

ISO/IEC JTC 1/SC 2/WG 2 N 3063

ISO/IEC JTC 1/SC 2/WG 2/IRG N 1180

**ISO/IEC JTC1/SC2/WG2/IRG****Ideographic Rapporteur Group  
(IRG)**

<b>TITLE:</b>	Proposed additions to the CJK Strokes block of the UCS
<b>SOURCE:</b>	IRG Rapporteur
<b>STATUS:</b>	Submission from the IRG to WG2
<b>DISTRIBUTION:</b>	Members of ISO/IEC JTC1/SC2/WG2
<b>DATE:</b>	2006-4-3
<b>REFERENCE:</b>	WG2 N2807R, N2808R, IRG N1174
<b>ATTACHMENTS:</b>	IRG N1181 (“Summary for Stroke submission”)

**Summary**

This document proposes the addition of twenty new CJK Strokes to the CJK Strokes block of the UCS.

WG2 resolution M45.34 (N2754R) expanded the scope of IRG work to include CJK Strokes. The CJK Strokes block (U+31C0..U+31EF) now contains a total of sixteen CJK Strokes (U+31C0..U+31CF), derived from HKSCS (ISO/IEC 10646:2003/Amd.1).

The IRG formed ad-hoc groups to complete the repertoire of common CJK Strokes, and finalized the repertoire in the IRG#25 meeting held in Berkeley, California, U.S.A. The twenty new CJK Strokes are proposed for code points in the range (U+31D0..U+31E3). Information on naming conventions and collation is also provided.

**List of the proposed characters, character names and code positions**

	31C	31D	31E				
0	一	一	乙	31C0	CJK STROKE T	31E0	CJK STROKE HXWG
				31C1	CJK STROKE WG	31E1	CJK STROKE HZZZG
1	丿	丨	㇇	31C2	CJK STROKE XG	31E2	CJK STROKE PG
				31C3	CJK STROKE BXG	31E3	CJK STROKE Q
2	㇇	ノ	㇇	31C4	CJK STROKE SW		
				31C5	CJK STROKE HZZ		
3	㇇	ノ	〇	31C6	CJK STROKE HZG		
				31C7	CJK STROKE HP		
4	㇇	、		31C8	CJK STROKE HZWG		
				31C9	CJK STROKE SZWG		
5	㇇	㇇		31CA	CJK STROKE HZT		
				31CB	CJK STROKE HZZP		
6	㇇	一		31CC	CJK STROKE HPWG		
				31CD	CJK STROKE HZW		
7	㇇	㇇		31CE	CJK STROKE HZZZ		
				31CF	CJK STROKE N		
8	㇇	㇇		31D0	CJK STROKE H		
				31D1	CJK STROKE S		
9	㇇	㇇		31D2	CJK STROKE P		
				31D3	CJK STROKE SP		
A	㇇	㇇		31D4	CJK STROKE D		
				31D5	CJK STROKE HZ		
B	㇇	㇇		31D6	CJK STROKE HG		
				31D7	CJK STROKE SZ		
C	㇇	㇇		31D8	CJK STROKE SWZ		
				31D9	CJK STROKE ST		
D	㇇	㇇		31DA	CJK STROKE SG		
				31DB	CJK STROKE PD		
E	㇇	㇇		31DC	CJK STROKE PZ		
				31DD	CJK STROKE TN		
F	㇇	㇇		31DE	CJK STROKE SZZ		
				31DF	CJK STROKE SWG		

## Stroke Type Naming Conventions

Stroke types have traditionally been named using combinations of Han characters meaning ‘horizontal’, ‘vertical’, ‘bent’, ‘curved’, etc.

These characters are conveniently represented by single letters of the alphabet, which are abbreviations for the Mandarin pronunciations of these characters. The following table shows each abbreviation, in alphabetical order, followed by the corresponding Hanyu Pinyin, Han character, and approximate meaning in English. In all cases, the final alphabetic string in the character name may be converted to the sequence of Han characters for that name, simply by substituting a Latin letter in the following list for the corresponding Han character.

For example, "HZZZG" becomes 横折折折钩.

