ISO/IEC JTC 1/SC 2

CODED CHARACTER SETS

SECRETARIAT: JAPAN (JISC)

DOC TYPE: Text for DCOR ballot or comment

TITLE: Draft Technical Corrigendum No. 3 to ISO/IEC 10646-1: 1993, Information technology -- Universal Multiple-Octet Coded Character Set (UCS) -- Part 1: Architecture and Basic Multilingual Plane

SOURCE: Mr. Bruce Paterson, Project Editor

PROJECT: JTC 1.02.18.01.00.03

STATUS: In accordance with Resolution M09.10 adopted at the 9th SC 2 plenary meeting, Fukuoka, Japan, 1999-03-16, this document is circulated to the SC 2 members for a three-month latter ballot. P-members of SC 2 are requested to submit their votes to the SC 2 Secretariat as soon as possible but not later than **1999-06-30.**

ACTION ID: LB

DUE DATE: 1999-06-30

DISTRIBUTION: P, O and L Members of ISO/IEC JTC 1/SC 2
WG Conveners and Secretariats
Secretariat, ISO/IEC JTC 1
ISO/IEC ITTF

NO. OF PAGES: 4

ACCESS LEVEL: Defined

WEB ISSUE #:   048

DRAFT

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

Part 1:

Architecture and Basic Multilingual Plane

TECHNICAL CORRIGENDUM 3

Replace Clause 25 "Code tables and lists of character names" with the following text.

25 Code tables and lists of character names

25.1 General

An overview of the Basic Multilingual Plane is shown in figure 3. Detailed code tables and lists of character names for the Basic Multilingual Plane are shown on the following pages and in applicable Amendments.

Guidelines to be used for constructing names of characters are given in annex K for information. In some cases, a name of a character is followed by additional explanatory statements not part of the name. These statements are in parentheses and not in capital letters except for the initials of the word, where required.

25.2 Character names and annotations for Hangul syllables

Names for the Hangul syllable characters in code positions (hex) 0000 AC00 - 0000 D7A3 are derived from their code position numbers by the numerical procedure described below. Lists of names for these characters are not provided.

1. Obtain the code position number of the Hangul syllable character. It is of the form $0000\ h_1h_2h_3h_4$ where h_1, h_2, h_3, and h_4 are hexadecimal digits; h_1h_2 is the Row number within the BMP and h_3h_4 is the cell number within the row. The number $h_1h_2h_3h_4$ lies within the range AC00 to D7A3.

2. Derive the decimal numbers d_1, d_2, d_3, d_4 that are numerically equal to the hexadecimal digits h_1, h_2, h_3, h_4 respectively.

3. Calculate the character index C from the formula:

$$C = 4096 \times (d_1 - 10) + 256 \times (d_2 - 12)$$
$$+ 16 \times d_3 + d_4$$

Note: If C < 0 or > 11,171 then the character is not in the HANGUL SYLLABLES block.

4. Calculate the syllable component indices I, P, F from the following formulae:

$I = C\ /\ 588$	(Note: $0 \le I \le 18$)
$P = (C\ \%\ 588)\ /\ 28$	(Note: $0 \le P \le 20$)
$F = C\ \%\ 28$	(Note: $0 \le F \le 27$)

where "/" indicates integer division (i.e. $x\ /\ y$ is the integer quotient of the division), and "%" indicates the modulo operation (i.e. $x\ \%\ y$ is the remainder after the integer division $x\ /\ y$).

5. Obtain the Latin character strings that correspond to the three indices I, P, F from columns 2, 3, and 4 respectively of Table 1 below (for $I = 11$ and for $F = 0$ the corresponding strings are null). Concatenate these three strings in left-to-right order to make a single string, the syllable-name.

6. The character name for the character at position $0000\ h_1h_2h_3h_4$ is then:

HANGUL SYLLABLE *s-n*

where "*s-n*" indicates the syllable-name string derived in step 5.

