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

AMENDMENT 2: N'Ko, Phags-pa, Phoenician and other characters

Page 1, Clause 1 Scope

In the note, update the Unicode Standard version from 4.1 to 5.0.

Page 2, Clause 3 Normative references

Update the reference to the Unicode Bidirectional Algorithm and the Unicode Normalization Forms as follows:

Unicode Standard Annex, UAX#9, The Unicode Bidirectional Algorithm, Version 5.0.0, March 2006.

Unicode Standard Annex, UAX#15, Unicode Normalization Forms, Version 5.0.0, March 2006.

Page 6, Figure 1 - Entire coding space of the Universal Multiple-Octet Coded Character set

Remove the note.

Page 7, Figure 2 – Group 00 of the Universal Multiple-Octet Coded Character set

Remove the second note and rename the first note "NOTE" instead of "NOTE 1".

Page 8, Sub-clause 6.4 Naming of characters

In the list item c), remove 'unified' in front of 'ideographs'.

Replace last paragraph starting with "Guidelines to be used" with the following text:

Additional rules to be used for constructing the names of characters are given in clause 28.1.

Page 9, Clause 7 General requirement for the UCS

Make the second note normative by removing the "NOTE 2" term and by setting it as requirement b. The first note is renamed "NOTE" and the following requirements become c. and d.

Page 10, Sub-clause 9.2 Other Planes reserved for future standardization

Replace text and note by the following:

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

Page 13, Clause 18 Block names

Replace title with "Block and Collection names". Add a sub-clause "18.1 Block names". Add the following paragraph after last.

Rules to be used for constructing the names of blocks are given in clause 28.1.

Add the following sub-clause:

18.2 Collection names

Collections are shown in Annex A.

Rules to be used for constructing the names of collections are given in clause 28.1.

Page 14, Clause 19 Characters in bidirectional context

In the Note, replace "Unicode Standard Version 4.1" with "Unicode Standard".

Page 14, Sub-clause 20.3 Format characters

In the list of format characters, remove the following entry:

 1D159   MUSICAL SYMBOL NULL NOTEHEAD

Page 15, Sub-clause 20.4 Variation selectors

In the paragraph before Note 2, replace "table" with "list", and remove the side borders around the description of the variant appearances for mathematical symbols.

Page 16, Sub-clause 20.4 Variation selectors

In the paragraph before Note 3, replace "table" with "list", remove the last sentence "These combinations … table", and remove the side borders around the

description of the variant appearances for Mongolian.

Page 17, Sub-clause 20.4 Variation selectors

Insert the following text after the list describing the Mongolian variation sequences:

The following list provides a description of the variant appearances corresponding to the use of appropriate variation selectors with all allowed base Phags-pa characters. These variation selector sequences do not select fixed visual representation; rather, they select a representation that is reversed from the normal form predicted by the preceding character.

Sequence (UID notation)	Description of variant appearance
<A856, FE00>	PHAGS-PA LETTER reversed shaping SMALL A
<A85C, FE00>	PHAGS-PA LETTER reversed shaping HA
<A85E, FE00>	PHAGS-PA LETTER reversed shaping I
<A85F, FE00>	PHAGS-PA LETTER reversed shaping U
<A860, FE00>	PHAGS-PA LETTER reversed shaping E
<A868, FE00>	PHAGS-PA SUBJOINED LETTER reversed shaping YA

Page 23, Clause 28 Character names and annotations, sub-clause 28.1 General

Replace paragraph with following new sub-clauses 28.1.1 to 28.1.4

28.1.1 Entity names

This standard specifies names for the following entity types:
- characters
- named UCS sequence identifiers (clause 29)
- blocks (clause 18, Annex A.2)
- collections (clause A.1)

The names given by this standard to these entities shall follow the rules for name formation and name uniqueness specified in this clause. This specification applies to the entity names in the English language version of this standard.

> NOTE 1 – In a version of such a standard in another language:
>
> a) these rules may be amended to permit names to be generated using words and syntax that are considered appropriate within that language;
>
> b) the entity names from this version of the standard may be replaced by equivalent unique names constructed according to the rules amended as in a) above.