B	biǎn	扁	‘flat’	Q	quān	圈	‘circle’
D	diǎn	點	‘dot’	S	shù	豎	‘vertical’
G	gōu	鉤	‘hook’	T	tí	提	‘rising’
H	héng	橫	‘horizontal’	W	wān	彎	‘curved’
N	nà	捺	‘right-falling’	X	xié	斜	‘slant’
P	piě	撇	‘left-falling’	Z	zhé	折	‘bent’
















## Variants and usage examples of CJK Strokes

The following table illustrates typical variant forms of each stroke type and Han characters encoded in ISO/IEC 10646 as examples of usage.

Stroke	Name	Variants	Usage examples
一	H (橫)		一, 三, 丁, 丞, 丈, 世, 不, 上, 十, 卅, 七
 U+31C0	T (提)		冰, 淋, 治, 冶, 冽, 暴, 汜, 录, 地, 虫
	S (豎)	/	巾, 中, 串, 冫, 乍, 上, 五, 丑

丿	SG (豎鉤)		爭, 事, 求, 水
ノ	P (撇)		乂, 爻, 禾, 毛, 乏, 乖, 采, 衣, 八, 行
㇇	SP (豎撇)		乃, 月, 用, 齊, 几, 人, 班, 大
丶	D (點)	ノ	丸, 义, 永, 冰, 凡, 丹, 主, 求, 火, 刃
㇏ U+31CF	N (捺)	㇏ ㇏ ㇏	大, 人, 天, 入, 走, 边, 廷, 尺
㇏	TN (提捺)	㇏	尗, 八, 入, 廻
㇇	HZ (橫折)		口, 冂, 田, 品, 吕, 申, 甲, 圆, 𠂇
㇇ U+31C7	HP (橫撇)	㇇	又, 双, 爻, 今
→	HG (橫鉤)		疋, 了, 予, 矛, 子, 字, 疏, 写, 冫
㇇	SZ (豎折)	㇇	断, 继, 山, 互, 彙, 牙, 乐, 东

┘	<b>SWZ</b> (豎彎左)		肅嘯蕭簫
└ U+31C4	<b>SW</b> (豎彎)		區，亡，妄，四
↓	<b>ST</b> (豎提)		食，良，艮，很，狠，鄉，民
┘	<b>PZ</b> (撇折)	ㄥ	弘，公，翁
<	<b>PD</b> (撇點)		巡，獵，災，笛，女，𠃉
ㄣ	<b>PG</b> (撇鉤)		メ
ㄩ U+31C1	<b>WG</b> (彎鉤)		狐，嶽，貓，家，逐
ㄣ U+31C2	<b>XG</b> (斜鉤)		戈，弋，戰，我
ㄣ U+31C3	<b>BXG</b> (扁斜鉤)		心，必，沁，恣，蕊
┘ U+31C5	<b>HZZ</b> (橫折折)		卍
ㄣ U+31CD	<b>HZW</b> (橫折彎)		爻，投，朶
┘ U+31CA	<b>HZT</b> (橫折提)		讠，计，鳩

 U+31C6	<b>HZG</b> (橫折鉤)		羽, 习, 包, 勻, 葡, 用, 青, 甫, 勺, 月, 也, 乚
 U+31C8	<b>HZWG</b> (橫折彎鉤)		飞, 风, 瘋, 九, 几, 气, 虱
	<b>SZZ</b> (豎折折)	  	亞, 鼎, 𠂇, 吳, 专, 設, 廨, 奠
	<b>SWG</b> (豎彎鉤)		乱, 己, 巳, 巳
 U+31CE	<b>HZZZ</b> (橫折折折)		凸
 U+31CB	<b>HZZP</b> (橫折折撇)		建, 及
	<b>HXWG</b> (橫斜彎鉤)		乙, 𠂇, 乞
 U+31CC	<b>HPWG</b> (橫撇彎鉤)		阝, 队, 邮
 U+31C9	<b>SZWG</b> (豎折彎鉤)		号, 亏, 弓, 强, 丐, 马, 𠂇
	<b>HZZZG</b> (橫折折折鉤)		乃, 孕, 仍
	<b>Q</b> (圈)		〇, 柁, 𠂇, 𠂇

## Unicode Character Properties

All proposed characters should have the following Unicode properties so that they match those of the encoded sixteen CJK Strokes.

