Date: **2012-05-09**

ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)

Doc. Type: Input to ISO/IEC 10646

Title: **ISO/IEC 10646: Requirements for CJK Font Provision**
Source: **ISO/IEC 10646 Project Editor**
Project: JTC1 02.10646
Status: For review by IRG
Date: 2012-05-09
Distribution: IRG
Reference:
Medium:

Summary: This document covers two topics: 1) requirements for CJK fonts covering new repertoires, 2) status of CJK fonts covering existing repertoires. It also gives guidelines to streamline coverage for existing characters.

Background: As a consequence of the review of the multi-column version of CJK Extension B, the mechanisms for displaying all characters for all sources have become much more complex. Not only can any characters can be expressed from any of the 9 sources (G, T, H, M, J, K, KP, V, and U), but also any content from those sources can be expressed using a very large number of fonts. Because of that complexity it has become more important to allow versioning for font resources, allowing side by side install. At the same time it is desirable to decrease the complexity by streamlining the font descriptions in the chart production tool. This can be mainly achieved by creating full updates for the various sources.

Versioning: Versioning is the process by which a unique name is given to a font file and its logical names. Note that the important part is the uniqueness of its logical names. Typically fonts may have multiple logical names: Family, style, PostScript (PS) that may be different. The uniqueness can be provided by appending a variable to a fixed part, the variable being a number, the date, etc… Examples from the K Source and T source lists:

Hg_CJK_EUC_KR_v50
CJK_B_Delta_TCA20111101

where 'v50' is a version number, and 20111101 represents a date. This allows a future version which could be called 'Hg_CJK_EUC_KR_v**51**' to be installed side by side with the previous version allowing characters to be used from both fonts at the same time if needed (and font with a new date for the second example).

Font encoding: In preliminary phases of the development of a CJK Extension it is somewhat difficult to affect a stable code point to a given glyph because the code point allocations from the various CJK sources are interrelated. The production tool has a definition file allowing remapping either individually or by range. Examples:

```
CMEX_ATSong,21 /Q=2A719 /R=4E23-4E23
CMEX_ATSong,21 /Q=2A71E /R=4E29-4E2B
```
The first example remaps the character 4E23 (font internal coding) to 2A719 (chart code point). The second example remaps the characters 4E29-4E2B to 2A71E-2A720.

This mechanism can be used for large repertoire (it was used extensively during the development of CJK Extension C). However it can become cumbersome to maintain and can introduce errors. Furthermore regression may occur when final fonts are produced and the mappings are removed.

Therefore it is recommended that usage of mapping should be minimized and font encoded to the current proposed code points as soon as possible.

Furthermore, should mapping be required, a machine readable document describing this mapping must be provided to the project editor.

Font streamlining: During the production of the 3rd edition of 10646, to avoid regression risks, it was decided to not replace large fonts when updates were required, but instead to use overlapping fonts replacing only the updated or added characters. Example for the K source:

```
$KSOURCE
CJK_BMP_KR-v46,21,1 /X=0000-10FFFF /I=3400-9FFF; URO and A
CJK_BMP_KR-v46,21,1 /X=0000-10FFFF /I=F900-FA0B; Compat BMP
CJK_BMP_KR-v46,21,1 /Q=FA2E /R=90DE-90DE; Compat BMP
CJK_BMP_KR-v46,21,1 /Q=FA2F /R=96B7-96B7; Compat BMP
Hg_CJK_EUC_KR_v50,21,1 /Q=2700E /R=2700E-2700E; 3 fixes for FDIS 10646 3rd ed
Hg_CJK_EUC_KR_v50,21,1 /Q=203DD /R=203DD-203DD
Hg_CJK_EUC_KR_v50,21,1 /Q=25566 /R=25566-25566
CJK_ExtB_KR-v45,21,1 /X=0000-10FFFF /I=20000-2A6DF; B
CJK_ExtC_KR-v45,21,1 /X=0000-10FFFF /I=2A700-2B73F ;C
```
Two CJK Compatibility characters were added in FA2E and FA2F, three CJK Extension B characters were updated.

