

Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: ISO/IEC JTC1/SC2/WG2/IRG

Title: Proposal to disunify U+5F50 into three separate ideographs—redux

Source: Ken Lunde

Status: Member Body Contribution

Action: To be considered by the IRG and UTC

Date: 2022-02-18

This document is a proposal to disunify U+5F50 𠄎 into three separate ideographs, and to treat the two disunified forms as Urgently Needed Characters (UNCs) for encoding in the next version of the Unicode Standard. Disunifications are treated as *de facto* UNCs.

Background

IRG N2509 (aka L2/21-152) from IRG #57 proposes that U+5F50 𠄎 be disunified into three separate ideographs, 𠄎, 𠄎, and 𠄎. The first form would remain encoded at U+5F50 𠄎. The second form would be assigned the U-Source source reference UTC-01005 that is currently assigned to U+5F50 𠄎. The third form was proposed in L2/21-179 as a new U-Source ideograph to be added to UAX #45 as UTC-03251, and was accepted during UTC #169.

IRG N2538 from IRG #58 proposes that the second disunified form of U+5F50 𠄎, 𠄎, be treated as a UNC.

Proposal

Given that there is already a proposal on the table to disunify U+5F50 𠄎 into three separate ideographs, which is in reference to IRG N2509 (aka L2/21-152), it doesn't make much sense to treat only one of the disunified forms as a UNC. The formal proposal is as follows: **Disunify U+5F50 𠄎 into three separate ideographs—𠄎, 𠄎, and 𠄎—and treat the two disunified forms—𠄎 and 𠄎—as Urgently Needed Characters.**

The next two available CJK Unified Ideograph code points are **U+2B739** and **U+2B73A** at the end of Extension C, and we further propose that they be assigned the following normative and informative Unihan database property values:

U+2B739	kIRG_MSource	MC-00140
U+2B739	kIRG_USource	UTC-01005
U+2B739	kRSUnicode	58.0
U+2B739	kTotalStrokes	3
U+2B73A	kIRG_USource	UTC-03251
U+2B73A	kRSUnicode	58.0
U+2B73A	kTotalStrokes	3

This entails moving UTC-01005 from U+5F50 to U+2B739, and changing the UAX #45 *USourceData.txt* records for UTC-01005 and UTC-03251 as follows:

```
UTC-01005;C;U+2B739;58.0;;𠄎𠄎;UTCDoc L2/12-333 37, UTCDoc L2/21-179 2;;3;0
UTC-03251;C;U+2B73A;58.0;;𠄎𠄎;UTCDoc L2/21-179 1;;3;0
```

That is all.

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/IRGN2543 and IRGN2538_TCA Feedback

Date: 2022-03-11

Doc. Type:	Member body contribution
Title:	TCA's Feedback to IRGN2543 and IRGN2538
Source:	TCA
Status:	Input to IRG meeting 58
Action:	For consideration by JTC1/SC2/WG2/IRG
Distribution:	IRG Members and Ideographic Experts
No. of pages:	2
Appendixes	None
Medium:	Electronic

TCA has some thoughts on the proposal to separate the following three character (see figure below) into separate codes as mentioned in this document, as described below.

SNOUT	𠂇
HAND	𠂆
LONG HAND	𠂇

Table 1. The characters

5F50 𠂇 58.0

G0-6566 𠂇 H-C6D0 𠂇 T3-2140 𠂇 J1-3C77 𠂇 K2-313A 𠂇 V1-5454

UTC-01005 𠂇

CJK Radicals Supplement

2E95 𠂇 CJK RADICAL SNOUT TWO

→ 5F50 𠂇

IRGN2509 page 4 & ISO/IEC 106464 code chart

For the three characters in the picture above, TCA considered that it is only necessary to split into two codepoint, because the meaning of both characters is the same (hand 𠂆 and long hand 𠂇), both are evolved from the ancient character "又(𠂇)" and the modern Chinese character "手".

U+5F50 was discussed in IRG #57, the decision was made to prefer separate coding, if China and Vietnam agree, then 𠂇(snout) and 𠂆(hand) will be coded separately (Refer to IRGN2500). William T. Nelson proposed that long hand (𠂇) is also used as a variant of snout(𠂇) mentioned in Japanese dictionaries, it can be handled by IVS because the two forms are too similar.