> NOTE 2 – Additional guidelines for constructing entity names are given in annex L for information.

28.1.2 Name formation

An entity names shall consist only of the following characters:
- LATIN CAPITAL LETTER A through LATIN CAPITAL LETTER Z,
- DIGIT ZERO through DIGIT NINE,
- SPACE,
- HYPHEN-MINUS, and
- FULL STOP if the entity being named is a collection.

The first character in an entity name shall be a Latin capital letter. The last character in an entity name shall be either a Latin capital letter or a Digit.

An entity name shall not contain two or more consecutive SPACE characters or consecutive HYPHEN-MINUS characters. A collection name shall not contain two or more consecutive FULL STOP characters.

A sequence of a SPACE followed by a HYPHEN-MINUS or a sequence of a HYPHEN-MINUS followed by a SPACE may appear only in character names or named UCS sequence identifiers.

EXAMPLE   Each of the following two character names contains a consecutive SPACE and HYPHEN-MINUS:
TIBETAN LETTER -A
TIBETAN MARK BKA- SHOG YIG MGO

FULL STOP may appear only in between two alphanumeric characters (LATIN CAPITAL LETTER A through LATIN CAPITAL LETTER Z, DIGIT ZERO through DIGIT NINE) in a collection name.

EXAMPLE 1 The following collection name contains FULL STOP in between two Digits, DIGIT FOUR and DIGIT ONE:
UNICODE 4.1

EXAMPLE 2 The following collection name contains FULL STOP in between one Latin letter, LATIN CAPITAL LETTER D, and a Digit, DIGIT SEVEN:
BMP-AMD.7

28.1.3 Single name

Each entity named in this standard shall be given only one name.

> NOTE – This does not preclude the informative use of name aliases or acronyms for the sake of clarity. However, the normative entity name will be unique.

28.1.4 Name uniqueness

Each entity name must also be unique within an appropriate name space, as specified here.

Block names

Block names constitute a name space. Each block name shall be unique and distinct from all other block names specified in the standard.

Collection names

Collection names constitute a name space. Each collection name shall be unique and distinct from all other collection names specified in the standard.

Character names and named UCS sequence identifiers

Character names and named UCS sequence identifiers, taken together, constitute a name space. Each character name or named UCS sequence identifier shall be unique and distinct from all other character names or named UCS sequence identifiers.

Determining uniqueness

For block names and collection names, two names shall be considered unique and distinct if they are different even when SPACE and medial HYPHEN-MINUS characters are ignored in comparison of the names.

> NOTE 1 – A medial HYPHEN-MINUS is a HYPHEN-MINUS character that occurs immediately after a character other than SPACE and immediately before a character other than SPACE.

EXAMPLE 1 The following hypothetical block names would be unique and distinct:
> LATIN-A
> LATIN-B

EXAMPLE 2 The following hypothetical block names would not be unique and distinct:
> LATIN-A
> LATIN A
> LATINA

For character names and named UCS sequence identifiers, two names shall be considered unique and distinct if they are different even when SPACE and medial HYPHEN-MINUS characters are ignored and even when the words "LETTER", "CHARACTER", and "DIGIT" are ignored in comparison of the names.

EXAMPLE 1 The following hypothetical character names would not be unique and distinct:
> MANICHAEAN CHARACTER A
> MANICHAEAN LETTER A

EXAMPLE 2 The following two actual character names are unique and distinct, because they differ by a HYPHEN-MINUS that is not a medial HYPHEN-MINUS:
> TIBETAN LETTER A
> TIBETAN LETTER -A

The following two character names shall be considered unique and distinct:

HANGUL JUNGSEONG OE
HANGUL JUNGSEONG O-E

> NOTE 2 – These two character names are explicitly handled as an exception, because they were defined in an earlier version of this International Standard before the introduction of the name uniqueness requirement. This pair is, has been, and will be the only exception to the uniqueness rule in this International Standard.

28.1.5 Annotations

A character name or a named UCS sequence identifier may be followed by an additional explanatory statement not part of the name, and separated by a single SPACE character. These statements are in parentheses and use the Latin lower case letters a-z, digits 0-9, SPACE and HYPHEN-MINUS. A capital Latin letter A-Z may be used for word initials where required.

