INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ORGANISATION INTERNATIONALE DE NORMALISATION

ISO/IEC JTC1/SC2/WG2

Universal Multiple-Octet Coded Character Set (UCS)

ISO/IEC JTC1/SC2/WG2 N 1934

Date: 1998-11-11

TITLE:Editorial Corrigendum on Zones and related features of 10646-1SOURCE:Bruce Paterson, project editorSTATUS:Response to Resolution M35.10ACTION:For review and confirmation by WG2DISTRIBUTION:JTC1/SC2/WG2

This paper provides a draft Editorial Corrigendum for ISO/IEC 10646-1 to remove the four zones that were defined in the First Edition, namely the A-zone, I-zone, O-zone, and R-zone, in reponse to Resolution M35.10 taken at the London meeting in September 1998. The S-zone, which is reserved for UTF-16, and the private use zone, which is never used for other purposes, are retained.

This paper also proposes the renaming of a few contiguous blocks of characters as single blocks, where the characters all belong to the same script and the block names merely duplicate the existing collection names. Such a change was recently made for the HEBREW blocks in Corrigendum no.2.

1. Zones.

The required amendments to clauses 8, 10, and 11, and to Figures 2, 3 and 4, are shown on the following pages. Some material has been swapped between clauses 10 and 11 so that all the requirements relating to private use characters are now brought together in a single new clause 10.

This draft takes WG2 N 1796 as the basis for the current text of 10646-1, but Figures 3 and 4 now include all Amendments and drafts up to PDAM.31.

2. Blocks

The following changes are proposed in Annex A.2 "Blocks in the BMP". Replace

- BASIC GREEK - GREEK SYMBOLS AND COPTIC	0370-03CF 03D0-03FF	by GREEK	0370-03FF
Replace - BASIC ARABIC - ARABIC EXTENDED	0600-065F 0660-06FF	by ARABIC	0600-06FF
Replace - GEORGIAN EXTENDED - BASIC GEORGIAN	10A0-10CF 10D0-10FF	by GEORGIAN	10A0-10FF

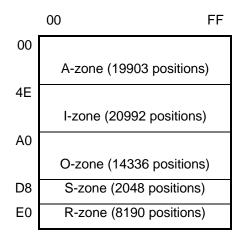
8 The Basic Multilingual Plane

[existing text]

Plane 00 of Group 00 shall be the Basic Multilingual Plane (BMP). The BMP can be used as a two-octet coded character set in which case it shall be called UCS-2 (see 14.1).

The Basic Multilingual Plane shall be divided into five zones:

A-zone:code positions 0000 0000 to 0000 4DFFI-zone:code positions 0000 4E00 to 0000 9FFFO-zone:code positions 0000 A000 to 0000 D7FFS-zone:code positions 0000 D800 to 0000 DFFFR-zone:code positions 0000 E000 to 0000 FFFD



Code positions 0000 0000 to 0000 001F in the BMP are reserved for control characters, and code position 0000 007F is reserved for the character DELETE (see clause 16). Code positions 0000 0080 to 0000 009F are reserved for control characters.

In the Basic Multilingual Plane, the A-zone is used for alphabetic and syllabic scripts together with various symbols. The I-zone is used for Chinese/Japanese/Korean (CJK) unified ideographs (unified East Asian ideographs). The O-zone is used for Korean Hangul syllables, and for various other scripts. The S-zone is reserved for the use of UTF-16 (see Annex Q). The R-zone shall be used for the restricted use zone in the BMP which contains private use characters, presentation forms, and compatibility characters (see clause 10).

8 The Basic Multilingual Plane

[proposed amended text; new text is shown underlined, deleted text is not shown]

Plane 00 of Group 00 shall be the Basic Multilingual Plane (BMP). The BMP can be used as a two-octet coded character set in which case it shall be called UCS-2 (see 14.1).

Code positions 0000 0000 to 0000 001F in the BMP are reserved for control characters, and code position 0000 007F is reserved for the character DELETE (see clause 16). Code positions 0000 0080 to 0000 009F are reserved for control characters.

<u>Code positions 0000 D800 to 0000 DFFF are</u> reserved for the use of UTF-16 (see Annex Q). <u>These positions are known as the S-zone.</u>

<u>Code positions 0000 E000 to 0000 F8FF are</u> reserved for private use (see clause 10). These positions are known as the private use zone.

Code postions FFFE and FFFF are reserved.