Based on a resolution adopted in IRG meeting #37, it was agreed to use the updated font (Hg_CJK_EUC_KR_v50) for all K sources for future editions of the standard. This results in a much simpler definition as follows:

```
$KSOURCE
Hg_CJK_EUC_KR_v50,21,1 /X=0000-10FFFF /I=3400-9FFF; URO and A
Hg_CJK_EUC_KR_v50,21,1 /X=0000-10FFFF /I=F900-FA2F; Compat BMP
Hg_CJK_EUC_KR_v50,21,1 /X=0000-10FFFF /I=20000-2A6DF; B
Hg_CJK_EUC_KR_v50,21,1 /X=0000-10FFFF /I=2A700-2B73F ;C
```

This definition is now in use for draft charts being processed after the issuance of 10646:2012 (3rd edition). Similar streamlining was done for the H source (Hong Kong SAR) where a consolidated font was available and the regression risk was minimal. The last update concerns the BMP CJK Compatibility Ideographs for the J (Japan) source where existing glyphs from the official J source are now used.

Today, many sources have complicated descriptions. As much as possible they should be simplified in a way similar to what was possible for the K source. The following sections describe all CJK sources with suggestion for concerned parties on how to create consolidated versions of their fonts if they desire to do so.

G source (China PRC)
Current definition:

```
$GSOURCE
ZDSUpdate11,21,1 /Q=3D34 /R=3D34-3D34
ZDSUpdate11,21,1 /Q=3EAC /R=3EAC-3EAC
ZDSUpdate11,21,1 /Q=4227 /R=4227-4227
ZDSUpdate11,21,1 /Q=4565 /R=4565-4565
ZDSUpdate11,21,1 /Q=4907 /R=4907-4907
ZDSUpdate11,21,1 /Q=49C8 /R=49C8-49C8
ZDSUpdate11,21,1 /Q=4C41 /R=4C41-4C41
ZDSUpdate11,21,1 /Q=5829 /R=5829-5829
ZDSUpdate11,21,1 /Q=5AB2 /R=5AB2-5AB2
;ZDSUpdate11,21,1 /Q=5FF9 /R=5FF9-5FF9 ; unwanted, see 225D6
ZDSUpdate11,21,1 /Q=6A37 /R=6A37-6A37
ZDS,21,1 /X=0000-10FFFF /I=3400-9FFF; URO and A
FZSongYi_EXB-Z13_IRG35, 21 /X=0000-10FFFF /I=20000-2A6DF; B Overlay for FDIS 10646 3rd (41 chars)
China_ExtB_Modified,21 /Q=20768 /R=20768-20768 ; IRG37 Review
China_ExtB_Modified,21 /Q=22903 /R=22903-22903 ; IRG37 Review
China_ExtB_Modified,21 /Q=2292A /R=2292A-2292A ; IRG37 Review
China_ExtB_Modified,21 /Q=240D5 /R=240D5-240D5 ; IRG37 Review
China_ExtB_Modified,21 /Q=26E82 /R=26E82-26E82 ; IRG37 Review
FZSongYi_EXB-Z13,21 /X=0000-10FFFF /I=20000-2A6DF; B
China_ExtC, 21 /X=0000-10FFFF /I=2A700-2B73F ;C
ChinaUNC,21 /X=0000-10FFFF /I=2B740-2B81F ;D
```

Explanation :
The ZDS font (file name CJK+A2009-12-21.ttf created 12/21/2009) covers the URO and Extension A. It is overlapped by the ZDSUpdate11 font (file name 11_G_Hanzi_2010-08-4.ttf created 8/4/2010) for 11 characters. Of these 11 characters, only 10 were applied. After further study, the change for 5FF9 was not done to avoid a duplicate with 225D6. This had some consequences for the T source for 5FF9.
For reference: 5FF9 忹, 225D6 忹 (issue concerns the middle stroke of the second component).
Ideally the ZDS font should be updated with the remaining 10 characters from the ZDSUpdate11 font.

The FZSongYi_EXB-Z13 font (File name FZSY_SIP.ttf created 12/12/2005) covers Extension B. It is overlapped by the FZSongYi_EXB-Z13_IRG35 font (file name CJK_B_Modified_G_Fonts20110704.ttf) for 41 characters. It is further overlapped by the China_ExtB_Modified font (file name ChinaExtBModifidOnIRG37-20111108.ttf created 2011/11/08) for 5 characters.
Ideally the FZSongYi_EXB font should be updated with the 46 characters from the two update fonts.