Such parenthetical annotations are not part of the entity names themselves, and the characters used in the annotations are not subject to the name uniqueness requirements.

A character name may also be followed by a single ASTERISK separated from the name by a single SPACE. If a parenthetical annotation is present, the ASTERISK follows the annotation and is separated from the closing parenthesis by a single SPACE.

The presence of the ASTERISK notes that additional information on the character is available in annex P of this standard.

Page 23, sub-clause 28.3 Character names and annotations for Hangul syllables

Replace the ending note as follows:

> NOTE – The annex R provides the names of Hangul syllables in two formats: syllable-name and full name/annotation, both available through linked files.

Page 25, New Clause 29

Insert this new clause after Clause 28 and renumber following clauses as appropriate.

29 Named UCS Sequence Identifiers

A named UCS Sequence Identifier (USI) is a USI associated to a name following the same construction rules as for character names. These rules are given in Clause 28.

> NOTE – The purpose of these named USIs is to specify sequences of characters that may be treated as single units, either in particular types of processing, in reference by standards, in listing of repertoires (such as for fonts or keyboards).

The following list provides a description of these named UCS sequence identifiers.

USI	USI name
<0100, 0300>	LATIN CAPITAL LETTER A WITH MACRON AND GRAVE
<0101, 0300>	LATIN SMALL LETTER A WITH MACRON AND GRAVE
<00E1, 0328>	LATIN SMALL LETTER A WITH ACUTE AND OGONEK
<0045, 0329>	LATIN CAPITAL LETTER E WITH VERTICAL LINE BELOW
<0065, 0329>	LATIN SMALL LETTER E WITH VERTICAL LINE BELOW
<00C8, 0329>	LATIN CAPITAL LETTER E WITH VERTICAL LINE BELOW AND GRAVE
<00E8, 0329>	LATIN SMALL LETTER E WITH VERTICAL LINE BELOW AND GRAVE
<00C9, 0329>	LATIN CAPITAL LETTER E WITH VERTICAL LINE BELOW AND ACUTE
<00E9, 0329>	LATIN SMALL LETTER E WITH VERTICAL LINE BELOW AND ACUTE
<00CA, 0304>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND MACRON
<00EA, 0304>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND MACRON
<00CA, 030C>	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND CARON
<00EA, 030C>	LATIN SMALL LETTER E WITH CIRCUMFLEX AND CARON
<012A, 0300>	LATIN CAPITAL LETTER I WITH MACRON AND GRAVE
<012B, 0300>	LATIN SMALL LETTER I WITH MACRON AND GRAVE
<0069, 0307, 0301>	LATIN SMALL LETTER I WITH DOT ABOVE AND ACUTE
<006E, 0360, 0067>	LATIN SMALL LETTER NG WITH TILDE ABOVE
<004F, 0329>	LATIN CAPITAL LETTER O WITH VERTICAL LINE BELOW
<006F, 0329>	LATIN SMALL LETTER O WITH VERTICAL LINE BELOW
<00D2, 0329>	LATIN CAPITAL LETTER O WITH VERTICAL LINE BELOW AND GRAVE
