

Source:	BAI Yi
Title:	Feedback on A few considerations on CJK Supplementary Components for IDS
Status:	Individual Contribution on IRG #56
Action:	To be considered by IRG

1. In Section 2.4, it is proposed that either Kawabata-san's program or a new program can be used in IDS checking. Currently I'm developing a new IDS check system that can fulfill requirements listed in Annex B of PnP, additionally,
1) it can do "fuzzy search" by applying UCV rules to IDS to find potential unifiable characters. We've found unifiable characters in the working set 2017 by this tool.

Unification

Sn	Image/Source	Comment Type	Description
00145	<p>00145 魚</p> <p>人 9.11 13 □ (3) V-F0B0C</p> <p>Trad 𩺰イ魚</p>	Unification	<p>29D51 𩺰 𩺰 𩺰</p> <p>魚 195.2 UCS2003 GKX-1465.12 T5-444F</p> <p>Unifiable with 𩺰 under UCV rule 322.</p> <p>322 魚 𩺰 𩺰</p>

Fig 1. Potential unifiable character found in ws2017 by applying UCV rules
<https://hc.jsecs.org/irg/ws2017/app/index.php?id=00145>
Also, about 200 possible unifiable characters are found in existing character sets.
<https://github.com/eisoch/irg/issues/28#issuecomment-781768239>

2) it can be extended to include Supplementary Components in checking.
I think, if possible, this system can work along with Kawabata-san's program in future working sets to provide better machine checking on IDS and additional advice to reviewers.

2. In Section 2.6, it is proposed that when Supplementary Components are updated, old IDSes for existing characters should also be updated. In addition to that, when a new extension set is published, of course, characters in that extension set can also be used in

IDSes. For example, 𩺰 (U+30009), which was encoded in extension G, has identical

shape of &H5-01. Thus all IDS including this component can be updated. After an update on the working set, or after a new extension set is released, working in batch mode, our IDS system can check the IDS of all encoded characters and characters working sets, to provide advice about optimizing the IDS.

03966	<p>03966 𩺰</p> <p>衣 145.12 18 □ (1) UTC-03143</p> <p>Trad 𩺰&H5-01;𩺰石衣</p>	IDS	As 𩺰 has been encoded the IDS can be optimized to 𩺰𩺰𩺰石衣.
-----------------------	--	-----	--

Fig 2. Advice on optimizing IDS after checking

(End of document).