in the sequence. If each of the code positions in such a sequence has a character allocated to it, the USI can

be used to identify the sequence of characters allocated at those code positions. The syntax for UID1, UID2, etc. is specified in clause 6.5. A COMMA character (optionally followed by a SPACE character) separates the UIDs. The UCS Sequence Identifier shall include at least two UIDs; it shall begin with a LESS-THAN SIGN and be terminated by a GREATER-THAN SIGN."

> NOTE – UCS Sequences Identifiers cannot be used for specification of subset and collection content. They may be used outside this standard to identify: composite sequences for mapping purposes, font repertoire, etc.

Page 8, clause 7

Remove the note.

Page 8, Clause 8

Insert the following note after the first paragraph which ends "UCS-2 (see 13.1)".

> NOTE 1 – Since UCS-2 only contains the repertoire of the BMP it is not fully interoperable with UCS-4, UTF-8 and UTF-16.

Insert the following paragraph and note after the second paragraph which ends "reserved for control characters."

Code positions 0000 2060 to 0000 206F, 0000 FFF0 to 0000 FFFC, and 000E 0000 to 000E 0FFF are reserved for Alternate Format Characters (see annex F).

> NOTE 2 – Unassigned code positions in those ranges may be ignored in normal processing and display.

Replace the last paragraph as follows:

Code positions 0000 FDD0 to 0000 FDEF, 0000 FFFE, and 0000 FFFF are permanently reserved.

> NOTE 3 – Code position 0000 FFFE is reserved for "signature" (see annex H). Code positions 0000 FDD0 to 0000 FDEF, and 0000 FFFF can be used for internal processing uses requiring numeric values which are guaranteed not to be coded characters, such as in terminating tables, or signaling end-of-text. Furthermore, since 0000 FFFF is the largest BMP value, it may also be used as the final value in binary or sequential searching index within the context of UCS-2 or UTF-16."

> NOTE 4 – A "permanently reserved" code position cannot be changed by future amendments.

Page 8, clause 9

Replace clause 9 to read as follows:

9. Other planes

9.1 Planes reserved for future standardization

Planes 11 to FF in Group 00 and all planes in any other groups (i.e. Planes 00 to FF in Groups 01 to 7F) are reserved for future standardization, and thus those code positions shall not be used for any other purpose.

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

> NOTE – To ensure continued interoperability between the UTF-16 form and other coded representations of the UCS, it is intended that no characters will be allocated to code positions in Planes 11 to FF in Group 00 or any planes in any other groups.

9.2 Planes accessible by UTF-16

Each code position in Planes 01 to 10 of Group 00 has a unique mapping to a four-octet sequence in accordance with the UTF-16 form of coded representation (see annex C). This form is compatible with the two-octet BMP form of UCS-2 (see 13.1).

Page 8, clause 10

Replace sub-clause 10.2 to read as follows:

10.2 Code positions for private use characters

The code positions of Plane 0F and Plane 10 of Group 00 shall be for private use.

The 6400 code positions E000 to F8FF of the Basic Multilingual Plane shall be for private use.

The contents of these code positions are not specified in ISO/IEC 10646 (see 10.1).

Page 9, clause 14

In the first paragraph, change the text "listed in annex B" *to* "listed in annex B of each part of ISO/IEC 10646".

Page 9, clause 14.1

Change the text "(see clause B.1)" *to* "(see clause B.1 of each part of ISO/IEC 10646)".

Page 9, clause 14.2

Change the text "listed in clause B.2" *to* "listed in clause B.2 of each part of ISO/IEC 10646".

Page 12, clause 19

Add a new clause 19.1 at the end of clause 19.

19.1 Directional of bi-directional text

The Unicode Bidirectional Algorithm describes the algorithm used to determine the directionality for bidirectional text.

Page 13, Clause 22 Compatibility characters

Add the following sub-clauses:

22.1 Compatibility Ideographs
The CJK compatibility ideographs (characters part of the CJK COMPATIBILITY IDEOGRAPHS-2001 collection) are ideographs that should have been unified with one of the CJK unified ideographs (characters part of the CJK UNIFIED IDEOGRAPHS-2001 collection), per the unification rule described in Annex S of this International Standard.

However, they are included in this International Standard as separate characters, because, based on various national, cultural, or historical reason for some specific country and region, some national and regional standards assign separate code positions for them.

> NOTE – For this reason, compatibility ideographs should only be used for maintaining and guaranteeing a round trip conversion with the specific national, regional, or other standard. Other usage is strongly discouraged.

22.2 Source references for CJK Compatibility Ideographs

ISO/IEC 10646-1 provides the source reference for CJK Compatibility Ideographs included in the Basic Multilingual Plane. ISO/IEC 10646-2 provides the source reference for CJK Compatibility Ideographs included in the Supplementary Planes.