<00F2, 0329>	LATIN SMALL LETTER O WITH VERTICAL LINE BELOW AND GRAVE
<00D3, 0329>	LATIN CAPITAL LETTER O WITH VERTICAL LINE BELOW AND ACUTE
<00F3, 0329>	LATIN SMALL LETTER O WITH VERTICAL LINE BELOW AND ACUTE
<0053, 0329>	LATIN CAPITAL LETTER S WITH VERTICAL LINE BELOW
<0073, 0329>	LATIN SMALL LETTER S WITH VERTICAL LINE BELOW
<016A, 0300>	LATIN CAPITAL LETTER U WITH MACRON AND GRAVE
<016B, 0300>	LATIN SMALL LETTER U WITH MACRON AND GRAVE
<10E3, 0302>	GEORGIAN LETTER U-BRJGU
<17D2, 1780>	KHMER CONSONANT SIGN COENG KA
<17D2, 1781>	KHMER CONSONANT SIGN COENG KHA
<17D2, 1782>	KHMER CONSONANT SIGN COENG KO
<17D2, 1783>	KHMER CONSONANT SIGN COENG KHO
<17D2, 1784>	KHMER CONSONANT SIGN COENG NGO
<17D2, 1785>	KHMER CONSONANT SIGN COENG CA
<17D2, 1786>	KHMER CONSONANT SIGN COENG CHA
<17D2, 1787>	KHMER CONSONANT SIGN COENG CO
<17D2, 1788>	KHMER CONSONANT SIGN COENG CHO
<17D2, 1789>	KHMER CONSONANT SIGN COENG NYO
<17D2, 178A>	KHMER CONSONANT SIGN COENG DA
<17D2, 178B>	KHMER CONSONANT SIGN COENG TTHA
<17D2, 178C>	KHMER CONSONANT SIGN COENG DO
<17D2, 178D>	KHMER CONSONANT SIGN COENG TTHO
<17D2, 178E>	KHMER CONSONANT SIGN COENG NA
<17D2, 178F>	KHMER CONSONANT SIGN COENG TA
<17D2, 1790>	KHMER CONSONANT SIGN COENG THA
<17D2, 1791>	KHMER CONSONANT SIGN COENG TO
<17D2, 1792>	KHMER CONSONANT SIGN COENG THO
<17D2, 1793>	KHMER CONSONANT SIGN COENG NO
<17D2, 1794>	KHMER CONSONANT SIGN COENG BA
<17D2, 1795>	KHMER CONSONANT SIGN COENG PHA
<17D2, 1796>	KHMER CONSONANT SIGN COENG PO
<17D2, 1797>	KHMER CONSONANT SIGN COENG PHO
<17D2, 1798>	KHMER CONSONANT SIGN COENG MO
<17D2, 1799>	KHMER CONSONANT SIGN COENG YO
<17D2, 179A>	KHMER CONSONANT SIGN COENG RO
<17D2, 179B>	KHMER CONSONANT SIGN COENG LO
<17D2, 179C>	KHMER CONSONANT SIGN COENG VO
<17D2, 179D>	KHMER CONSONANT SIGN COENG SHA
<17D2, 179E>	KHMER CONSONANT SIGN COENG SSA
<17D2, 179F>	KHMER CONSONANT SIGN COENG SA
<17D2, 17A0>	KHMER CONSONANT SIGN COENG HA
<17D2, 17A1>	KHMER CONSONANT SIGN COENG LA
<17D2, 17A2>	KHMER VOWEL SIGN COENG QA
<17D2, 17A7>	KHMER INDEPENDENT VOWEL SIGN COENG QU
<17D2, 17AB>	KHMER INDEPENDENT VOWEL SIGN COENG RY
<17D2, 17AC>	KHMER INDEPENDENT VOWEL SIGN COENG RYY
<17D2, 17AF>	KHMER INDEPENDENT VOWEL SIGN COENG QE
<17BB 17C6>	KHMER VOWEL SIGN OM
<17B6, 17C6>	KHMER VOWEL SIGN AAM
<31F7, 309A>	KATAKANA LETTER AINU P
<02E5, 02E9>	MODIFIER LETTER EXTRA-HIGH EXTRA-LOW CONTOUR TONE BAR

All the allowed named UCS sequence identifiers are shown in this clause; all other such named sequences are undefined.

Page 30-1348 Clause 33, Code Tables and list of character names

1. 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 first edition of ISO/IEC 10646:2003 are mentioned in parenthesis.)*

Some of these tables also contain updated character glyphs.

