Title:	HKSCS and GB 18030 PUA characters, request for additional characters and related information including the proposal summary form
Source:	UTC/US
Authors:	Michel Suignard
Action:	For consideration by UTC and WG2
Reference:	WG2 N2807

Summary

This documents describes characters still encoded in the Private Use Area of ISO/IEC 10646/Unicode as commonly found in the mapping information for Chinese coded characters such as HKSCS and GB-18030 that are not yet encoded and proposes code position for these characters. The supporting document WG2 N 2807 provides additional information for these PUA characters including those which are not proposed for encoding. The Proposal summary form is appended at the end of the document.

Miscellaneous Technical (2300-23FF)

Proposed Code	Glyph	Name	HKSCS Code	PUA Code
23DA		EARTH GROUND	88A9	F34A
23DB		FUSE	88AA	F34B

CJK Symbols Supplement (31C0-31EF)

This list is part of a larger set of CJK Basic Stroke that could be encoded as well. Because of that, a block of 48 code positions should be allocated to make room for that full set (31C0-31EF). The block name may also be renamed accordingly.

Proposed Code	Glyph	Name	HKSCS Code	PUA Code
31C0		CJK UNIFIED BASIC STROKE T	8840	F303
31C1		CJK UNIFIED BASIC STROKE WG	8841	F304
31C2		CJK UNIFIED BASIC STROKE XG	8842	F305
31C3		CJK UNIFIED BASIC STROKE SZ	8844	F307
31C4		CJK UNIFIED BASIC STROKE HZZ	8846	F309
31C5		CJK UNIFIED BASIC STROKE HZG	8849	F30C

31C6	㇆	CJK UNIFIED BASIC STROKE HP	884A	F30D
31C7	㇇	CJK UNIFIED BASIC STROKE HZWG	884D	F310
31C8	㇈	CJK UNIFIED BASIC STROKE SZZG	884F	F312
31C9	㇉	CJK UNIFIED BASIC STROKE HZT	8850	F313
31CA	㇊	CJK UNIFIED BASIC STROKE HZZP	8851	F314
31CB	㇋	CJK UNIFIED BASIC STROKE HPWG	8852	F315
31CC	㇌	CJK UNIFIED BASIC STROKE HZW	8854	F317
31CD	㇍	CJK UNIFIED BASIC STROKE HZZZ	8855	F318
31CE	㇎	CJK UNIFIED BASIC STROKE N	C879	F7E5

CJK Unified Ideographs (4E00-9FFF)

The following list of characters is the CJK Unified Ideographs addition with source reference information as Hanzi H source (Hong Kong Supplementary Character Set) and can be documented as such in the file CJKU_SR.txt. The reference is noted as 'H-' followed by the HKSCS Code in hexadecimal notation.

Proposed Code	Glyph	HKSCS Code	PUA Code
9FA6	燸	8C43	F57A
9FA7	曙	8C6D	F5A4
9FA8	匯	8C74	F5AB
9FA9	蘄	8CB7	F5CC
9FAA	赳	8CB9	F5CE
9FAB	幹	8CBB	F5D0
9FAC	鋼	8CC0	F5D5
9FAD	驖	8CD7	F5EC
9FAE	驥	8CD8	F5ED
9FAF	鉞	8CDA	F5EF
9FB0	𠯏	C8A1	F7EB
9FB1	図	C8A3	F7ED
9FB2	㇏	8843	F306
9FB3	鏈	8CED	F602
9FB4	㙫	8D48	F61C

The following character which is already encoded needs a new source reference as follows:

Code position	Glyph	Source reference (H Column)
4491	䒑	H-C87E
23551	椻	H-8CEB

As noted in WG2 N2807, the character U+23551 椻 differs slightly from H-8CEB as represented above and therefore a new character for H-8CEB could be added instead.

The following list of characters is the CJK Unified Ideographs addition with source reference information as Hanzi G9 source for GB18030-2000 and can be documented as such in the file CJKU_SR.txt. The reference is noted as 'G9-' followed by the GB18030 Code in hexadecimal notation.

Proposed Code	Glyph	GB18030 Code	PUA code
9FB5	龵	FE59	E81E
9FB6	龶	FE61	E826
9FB7	龷	FE66	E82B
9FB8	龸	FE67	E82C
9FB9	龹	FE6D	E832
9FBA	龺	FE7E	E843
9FBB	龻	FE90	E854
9FBC	龼	FEA0	E864

The following characters which are already encoded needs a new source reference as follows:

Code position	Glyph	Source reference (G column)
20087	𠂇	G9-FE51
20089	𠂉	G9-FE52
200CC	𠃌	G9-FE53
215D7	𡗗	G9-FE6C
2298F	𢦏	G9-FE76
241FE	𤇾	G9-FE91

Vertical Forms (FE10-FE1F)
This is a newly created block to contain vertical forms variants required to encoding the corresponding characters contained in GB18030.