The source references for the CJK Compatibility Ideographs included in the BMP are provided in a machine-readable format that is accessible as a link to this document. The content pointed to by this link is also normative.

> NOTE 1 – The referenced file is only available to users who obtain their copy of the standard in a machine-readable form. However, the file format makes it printable.

The Kanji J sources are:

J3   JIS X 213:2000 level-3
J4   JIS X 213:2000 level-4

The Hanja K source is:

K0   KS C 5601-1987

The Unicode U source is:

U0   The Unicode Standard 3.0-2000

The content linked to is a plain text file that specifies, after a 8-line header, as many lines as CJK Compatibility Ideographs; each containing the following information organized in fixed width fields:
- 01-06 octet: BMP code value (`0hhhh`)
- 07-12 octet: Corresponding CJK Unified Ideograph (`0hhhh`).

- 13-20 octet: Kanji J sources (`J3-hhhh`), (`J4-hhhh`).
- 21-28 octet: Hanja K sources (`K0-hhhh`).
- 29-36 octet : Unicode U sources (`U0-hhhh`)

The format definition uses 'h' as a hexadecimal unit. Uppercase characters and all other symbols including the space character between parentheses appear as shown.

Click on this highlighted text to access the reference file.

> NOTE 2 – The content is also available as a separate viewable file in the same file directory as this document. The file is named: "CJKC0SR.txt".

Page 13, clause 24

In the first paragraph, change the text "clause B.1" *to* "clause B.1 of each part of ISO/IEC 10646"*, and change the text* "clause B.2" *to* "clause B.2 of each part of ISO/IEC 10646"*.*

Page 14, clause 24

A sub-clause 24.5 is added to the clause 24 Combining Characters as follows:

24.5 Variation selectors

Variation selectors are combining characters following immediately a specific base character to indicate a specific variant form of graphic symbol for that character. Some variation selectors are specific to a script, such as the Mongolian free variation selectors, others are used with various other base characters such as the mathematical symbols. Variations selectors following other characters have no effect on the selection of the graphic symbol for that character.

No sequences using characters from VARIATION SELECTOR-2 to VARIATION SELECTOR-16 are defined at this time.

The following table provides a description of the variant appearances corresponding to the use of appropriate variation selectors with all allowed base mathematical symbols.

> NOTE 1 – The VARIATION SELECTOR-1 is the only variation selector used with mathematical symbols.

Sequence (UID notation)	Description of variant appearance
<2229, FE00>	INTERSECTION with serifs
<222A, FE00>	UNION with serifs
<2268, FE00>	LESS-THAN BUT NOT EQUAL TO with vertical stroke
<2269, FE00>	GREATER-THAN AND NOT DOUBLE EQUAL with vertical stroke

Sequence (UID notation)	Description of variant appearance
<2272, FE00>	LESS-THAN OR EQUIVALENT TO following the slant of the lower leg
<2273, FE00>	GREATER-THAN OR EQUIVALENT TO following the slant of the lower leg
<2278, FE00>	NEITHER LESS-THAN NOR GREATER-THAN with vertical stroke
<2279, FE00>	NEITHER GREATER-THAN NOR LESS-THAN with vertical stroke
<228A, FE00>	SUBSET OF WITH NOT EQUAL TO with stroke through bottom members
<228B, FE00>	SUPERSET OF WITH NOT EQUAL TO with stroke through bottom members
<2293, FE00>	SQUARE CAP with serifs
<2294, FE00>	SQUARE CUP with serifs
<2295, FE00>	CIRCLED PLUS with white rim
<2297, FE00>	CIRCLED TIMES with white rim
<229C, FE00>	CIRCLED EQUALS equal sign touching the circle
<22DA, FE00>	LESS-THAN EQUAL TO OR GREATER-THAN with slanted equal
<22DB, FE00>	GREATER-THAN EQUAL TO OR LESS-THAN with slanted equal
<2A3C, FE00>	INTERIOR PRODUCT tall variant with narrow foot
<2A3D, FE00>	RIGHTHAND INTERIOR PRODUCT tall variant with narrow foot
<2A9D, FE00>	SIMILAR following the slant of the upper leg OR LESS-THAN
<2A9E, FE00>	SIMILAR following the slant of the upper leg OR GREATER-THAN
<2AAC, FE00>	SMALLER THAN OR EQUAL TO with slanted equal
<2AAD, FE00>	LARGER THAN OR EQUAL TO with slanted equal
<2ACB, FE00>	SUBSET OF ABOVE NOT EQUAL TO with stroke through bottom members
<2ACC, FE00>	SUPERSET OF ABOVE NOT EQUAL TO with stroke through bottom members

The following table provides a description of the variant appearances corresponding to the use of appropriate variation selectors with all allowed base Mongolian characters. Only some presentation forms of the base Mongolian characters used with the Mongolian free variation selectors produce variant appearances. These combinations are described in the following table.