Plane 00

Table 6 - Rows 01-02: Latin Extended-B (6)
Table 10 - Row 03: Greek and Coptic (10)
Table 12 - Row 04: Cyrillic (12)
Table 13 - Row 05: Cyrillic Supplement (13)
Table 15 - Row 05: Hebrew (15)
Table 17 - Row 06: Arabic (17)
Table 29 - Row 0C: Kannada (27)
Table 30 - Row 0D: Malayalam (28)
Table 40 - Row 12: Ethiopic (38)
Table 57 - Row 17: Khmer (52)
Table 65 - Row 1D: Combining Diacritical Marks Supplement
Table 73 - Row 20: Combining Diacritical Marks for Symbols (65)
Table 74 - Row 21: Letterlike Symbols (66)
Table 75 - Row 21: Number Forms (67)
Table 80 - Row 23: Miscellaneous Technical (72)
Table 88 - Row 26: Miscellaneous Symbols (80)
Table 91 - Row 27: Miscellaneous Mathematical Symbols-A (83)
Table 99 - Row 2B: Miscellaneous Symbols and Arrows (91)
Table 133 - Row A7: Modifier Tone Letters

Plane 01

Table 206 - Row 0A: Kharoshti
Table 208 - Row D0: Byzantine Musical Symbols (176)
Table 219 - Row D7: Mathematical Alphanumeric Symbols (185)

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

0242-024F, 037B-037D, 04CF, 04FA-04FF, 0510-0513, 05BA, 0CE2-0CE3, 0CF1-0CF2, 1DC4-1DCA, 1DFE-1DFF, 20EC-20EF, 214D-214E, 2184, 23DC-23E7, 26B2, 27C7-27CA, 2B14-2B1A, 2B20-2B23, A717-A71A, 1D7CA-1D7CB

and updated graphic symbols at the following code positions (the graphic symbol for U+1234 was inadvertently modified by ISO/IEC 10646:2003/Amd.1):

06DF-06E1, 0D66, 1234, 17D2, 33AC, 10A3F, 1D09C, 1D09F

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.

Plane 00

Table 21 - Row 07: NKo
Table 62 - Row 1B: Balinese
Table 101 - Row 2C: Latin Extended-C
Table 134 - Row A7: Latin Extended-D
Table 136 - Row A8: Phags-pa

Plane 01

Table 198 - Row 09: Phoenician
Table 200-205 - Rows 20-23: Cuneiform
Table 206 - Row 24: Cuneiform Numbers and Punctuation
Table 213 - Row D3: Counting Rod Numerals

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

07C0-07FA, 1B00-1B4B, 1B50-1B7C, 2C60-2C6C, 2C74-2C77, A720-A721, A840-A877, 10900-10919, 1091F, 12000-1236E, 12400-12462, 12470-12473 1D360-1D371

Page 960, Table 179 - Row D3: Tai Xuan Jing Symbols

In the name list, replace the entries for 1D300-1D305 with the following:

1D300   MONOGRAM FOR EARTH (ren) *
1D301   DIGRAM FOR HEAVENLY EARTH (tian ren) *
1D302   DIGRAM FOR HUMAN EARTH (di ren) *
1D303   DIGRAM FOR EARTHLY HEAVEN (ren tian) *
1D304   DIGRAM FOR EARTHLY HUMAN (ren di) *
1D305   DIGRAM FOR EARTH (ren ren) *

Page 1349, Annex A.1

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

　　4 LATIN EXTENDED-B

In the list of collection numbers and names, after

　　127 VERTICAL FORMS

insert new entries as follows:

128 NKO	07C0-07FF
129 BALINESE	1B00-1B7F
130 LATIN EXTENDED-C	2C60-2C7F
131 LATIN EXTENDED-D	A720-A7FF
132 PHAGS-PA	A840-A87F

after

　　1018 ANCIENT GREEK MUSICAL NOTATION

insert new entries as follows:

1019 PHOENICIAN	10900-1091F
1020 CUNEIFORM	12000-123FF

1021 CUNEIFORM NUMBERS AND
    PUNCTUATION   12400-1247F
1022 COUNTING ROD NUMERALS   1D360-1D37F

Page 1351, annex A.1

In the list of collections numbers and names, after

306 UNICODE 4.1   See A6.4 *

insert the new entry:

307 UNICODE 5.0   See A6.5 *

Under Note 2, after "126", add ",128".