So;0;ON;;;;;N;;;;;

## Collation of CJK Strokes

Because the HKSCS strokes were added to the block before the more complete repertoire was developed, the strokes in the block are out of order, in terms of the traditional stroke classification. CJK Strokes are divided into five broad types, in Mandarin called the five 札 zhá types, since the Han character “札” exhibits each of the types: 一 | 丿 丶 乙 ( héng shù piě diǎn zhé ). The collation order proposed here assigns each stroke to one and only one of the five types, and further sub-classifies each stroke according to its specific features (see IRG N987). The following is an XML representation of the collation data for the proposed CJK Strokes block, in accordance with CLDR guidelines.

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Idml SYSTEM "http://www.unicode.org/cldr/dtd/1.4/ldml.dtd">
<ldml>
  <identity>
    <version number="$Revision: 1.00 $"/>
    <generation date="$Date: 2005/12/12 12:12:12 $"/>
    <language type="root" /> <!-- since this is a pan-language ordering of CJK -->
  </identity>
  <collations> <!-- removing the valid sublocales until a complete list is available -->
    <collation type="IRG-Strokes" > <!-- or whatever name would be most descriptive -->
      <rules>
        <!-- 札 zhá type 1 -->
        <!-- h; t -->
        <pc> 一 丿 </pc>
        <!-- 札 zhá type 2 -->
        <!-- s; sg -->
        <pc> | 丨 </pc>
        <!-- 札 zhá type 3 -->
        <!-- p; sp -->
```