> NOTE 2 – The Mongolian characters have various presentation forms depending on their position in a CC-data element. These presentations forms are called isolate, initial, medial and final.

Sequence (UID notation)	position	Description of variant appearance
<1820, 180B>	isolate, medial, final	MONGOLIAN LETTER A second form
<1820, 180C>	medial	MONGOLIAN LETTER A third form
<1821, 180B>	initial, final	MONGOLIAN LETTER E second form

Sequence (UID notation)	position	Description of variant appearance
<1822, 180B>	medial	MONGOLIAN LETTER I second form
<1823, 180B>	medial, final	MONGOLIAN LETTER O second form
<1824, 180B>	medial	MONGOLIAN LETTER U second form
<1825, 180B>	medial, final	MONGOLIAN LETTER OE second form
<1825, 180C>	medial	MONGOLIAN LETTER OE third form
<1826, 180B>	isolate, medial, final	MONGOLIAN LETTER UE second form
<1826, 180C>	medial	MONGOLIAN LETTER UE third form
<1828, 180B>	initial, medial	MONGOLIAN LETTER NA second form
<1828, 180C>	medial	MONGOLIAN LETTER NA third form
<1828, 180D>	medial	MONGOLIAN LETTER NA separate form
<182A, 180B>	final	MONGOLIAN LETTER BA alternative form
<182C, 180B>	initial, medial	MONGOLIAN LETTER QA second form
<182C, 180B>	isolate	MONGOLIAN LETTER QA feminine second form
<182C, 180C>	medial	MONGOLIAN LETTER QA third form
<182C, 180D>	medial	MONGOLIAN LETTER QA fourth form
<182D, 180B>	initial, medial	MONGOLIAN LETTER GA second form
<182D, 180B>	final	MONGOLIAN LETTER GA feminine form
<182D, 180C>	medial	MONGOLIAN LETTER GA third form
<182D, 180D>	medial	MONGOLIAN LETTER GA feminine form
<1830, 180B>	final	MONGOLIAN LETTER SA second form
<1830, 180C>	final	MONGOLIAN LETTER SA third form
<1832, 180B>	medial	MONGOLIAN LETTER TA second form
<1833, 180B>	initial, medial, final	MONGOLIAN LETTER DA second form
<1835, 180B>	final	MONGOLIAN LETTER JA second form
<1836, 180B>	initial, medial	MONGOLIAN LETTER YA second form
<1836, 180C>	medial	MONGOLIAN LETTER YA third form
<1838, 180B>	final	MONGOLIAN LETTER WA second form
<1844, 180B>	medial	MONGOLIAN LETTER TODO E second form
<1845, 180B>	medial	MONGOLIAN LETTER TODO I second form
<1846, 180B>	medial	MONGOLIAN LETTER TODO O second form

<1847, 180B>	isolate, medial, final	MONGOLIAN LETTER TODO U second form
<1847, 180C>	medial	MONGOLIAN LETTER TODO U third form
<1848, 180B>	medial	MONGOLIAN LETTER TODO OE second form
<1849, 180B>	isolate, medial	MONGOLIAN LETTER TODO UE second form
<184D, 180B>	initial, medial	MONGOLIAN LETTER TODO QA feminine form
<184E, 180B>	medial	MONGOLIAN LETTER TODO GA second form
<185D, 180B>	medial, final	MONGOLIAN LETTER SIBE E second form
<185E, 180B>	medial, final	MONGOLIAN LETTER SIBE I second form
<185E, 180C>	medial, final	MONGOLIAN LETTER SIBE I third form
<1860, 180B>	medial, final	MONGOLIAN LETTER SIBE UE second form
<1863, 180B>	medial	MONGOLIAN LETTER SIBE KA second form
<1868, 180B>	initial, medial	MONGOLIAN LETTER SIBE TA second form
<1868, 180C>	medial	MONGOLIAN LETTER SIBE TA third form
<1869, 180B>	initial, medial	MONGOLIAN LETTER SIBE DA second form
<186F, 180B>	initial, medial	MONGOLIAN LETTER SIBE ZA second form
<1873, 180B>	medial, final	MONGOLIAN LETTER MANCHU I second form
<1873, 180C>	medial, final	MONGOLIAN LETTER MANCHU I third form
<1873, 180D>	medial	MONGOLIAN LETTER MANCHU I fourth form
<1874, 180B>	medial	MONGOLIAN LETTER MANCHU KA second form
<1874, 180B>	final	MONGOLIAN LETTER MANCHU KA feminine first form
<1874, 180C>	medial	MONGOLIAN LETTER MANCHU KA feminine first form
<1874, 180C>	final	MONGOLIAN LETTER MANCHU KA feminine second form
<1874, 180D>	medial	MONGOLIAN LETTER MANCHU KA feminine second form
<1876, 180B>	initial, medial	MONGOLIAN LETTER MANCHU FA second form
<1880, 180B>	all	MONGOLIAN LETTER ALI GALI ANUSVARA ONE second form
<1881, 180B>	all	MONGOLIAN LETTER ALI GALI VISARGA ONE second form
<1887, 180B>	isolate, final	MONGOLIAN LETTER ALI GALI A second form
<1887, 180C>	final	MONGOLIAN LETTER ALI GALI A third form

