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

Doc. Type: **Project Editor Input to ISO/IEC 10646:2003**

Title: **CJK Ideograph source references corrections**

Source: **Project editor (Michel Suignard) and Eric Muller**
Project: JTC1 02.18
Status: For review by WG2
Date: 2003-10-22
Distribution: WG2
Reference:
Medium:

The following text proposes fixes to the CJK Ideographs source references (both unified and compatibility characters) and seeks further clarification for existing references.

1. Clarification

Currently the CJK Unified Ideograph source reference file (CJKU_SR.txt) contains global references for dictionary input: G_KX, G_HZ, G_CY, G_CH, G_HC, G_BK, G_FZ and G_4K. It would be extremely useful to add such information as page number, column, etc. for easier referencing, especially for the characters for which there is no other source reference information (such as U+20001, U+2000A, etc.).

2. Corrections

a) IRG input

Source: http://www.cse.cuhk.edu.hk/~irg/irg/CJK/IRGN975_SingleCJKMappingTable_Notes_E_July11.pdf proposes the following change to CJKU_SR.txt:

Addition:
```
04443  V_SOURCE "V0-417A"
04695  T_SOURCE "T4-6E3B"
058F3  K_SOURCE "K2-2B3E"
06B76  KP_SOURCE "KP1-61A9"
09376  J_SOURCE "J1-647E"
0FA1F  J_SOURCE "JA-264B"
0FA23  T_SOURCE "TF-3862" J_SOURCE "JA-2728"
```

Change
```
04D56 from "T5-6F54" to "T4-684F"
```

Editor comment:
The additions for 06B76, 0FA1F and 0FA23 are already part of ISO/IEC 10646:2003. The other sources are already described in the charts pages for these characters. Fixing the source reference file will make the standard consistent.

In addition, many CJK Unified Ideographs located in the CJK Compatibility block are still missing source references. This should be fixed as soon as possible. The characters in question are listed below (all characters are listed in case additional references need to be added to the ones with existing source information):

```
0FA0E;;;;;;;;
0FA0F;;;;;;;;
0FA11;;;;;;;;
0FA13;;;;;;;;
0FA14;;;;;;;;
0FA1F;;;;JA-264B;;;;
0FA21;;;;;;;;
0FA23;;TF-3862;JA-2728;;;;
0FA24;;;;;V2-8544;;
0FA27;;;;;;;;
0FA28;;TF-584C;;;;;
0FA29;;;;;;;;
```

b) JPN source

Three entries in the CJK Unified Ideographs source reference file seem to have erroneous J source information:

```
041B4;;;;JPN171;K3-2D78;;;
0517B;G0-5178;;JPN200;;;;
08980;G8-2F79;;JPN246;;;;
```

Because JPN is not a recognized format for the J source, they should be either removed or updated.

b) K4 source

Six entries in the CJK Unified Ideographs source reference file seem to have erroneous K source information:

```
03614;G_HZ;T3-5C35;;K4-0040;;;
03B20;G5-4A32;T3-5473;;K4-0106;;;
05579;G1-7C2E;;;K4-0031;;;
05C21;GE-277C;T3-6525;;K4-0079;;;
06725;GE-2D3E;T3-6633;;K4-0181;;;
06DA4;G0-3553;;;K4-0098;;;
```

- These K4 values have kept their decimal value unlike the K4 values referenced by CJK B code positions. They should be converted to hexadecimal values.
- The corresponding characters have empty K4 column in the merged standard (ISO/IEC10646:2003). This is obviously not consistent with the source information. Either they should have a representative glyph, or the references should be removed from CJKU_SR.txt
- Finally because the K4 original index is decimal in value and has been converted in hexadecimal notation for user convenience, a note should be added to ISO/IEC 10646 clarifying the relationship between the original K4 decimal index and the hexadecimal notation used in the standard. It may also be mentioned that the hex code is not bases on a row/column notation like the other hex indexes.

d) HKSCS references
The Hong Kong Supplementary Character Set (HKSCS) source reference data is outdated. They should be updated with the latest known HKSCS standard (2001). The authoritative HKSCS-2001 source is:
http://www.info.gov.hk/digital21/eng/hkscs/download/big5-iso.txt

This would result in the following changes:

Removal:
```
U+8728   HKSCS      957A
```

Addition to CJKU_SR.txt
(too many to list all of them here, about 3270 additions, this will be programmatically extracted by comparing the authoritative HKSCS 2001 source and the CJKU_SR.txt file)

```
U+3435   HKSCS      9277
U+3440   HKSCS      96DF
U+344C   HKSCS      89D5
U+3464   HKSCS      93CD
```

```
…
U+2A64A  HKSCS      9FEC
U+2A65B  HKSCS      8AE2
```

Note concerning transition from ISO/IEC 10646-1:2000 Amd.1 to ISO/IEC 10646:2003

The following changes were made unbeknownst to the editor between these two versions of the standard. These changes are inline with HKSCS 2001 source.

```
U+23FF0 HKSCS  from 8E6A to HKSCS FDD8
U+25D20 HKSCS  from FDD8 to HKSCS 8E6A
```

Addition to CJKC_SR.txt (CJK Compatibility Ideographs)

```
U+F907   HKSCS      8BF8
U+2F825  HKSCS      9874
U+2F83B  HKSCS      9AC8
U+2F878  HKSCS      8BC3
U+2F8A6  HKSCS      FC77
U+2F9BC  HKSCS      957A
```

Editor's comment:
All these modifications should be brought into the next amendment to ISO/IEC 10646:2003.

Finally it should be noted that there are still many HKSCS ideographs still encoded in PUA area. This is not acceptable from a WG2 point of view. These characters should either be part of the CJK unification process or added to the CJK compatibility area.
 [end]