

Universal Multiple-Octet Coded Character Set
UCS

Source:	IRG Rapporteur
Title:	Response to WG2 on related issues in SC2 N3814
Status :	IRG contribution
Distribution:	SC2/WG2 members and Liaison organizations
Medium :	Electronic

Summary

In response to WG2 regarding the resolution 48.40 (CJK comments on FDAM1), IRG reviewed all CJK related NB comments and made recommendations as follows.

1. On page 6 of FDAM1 text, HKSCS source reference of character U+ 3978 should be H-8CF3.

- Proposed changes: Change H-8CF4 to H-8CF3.

The picture shown below is a visual representation of U+3978 on FDAM1

057/120	𢄎	𢄎
3978	H-8CF4	4-4350 4-3548

FDAM1 U+3978

where as in HKSCS-2004, H-8CF4 refers to a different character as shown in the following picture:

344A	𢄎
344A F609 (8CF4)	

HKSCS-2004 8CF4

The correct source reference should be H-8CF3 as shown below:

3978	𢄎
3978 F608 (8CF3)	

HKSCS-2004 8CF3

IRG recommendation:

The source standard HKSCS-2004 indicates that it is actually H-8CF3 which should be printed as a source reference on U+3978. CJKU_SR.txt also supports this source reference. As in the ballot comment, IRG also considers it as a typo on the FDAM1 text and recommends correcting it in the future amendment.

2. On CJKU_SR.txt accompanied with FDAM1.txt, HKSCS source reference H-9EE5 should be associated with U+27F2E instead of U+47B6.

● Proposed changes:

a) Delete H-9EE5 from the following line of the CJKU_SR.TXT:

047B6;G_KX;T3-4F24;;;H-9EE5;KP1-7A0C;



b) Add H-9EE5 as shown in the following line to CJKU_SR.TXT:

27F2E;G_KX;T4-5422;;; H-9EE5;;

The following picture shows the glyph shape of H-9EE5 in HKSCS-2004

HKSCS-2004 9EE5 ( 走易) 

and the following pictures are glyphs of U+047B6 and U+27F2E respectively printed on ISO/IEC 10646:2003.

U+047B6 ( 走易) | 071/182
47B6 
3-4F24
3-4704

U+27F2E ( 走易) 

IRG recommendation:

Based on the justifications provided by HKSAR shown below, IRG recommends WG2 to accept proposed changes in the earliest opportunity to minimize the impact for implementations.

Justifications by HKSAR

1. The correction to the source reference information of (HKSCS) character H-9EE5 had been announced publicly on the HKSAR Government website in May 2005 at the

following URL:

<http://www.info.gov.hk/digital21/chi/hkscs/download/123for2004.pdf>,

Relevant extract is attached below for easy reference.

In HKSCS-2001, the Big-5, ISO/IEC 10646-1:2000 and ISO/IEC 10646-2:2001 code points for the character “𪛗” are 9EE5, 47B6 and 47B6 respectively. Since the internal code of this character in the ISO/IEC 10646:2003 is 27F2E, the ISO 10646 code points of this character are revised as ED64 and 27F2E in ISO/IEC 10646-1:2000 and ISO/IEC 10646:2003 respectively in the HKSCS-2004.

2. This is a correction assessed and approved by the Chinese Language Interface Advisory Committee¹ (CLIAC) in Hong Kong SAR, after serious and cautious assessment on the potential impact.
3. Each of the following 4 source references appears twice in CJKU_SR.txt:
 - 1) [T4-6E3B]
04695;G3-4862;T4-6E3B;;K3-322D;;;KP1-752A;
278AE;G_HZ;T4-6E3B,,,,;

- Proposed change: Remove T4-6E3B from the source reference for U+04695.

The following picture provided by TCA explains the intended glyph shape of the source character T4-6E3B and relevant UCS code positions and representative glyphs in question.

Code	Glyph	UCS	Glyph	Code
1. T4-6E3B	T4-6E3B 𪛗	04695	𪛗	T5-7A38
	T4-6E3B 𪛗	278AE	𪛗	Correct.

IRG recommendation:

The top left component of T4-6E3B is 𪛗 as in U+278AE, not 𪛗 as in U+04695. Therefore, IRG recommends removing T4-6E3B from the source reference for U+04695.

¹ Members of the CLIAC come from various sectors including the academia, language and linguistics associations, the publishing industry and the information technology industry. One of the main functions of the CLIAC is the development and publication of the HKSCS.

2) [TF-3862]

0FA23;;TF-3862;JA-2728;;;;;U0-FA23

27EAF;;TF-3862;;;;;

- Proposed change: Remove TF-3862 from the source reference for U+0FA23.