In the alphabetical list of keywords in Note 4, insert the following entries:

Balinese	129
Counting Rod Numerals	1022
Cuneiform	1020 1021
NKo	128
Phags-pa	132
Phoenician	1019

In the alphabetical list of keywords in Note 4, replace:

Latin   1 2 3 4 30
with:
Latin   1 2 3 4 30 130 131

Page 1352, Annex A.2.1

In the list of blocks in the BMP, insert the following new entries:

NKO	07C0-07FF
BALINESE	1B00-1B7F
LATIN EXTENDED-C	2C60-2C7F
LATIN EXTENDED-D	A720-A7FF
PHAGS-PA	A840-A87F

Page 1353, Annex A.2.2

In the list of blocks in the SMP, insert the following new entries:

PHOENICIAN	10900-1091F
CUNEIFORM	12000-123FF
CUNEIFORM NUMBERS AND PUNCTUATION	12400-1247F
COUNTING ROD NUMERALS	1D360-1D37F

Page 1356, annex A.5.1

In the definition of collection 340 COMBINED FIRST EDITION, delete the following line:

D4   C1

Page 1357, annex A.6   Unicode Collections

At the end of Annex A.6 add new clause A.6.5 as follows.

A.6.5   307 UNICODE 5.0

307   The fixed collection UNICODE 5.0 consists of a fixed collection from A.6 and several ranges of code positions. The collection list is arranged by planes as follows.

Plane 00-10

Collection number and name
306   UNICODE 4.1

Plane 00

Row	Positions (cells)
02	42-4F
03	7B-7D
04	CF FA-FF
05	10-13 BA
07	C0-FA
0C	E2-E3 F1-F2
1B	00-4B 50-7C
1D	C4-CA FE-FF
20	EC-EF
21	4D-4E 84
23	DC-E7
26	B2
27	C7-CA
2B	14-1A 20-23
2C	60-6C 74-77
A7	17-1A 20-21
A8	40-77

Plane 01

Row	Positions (cells)
09	00-19 1F
20-22	00-FF
23	00-6E
24	00-62 70-73
D3	60-71
D7	CA-CB

Page 1358, Annex B.1 List of all combining characters

Insert the following new entries:

05BA HEBREW POINT HOLAM HASER FOR VAV
07EB NKO COMBINING SHORT HIGH TONE
07EC NKO COMBINING SHORT LOW TONE
07ED NKO COMBINING SHORT RISING TONE
07EE NKO COMBINING LONG DESCENDING TONE
07EF NKO COMBINING LONG HIGH TONE
07F0 NKO COMBINING LONG LOW TONE
07F1 NKO COMBINING LONG RISING TONE
07F2 NKO COMBINING NASALIZATION MARK
07F3 NKO COMBINING DOUBLE DOT ABOVE
0CE2 KANNADA VOWEL SIGN VOCALIC L
0CE3 KANNADA VOWEL SIGN VOCALIC LL
1B00 BALINESE SIGN ULU RICEM
1B01 BALINESE SIGN ULU CANDRA
1B02 BALINESE SIGN CECEK
1B03 BALINESE SIGN SURANG
1B04 BALINESE SIGN BISAH
1B34 BALINESE SIGN REREKAN
1B35 BALINESE VOWEL SIGN TEDUNG
1B36 BALINESE VOWEL SIGN ULU