[existing text]

10 The restricted use zone

Sets of graphic characters that are used in particular ways are provided in the restricted use zone. These sets include:

- a) Private use characters,
- b) Presentation forms of characters,
- c) Compatibility characters.

10.1 Private use characters

Private use characters are not restrained in any way by ISO/IEC 10646. Private use characters can be used to provide user-defined characters. For example, this is a common requirement for users of ideographic scripts.

NOTE 1 - For meaningful interchange of private use characters, an agreement, independent of ISO/IEC 10646, is necessary between sender and recipient.

Private use characters can be used for dynamicallyredefinable characters applications.

NOTE 2 - For meaningful interchange of dynamicallyredifinable characters, an agreement, independent of ISO/IEC 10646 is necessary between sender and recipient. ISO/IEC 10646 does not specify the techniques for defining or setting up dynamically-redefinable characters.

10.2 Presentation forms of characters

Each presentation form of character provides an alternative form, for use in a particular context, to the nominal form of the character or sequence of characters from the other zones of graphic characters. The transformation from the nominal form to the presentation forms may involve substitution, superimposition, or combination.

The rules for the superimposition, choice of differently shaped characters, or combination into ligatures, or conjuncts which are often of extreme complexity are not specified in ISO/IEC 10646.

In general, presentation forms are not intended to be used as a substitute for the nominal forms of the graphic characters specified elsewhere within this coded character set. However, specific applications may encode these presentation forms instead of the nominal forms for specific reasons among which is compatibility with existing devices. The rules for searching, sorting, and other processing operations on presentation forms are outside the scope of ISO/IEC 10646.

[proposed amended text]

10 Private use groups, planes, and zones

10.1 Private use characters

Private use characters are not restrained in any way by ISO/IEC 10646. Private use characters can be used to provide user-defined characters. For example, this is a common requirement for users of ideographic scripts.

NOTE 1 - For meaningful interchange of private use characters, an agreement, independent of ISO/IEC 10646, is necessary between sender and recipient.

Private use characters can be used for dynamicallyredefinable characters applications.

NOTE 2 - For meaningful interchange of dynamicallyredefinable characters, an agreement, independent of ISO/IEC 10646 is necessary between sender and recipient. ISO/IEC 10646 does not specify the techniques for defining or setting up dynamically-redefinable characters.

10.2 Code positions for private use characters

The code positions of the 32 groups from Group 60 to Group 7F shall be for private use.

The code positions of Plane 0F and Plane 10, and of the 32 planes from Plane E0 to Plane FF, 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).

1<u>1</u> <u>Sets of characters with particular</u> uses

Sets of graphic characters that are used in particular ways are provided in <u>ISO/IEC 10646</u>. These sets include:

a) Presentation forms of characters,

b) Compatibility characters.

11.1 Presentation forms of characters

Each presentation form of a character provides an alternative form, for use in a particular context, to the

..... [unchanged from 10.2]

•••

searching, sorting, and other processing operations on presentation forms are outside the scope of ISO/IEC 10646.

Within the BMP these characters are mostly allocated to positions in rows FB to FF.

[existing text, continued]

10.3 Compatibility characters

Compatibility characters are included in ISO/IEC 10646 primarily for compatibility with existing coded character sets to allow two-way code conversion without loss of information.

11 Private use groups, planes, and zones

The code positions of the 32 groups from Group 60 to Group 7F shall be for private use.

The code positions of Plane 0F and Plane 10, and of the 32 planes from Plane E0 to Plane FF, 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).

[proposed amended text, continued]

11.2 Compatibility characters

Compatibility characters are included in ISO/IEC 10646 primarily for compatibility with existing coded character sets to allow two-way code conversion without loss of information.

Within the BMP many of these characters are allocated to positions within rows F9, FA, FE, and FF, and within rows 31 and 33. Some compatibility characters are also allocated within other rows.