The following picture provided by TCA explains the intended glyph shape of the source character T4-6E3B and relevant UCS code positions and representative glyphs in question.

2. TF-3862 (JA-2728)		0FA23	A unified CJK ideograph, not a compatibility ideograph, despite its name.	
		27EAF		Correct.

IRG recommendation:

IRG recognizes U+27EAF should have been unified with U+0FA23 which is one of CJK Unified Ideographs. However, IRG recommends removing TF-3862 from the source reference for U+0FA23 because of the following reasons:

- U+27EAF has been encoded with a single source TF-3862.
- U+27EAF is already being used in existing implementations for mapping to TF-3862 in the source standard CNS 11643.

3) [V0-417A]

04443;G_KX;T3-5866;;K3-3039;V0-417A;;KP1-6C04;

06726;G0-6B7C;T1-764D;J0-5B2F;K0-5954;V0-417A;;KP0-DBC2;

- Proposed change: Remove V0-417A from the source reference for U+04443.

The following pictures from SuperCJK explain two UCS code positions in question and information on radical.

U+04443	04443 44171 44171 0997.080 32115.080 肉 14	G_KX T3-5866 K3-3039 KP1-6C04		(Radical meat)
---------	---	--	--	----------------

U+06726	06726 21518 21518 0507.030 32115.080 月 14	G0-6B7C T1-764D J0-5B2F K0-5954 V0-417A KP0-DBC2		(Radical moon)
---------	---	---	--	----------------

And the following picture provided by Vietnam tells about the source character V0-417A,

glyph and references (UCS code position, source id, KX index, source dictionary index) associated with it.

V0-417A	朦	0-6726	V0-417A	0507.030	K493.04
---------	---	--------	---------	----------	---------

IRG recommendation:

According to the source dictionary and Kangxi index of V0-417A by Vietnam, V0-417A should belong to a character of radical moon. IRG recommends removing the reference V0-417A from U+04443.

4) [V2-8D4D]

06DFB;G0-4C6D;T1-5B50;J0-453A;K0-7455;V2-8D4D;;KP0-EDC7;

06F58;G0-454B;T1-6D59;J0-5F2F;K0-5A6B;V2-8D4D;;KP0-DCF3;

- Proposed change: Remove V2-8D4D from the source reference for U+06F58.

The following pictures from SuperCJK explain two UCS code positions in question and information on radical.

U+06DFB	06DFB 26253 26253 0632.190 31639.040 水 8	G0-4C6D T1-5B50 J0-453A K0-7455 V2-8D4D KP0-EDC7	添 添
---------	--	---	--------

U+06F58	06F58 27113 27113 0649.050 31745.100 水 12	G0-454B T1-6D59 J0-5F2F K0-5A6B V2-8D4D KP0-DCF3	潘 潘
---------	---	---	--------

And the following picture provided by Vietnam tells about the source character V2-8D4D, glyph and references (UCS code position, source id, KX index, source dictionary index) associated with it.

V2-8D4D	添	0-6DFB	V2-8D4D	0632.190	K746.08
---------	---	--------	---------	----------	---------

IRG recommendation:

As shown in the visual representations of U+06DFB, U+06F58 and V2-8D4D as well as its source references by Vietnam, it is obvious that U+06F58 should not refer to V2-8D4D. Therefore, IRG recommends removing V2-8D4D from the source reference of U+06F58.

4. There's a source reference "G_FZ_BK" for U+24319 in CJKU_SR.txt, which is not explained in

the standard. Is "G_FZ_BK" a valid reference?

- Proposed change: Change the source reference G_FZ_BK for U+24319 to G_FZ.

The following picture from SuperCJK explains the UCS code position in question and information on source reference.

24319 28820 28820 0680.101 火 10	G_FZ_BK	焗 焗 焗 焗
--	---------	------------

IRG recommendation:

Original intention of the source reference G_FZ_BK was to indicate that the character came from both G_FZ and G_BK. Since China thinks it appropriate to choose one and it should be G_FZ, IRG would recommend correcting the source reference G_FZ for U+24319. The same change has already been made in SuperCJK database maintained by the IRG.

5. There's a line 055ECG0CTEB "TEB" in IICORE.txt file. is it a typo of "TEF"?

- Proposed change: No action.

IRG recommendation:

No change is required on this. As explained in IRG N1094R

(http://www.cse.cuhk.edu.hk/~irg/irg/IICore/IRGN1094R_IICoreSourceUseIdentifier.doc),

TEB is a valid identifier and it is not a typo. The source identifier "TEB" indicates it is from CNS 11643 plane 14 (TE) and the usage of the character is 'for education' (B).

-- end of document ---