```

    <pc> 丿 </pc>
    <!-- 札 zhá type 4 -->
    <!-- d; n; tn -->
    <pc> 丶 丶 丶 </pc>
    <!-- 札 zhá type 5 -->
    <!-- hz; hp; hg; sz; swz; sw; st; pz; pd; pg; wg; xg; bxg; hzz; hzw; hzt; hzg; hxg; szz; swg; hzzz; hzzp;
    hxwg; hpwg; szwg; hzzzg; q -->
    <pc> ㇀ ㇁ ㇂ ㇃ ㇄ ㇅ ㇆ ㇇ ㇈ ㇉ ㇊ ㇋ ㇌ ㇍ ㇎ ㇏ ㇐ ㇑ ㇒ ㇓ ㇔ ㇕ ㇖ ㇗ ㇘ ㇙ ㇚ ㇛ ㇜ ㇝ ㇞ ㇟ </pc>
    </rules>
    </collation  >
</collations>
</ldml>

```

## Rationale for proposed repertoire

The repertoire of strokes proposed for addition to the existing CJK Strokes block is derived from the stroke types occurring in representative forms of currently encoded UCS Ideographs. All of these proposed strokes are currently missing from the CJK Strokes block. Representative forms of some proposed CJK Strokes are similar in appearance to representative forms of some single-stroke CJK Ideographs currently encoded in various UCS blocks (URO, Ext A, Ext B, Kangxi Radicals, CJK Radicals Supplement). However, single-stroke CJK Ideographs do not have the properties of CJK Strokes, and single-stroke CJK Ideographs may in some cases exhibit a range of variation in their representative glyphs which conflates necessary distinctions for the CJK Strokes block. For example, the proposed CJK Strokes 丿 “CJK STROKE P” (U+31D2) vs. 丿 “CJK STROKE SP” (U+31D3) are conflated in the representative forms currently used for 丿 “KANGXI RADICAL SLASH” (U+2F03) and 丿 “CJK UNIFIED IDEOGRAPHS-4E3F” (U+4E3F). The proposed strokes complete the set of most commonly seen stroke types. The precise definition of the CJK Strokes block and clear differentiation of it from the blocks of CJK Ideographs and Radicals serves an essential purpose in the indexing and collation of encoded and unencoded CJK Ideographs and Radicals. It is anticipated that some very rare strokes may be proposed for future addition to this set.



**ATTACHEMENTS: IRG N1181 "Summary for Stroke submission"**

**A. Administrative**

1. Title:	<b>Summary for Stroke submission</b>
2. Requester's name:	<i>IRG Rapporteur</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>IRG</i>
4. Submission date:	<i>2006-4-3</i>
5. Requester's reference (if applicable):	<i>IRG N1174</i>
6. Choose one of the following:	
This is a complete proposal:	<i>Yes</i>
(or) More information will be provided later:	

**B. Technical – General**

1. Choose one of the following:					
a. This proposal is for a new script (set of characters):					
Proposed name of script:					
b. The proposal is for addition of character(s) to an existing block:	<i>Yes</i>				
Name of the existing block:	<i>CJK Strokes</i>				
2. Number of characters in proposal:	<i>20</i>				
3. Proposed category (select one from below - see section 2.2 of P&P document):					
A-Contemporary	<input type="checkbox"/>	B.1-Specialized (small collection)	<input checked="" type="checkbox"/>	B.2-Specialized (large collection)	<input type="checkbox"/>
C-Major extinct	<input type="checkbox"/>	D-Attested extinct	<input type="checkbox"/>	E-Minor extinct	<input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic	<input type="checkbox"/>	G-Obscure or questionable usage symbols			
4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document):	<i>1</i>				
Is a rationale provided for the choice?	<i>No</i>				
If Yes, reference:					
5. Is a repertoire including character names provided?	<i>Yes</i>				
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>Yes</i>				
b. Are the character shapes attached in a legible form suitable for review?	<i>Yes</i>				
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?	<i>Beijing Founder Electronic Co.,Ltd.</i>				
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	<i>9, No. 5 street, Shangdi, Information Industry Base, Haidian District, Beijing 100085, P.R. China; yjh@founder.com.cn; http://www.foundertype.com</i>				
7. References:					
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>Yes (IRG N1174)</i>				
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>Yes; see "Variants and usage examples of CJK Strokes" section in IRG N1180.</i>				
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)?	<i>Yes</i>				

**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.

**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before?	<i>No</i>
If YES explain	
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.)?	<i>Yes</i>
If YES, with whom?	<i>IRG and Ideographic experts</i>
If YES, available relevant documents:	<i>IRG N987, IRG N1081, IRG N1086A, IRG N1096, IRG</i>

- |  |                                |
|--|--------------------------------|
| 3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? | <i>Users of Han characters</i> |
| Reference: _____   |                                |
| 4. The context of use for the proposed characters (type of use; common or rare)  | <i>common</i>                  |
| Reference: _____   |                                |
| 5. Are the proposed characters in current use by the user community?   | <i>Yes</i>                     |
| If YES, where? Reference: _____  |                                |
| Yes; see the representative glyphs in the published ISO/IEC 10646 Standard.  |                                |
| 6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?                                  | <i>Yes</i>                     |
| If YES, is a rationale provided?   | <i>Yes</i>                     |
| If YES, reference: _____   |                                |
| <i>IRG N1180</i>   |                                |
| 7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?  | <i>Yes</i>                     |
| 8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?  | <i>No</i>                      |
| If YES, is a rationale for its inclusion provided?   |                                |
| If YES, reference: _____   |                                |
| 9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?               | <i>No</i>                      |
| 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?                                     | <i>Yes</i>                     |
| If YES, is a rationale for its inclusion provided?   | <i>Yes</i>                     |
| If YES, reference: _____   |                                |
| <i>See "Rationale for proposed repertoire" section in IRG N1180</i>  |                                |
| 11. Does the proposal include use of combining characters and/or use of composite sequences?   | <i>No</i>                      |
| 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?  |                                |
| If YES, reference: _____   |                                |
| 12. Does the proposal contain characters with any special properties such as control function or similar semantics?  | <i>No</i>                      |
| If YES, describe in detail (include attachment if necessary)   |                                |
| _____  |                                |
| _____  |                                |
| 13. Does the proposal contain any Ideographic compatibility character(s)?  | <i>No</i>                      |
| If YES, is the equivalent corresponding unified ideographic character(s) identified?   |                                |
| If YES, reference: _____   |                                |