T source (Taiwan)
Current definition:

```
$TSOURCE
Uni2F800Cjkcompatideosup,21 /Q=5FF9 /R=2F89F-2F89F ; temp fix for T source at 5FF9
MingLiu For UTC,21 /Q=488C /R=488C-488C ; missing from TW-Sung
TW-Sung,21 /X=0000-10FFFF /I=3400-9FFF; URO and A
TW-Sung,21 /Q=FA28 /R=FA28-FA28; one cjk unified among compat
CJK_B_Delta_TCA20111101,21 /X=0000-10FFFF /I=20000-2A6DF ;B Overlay FDIS 10646 3rd ed (23 chars)
CJK_B_Delta_TCA20110701,21 /X=0000-10FFFF /I=20000-2A6DF ;B Overlay FDIS 10646 3rd ed (207 chars)
TW-Sung-Ext-B,21 /Q=21F2C /R=21F12-21F12 ;Move 21F12 to 21F2C
CJK_B_Delta_TCA20111109,21 /Q=24503 /R=24503-24503
CJK_B_Delta_TCA20111109,21 /Q=2462C /R=2462C-2462C
TW-Sung-Ext-B,21 /Q=24C53 /R=24C36-24C36 ;Move 24C36 to 24C53
CJK_B_Delta_TCA20111109,21 /Q=26F75 /R=26F75-26F75
CJK_B_Delta_TCA20111109,21 /Q=28453 /R=28453-28453
TW-Sung-Ext-B,21 /X=0000-10FFFF /I=20000-2A6DF ;B
TW-Sung-Ext-C,21 /X=0000-10FFFF /I=2A700-2B73F ;C
ATBZ33-1215,21 /X=0000-10FFFF /I=2B740-2B81F ;D
;TW-Sung-Ext-B,21 /X=0000-10FFFF /I=2F800-2F850 (compat) ; unused for now, see coverage below
Uni2F800Cjkcompatideosup,22 /X=0000-10FFFF /I=2F800-2FA1F
```

Explanation :
The TW-Sung font (file name TW-Sung-98.ttf created 10/1/2010) covers the URO and Extension A. It is overlapped by the Uni2F800Cjkcompatideosup font for one character: 5FF9 and by MingLiu For UTC font for one missing character: 488C.
Ideally the TW-Sung font should be updated with these two characters.

The TW-Sung-Ext-B font (file name TW-Sung-ExtB.ttf created 6/17/2009) covers Extension B. It was extensively amended:
- The font CJK_B_Delta_TCA20111101 (font name CJK_B_Delta_TCA20111101.ttf created 2011/11/01) updates 23 characters.
- The font CJK_B_Delta_TCA20110701 (font name CJK_B_Delta_TCA20110701.ttf created 2011/07/06) updates 207 characters.
- The font CJK_B_Delta_TCA20111109 (font name CJK_B_Delta_TCA20111109.ttf created 2011/11/09) updates 4 characters.
- The character 21F12 from TW-Sung-Ext-B was moved to 21F2C.
- The character 24C36 from TW-Sung-Ext-B was moved to 24C53.

Ideally the TW-Sung-Ext-B font should be updated with these 236 characters.

The TW-Sung-Ext-B font (file name TW-Sung-ExtB.ttf created 6/17/2009) partially covers the SIP CJK Compatibility but is not used because another font Uni2F800Cjkcompatideosup contains the whole set covered by the T source.
Ideally the TW-Sung-Ext-B font should be updated to contain all T source from that block. It would also be a better choice for representing the T source glyph than the generic Uni2F800Cjkcompatideosup font.

J source (Japan)

Current definition:

```
$JSOURCE
Japan BMP Sup-1006,21,1 /Q=4148 /R=4148-4148
Japan BMP Sup-1006,21,1 /Q=4165 /R=4165-4165
JapanBMPS,21,1  /Q=8362 /R=8362-8362 ; pre-empt RgHeiseiM for 8362
JapanBMPS,21,1  /Q=83C2 /R=83C2-83C2 ; pre-empt RgHeiseiM for 83C2
RgHeiseiM-W3, 21,1 /X=0000-10FFFF /I=3400-9FFF;; URO, A
JapanBMPS,21,1 /X=0000-10FFFF /I=3400-9FFF; ; URO, A
Japan ARIB-K,21 /Q=9FC4 /R=9FC4-9FC5 ; ARIB set in URO
Japan ARIB-K,21 /Q=9FC6 /R=FA6D-FA6D ; ARIB set in URO
RgHeiseiM-W3,21 /X=0000-10FFFF /I=F900-FA6A ; compat except arib
MS Mincho For UTC,22 /Q=F900 /R=F900-FA6A ; compat except arib
Japan ARIB-K,22 /Q=FA6B /R=FA6B-FA6C ; arib
Japan ARIB-K,22 /Q=FA6D /R=FA6E-FA6E ; arib
RgHeiseiM-W3, 21,1 /Q=20B9F /R=E3C4-E3C4 ; the one missing from Japan B
Japan B,21 /X=0000-10FFFF /I=20000-2A6DF ; B
Japan C 0903,21 /Q=2A708 /R=4708-5724 ;C
Japan UNC 0903,21 /X=0000-10FFFF /I=2B740-2B81F ;D
```

