

Doc Type: Ideographic Rapporteur Group Document
Title: Feedback on Proposals of 4 IDCs (IRG N2273R)
Source: Kushim Jiang (姜兆勤)
Status: Individual Contribution to IRG #55
Action: For consideration by IRG
Date: 2020-3-18

IRG N2273R shows 4 IDCs which are indeed necessary to be encoded. However, the proposed code points and blocks may need to be changed.

The security of generating IDS programmatically requires that an IDS must be structurally finite. This needs to be ensured by two factors:

- Characters are finite in stroke count;
- IDS are finite by definition.

The first factor is obvious. However, due to the recursive definition of IDS, any unary, binary and ternary operators must be not idempotent. Conversely, an idempotent operator cannot be encoded as an IDC, but only as an IDM (Ideographic Description Mark).

For proposition always holds that

$$\text{GetStrokeCount}(\text{IDS}) > 0$$

and

$$\text{GetStrokeCount}(\text{BinaryOperator IDS1 IDS2}) = \text{GetStrokeCount}(\text{IDS1}) + \text{GetStrokeCount}(\text{IDS2}),$$

then

$$\text{GetStrokeCount}(\text{BinaryOperator IDS1 IDS2}) > \max \{ \text{GetStrokeCount}(\text{IDS1}), \text{GetStrokeCount}(\text{IDS2}) \},$$

and similarly

$$\text{GetStrokeCount}(\text{TernaryOperator IDS1 IDS2 IDS3}) > \max \{ \text{GetStrokeCount}(\text{IDS1}), \text{GetStrokeCount}(\text{IDS2}), \text{GetStrokeCount}(\text{IDS3}) \},$$

which shows that in IDS definition, any binary and ternary operators are not idempotent.

For any sequences in the form of “ IDS ”, there is always:

$$\text{IDS}(\text{IDS}) = \text{IDS}.$$

That is why U+303E IDEOGRAPHIC VARIATION INDICATOR (IVI) is not considered a part of the IDS and does not invalidate the sequence (*Core Specification* v13.0, p. 734).

For any IDSes, there is always:

$$\text{IDS} \text{ IDS} = \text{IDS}, \text{IDS} \text{ IDS} = \text{IDS}.$$

Then the operator “ IDS ” and “ IDS ” are idempotent, this will lead to insecurity in generating IDS. So IDS and IDS cannot be encoded as an IDC, but an IDM like IDS .

For “U+303E IDS ” is encoded separate from “U+2246 IDS ”, and to make semantic distinctions, IDS and IDS had better be encoded separate from “U+2194 IDS ” and “U+21B7 IDS ”.

The IVI IDS need to be put ahead of an IDS. Due to the recursive definition of IDS, the IVI can also be put inside an IDS if the IDS is composed of sub-IDSes. Therefore, IDS and IDS can also be put ahead of an IDS or a sub-IDS, which means that IDS 虎 IDS 虎具 is a legal pseudo-IDS expression of “U+27E42 贊”.

In conclusion, IDS and IDS can be encoded as IDCs in maybe U+2FFC and U+2FFD, but IDS and IDS should be encoded as IDMs in maybe U+16FE5 and U+16FE6.

Eiso Chan (陈永聪) and Tao Yang (陶扬) also contributed to this article. #