In addition, for IRGN2538, the name urgent character(≡) applied for in Macao should not be a new character, it can be solved by applying for horizontal expansion, either with the new code or with U+5F50.

(End of document)

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N2543

Date: 2022-03-13


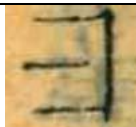


Source:	China
Meeting:	IRG#58
Title:	Feedback on IRGN2543 by China
Status:	Member's submission
Actions required:	To be considered by IRG
Distribution:	IRG
Author:	TAO Yang
Medium:	Electronic
Pages:	2
Appendixes:	0

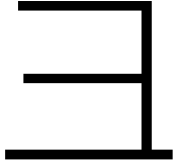
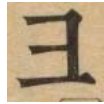


For 扌 has been encoded independent, so the dispute focus on 冫/ 冫/ 冫.

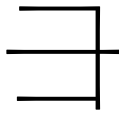


First, it's needed to clarify the actual use of the three characters. I tried to distinguish the meaning of three words according to the information I could find. The evidences show that, 冫 and 冫 share the meaning of snow, 冫 and 冫 share the meaning of hand. That further proofed why G-source chose 冫 as the head character in modern dictionaries.

Generally speaking, these three characters cannot completely replace each other, each has its own use case. Encode them separately should be a good solution.

Because China has adopted the reform rules to determine the glyph on the code point U+5F50, the separation will produce two identical glyphs. Maybe China could put 冫 on the new code point, while the other sources deliver the glyphs of 冫.

Glyph	Function	Character Example	Examples	Sources
冫	Right hand, component of 白		白，又手也，从 扌冫，凡白之属 皆从白。	说文解字
	Hand		蓋冫耳目口鼻之 欲，雖生而有之， 之性乎然。	刘戴山先生集
	Snow		雪 冫	第二次汉字简化 方案（草案）
	Katakana (yo)		西藏语谓羚羊 为、二ヨール， 其音即ノール	宋教仁集

Glyph	Function	Character Example	Examples	Sources
	Head of 彙		《唐韵》《集韵》： 竝居例切，音尉。 《广韵》：彙头 也。	康熙字典
	Snow		雨雪之雪假作 彙	建德縣志
	Head of 录		说文作𠄎，刻木 录，录也……变 彙为𠄎	隶辨

Glyph	Function	Character Example	Examples	Sources
	Again(Right hand)		彙即又字，偏旁 也	十三经文字偏旁 考略
	Component, Again(Right hand)		偏旁作彙，帚尋 聿肅妻秉庚急雪 彗丑尹等字从之	十三经文字偏旁 考略

(End of document)

ISO/IEC JTC1/SC2/WG2/IRG
Ideographic Research Group
(IRG)

Source/Contribution Identifier: Hong Kong Special Administrative Region of the People's Republic of China

Meeting: IRG Meeting No. 59, Online Meeting

Title: Feedback on IRGN2543 on Disunification of U+5F50

Status: Member Body Submission

Action: For consideration by IRG

Feedback on IRGN2543 on Disunification of U+5F50

1. The Hong Kong Special Administrative Region (HKSAR) is fine with the disunification of U+5F50 as proposed in IRGN2543.
2. The character 彡 with H-source glyph (H-C6D0) should remain in U+5F50 as it is commonly used in HKSAR and is included in the Hong Kong Supplementary Character Set (HKSCS) as well as the Reference Glyphs for Chinese Computer Systems in Hong Kong. 彡 is also used as a Kangxi radical in HKSAR as found in some local publications and dictionaries.
3. As a variant form of the radical character 彡, the ideograph 彡 (U+2E95 in the CJK Radicals Supplement block) is also commonly used in HKSAR and is included in the HKSCS with H-source glyph (H-C8DE) as well. It is also found in some local publications and dictionaries.

