ISO/IEC JTC1/SC2/WG2 IRG N2124

Title: **Report of very preliminary IDS check by Taichi Kawabata**
Author: **suzuki toshiya**
Type: **individual contribution**
Action: **Reviewed in IRG#45**

I extracted IDS columns of the attribute Excel files for IRG Collection 2015 submission and asked Taichi Kawabata for preliminary IDS check. His checking is still ongoing, but he commented that he found many syntax errors in IDS data (I attached Excel file). Here, syntax error in IDS means the IDS data is not parsed correctly, it does not mean that the described glyph shape is different from the submitted glyph image. Kawabata-san commented that it is too early to start the estimation of the possibly unifiable candidates, because there might be many incorrect descriptions in IDS without syntax errors. Kawabata-san commented that 6 months or so would be needed for him to check the submitted IDS data could fit with existing data, before the real IDS checking.

Considering the syntax errors in IDS, in my personal opinion, it is expected that each submitters check their own submissions, for smooth reviewing. In fact, there is a possibility that strictly applying 5% rules to current submissions, some submissions would be refused by the unstable quality (over 1000 error means 10%). It does not mean that the rescheduling of the deadline to submit IRG Collection 2015. No new characters should be submitted to IRG Collection 2015 anymore, however, the problematic parts (e.g. wrong attribute, wrong IDS, wrongly designed glyph shape, etc) should be detected by the submitter before spending the human resource by IRG members.

(end of document)

Addition to IRGN2124 by John Knightley 18th November 2015

Further observations on 2015 submissions from IDS data

For approximately 85% of existing UCS characters have IDS 3 characters long, and a similar percentage would be expected IDS data of IRG submissions.

	IDS 3 long	Characters	Percentage
IRGN2091	238	313	76%
IRGN2107	1503	1656	91%
IRGN2113	335	3472(2261)	10%(15%)
IRGN2114	394	485	81%
IRGN2115	1903	2342	81%
IRGN2116	1113	2005	56%