Universal Multiple-Octet Coded Character Set International Organization for Standardization Organisation Internationale de Normalisation Международная организация по стандартизации

Doc Type: Working Group Document

Title: Review of Tangut repertoire in DAM ballot

Source: Andrew West, Michael Everson, Viacheslav Zaytsev

**Status:** Individual Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2015-07-16

#### 1. Introduction

We have carried out a review of the Tangut repertoire in the ISO/IEC 10646:2014 DAM2 ballot (SC2 N4399), in particular checking the stroke count of every Tangut ideograph and component.

#### 2. Tangut Ideographs

### 

The middle component of this character is given as  $\frac{2}{3}$  (18A35) in the DAM code chart, but as can be seen from the IDS sequence and source glyph image from Sofronov 1968 shown in this extract from N4522 p. 398, the middle component should be  $\frac{2}{3}$  (18A34).

17CF0	毲	Ⅲ 泵 泵 芘	R113	17	GCCQGABCC CQDCAMCD								<b>5397</b>		
-------	---	---------	------	----	-----------------------	--	--	--	--	--	--	--	-------------	--	--

Therefore the glyph for 17CF0 should be changed from  $\mathfrak{A}$  to  $\mathfrak{A}$ . This does not entail any change in stroke count or code point.

## 17F24 貕 (L2008-3438) [260.13]

The glyph is correct, but the IDS sequence and stroke count given in N4522 p. 459 is incorrect. The IDS sequence should be 3 %, and the stroke count 12.



As a result, the code point for this character should be changed from 17F24 to 17F20, and 17F20..17F23 moved to 17F21..17F24.

## 

18317	骸	□旹□□□炎	R257	14	DCBABEEACC	骸	骸	<b>談</b> 5928	馥	骸	<b>形</b> 4533	馥	
10017	IIX	UIIU X	14207		QCCQ	5928	5928	5928	4983	3464	4533 4685	1703	

As a result, the code point for this character should be changed from 18316 to 18314, and 18314...18315 moved to 18315...18316.

# 

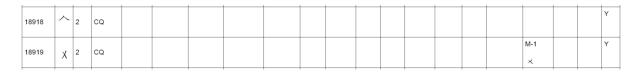
186A2	殿	□ 幂 灸	R382	14	ABEABEAACC CQCQ	<b>殿</b> 0523	展 0523	展0523		<b>殿</b> 3500	<b>長</b> 5464 5645	<b>廣发</b> 0290	展 2-122
186A3	殿	四	R382	14	ABEABEAACC CQCQ	<u></u> 4565	<b>殿</b> 4565	<b>殿</b> 4565	<b>殿</b> 0186				

There is only a minor, non-significant glyph difference between these two characters (whether the second stroke is straight as in 186A1 or slightly bent as in 186A2). According to the unification principles used in the proposal document these two characters should have been unified (cf. the unifications of L2008-0042/4537, L2008-0057/4548, L2008-0184/4510, and L2008-1134/4624 discussed in N4522 p. 10 which show the same insignificant glyph differences).

Therefore, 186A2 should be removed, and 186A3..187ED moved to 186A2..187EC. 186A1 should use the 186A2 glyph for consistency with 186A0 and its source reference should be changed to L2008-0523-4565.

#### 3. Tangut Components

In N4326 ("Proposal to encode Tangut radicals in the UCS") an inverted-V-shaped component  $^{^{^{^{^{^{\prime}}}}}}$  and an X-shaped component  $^{^{^{\prime}}}$  were proposed for encoding, but these two were removed from the final Tangut components proposal N4636.



It has been pointed out to us that the removal of these two characters was a mistake as they are required to represent the Tangut stroke types given in Li Fanwen 1997, 2006 and 2008, and reproduced as Fig. 5 in N4326.

筆	名	號碼	筆		形	与	Z	1	列	説明
單	横	1		7	L	。菠	。	。	恒。	横和横上下鈎
	垂	2	1	1	-	浅	颏	猦	報	直和撇
筆	點	3	,	1	VIE SPE	훍		姚	乘。	點和捺
	叉	4	乂	+	++	貒	訛	'莨'	继	兩筆交叉和横穿一筆
復	+1+	5	‡	丰	+++	浅	*幇	解	都	一筆横穿三筆,直穿二至三筆
21	***	6	#	1111	10	雁.	置	0071	200	一筆横、直穿四筆
	角	7	Γ	八	ţ	派	°Œ	觙	砰	一筆轉折和兩筆相接而成的角
筆	八	8	y		^	溪	췙	創	衸	類似漢字八和八的變形
	小	9	110	***	11	黹	辦	辨	帯。	類似漢字小和小的變形

Table of Tangut Stroke Types in Lǐ Fànwén 2008 page 23

All the stroke types in this table can be represented using either the current set of Tangut components or CJK stroke characters, except for  $^{^{\prime}}$  (row 8) and  $^{\prime}$  (row 4). These two components are also very useful for Tangut IDS sequences,  $^{^{\prime}}$  being used to describe  $^{\prime}$  6,  $^{\prime}$  7,  $^{\prime}$  3,  $^{\prime}$  3 and  $^{\prime}$  8; and  $^{\prime}$  8 being used to describe  $^{\prime}$  2 and  $^{\prime}$  8.

We therefore recommend adding  $^{\wedge}$  at 18817 and  $\chi$  at 18818, and move 18817..18AF0 to 18819..18AF2.