<p><i>Zhongwen Xiangjie Zidian</i> (《現代中文詳解字典》) published by the Commercial Press (HK) Ltd in 2002:</p>	<p><i>Longman Advanced Chinese Dictionary</i> (《朗文中文高級新辭典》)(2nd edition) published by Pearson Education Asia Limited in 2003:</p>
<p style="text-align: center;">彡 (彡 彡) 部</p> <p>彡 [tuàn 去ㄨㄢˋ ㄩˋ] ㄊㄨㄢˋ ㄩˋ 《周易》中說明卦義的文字，也稱“卦辭”。</p> <p>彡 [huì ㄏㄨㄟˋ ㄨㄟˋ] ㄏㄨㄟˋ ㄨㄟˋ ① 掃帚。② 彡星，俗稱掃帚星。</p> <p>彡 [zhì ㄓˋ ㄩˋ] ㄓˋ ㄩˋ 豬 ◆ 狗彡不若。</p> <p>彡 (彡) [huì ㄏㄨㄟˋ ㄨㄟˋ] ㄏㄨㄟˋ ㄨㄟˋ ① 聚集；集合 ◆ 彡彡 彡編 彡總 彡報。② 聚集而成的東西 ◆ 詞彡 信息總彡。</p>	<p style="text-align: center;">彡部</p> <p>(3) 彡 ㄏㄨㄟˋ ㄨㄟˋ ① gei² 計 ② ji ③ 4 豬的頭。也作「彡」。</p> <p style="text-align: center;">六至二十三畫</p> <p>(9) 彡 ㄊㄨㄢˋ ㄩˋ ① tən³ 通信勿 ② tuàn ③ 士ㄨㄢˋ 【彡辭】 tuàn cí 《易經》中講卦義的文字。</p> <p>(11) 彡 (彡) ㄏㄨㄟˋ ㄨㄟˋ ① sœy⁶ 睡 ② hui ③ ㄏㄨㄟˋ ㄨㄟˋ 彡星，俗稱「掃帚星」，環繞太陽運行的一種天體，由石塊、氣體、塵埃等組成。在接近太陽時，其背向太陽的一面常拖着一條掃帚狀的尾巴。著名的有哈雷彡星。</p>

4. The ideograph 𠄎 is not commonly used in HKSAR and is not included in the HKSCS.

5. To conclude, 𠄎 (H-C6D0) is commonly used in HKSAR and it will remain in the code point U+5F50. We have no objection to encode the two disunified forms 𠄎 and 𠄎 as new characters.

End of document

Reference:

1. Hong Kong Supplementary Character Set – 2016 (《香港增補字符集—2016》):

5F50	2E95
𠄎	𠄎
H-C6D0	H-C8DE

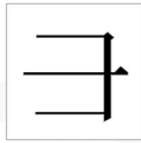
2. Reference Glyphs for Chinese Computer Systems in Hong Kong (《香港電腦漢字參考字形》), May 2017



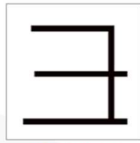
3. Glyph Comparison of U+2E95 and U+5F50 in 4 fonts :

Code Point	Adobe 明體 Std L	新細明體	微軟正黑體	華康香港 新標準宋體
U+2E95	𠄎	𠄎	𠄎	𠄎
U+5F50	𠄎	𠄎	𠄎	𠄎

4. 𠄎 (https://zi-hi.com/sp/uni/2E95)



GlyphWiki



裝置上的字型

U+2E95

CJK RADICAL SNOOT TWO

文字區域 Block	中日韓漢字部首補充 / CJK Radicals Supplement (U+2E80–U+2EFF) PDF表格
初始版本 Since Version	Unicode 3.0 發布於 1999/09
文字語系 Script	Han
一般分類 General Category	So / 其它符號 (Other_Symbol)
書寫方向 Bidirectional Class	ON / most other symbols and punctuation marks
書寫鏡射 Bidirectional Mirrored	N / 否
組合類別 Combining Class	0 / Not_Reordered; 無
東亞寬度 East Asian Width	W / 全形 UAX#11
直排方向 Vertical Orientation	U / 正立 UAX#50
詳細附註 Notes	參見 → U+5F50 𠄎

編碼資訊

HTML Entity	&#11925; &#x2E95;
UTF-8	E2 BA 95
UTF-16	2E95
UTF-32	00002E95
URL Encode	%e2%ba%95

東亞編碼資訊

Big5	[Big5-HKSCS] C8DE [Big5-UAO] 9CDB
Adobe-CNS1	[CNS1-3] 17699
Adobe-Japan1	[Pro] (aj4) 15391