<1887, 180D>	final	MONGOLIAN LETTER ALI GALI A fourth form
<1888, 180B>	final	MONGOLIAN LETTER ALI GALI I second form
<188A, 180B>	initial, medial	MONGOLIAN LETTER ALI GALI NGA second form

NOTE 3 – The variation selector only selects a different *appearance* of an already encoded character. It is not intended as a general code extension mechanism. Only the sequences specifically defined in this annex are sanctioned for standard use; all other sequences are undefined. No sequences containing combining characters or composite characters will be defined.

NOTE 4 – The exhaustive list of standardized variants is also described as *StandardizedVariants.html* in the Unicode character database.

Page 880, Annex A.1

Add a '' (for fixed collection) to the following collection:*

38 ARROWS
39 MATHEMATICAL OPERATORS,
45 BLOCK ELEMENTS,
46 GEOMETRIC SHAPES,
49 CJK SYMBOLS AND PUNCTUATION
51 KATAKANA

In the list of collection numbers and names, after

91 TIBETAN

insert new entries as follows:

92 CYRILLIC SUPPLEMENT	0500-052F
93 TAGALOG	1700-171F
94 HANUNOO	1720-173F
95 BUHID	1740-175F
96 TAGBANWA	1760-177F
97 MISCELLANEOUS MATHEMATICAL SYMBOLS-A	27C0-27EF
98 SUPPLEMENTAL ARROWS-A	27F0-27FF *
99 SUPPLEMENTAL ARROWS-B	2900-297F *
100 MISCELLANEOUS MATHEMATICAL SYMBOLS-B	2980-29FF *
101 SUPPLEMENTAL MATHEMATICAL OPERATORS	2A00-2AFF *
102 KATAKANA PHONETIC EXTENSIONS	31F0-31FF *
103 VARIATION SELECTORS	FE00-FE0F *

after

271 COMBINING CHARACTERS B-2

insert new entries as follows:

281 MES-1	SEE A.4.1 *
282 MES-2	SEE A.4.2 *

Page 881, annex A.1

In the list of collections numbers and names, after

302 BMP SECOND EDITION   See A3.3   *

insert the following text:

The following collections contain characters both inside and outside the Basic Multilingual Plane.

```
303 UNICODE 3.1            See A5.1   *
304 UNICODE 3.2            See A5.2   *
380 CJK UNIFIED IDEOGRAPHS-2001 *
                           3400-4DB5
                           4E00-9FA5
                           FA0E-FA0F
                           FA11
                           FA13-FA14
                           FA1F
                           FA21
                           FA23-FA24
                           FA27-FA29
                           20000-2A6D6
381 CJK COMPATIBILITY IDEOGRAPHS-2001*
                           F900-FA0D
                           FA10
                           FA12
                           FA15-FA1E
                           FA20
                           FA22
                           FA25-FA26
                           FA2A-FA6A
                           2F800-2FA1D
10646 UNICODE              0000-FDCF
                           FDF0-FFFD
                           10000-1FFFD
                           20000-2FFFD
                           30000-3FFFD
                           40000-4FFFD
                           50000-5FFFD
                           60000-6FFFD
                           70000-7FFFD
                           80000-8FFFD
                           90000-9FFFD
                           A0000-AFFFD
                           B0000-BFFFD
                           C0000-CFFFD
                           D0000-DFFFD
                           E0000-EFFFD
                           F0000-FFFFD
                           100000-10FFFD
```

NOTE – The UNICODE collection incorporates all characters currently encoded in the standard.

Page 881, annex A.1

Under collection 400, change "PRIVATE USE PLANES G=00, P=0F, 10, & E0-FF" *to* "(This collection number shall not be used, see Note 2.)".

Add the following collection after 400:

401   PRIVATE USE PLANES-0F-10   G=00, P=0F-10

Under collection 500, change "PRIVATE USE GROUPS G=60-7F" *to* "(This collection number shall not be used, see Note 2.)".

Under Note 1, change the text "90, and 91" *to* "90, 91, 93, 94, 95, and 96".

Replace the Note 2 as follows:

NOTE 2 – Collections numbered 57, 58, and 59 were specified in the First Edition of this International Standard but have now been deleted. Collections numbered 400 and 500 were specified in the First and Second Editions of this Standard but have now been deleted.