Explanation :

Two fonts: JapanBMPS (file name JapanBMPS.ttf created 1/29/2009) and RgHeiseiM-W3 (file name TChemw3.ttc created 5/28/2006) cover the URO and Extension A. In case of dual coverage RgHeiseiM-W3 is given preference except for two cases: 8362 and 83C2. Furthermore, they are overlapped by the font 'Japan BMP Sup-1006' (file name JapanBMP1006.ttf created 8/2/2010) for two characters (4148 and 4165). Finally the font 'Japan ARIB-K' (file name jarib_k.ttf created 8/28/2008) adds three characters: 9FC4-9FC6.
This is a fairly complicated construction and should ideally be replaced by a single font.

The RgHeiseiM-W3 and 'MS Mincho For UTC' fonts (with priority for the first font) covers the BMP CJK Compatibility characters. This is augmented by the font 'Japan ARIB-K' (file name jarib_k.ttf created 8/28/2008) for three characters (FA6B-FA6D). As mentioned earlier, this is different from the 3rd edition of ISO/IEC10646 where RgHeiseiM-W3 was not used at all for this block.
Ideally a single font, preferably based on RGHeiseiM-W3 should be updated to contain all these characters.

The 'Japan B' font (file name japanB.ttf created 4/24/2009) covers the Extension B characters except for 20B9F which is provided by RgHeiseiM-W3 (file name TChemw3.ttc created 5/28/2006).
Ideally the 'Japan B' font should be updated to include the missing character.

K and KP Korean sources (ROK and DPRK respectively)

Current definitions:

```
; KSOURCE section
$KSOURCE
Hg_CJK_EUC_KR_v50,21,1 /X=0000-10FFFF /l=3400-9FFF; URO and A
Hg_CJK_EUC_KR_v50,21,1 /X=0000-10FFFF /l=F900-FA2F; Compat BMP
Hg_CJK_EUC_KR_v50,21,1 /X=0000-10FFFF /l=20000-2A6DF; B
Hg_CJK_EUC_KR_v50,21,1 /X=0000-10FFFF /l=2A700-2B73F ;C

; KPSOURCE  section
$KPSOURCE
Batang For UTC, 21 /Q=F936 /R=F936-F936 ; compat
UniFA30Cjkcompatideographs,22 /X=0000-10FFF /l=FA70-FAD9 ; compat
KP-C1-TTF,21 /X=0000-10FFFF /l=2A700-2B73F ; C
Uni2F800Cjkcompatideosup,22 /X=0000-10FFFF /l=2F800-2FA1F ; SIP compat
```

Explanation :
The K sources are already streamlined (see prior text in this document).

The KP source coverage is incomplete. Only Extension C and the compatibility blocks (BMP and SIP) are covered. No font coverage is done currently for URO, Extension A and B in ISO/IEC 10646 charts. Note that Extension D has no KP sources. There is an available font: KP CheongPong that supports KP0 sources, but unfortunately not KP1 sources. It is still worth discussing whether or not 'KP CheongPong' should be used to represent the KP0 sources in the charts. KP0 sources are only used for the URO and Extension A coverage (i.e. there are no KP0 sources in Extension B and Extension C). At the same time, URO and Extension A contain KP1 sources, so a KP0-only coverage for these blocks would still be incomplete.

V source (Vietnam)