1B37   BALINESE VOWEL SIGN ULU SARI
1B38   BALINESE VOWEL SIGN SUKU
1B39   BALINESE VOWEL SIGN SUKU ILUT
1B3A   BALINESE VOWEL SIGN RA REPA
1B3B   BALINESE VOWEL SIGN RA REPA TEDUNG
1B3C   BALINESE VOWEL SIGN LA LENGA
1B3D   BALINESE VOWEL SIGN LA LENGA TEDUNG
1B3E   BALINESE VOWEL SIGN TALING
1B3F   BALINESE VOWEL SIGN TALING REPA
1B40   BALINESE VOWEL SIGN TALING TEDUNG
1B41   BALINESE VOWEL SIGN TALING REPA
          TEDUNG
1B42   BALINESE VOWEL SIGN PEPET
1B43   BALINESE VOWEL SIGN PEPET TEDUNG
1B44   BALINESE ADEG ADEG
1B6B   BALINESE MUSICAL SYMBOL COMBINING
          TEGEH
1B6C   BALINESE MUSICAL SYMBOL COMBINING
          ENDEP
1B6D   BALINESE MUSICAL SYMBOL COMBINING
          KEMPUL
1B6E   BALINESE MUSICAL SYMBOL COMBINING
          KEMPLI
1B6F   BALINESE MUSICAL SYMBOL COMBINING
          JEGOGAN
1B70   BALINESE MUSICAL SYMBOL COMBINING
          KEMPUL WITH JEGOGAN
1B71   BALINESE MUSICAL SYMBOL COMBINING
          KEMPLI WITH JEGOGAN
1B72   BALINESE MUSICAL SYMBOL COMBINING
          BENDE
1B73   BALINESE MUSICAL SYMBOL COMBINING
          GONG

Page 1369, Annex D.4 Mapping from UCS-4 form to UTF-8 form

In second paragraph replace 'character' with 'code position'.

In second note, replace "coded character" with "code position".

Page 1369, Annex D.5 Mapping from UTF-8 form to UCS-4 form

In the note, second paragraph, replace "coded character" with "octet".

Page 1371, Annex E Mirrored characters in bidi-rectional context

In the list of characters, insert the following new entries:

27C8   REVERSE SOLIDUS PRECEDING SUBSET
27C9   SUPERSET PRECEDING SOLIDUS

Page 1379, 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 ma-

chine-readable format that is accessible as a link to this document.

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

> NOTE – The content is also available as a separate viewable file in the same file directory as this document. The file is named: "Am2names.txt".

Page 1379, Annex L

Replace introductory text and the rules 1 to 12 with the following:

The clause 28 of this standard specifies rules for name formation and name uniqueness. These rules are used in other information technology coded character set standards such as ISO/IEC 646, ISO/IEC 6937, ISO/IEC 8859, and ISO/IEC 10367. This annex provides additional guidelines for the creation of these entity names.

> NOTE – These guidelines do not apply to the names of CJK Ideographs and Hangul syllables which are formed using rules specified in clause 28.2 and 28.3 respectively.

Guideline 1

The name of an entity wherever possible denotes its customary meaning (for example, the character name: PLUS SIGN or the block name: BENGALI).

Some entities, such as characters, may have a name describing shapes, not usage, (for example, the character name: UPWARDS ARROW).

The name on an entity is not intended to identify its properties or attributes, or to provide information on its linguistic characteristics, except as defined in guideline 4 below.

Guideline 2

An acronym consists of Latin capital letters A to Z and digits and is associated with a name.

Acronyms may be used in entity names where usage already exists and clarity requires it. For example, the names of control functions are coupled with an acronym.

Examples:

Name: LOCKING-SHIFT TWO RIGHT
Acronym: LS2R
Name: SOFT HYPHEN
Acronym: SHY
Name: INTERNATIONAL PHONETIC ALPHABET
Acronym: IPA

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

Guideline 3

Character names and named UCS Sequence Identifiers only include digits 0 to 9 if spelling out the name of the corresponding digits(s) would be inappropriate.

> NOTE – As an example the name of the character at position 201A is SINGLE LOW-9 QUOTATION MARK; the symbol for the digit 9 is included in this name to illustrate the shape of the character, and has no numerical significance.

Guideline 4

Character names and named UCS Sequence Identifiers are 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 guidelines 9 to 11. The words WITH and AND may be included for additional clarity when needed.

1	Script	5	Attribute
2	Case	6	Designation
3	Type	7	Mark(s)
4	Language	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 1 – A ligature is a graphic symbol in which two or more other graphic symbols are imaged as a single graphic symbol.

> NOTE 2 – Where a character comprises a base letter with multiple marks, the sequence of those in the name is the order in which the marks are positioned relative to the base letter, starting with the marks above the letters taken in upwards sequence, and followed by the marks below the letters taken in downwards sequence.