Example.
For the character in code position D4DE:
$d_1 = 13$, $d_2 = 4$, $d_3 = 13$, $d_4 = 14$.
$C = 10462$
$I = 17$, $P = 16$, $F = 18$.
The corresponding Latin character strings are:
P , WI, BS.
The syllable-name is PWIBS, and the character name is:
HANGUL SYLLABLE PWIBS

Annotations for the Hangul syllable characters in code positions (hex) 0000 AC00 - 0000 D7A3 are also derived from their code position numbers by a similar numerical procedure described below.

7. Carry out steps 1 to 4 as described above.

8. Obtain the Latin character strings that correspond to the three indices I, P, F from columns 5, 6, and 7 respectively of Table 1 below (for $I = 11$ and for $F = 0$ the corresponding strings are null). Concatenate these three strings in left-to-right order to make a single string, and enclose it within parentheses to form the annotation.

Example.
For the character in code position D4DE:

$d_1 = 13$, $d_2 = 4$, $d_3 = 13$, $d_4 = 14$.
$C = 10462$
$I = 17$, $P = 16$, $F = 18$.

The corresponding Latin character strings are:

ph, wi, ps,

and the annotation is (phwips).

Table 1: Elements of Hangul syllable names and annotations

Index number	Syllable name elements			Annotation elements		
	I string	P string	F string	I string	P string	F string
0	G	A	▨	k	a	▨
1	GG	AE	G	kk	ae	k
2	N	YA	GG	n	ya	kk
3	D	YAE	GS	t	yae	ks
4	DD	EO	N	tt	eo	n
5	R	E	NJ	r	e	nc
6	M	YEO	NH	m	yeo	nh
7	B	YE	D	p	ye	t
8	BB	O	L	pp	o	l
9	S	WA	LG	s	wa	lk
10	SS	WAE	LM	ss	wae	lm
11	▨	OE	LB	▨	oe	lp
12	J	YO	LS	c	yo	ls
13	JJ	U	LT	cc	u	lth
14	C	WEO	LP	ch	weo	lph
15	K	WE	LH	kh	we	lh
16	T	WI	M	th	wi	m
17	P	YU	B	ph	yu	p
18	H	EU	BS	h	eu	ps
19	▨	YI	S	▨	yi	s
20	▨	I	SS	▨	i	ss
21	▨	▨	NG	▨	▨	ng
22	▨	▨	J	▨	▨	c
23	▨	▨	C	▨	▨	ch
24	▨	▨	K	▨	▨	kh
25	▨	▨	T	▨	▨	th
26	▨	▨	P	▨	▨	ph
27	▨	▨	H	▨	▨	h

Clause 25, Tables 124 to 211 (see Amendment 5)

In Amendment 5 delete all odd-numbered pages from page 7 to page 181 inclusive (the tables of character names numbered Table 124 to Table 211).

Page 700, Annex A (see Technical Corrigendum 2)

After the title of clause A.3 ".Fixed collections of the whole BMP" add the following new title:

A.3.1   301 BMP-AMD.7

After the entry for row FF in the list of code position ranges, add the following new title:

A.3.2   299 BMP FIRST EDITION

At the end of clause A.3 ".Fixed collections of the whole BMP" add the following new text [overleaf].