Current definition:

```
$VSOURCE
BMP&ExtB-Vietnam,20 /X=0000-10FFFF /I=3400-9FFF;URO, A
BMP&ExtB-Vietnam,20 /Q=FA24 /R=FA24-FA24; one unified among compat
BMP&ExtB-Vietnam,20 /Q=210F3 /R=2105D-2105D ;move 2105D to 210F3
zyksun,21 /Q=2008E /R=2008E-2008E ; for now we use UCS-2003 glyph for 2008E
zyksun,21 /Q=207E4 /R=207E4-207E4 ; for now we use UCS-2003 glyph for 207E4
zyksun,21 /Q=20CA6 /R=20CA6-20CA6 ; for now we use UCS-2003 glyph for 20CA6
Nom Na Tong,20 /Q=20FBA /R=20FBA-20FBA ; use old font for 20FBA
zyksun,21 /Q=21024 /R=21024-21024 ; for now we use UCS-2003 glyph for 21024
Nom Na Tong,20 /Q=21130 /R=21130-21130 ; use old font for 21130
Nom Na Tong,20 /Q=21150 /R=21150-21150 ; use old font for 21150
zyksun,21 /Q=21498 /R=21498-21498 ; for now we use UCS-2003 glyph for 21498
zyksun,21 /Q=2339E /R=2339E-2339E ; for now we use UCS-2003 glyph for 2339E
zyksun,21 /Q=24790 /R=24790-24790 ; for now we use UCS-2003 glyph for 24790
zyksun,21 /Q=25E3F /R=25E3F-25E3F ; for now we use UCS-2003 glyph for 25E3F
zyksun,21 /Q=25F44 /R=25F44-25F44 ; for now we use UCS-2003 glyph for 25F44
zyksun,21 /Q=25F8B /R=25F8B-25F8B ; for now we use UCS-2003 glyph for 25F8B
zyksun,21 /Q=268B6 /R=268B6-268B6 ; for now we use UCS-2003 glyph for 268B6
zyksun,21 /Q=28C7E /R=28C7E-28C7E ; for now we use UCS-2003 glyph for 28C7E
Nom Na Tong,20 /Q=28C96 /R=28C96-28C96 ; use old font for 28C96
BMP&ExtB-Vietnam,20 /X=0000-10FFFF /I=20000-2A6DF; B
BMP&ExtB-Vietnam,20 /Q=2ABAB /R=2ABAB-2ABAB; C char in B block font
VNFontsForC,21 /X=0000-10FFFF /I=2A700-2B73F ; C
```

Explanation :

The BMP&ExtB-Vietnam font (file name CJK_B-Vietnam.ttf created 7/2/2011) covers the URO, Extension A and Extension B.

For Extension B, the updates are as follows:

- Character from BMP&ExtB-Vietnam is moved from 2105D to 21F3
- Characters from zyksun (reference font used for ISO 10646:2003 single column layout for Extension B) for the following values: 2008E, 207E4, 20CA6, 21024, 21498, 2339E, 24790, 25E3F, 25F44, 25F8B, 268B6, and 28C7E.
- Characters from 'Nom Na Tong' (file name NonNaTongLight.ttf created 8/18/2006) for the following values: 20FBA, 21130, 21150, and 28C96.

Ideally, the BMP&ExtB-Vietnam font should be updated to include these changes.

The VNFontsForC font (file name VietNamFonts_For_Ext_C.ttf created 1/8/2009) covers Extension C. One missing character: 2ABAB is provided by BMP&ExtB-Vietnam .

Ideally the VNFontsForC font should be updated to include 2ABAB.

Other sources (Hong Kong SAR, Macau SAR, Unicode Consortium)

Current definitions:

```
; HSOURCE section
$HSOURCE
Ming(for ISO10646),21 /X=0000-10FFFF /I=3400-4DBF; A
Ming(for ISO10646),21 /X=0000-10FFFF /I=4E00-9FFF; URO
MingLiU_HKSCS For UTC,21 /X=0000-10FFFF /I=4E00-9FFF; provide Big5 for URO
Ming(for ISO10646),21 /X=0000-10FFFF /I=F900-FAFF; BMP Compat
Ming(for ISO10646),21 /X=0000-10FFFF /I=20000-2FA1F; B, and C (one character), SIP compat

; MSOURCE section
$MSOURCE
MacaoExtendedC1,21 /X=0000-10FFFF /I=2A700-2B73F ; C

; USOURCE section
$USOURCE
UTCHan Medium,23,1 /X=0000-10FFFF /I=3400-9FFF ; URO, A
UTCHan Medium,23,1 /X=0000-10FFFF /I=F900-FAFF ; BMP compat
zyksun,21 /Q=2105D /R=2105D-2105D ; for now we use UCS-2003 glyph for 2105D
UTCHan Medium,23,1 /Q=2A8A7 /R=2A8AB-2A8AB ; Fix for 2A8A7 mistake in UTR 45
UTCHan Medium,23,1 /X=0000-10FFFF /I=20000-2B81F ; B, C, D
UTCHan Medium,23 /X=0000-10FFFF /I=2F800-2FA1F ; SIP compat
```

Explanation :

Most of these source definitions are already streamlined. Few details:

Hong Kong: Possibly the BIG5 characters could be added to the main HK font Ming(for ISO10646), but the current solution is acceptable.

Unicode Consortium: there are one update in Extension B (2105D) and one update in Extension C (2A8A7), ideally the 'UTCHan Medium' should be updated to include these two changes.