Guideline 5

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

Examples:

К	LATIN CAPITAL LETTER K
Ю	CYRILLIC CAPITAL LETTER YU

Guideline 6

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

Guideline 7

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

Examples:

A	LATIN CAPITAL LETTER A
A	GREEK CAPITAL LETTER ALPHA
A	CYRILLIC CAPITAL LETTER A

Guideline 8

Where possible, named UCS Sequence Identifiers are constructed by appending the names of the constituent elements together while eliding duplicate elements. Should this process result in a name that already exists, the name is modified suitably to guarantee uniqueness among character names and named UCS Sequence Identifiers. The words WITH and AND may be included for additional clarity when needed.

Guideline 9

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

Guideline 10

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

Guideline 11

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 is given to the character. The customary name may be followed in the list of characters of a particular standard by an explanatory statement containing the name which this character has in the script specified by this particular standard.

Example:

⌣   UNDERTIE (Enotikon)

Page 1394, Annex P, Additional information on characters

Insert the following new entry for 1D300-1D305 after the FFE3 entry:

1D300 MONOGRAM FOR EARTH,
1D301 DIGRAM FOR HEAVENLY EARTH,
1D302 DIGRAM FOR HUMAN EARTH,
1D303 DIGRAM FOR EARTHLY HEAVEN,
1D304 DIGRAM FOR EARTHLY HUMAN,
1D305 DIGRAM FOR EARTH

> A Tai Xuan Jing symbol comprises a combination of three elements: tian, di and ren, and these three Chinese words usually translate to heaven, earth and human, respectively. The character names of the six Tai Xuan Jing symbols in this International Standard, however, are based on an uncommon mapping; tian for heaven, di for human, and ren for earth. Users are advised to identify these symbols by their representative glyphs or Chinese annotations but not character names.

Page 1396, Annex R

Replace text and tables with the following text and links to files:

This annex provides the names of Hangul syllables in two formats, both available through linked files:

1. Tabular arrangement showing the syllable-name of each character in the block HANGUL SYLLABLES (AC00 - D7A3). The syllable-name is the final component of the full character name, and is derived as described in clause 28.3, steps 1 to 5, which is the definitive specification of the names in that block.
    The leftmost column of the table shows the cell numbers (00 - FF) of the corresponding characters. The headings of the other columns of the table show the row numbers of the characters.

    > NOTE 1 – The content linked to is a PDF file, using a format similar to this standard containing the tabular arrangement.

Click on this highlighted text to access the file containing the Hangul syllable names in tabular arrangement.

> The content is also available as a separate viewable file in the same directory as this document. The file is named: "HangulTb.pdf".

2. The full name and annotation of the Hangul syllables are also provided in a machine-readable format that is accessible as a link to this document.

    > NOTE 2 – The content linked to is a plain text file, using ISO/IEC 646-IRV characters with LINE FEED as end of line mark that specifies, after a 5-lines header, as all the Hangul syllables, each line specified as follows:

    - 01-04 octet: UCS-2 code position in hexadecimal notation,
    - 05 octet: SPACE character,
    - 06 octet until end of line: Hangul syllable with the annotation between parentheses.

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

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

Because the file "HangulSy.txt" is unchanged by this amendment, the link above is not active.

Page 1418, Annex U

*Create a new informative Annex U (the former one was removed by 10646:2003/Amd1:2005) with the following title "**Annex U (Informative)**" and the following text:*

Characters in Identifiers

A common task facing an implementer of UCS is the provision of a parsing and/or lexing engine for identifiers. Each programming language standard has its own identifier syntax; different programming languages have different conventions for the use of certain characters from the ASCII (ISO 646-IRV) range ($, @, #, _) in identifiers. Questions as to which characters to use for syntactic purposes versus which to be allowed in identifiers, whether case-pairing should be included, normalization should be performed, and other factors enter into the picture when defining the set of permitted characters for a given identification purpose.

Unicode Consortium publishes a document "UAX 31 – Identifier and Pattern Syntax" to assist in the standard treatment of identifiers in UCS character-based parsers. Those specifications are recommended for determining the list of UCS characters suitable for use in identifiers. The document is available at http://www.unicode.org/reports/tr31/.

(This page left intentionally blank)