Source:	Taichi Kawabata
Title:	**Regards to L2/18-012**
Distribution:	UTC for discussion
Medium:	Electronic
Pages:	2

Dear people at UTC,

Regarding to the proposed document of L2/18-012 [1], I have the following four concerns (and proposals). The main subject is on proposal of introducing of unary operators. I really appreciate if they are taken into the account during the discussion on this proposal.

i. Unary Operators may induce unnecessary complexity and ambiguity on IDS.

    Current IDS scheme already has some ambiguities for identifying the ideograph. For example, ▯▯ABC can also be described as "▯A▯ BC" or "▯AB▯C".

    Introduction of Unary operator would further worsen and complicate the situation. For example, suppose a=↔A, b=↔B, c=↔C, then, ▯▯ ABC=↔▯▯cba=▯▯A↔▯▯cb=↔▯▯b▯▯ aC. Same situation may happen to rotation operator for symmetric ideographs.

    In actual example, "興" may be described as 曰田↻Ⳝ or ↻ 曰Ⳝ田, or 壯 may be described as ↔▯↻ 干片. This may be an extreme, but indeed possible way to describe ideograph if such operators are introduced, inducing complex computation and symmetricity database to properly handle them.

ii. Unary Operators are only beneficial to quite limited unencoded components.

    Many ideographs are composed by placing multiple ideographic components in order, where the merit of IDS comes. However, when the ideograph becomes components, it sometimes mutates, mangles, or reduces its strokes, resulting to become "uncoded components".

    There are nearly a thousand of such components in already encoded ideographs. (This is roughly counted from the IDS data taken from [2].)

    However, among them, unencoded components that can be described by the proposed unary operators are quite limited. Here is the list of unencoded components that I could find so far.

    rotation
    鬼（鬼 ），虎（覽 ），或（戜），止（㐃）

---

[1]: http://www.unicode.org/L2/L2018/18012-irgn2273-four-new-idcs.pdf
[2]: https://github.com/cjkvi/cjkvi-ids/blob/master/ids.txt

mirror

弓（� 　）, 臣（䆨）, 犭（爘）, 才（吪）, 止（㞍）, 臣（囜）自（䶃）

Furthermore, it should be noted that only very few ideographs contain such components. For example, "↻鬼" is only used in "魑　" and is rarely used.

In addition, "↔𦣹" (mirror image of "𦣹") component is always used with "𦣹" as a pair. Thus we can describe them by using "▢" operator, e.g. ˮ䶃ˮ as "▢𦣹吏".

iii.   Instead of unary operator, I may propose to use ①- ⑯ as the DC of unencoded components, where each number denotes its stroke count.

That can be used in all other unencoded components, useful to narrow down the candidates of searching ideograph, and can also be used to calculate the total strokes. If the unencoded components really worth encoding, then let them be so[3].

iv.   Or, we may use existing character as a delegate, as an tentative description until component itself is encoded.

For example, "↔" (U+2194) or "↻" (U+21b7), (in addition to "？" (U+FF1F)), may serve as an unary operator by defining them so in appendix J of the UCS.

---

[3]:IRGN2225 (http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg48/IRGN2225.pdf)
With best regards,