Proposed Code	Glyph	Name	GB18030 Code	PUA Code
FE10	，	PRESENTATION FORM FOR VERTICAL COMMA	A6D9	E78D

FE11	、	PRESENTATION FORM FOR VERTICAL IDEOGRAPHIC FULL STOP	A6DB	E78F
FE12	○	PRESENTATION FORM FOR VERTICAL IDEOGRAPHIC COMMA	A6DA	E78E
FE13	:	PRESENTATION FORM FOR VERTICAL COLON	A6DC	E790
FE14	;	PRESENTATION FORM FOR VERTICAL SEMICOLON	A6DD	E791
FE15	!	PRESENTATION FORM FOR VERTICAL EXCLAMATION MARK	A6DE	E792
FE16	?	PRESENTATION FORM FOR VERTICAL QUESTION MARK	A6DF	E793
FE17	⎯	PRESENTATION FORM FOR VERTICAL LEFT WHITE LENTICULAR BRACKET	A6EC	E794
FE18	⎯	PRESENTATION FORM FOR VERTICAL RIGHT WHITE LENTICULAR BRACKET	A6ED	E795
FE19	⋮	PRESENTATION FORM FOR VERTICAL HORIZONTAL ELLIPSIS	A6F3	E796

UCS Sequence Identifiers (USI)
Should ISO/IEC 10646 in the future contain a list of USIs for well identified composite sequences, the following ones from the HKSCS set should be included:

Glyph	USI	Name
Ê̄	<00CA,0304>	LATIN CAPITAL LETTER WITH CIRCUMFLEX AND MACRON
Ê̌	<00CA,030C>	LATIN CAPITAL LETTER WITH CIRCUMFLEX AND CARON
ê̄	<00EA,0304>	LATIN SMALL LETTER WITH CIRCUMFLEX AND MACRON
ê̌	<00EA,030C>	LATIN SMALL LETTER WITH CIRCUMFLEX AND CARON

A. Administrative

1. **Title:** **Proposal to encode HKSCS and GB18030 PUA characters**
2. Requester's name: *UTC and US national member body (INCITS/L2)*
3. Requester type (Member body/Liaison/Individual contribution): *Liaison and member*
4. Submission date: *2004-06-16*
5. Requester's reference (if applicable): *NA*
6. Choose one of the following: *This is a complete proposal:*

B. Technical - General

1. Choose one of the following:
     a. This proposal is for a new script (set of characters): *Yes*
          Proposed names of script: *CJK Symbols Supplement*
                                                    Vertical Forms
.   b. The proposal is for addition of character(s) to an existing block: *Yes*
          Name of the existing block: *Miscellaneous Technical*
                                                    CJK Unified Ideographs
2. Number of characters in proposal: *50*
3. Proposed category (select one from below - see section 2.2 of P&P document): *A-Contemporary*
4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): 1
          Is a rationale provided for the choice? *Yes*
               If Yes, reference: *Spacing characters*
5. Is a repertoire including character names provided? *Yes*
          a. If YES, are the names in accordance with the "character naming guidelines"
               in Annex L of P&P document? *Yes*
          b. Are the character shapes attached in a legible form suitable for review? *Yes*
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
          publishing the standard? *Michel Suignard and Asmus Freytag*
          If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
          used: *see above*
7. References:
          a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
          b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
               of proposed characters attached? *No* (Government standards)
8. Special encoding issues:
          Does the proposal address other aspects of character data processing (if applicable) such as input,
          presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *Yes*
9. Additional Information:
          Submitters are invited to provide any additional information about Properties of the proposed Character(s) or
Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or
script. Examples of such properties are: Casing information, Numeric information, Currency information, Display
behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional
behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode
normalization related information. See the Unicode standard at http://www.unicode.org for such information on other
scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for
information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

---

[1] Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	*No*
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	*Yes*
If YES, with whom?	*INCITS*
If YES, available relevant documents:	L2 161R
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	*Yes*
Reference:	*Government Standard*
4. The context of use for the proposed characters (type of use; common or rare)	*Common*
Reference:	*See 3*
5. Are the proposed characters in current use by the user community?	*Yes*
If YES, where? Reference:	*See 3*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	*Yes*
If YES, is a rationale provided?	*Yes*
If YES, reference:	*WG2 N2807*
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	*NA*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	*Yes*
If YES, is a rationale for its inclusion provided?	*See 3*
If YES, reference:	*NA*
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	*No*
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	*Yes*
If YES, is a rationale for its inclusion provided?	*See 3*
If YES, reference:	*WG2 N2807*
11. Does the proposal include use of combining characters and/or use of composite sequences?	*No*
If YES, is a rationale for such use provided?	
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	*No*
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	*No*
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility character(s)?	*No*
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	