In the alphabetical list of keywords, add collections "98" *and* "99" *to the entry* "Arrows";

Add collection "92" *to the entry* "Cyrillic";

Add collection "102" *to the entry* "Katakana";

Add collection "101" *to the entry* "Mathematical operators";

Add collection "401" *and remove* "400 500" *from the entry* "Private use";

Add collections "97" *and* "100" *to the entry* "Symbols".

In the alphabetical list of keywords, after
    Braille patterns
insert
    Buhid   95
After
    Hangul
insert
    Hanunoo   94
After
    Mathematical operators
insert
    Mathematical symbols 97 and 100
    MES   281 282
After
    Syriac
insert
    Tagalog   93
    Tagbanwa   96
After
    Tibetan
insert
    Unicode   303, 304, 10646
    Variation selectors 103

In the alphabetical list of keywords, add a tab between "Yi" *and the collection numbers* "76 77".

Page 881, Annex A.2

In the list of blocks in the BMP, after
    CYRILLIC
insert a new *entry* as follows:
    CYRILLIC SUPPLEMENT   0500-052F

after
    RUNIC
insert new entries as follows:
    TAGALOG   1700-171F
    HANUNOO   1720-173F

BUHID 1740-175F
TAGBANWA 1760-177F

after
DINGBATS
insert new entries as follows:
MISCELLANEOUS MATHEMATICAL
SYMBOLS-A 27C0-27EF
SUPPLEMENTAL ARROWS-A 27F0-27FF

after
BRAILLE PATTERNS
insert new entries as follows:
SUPPLEMENTAL ARROWS-B 2900-297F
MISCELLANEOUS MATHEMATICAL
SYMBOLS-B 2980-29FF
SUPPLEMENTAL MATHEMATICAL
OPERATORS 2A00-2AFF

after
BOPOMOFO EXTENDED
insert a new entry as follows:
KATAKANA PHONETIC
EXTENSIONS 31F0-31FF

after
ARABIC PRESENTATION FORMS-B
insert a new entry as follows:
VARIATION SELECTORS FE00-FE0F

Page 882, annex A.3.1 301 BMP-AMD.7

Within the row 0B entry, replace the following sequence:
.. 8E-90 92-25 99-9A ..
by:
.. 8E-90 92-95 99-9A ..

Page 883, annex A.3.3 302 BMP Second Edition

Row 02 entry::
00-33 50-AD B0-EE
is changed to:
00-1F 22-33 50-AD B0-EE
Row 07 entry:
00-0D 0F-2C 30-4A 80-BF
is changed to:
00-0D 0F-2C 30-4A 80-B0
Row 0B entry:
.. 8E-90 92-25 99-9A ..
is changed to:
.. 8E-90 92-95 99-9A ..
Row 12 entry:
20-26 28-46 ..
is changed to:
00-06 08-46 ..
Row 34-4D entry:

3400-4DBF
is changed to:
3400-4DB5

Page 884, annex A

At the end of Annex A add new clauses A.4 and A.5 as follows.

A.4 Other collections within the BMP

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

NOTE – The acronym MES indicates Multilingual European Subset.

A.4.1 281 MES-1

281 MES-1 is specified by the following ranges of code positions as indicated for each row.

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

A.4.2 282 MES-2

282 MES-2 is specified by the following ranges of code positions as indicated for each row.

Rows	Positions (cells)
00	20-7E A0-FF
01	00-7F 8F 92 B7 DE-EF FA-FF
02	18-1B 1E-1F 59 7C 92 BB-BD C6-C7 C9 D8-DD EE
03	74-75 7A 7E 84-8A 8C 8E-A1 A3-CE D7 DA-E1
04	00-5F 90-C4 C7-C8 CB-CC D0-EB EE-F5 F8-F9
1E	02-03 0A-0B 1E-1F 40-41 56-57 60-61 6A-6B 80-85 9B F2-F3
1F	00-15 18-1D 20-45 48-4D 50-57 59 5B 5D 5F-7D 80-B4 B6-C4 C6-D3 D6-DB DD-EF F2-F4 F6-FE
20	13-15 17-1E 20-22 26 30 32-33 39-3A 3C 3E 44 4A 7F 82 A3-A4 A7 AC AF
21	05 16 22 26 5B-5E 90-95 A8
22	00 02-03 06 08-09 0F 11-12 19-1A 1E-1F 27-2B 48 59 60-61 64-65 82-83 95 97
23	02 10 20-21 29-2A
25	00 02 0C 10 14 18 1C 24 2C 34 3C 50-6C 80 84 88 8C 90-93 A0 AC B2 BA BC C4 CA-CB D8-D9
26	3A-3C 40 42 60 63 65-66 6A-6B
FB	01-02
FF	FD

A.5 Unicode collections

These collections correspond to Unicode 3.1 and 3.2. They include characters from the BMP as well as Supplementary Planes.