A.3.3 302 BMP SECOND EDITION

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

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

Rows	Positions (cells)
00	20-7E A0-FF
01	00-FF
02	00-33 50-AD B0-EE
03	00-4E 60-62 74-75 7A 7E 84-8A 8C 8E-A1 A3-CE D0-D7 DA-F3
04	00-86 88-89 8C-CE D0-F5 F8-F9
05	31-56 59-5F 61-87 89-8A 91-A1 A3-B9 BB-C4 D0-EA F0-F4
06	0C 1B 1F 21-3A 40-55 60-6D 70-ED..F0-FE
07	00-0D 0F-2C 30-4A 80-BF
09	01-03 05-39 3C-4D 50-54 58-70 81-83 85-8C 8F-90 93-A8 AA-B0 B2 B6-B9 BC BE-C4 C7-C8 CB-CD D7 DC-DD DF-E3 E6-FA
0A	02 05-0A 0F-10 13-28 2A-30 32-33 35-36 38-39 3C 3E-42 47-48 4B-4D 59-5C 5E 66-74 81-83 85-8B 8D 8F-91 93-A8 AA-B0 B2-B3 B5-B9 BC-C5 C7-C9 CB-CD D0 E0 E6-EF
0B	01-03 05-0C 0F-10 13-28 2A-30 32-33 36-39 3C-43 47-48 4B-4D 56-57 5C-5D 5F-61 66-70 82-83 85-8A 8E-90 92-25 99-9A 9C 9E-9F A3-A4 A8-AA AE-B5 B7-B9 BE-C2 C6-C8 CA-CD D7 E7-F2
0C	01-03 05-0C 0E-10 12-28 2A-33 35-39 3E-44 46-48 4A-4D 55-56 60-61 66-6F 82-83 85-8C 8E-90 92-A8 AA-B3 B5-B9 BE-C4 C6-C8 CA-CD D5-D6 DE E0-E1 E6-EF
0D	02-03 05-0C 0E-10 12-28 2A-39 3E-43 46-48 4A-4D 57 60-61 66-6F 82-83 85-96 9A-B1 B3-BB BD C0-C6 CA CF-D4 D6 D8-DF F2-F4
0E	01-3A 3F-5B 81-82 84 87-88 8A 8D 94-97 99-9F A1-A3 A5 A7 AA-AB AD-B9 BB-BD C0-C4 C6 C8-CD D0-D9 DC-DD
0F	00-47 49-6A 71-8B 90-97 99-BC BE-CC CF
10	00-21 23-27 29-2A 2C-32 36-39 40-59 A0-C5 D0-F6 FB
11	00-59 5F-A2 A8-F9
12	20-26 28-46 48 4A-4D 50-56 58 5A-5D 60-86 88 8A-8D 90-AE B0 B2-B5 B8-BE C0 C2-C5 C8-CE D0-D6 D8-EE F0-FF
13	00-0E 10 12-15 18-1E 20-46 48-5A 61-7C A0-F4
14-15	1401-15FF
16	00-76 80-9C A0-F0
17	80-DC E0-E9
18	00-0E 10-19 20-77 80-A9
1E	00-9B A0-F9
1F	00-15 18-1D 20-45 48-4D 50-57 59 5B 5D 5F-7D 80-B4 B6-C4 C6-D3 D6-DB DD-EF F2-F4 F6-FE
20	00-46 48-4D 4F 6A-70 74-8E A0-AF D0-E3
21	00-3A 53-83 90-F3
22	00-F1
23	00-7B 7D-9A
24	00-26 40-4A 60-EA
25	00-95 A0-F7
26	00-13 19-71
27	01-04 06-09 0C-27 29-4B 4D 4F-52 56 58-5E 61-67 76-94 98-AF B1-BE
28	00-FF
2E	80-99 9B-F3
2F	00-D5 F0-FB
30	00-3A 3E-3F 41-94 99-9E A1-FE
31	05-2C 31-8E 90-B7
32	00-1C 20-43 60-7B 7F-B0 C0-CB D0-FE
33	00-76 7B-DD E0-FE
34-4D	3400-4DBF
4E-9F	4E00-9FA5
A0-A3	A000-A3FF
A4	00-8C 90-A1 A4-B3 B5-C0 C2-C4 C6
AC-D7	AC00-D7A3
E0-F8	E000-F8FF
F9-FA	F900-FA2D
FB	00-06 13-17 1D-36 38-3C 3E 40-41 43-44 46-B1 D3-FF
FC	00-FF
FD	00-3F 50-8F 92-C7 F0-FB
FE	20-23 30-44 49-52 54-66 68-6B 70-72 74 76-FC FF
FF	01-5E 61-BE C2-C7 CA-CF D2-D7 DA-DC E0-E6 E8-EE F9-FD

[Editor's note: The details of the above entries will be adjusted as necessary when the exact character repertoire of ISO/IEC 10646-1 Second Edition is finalised.]