

ISO/IEC JTC 1/SC2/WG2/IRG
Ideographic Rapporteur Group

Source: Tatsuo L. Kobayashi / Yasuhiro Anan

Meeting: IRG#20 @ Hanoi

Title: Clarification of unification rule needed for Ext-C development

Keywords: (none)

Status: For your information

Short Description:

This paper discusses some issues on unification principles for CJK Unified Ideographs.

Proposed Conclusion / Requested Action: IRG to discuss and to provide better guideline of unification rule for further development of CJK Unified Ideographs repertoire.

Arguments / Text of Contribution: See attachment.

Clarification of unification rule needed for Ext-C development

0. Preface

During two rounds of C1 review, we encountered not a few cases where we could not find direct answers from the current unification rule (Annex S), without which we will be unable to produce reliable Ext-C1 repertoire and to develop further extensions of CJK Unified Ideographs.

This document asks to the IRG experts to examine examples shown below and to provide better guideline for Ext-C1 project.

The following four issues are presented with some C1v2 examples:

1. How to handle submissions of traditional v.s. simplified forms,
2. S.1.4.3,
3. S.1.5., and
4. Variations in radical part.

1. How to handle submissions of traditional v.s. simplified forms

Currently we have no rule to unify traditional and simplified forms automatically, so we would suggest discussing the IRG policy how to deal with the two forms in Ext-C project. One outstanding problem of encoding them without policy is that there can be isolated simplified forms without corresponding traditional forms.

1) Example of submission of simplified form for which corresponding traditional form is already encoded.

06458	部首	總筆劃	部首外筆劃	部首外第一筆筆順
	土	--	8	2
	Unicode	TCA	China	Japan
			CYY00347	
	Korea	Vietnam	HongKong	DPRK
	Singapore	Macau		KXI Index
				0232.181

➤ IDS:=  土 [Simplified form of 敗]

墩

Existing traditional form : U+213EF

2) Example of submission of simplified form for which corresponding traditional form cannot be found.

06454	部首	總筆劃	部首外筆劃	部首外第一筆筆順
	土	--	8	2
	Unicode	TCA	China	Japan
			CYY00343	
	Korea	Vietnam	HongKong	DPRK
	Singapore	Macau		KXI Index
				0232.181

- IDS:= 垠土[Simplified form of 電]
- U+6E11(垠) v.s. U+6FA0(垠); U+42F2(纒) v.s. U+7E69(纒)

2. S.1.4.3(different structures of a corresponding component)

There's another example of submissions of simplified form and traditional form.





Simplified form c1v2-05864


06451	部首	總筆劃	部首外筆劃	部首外第一筆筆順
	土	--	8	2
	Unicode	TCA	China	Japan
			CYY00340	
	Korea	Vietnam	HongKong	DPRK
	Singapore	Macau		KXI Index
				0232.181

Traditional form c1v2-06192

06822	部首	總筆劃	部首外筆劃	部首外第一筆筆順
	土	--	11	2
	Unicode	TCA	China	Japan
		TD-496A		
	Korea	Vietnam	HongKong	DPRK
	Singapore	Macau		KXI Index
				0238.191

Both are in the form of IDS:=  土[Country] . Please note that there are known variations of a component “country” which are considered different structure in terms of S.1.4.3. That is,

U+56EF(7 ) , U+56F6(8 ) , U+56FB(8 ) , U+5700(9 )

The question is, if those variations of  土[Country] are proposed, do we accept all of them?

3. S.1.5 (differences of actual shapes)

It's generally very hard to specify differences of actual shapes. The IRG experts are encouraged to develop / provide better examples and guidelines in order for users in the world to know what glyphs can be represent at a given code position.






The following is an example of submitted glyphs which could be considered as different in actual shapes each other.







c1v2-05875

c1v2-05876





c1v2-05988


塙 塙 塙

They are all in the form of IDS:=    [] [].

Some expert study shows that they are all glyphs that follow to  (U+5808) which origin is     












4. Variations in radical part

The question in this section is whether to encode glyphs with modified forms of known radicals. S.1.5 shows some cases that we unify those glyphs with variations of the same radical, for example, radical walk(), radical spirit(), radical eat(), radical grass() and so on.

The following is an example of a glyph with a variation of ()

C1v2-05899



By the way, we would unify  and  according to Annex S. Do we unify the glyph with the form of    as well as in the form of    and   ?

----end----