A.5.1 303 UNICODE 3.1

303   The fixed collection UNICODE 3.1 consists of collections from A.3 above and several ranges of code positions. The collection list is arranged by planes as follows.

Plane 0

Collection number and name
302   BMP SECOND EDITION

Row	Positions (cells)
03	F4-F5

Plane 1

Row	Positions (cells)
03	00-1E 20-23 30-4A
04	00-25 28-4D
D0	00-F5
D1	00-26 2A-DD
D4	00-54 56-9C 9E-9F A2 A5-A6 A9-AC AE-B9 BB BD-C0 C2-C3 C5-FF
D5	00-05 07-0A 0D-14 16-1C 1E-39 3B-3E 40-44 46 4A-50 52-FF
D6	00-A3 A8-FF
D7	00-C9 CE-FF

Plane 2

Row	Positions (cells)
00-A6	0000-A6D6
F8-FA	F800-FA1D

Plane 0E

Row	Positions (cells)
00	01 20-7F

Plane 0F

Row	Positions (cells)
00-FF	0000-FFFD

Plane 10

Row	Positions (cells)
00-FF	0000-FFFD

A.5.2 304 UNICODE 3.2

304   The fixed collection UNICODE 3.2 consists of fixed collections from A.3 and A.5.1 above and several ranges of code positions arranged by planes as follows.

Plane 0-10

Collection number and name
303   UNICODE 3.1

Plane 0

Collection number and name
98   SUPPLEMENTAL ARROWS-A
99   SUPPLEMENTAL ARROWS-B
100   MISCELLANEOUS MATHEMATICAL SYMBOLS-B
101   SUPPLEMENTAL MATHEMATICAL OPERATORS
102   KATAKANA PHONETIC EXTENSIONS
103   VARIATION SELECTORS

Rows	Positions (cells)
02	20
03	4F 63-6F D8-D9 F6
04	8A-8B C5-C6 C9-CA CD-CE
05	00-0F
06	6E-6F
07	B1
10	F7-F8
17	00-0C 0E-14 20-36 40-53 60-6C 6E-70 72-73
20	47 4E-52 57 5F-63 71 B0-B1 E4-EA
21	3D-4B F4-FF
22	F2-FF
23	7C 9B-CE
24	EB-FE
25	96-9F F8-FF
26	16-17 72-7D 80-89
27	68-75 D0-EB
30	3B-3D 95-96 9F-A0 FF
32	51-5F B1-BF
A4	A2-A3 B4 C1 C5
FA	30-6A
FE	45-46 73
FF	5F-60

Page 888, Annex B.1 List of all combining characters

Remove the following entry:

0B83 TAMIL SIGN VISARGA

After the entry

  1059 MYANMAR VOWEL SIGN VOCALIC LL,

insert new entries, as shown below:

1712 TAGALOG VOWEL SIGN I
1713 TAGALOG VOWEL SIGN U
1714 TAGALOG VIRAMA
1732 HANUNOO VOWEL SIGN I
1733 HANUNOO VOWEL SIGN U
1734 HANUNOO PAMUDPOD
1752 BUHID VOWEL SIGN I
1753 BUHID VOWEL SIGN U
1772 TAGBANWA VOWEL SIGN I
1773 TAGBANWA VOWEL SIGN U

After the entry

17D3 KHMER SIGN BATHMASAT

insert new entries, as shown below:

180B MONGOLIAN FREE VARIATION SELECTOR ONE
180C MONGOLIAN FREE VARIATION SELECTOR TWO

180D MONGOLIAN FREE VARIATION SELECTOR THREE

Insert the following new entries at the end of the list (after FB1E HEBREW POINT JUDEO-SPANISH VARIKA):

FE00 VARIATION SELECTOR-1
FE01 VARIATION SELECTOR-2
FE02 VARIATION SELECTOR-3
FE03 VARIATION SELECTOR-4
FE04 VARIATION SELECTOR-5
FE05 VARIATION SELECTOR-6
FE06 VARIATION SELECTOR-7
FE07 VARIATION SELECTOR-8
FE08 VARIATION SELECTOR-9
FE09 VARIATION SELECTOR-10
FE0A VARIATION SELECTOR-11
FE0B VARIATION SELECTOR-12
FE0C VARIATION SELECTOR-13
FE0D VARIATION SELECTOR-14
FE0E VARIATION SELECTOR-15
FE0F VARIATION SELECTOR-16

Page 895, Annex D.4

Remove the second sentence of the NOTE 1 in the clause D.4 starting by "The values 0000 FFFE…(see clause 8).".

Page 899, Annex F

In clause F.1.1 (Zero-width boundary indicators) replace the first paragraph by the following text:

The following characters are used to indicate whether or not the adjacent characters are separated by a word or hyphenation boundary. Each of these zero-width boundary indicators has no width in its usual own presentation.

