Subject: A few considerations on CJK Supplementary Components for IDS (Part 2)
Date: 2021.03.12.
Author: KIM, Kyongsok (Republic of Korea)
Status: Individual Contribution

After submitting IRG N2464 dated on 2021.03.08. to IRG, I dug old IRG N documents related with Components used in IDS and found new facts, which explains why I write part 2 of IRG N2464 (the document number of this document is IRG N2464_2).

The last Subsection number in Section 2 in IRG N2464 is 2.8.
This document starts with Subsection number 2.9 following 2.8 in IRG N2464.

2.9 About 700 CDP glyphs (components)

- In IRG N1939, "Regarding to CDP characters used in IDS database", KAWABATA Taichi, 2013-05-17, CDP components are introduced in response to IRG Resolution M39.9. Some parts of the document are quoted below:

*This document is a **response to IRG M39.9 2.** This document describes the **CDP glyphs** used in IDS database [3] as an "alternate" for **unencoded component**, and propose several ways to handle unencoded components in the IDS.*

CDP glyphs in IDS data
*Attached glyph list shows the list of glyphs used in the IDS data [2] that are used in IDS data [3]. These glyphs are taken from **EUDC (End User Defined Characters) font** provided from **Chinese Document Processing (CDP) Laboratory of Academia Sinica (http://cdp.sinica.edu.tw/).** They are denoted as **`&CDP-XXXX;' (an entity reference style** notation) in IDS data 3, that is a practice of CHISE Project (http://www.chise.org/), where original IDS data is created and distributed.*

A list of about 700 CDP components is attached to that document. Some CDP components are shown below:

UCS	Big5	char	IDS	num	subtraction
F137	854B	冈	⿴冂从	1	=两-一
F138	854C	垔	⿴冂土	0	
F139	854D	爾	⿴市犮	2	=薾-工
F13A	854E	亜	⿴	1	=關-冂-吞
F13D	8551	隶	⿴木 ⿴丰八	1	=劃-田-一一 刂

I am not an Hanja expert to evaluate CDP components. I just wonder if CDP could be a good starting basis for Supp. Comps. I don't know the IRG conclusion regarding whether to allow CDP chars in IDS.

2.10 A list of PUA chars (UCS cp. and corresponding glyph) ?

In WG2 N4241, "Information in support of N4234 (L2/12-087) to demonstrate extensive use of PUA in common IDS data" by Dr. Ken Lunde, 2012-02-14, we can see the following statement:

The list below indicates how many characters in each CJK Unified Ideograph block use private use characters in their IDSes, based on the current version of the IDS database:

URO: 480
Extension A: 107
Extension B: 1,972
Extension C: 55
Extension D: 14

If IRG allowed MB to use some designated PUA chars in IDS within IRG, the list of these PUA chars need be maintained somewhere. I don't know if there is a set of such PUA chars.

2.11 Circled digits in IDS ?

Circled digits are found in ids.txt at https://github.com/cjkvi/cjkvi-ids, as shown below:

U+5DE4 巟　⿱巛囚⑦[GJK]　⿱巛囚⑦[T]
U+689F 梟　⿰⑦木
U+83EF 華　⿱艹⑦
U+88CA 裊　⿰⑦衣
U+9115 鄉　⿰乡皀阝[GTKV]　⿰乡⑦阝[J]
U+9B1B 鬛　⿱髟囚⑦

Actually Mr. Kawabata mentioned the possibility of using circled digits, ① ~ ㉚, in IDS in his document IRG N1939 (see 2.9 above for details).
I wonder if IRG allowed/allows MB to use circled digits in IDS (within IRG only?).

2.12 A full-width question mark, PUA chars, circled digits, CDP components, Supp. Comps., etc.

As we saw above, at least five possibilities were proposed to represent unencoded components: 1) a full-width question mark, 2) PUA chars, 3) CDP Components, 4) circled digits, 5) Supp. Comps..
It is suggested that IRG discuss whether to allow to use only one method OR two or more methods in MB submissions and in old IDSes.
A full-width question mark and PUA chars conform to Informative Annex I whereas CDP components, circled digits, and Supp. Comps. do not.

2.13 More than one IDS for one char.

- It is suggested that IRG discuss whether to allow more than one IDS for one CJK char and, if the answer is yes, specify a syntax to express two or more IDSes for one char.

* * *