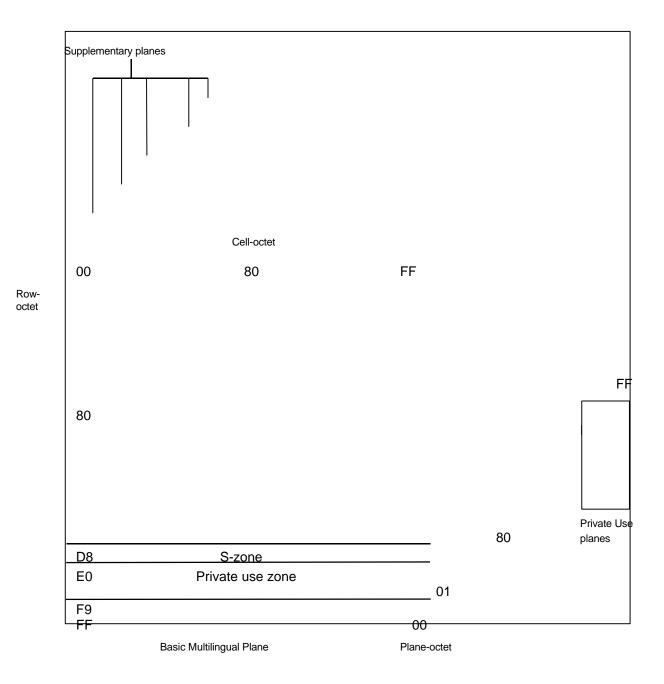
Editor's note: It would be preferable to move this new clause 11 to a later position in the text, for example following:

- 19 Block names

and before:

- 20 Characters in bi-directional context
- -21 Special characters
- 22 Order of characters

The intervening clauses 12 to 19 would then become 11 to 18.



Labels "S-zone" and "Private use zone" are specified in clause 8.

Figure 2 - Group 00 of the Universal Multiple-Octet Coded Character Set

Row-octet

)						
			David 00 to 00			
			Rows 00 to 33			
			(see Figure 4)			
3						
4						
		C	CJK Unified Ideographs Exte	ension A		
5						
			CJK Unified Ideograph	าร		
F						
.0			Yi Syllables			
\3 \4					Yi Radicals	
.5						
в С						
			Hangul Syllables			
07						
8			S-zone (for use in UTF-16 only)			
0F						
0			Private Use Zone			
8						
9			CJK Compatibility Ideographs			
А В	Alphabetic Presenta	ation Forms				
c –			Arabic Presentation Forms-A			
D	_			-		
E	Comb. Half M'ks	CJK Compat. F'm		Arabic	Presentation Forms-I	
F			Halfwidth And Fullwidth Forms			Special
		= not graphic char	racters		= reserved for future	е

NOTE: Vertical boundaries within rows are indicated in approximate positions only.

Figure 3 - Overview of the Basic Multilingual Plane

Row-octet

_									
0	Basic Latin				Latin-1 Supplement				
	Latin Extend	Latin Extended-A			Latin Extended-B				
2	Latin Extended-B		IPA (Int. Ph	on. Alph.)	lph.) Extensions			Spacing Modifier Letters	
3	Combining Diacritical	Marks			Basi	c Greek		Gree	k Symbols and Coptic
4				Cyrillic					
5		Ai	rmenian					Hebrew	
6	Basic Arabic						abic Extended		
7	Syriac					Thaana	a		
8									
9	Devanag							ngali	
A	Gurmuk	hi			Gujarati				
в	Oriya				Tamil				
С	Telugu				Kannada				
D	Malayala	m						hala	
E	Thai						L	ao	
F				Tibetan					
0		Myanmar						Georg	jian
1			Н	angul Jam	0				
2				Ethiopic					
3								Chero	kee
4			Unified Canac	lian Aborig	inal Syll	labics			
6					Ogham			Run	ic
7							Kh	mer	
8				Mongolian					
9									
D									
E				tended Ad					
F			Gre	ek Extend					
0				Super-/S	Subscrip	ots	· · · · ·		Comb. Mks. Symb.
1	Letterlike Symbols		Number Fo					Arrows	
2				natical Ope					
3			Miscella	aneous Te	chnical				
4	Control Pictures	O.C.R.			Enclosed Alphanumerics				
5	Box Draw	ling			ck Elen	nents		Geometri	ic Shapes
6				aneous Sy	mbols				
7		Di	ingbats						
8			Bra	aille Patter	ns				
8									
D									
E							CJK Radical	s Supplem	ent
F			Kangxi Radicals	3					Ideographic Descr.
0	CJK Symbols And Punctuation					Katakana			
1	-	Hangul Com	patibility Jamo		CJK I	Misc.	Bopomofo E	xt.	
2	· •	-	Enclosed CJ	K Letters	And Mo	nths			
3	CJK Compatibility								
-									
	= not graphic	characters					= reserved f	or future sta	andardization
-									

NOTE: Vertical boundaries within rows are indicated in approximate positions only.