SOFT HYPHEN (00AD): SOFT HYPHEN (SHY) is a graphic character, the visual representation of which is identical to that of HYPHEN, for use when an allowable automatic hyphenation line-break after it is to be indicated. Unless the SOFT HYPHEN occurs at the very end of a rendered line, the SOFT HYPHEN normally has zero width and no visible representation, and may also suppress the rendering of the following character.

> NOTE – For example, for Swedish, "biljett<SHY>tång" should be rendered as "biljettång" when there is no line-break after the SHY.

Page 901, clause F.2.5

Replace the clause F.2.5 as follows:

F.2.5 Mongolian vowel separator

MONGOLIAN VOWEL SEPARATOR (180E): This character may be used between the MONGOLIAN LETTER A or the MONGOLIAN LETTER E at the end of a word and the preceding consonant letter. It indicates a special form of the graphic symbol for the letter A or E and the preceding consonant. When rendered in visible form it is generally shown as a narrow space between the letters, but it may sometimes be shown as a distinct graphic symbol to assist the user.

Page 904, Annex G

Insert each of the new character name entries at the appropriate position, ordered alphabetically by the character name, in the list of character names in Annex G. These new names are provided in a machine-readable format that is accessible as a link to this document.

> NOTE 1 – The referenced file is only available to users who obtain their copy of the amendment in a machine-readable form. However, the file format makes it printable.

Click on this highlighted text to access the file containing the new names.

> NOTE 2 – The content is also available as a separate viewable file in the same file directory as this document. The file is named: "Am1names.txt".

Page 956, Annex M

Add to Annex M the following references.

After ISO/IEC 10367:

ISO 10754:1984 *Information and documentation – Extension of the Cyrillic alphabet coded character set for non-Slavic languages for bibliographic information interchange*

After JIS X 0212-1990:

JIS X 0213:2000, Japanese Standards Association. *7-bit and 8-bit double byte coded extended KANJI sets for information interchange, 2000-01-20.*

After Roop 1972:

Santos, Hector. 1994. *The Tagalog script.* (Ancient Philippine Scripts Series; 1). Los Angeles: Sushi Dog Graphics.

Santos, Hector. 1995. *The living scripts.* (Ancient Philippine Scripts Series; 2). Los Angeles: Sushi Dog Graphics.

Page 962. Annex P

Insert the following entries in annex P.

03D8 GREEK LETTER ARCHAIC KOPPA
    The name of this character distinguishes it from 03DE GREEK LETTER KOPPA, which is most commonly used with its numeric value, such as

in the dating of legal documentation. GREEK LETTER ARCHAIC KOPPA is primarily used alphabetically to represent the letter used in early Greek inscriptions.

03D9 GREEK SMALL LETTER ARCHAIC KOPPA
The name of this character distinguishes it from 03DF GREEK SMALL LETTER KOPPA, which is most commonly used with its numeric value, such as in the dating of legal documentation. GREEK SMALL LETTER ARCHAIC KOPPA is primarily used alphabetically to represent the letter used in early Greek inscriptions.

FF5F FULLWIDTH LEFT WHITE PARENTHESIS
This character has a common glyph variation that looks like a double left parenthesis

FF60 FULLWIDTH RIGHT WHITE PARENTHESIS
This character has a common glyph variation that looks like a double right parenthesis

Page 985, Annex S

Replace all occurrences of 'Source code' by 'Source'.

Replace the end of the first note

- CJK COMPATIBILITY IDEOGRAPHS (F900 – FAFF with the exception of FA1F and FA23).

By

- CJK COMPATIBILITY IDEOGRAPHS (F900 – FAFF with the exception of FA0E, FA0F, FA11, FA13, FA14, FA1F, FA21, FA23, FA24, FA27, FA28, and FA29).

Page 987, clause S.1.6

Add the following paragraph and note at the end of clause S.1.6:

The source separation rule described in this clause only applies to the CJK UNIFIED IDEOGRAPHS block specified in the Basic Multilingual Plane.

NOTE – CJK Compatibility Ideographs are created following a rule very similar to the source separation rule. However, the end result is the combination of a single CJK Unified Ideograph and one or several CJK Compatibility Ideographs. When the source separation rule is applied, all 'similar' source CJK Ideographs result in separate CJK Unified Ideographs.

Page 20-303 Code Tables and list of character names

1. Corrigenda and modifications to existing tables

Insert the additional character glyphs and names at the indicated positions in the tables given below, the character names replacing the existing entries which

read *"(This position shall not be used)". The table numbers are affected by the insertion of new tables (see below) preceding these modified tables. (The table numbers corresponding to the 2nd edition of ISO/IEC 10646-1 are mentioned in parenthesis.)*

Table 6 - Row 01-2: Latin Extended B (6)
Table 9 - Row 03: Combining Diacritical Marks (9)
Table 10 - Row 03: Greek and Coptic (10)
Table 12 - Row 04: Cyrillic (12)
Table 16 - Row 06: Arabic (15)
Table 19 - Row 07: Thaana (18)
Table 25 - Row 0B: Tamil (24)
Table 35 - Row 10: Georgian (34)
Table 38 - Row 12: Ethiopic (37)
Table 52 - Row 17: Khmer (47)
Table 58 - Row 20: General Punctuation (53)
Table 59 - Row 20: Superscripts and Subscripts (54)
Table 60 - Row 20: Currency Symbols (55)
Table 61 - Row 20: Combining Diacritical Marks for Symbols (56)
Table 62 - Row 21: Letterlike Symbols (57)
Table 64 - Row 21: Arrows (59)
Table 65 - Row 22: Mathematical Operators (60)
Table 66 - Row 22: Mathematical Operators (61)
Table 67 - Row 23: Miscellaneous Technical (62)
Table 68 - Row 23: Miscellaneous Technical (63)
Table 71 - Row 24: Enclosed Alphanumerics (66)
Table 73 - Row 25: Block Elements (68)
Table 74 - Row 25: Geometric Shapes (69)
Table 75 - Row 26: Miscellaneous Symbols (70)
Table 78 - Row 27: Dingbats (72)
Table 91 - Row 30: CJK Symbols and Punctuation (79)
Table 92 - Row 30: Hiragana (80)
Table 93 - Row 30: Katakana (81)
Table 99 - Row 32: Enclosed CJK Letters and Months (86)
Table 100 - Row 32: Enclosed CJK Letters and Months (87)
Table 110 - Row A4: Yi Radicals (97)
Table 144 - Row F9-FA: CJK Compatibility Ideographs (131)
Table 150 - Row FD: Arabic Presentation Forms-A (136)
Table 153 - Row FE: CJK Compatibility Forms (138)
Table 155 - Row FE: Arabic Presentation Forms-B (140)
Table 156 - Row FF: Halfwidth and Fullwidth Forms (141)
Table 158 - Row FF: Specials (143)

These tables contain updated graphic symbols at the following code positions:

066B-066C, 0B83, 125C, 17DB, 2114, 2216, 224C, 2278-2279, 2329-232A, 2380-238C, 25AA-25AB, 262B, 3018-301B, FEFF.

and the new characters and names at the following code positions:

0220, 034F, 0363-036F, 03D8-03D9, 03F4-03F6, 048A-048B, 04C5-04C6, 04C9-04CA, 04CD-04CE, 066E-066F, 07B1, 10F7-10F8, 2047, 204E-2052, 2057, 205F-2063, 2071, 20B0-20B1, 20E4-20EA, 213D-214B, 21F4-21FF, 22F2-22FF, 237C, 239B-23CE, 24EB-24FE, 2596-259F, 25F8-25FF, 2616-2617, 2672-267D, 2768-2775, 303B-303D, 3095-3096, 309F, 30A0, 30FF, 3251-325F, 32B1-

32BF, A4A2-A4A3, A4B4, A4C1, A4C5, FA30-FA5F, FDFC, FE45-FE46, FE73, FF5F-FF60.

The annotations for the characters at FDD0-FDEF, FFFE, and FFFF are changed to "(This position is permanently reserved)".

2. New tables.

Insert the following additional tables and adjust the numbering of the existing tables that follow. When correctly applied, all tables will be arranged by ascending code position.

Table 13 - Row 05: Cyrillic Supplement
Table 48 - Row 17: Tagalog
Table 49 - Row 17: Hanunoo
Table 50 - Row 17: Buhid
Table 51 - Row 17: Tagbanwa
Table 76 - Row 26: Miscellaneous Symbols
Table 79 - Row 27: Miscellaneous Mathematical
*   Symbols-A*
Table 80 - Row 27: Supplemental Arrows-A
Table 83 - Row 29: Supplemental Arrows-B
Table 84 - Row 29: Miscellaneous Mathematical
*   Symbols-B*
Table 85 - Row 2A: Supplemental Mathematical
*   Operators*
Table 86 - Row 2A: Supplemental Mathematical
*   Operators*
Table 98 - Row 31: Katakana Phonetic Extensions
Table 145 - Row FA: CJK Compatibility Ideographs
Table 151 - Row FE: Variation Selectors

These tables add new characters and names at the following code positions:

0500-050F, 1700-170C, 170E-1714, 1720-1736, 1740-1753, 1760-176C, 176E-1770, 1772-1773, 2680-2689, 27D0-27EB, 27F0-27FF, 2900-297F, 2980-29FF, 2A00-2AFF, 31F0-31FF, FA60-FA6A, FE00-FE0F