

Feedback on Proposals to Encode New Ideographic Description Characters


Andrew West



8 July 2022

Documents L2/18-012 “Proposal of 4 IDCs” by Tao Yang, Eiso Chan, Wang Yifan (replacing L2/17-386), and L2/21-118R “Preliminary proposal to add a new provisional kIDS property (Unihan)” by Ken Lunde and John H. Jenkins together propose to encode five additional ideographic description characters:

IDC	Type	Character Name
	Binary	IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM RIGHT
	Binary	IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM LOWER RIGHT
	Unary	IDEOGRAPHIC DESCRIPTION CHARACTER HORIZONTAL REFLECTION
	Unary	IDEOGRAPHIC DESCRIPTION CHARACTER ONE HUNDRED EIGHTY DEGREE ROTATION
	Binary	IDEOGRAPHIC DESCRIPTION CHARACTER STROKE SUBTRACTION

I maintain one of the most widely-used IDS databases for CJK unified ideographs (<https://babelstone.co.uk/CJK/IDS.TXT>), and I can confirm that all five of the above characters would be used in my database if they were encoded.

At present, IDS sequences in my database which could use IDC Surround from Right and IDC Surround from Lower Right have to make use of the IDC Overlaid operator () which makes these sequences inexact, and less useful to users. I would very much welcome the encoding of these two binary operators as they would improve the accuracy and usefulness of my database.

I consider that IDC Horizontal Reflection and IDC One Hundred Eighty Degree Rotation are also essential additions as they may be used to describe characters and components that would otherwise be practically indescribable. In my database I use a non-conformant syntax for mirrored and rotated characters or components, using U+2194 ↔ to indicate horizontal reflection (e.g.  ↔ 臣臣 for U+81E6 𠄎) and U+21B7 ↷ to indicate rotation by 180° (e.g.  ↷ 或↷或 for U+22A0B 𠄎). It would be very desirable to use a conformant syntax for such IDS sequences, which could only be achieved by encoding the two proposed unary operators. For the Unicode v. 15.0 version of my database, there are fifteen examples of IDS with ↔, and fourteen examples of IDS with ↷.

The IDC Stroke Subtraction operator is a very useful operator in those cases where an ideograph is the same as some other character but missing one stroke. It is usually possible to describe such a character without a subtraction operator, but using a subtraction operator usually makes the IDS sequence more compact and easier for the user to understand. In my

database I usually provide two IDS sequences for ideographs which could benefit from a subtraction operator: a conformant sequence (e.g. 𠄎丘 / for U+4E52 兵), and a non-conformant sequence using U+2296 ⊖ as a temporary subtraction operator (e.g. ⊖兵 \ for U+4E52 兵). If the subtraction operator is encoded, then I only need to provide the sequence with the subtraction operator in those cases where the stroke to be subtracted is unambiguous. In some cases, the subtraction operator would be ambiguous (e.g. U+382A 正 could be described as ⊖正 | , but as there are two vertical strokes in 正 the subtraction operator is ambiguous in this case). For the Unicode v. 15.0 version of my database, there are 107 examples of IDS with ⊖.

Tangut Usage

Tangut IDS sequence would also benefit from the encoding of a surround from lower right operator. For example, U+170DB 𐞇 is currently described in my Tangut character database as 𐞇 𐞇 𐞇, but this is misleading as the 𐞇 component is not placed on the left side of 𐞇 (cf. U+170D8 𐞇 which is correctly described as 𐞇 𐞇 𐞇).

Yi Usage


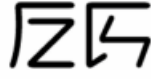

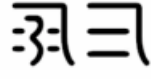





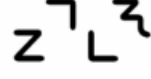
L2/08-193 (WG2 N3288) “Preliminary Proposal to encode Classical Yi characters” submitted by China NB in 2007 shows a table of sixteen description characters that were used by the authors of the proposal to help order the proposed set of characters (see Section 8 “Ideograph Structure Description”; also reproduced in the Appendix below). Although the proposal does not include any actual IDS sequences for Classical Yi characters, it is clear from the statement on page 10 (“The ordering of the proposed new set ought to be in accordance with the order of Classical Yi character structure description, and the ordering of Classical Yi strokes”) that these description characters must have been used to create IDS sequences for the proposed character set. When the Classical Yi proposal is revised and resubmitted such Yi IDS sequences will be invaluable for helping evaluate the proposed character repertoire and ordering. The ideographic description characters used for Classical Yi include four that are unencoded, including IDC Surround from Right. This is further evidence in support of encoding the IDC Surround from Right operator. The other three unencoded operators can be considered for encoding at a future date if further evidence for their use is provided.

Conclusion

The new ideographic description characters proposed in L2/18-012 and L2/21-118R are required for use in IDS databases for CJK ideographs, Tangut ideographs, and Yi syllables, and the sooner they are encoded the sooner I can convert non-conformant IDS sequences to conformant IDS sequences in my CJK IDS database. The first proposal for these new ideographic description characters was submitted almost five years ago (L2/17-386), and I do not know of any published objections to encoding them since then. The authors of these documents are all established IRG experts, so it is clear that IRG experts believe that these operators would be beneficial to IRG work. Therefore, I recommend that the UTC accept these five characters for encoding in Unicode version 16.0.

Appendix: Ideograph Structure Description (WG2 N3288)

Number	DESCRIPTION	DC	IDS Acronym & Syntax	DC	IDS Presentation e.g.
1.	INDEPENDENT	1	IDC-IDP D ₁		
2.	LEFT TO RIGHT	2	IDC-LTR D ₁ D ₂		
3.	ABOVE TO BELOW	2	IDC-ATB D ₁ D ₂		
4.	LEFT TO MIDDLE AND RIGHT	3	IDC-LMR D ₁ D ₂ D ₃		
5.	ABOVE TO MIDDLE AND BELOW	3	IDC-AMB D ₁ D ₂ D ₃		
6.	FULL SURROUND	2	IDC-FSD D ₁ D ₂		
7.	SURROUND FROM ABOVE	2	IDC-SAV D ₁ D ₂		
8.	SURROUND FROM BELOW	2	IDC-SBL D ₁ D ₂		
9.	SURROUND FROM LEFT	2	IDC-SLT D ₁ D ₂		
10.	SURROUND FROM RIGHT	2	IDC-SLT D ₁ D ₂		

Number	DESCRIPTION	DC	IDS Acronym & Syntax	DC	IDS Presentation e.g.
11.	SURROUND FROM UPPER LEFT	2	IDC-SUL D ₁ D ₂		
12.	SURROUND FROM UPPER RIGHT	2	IDC-SUR D ₁ D ₂		
13.	SURROUND FROM LOWER LEFT	2	IDC-SLL D ₁ D ₂		
14.	DIAGONAL FROM LEFT TO RIGHT	2	IDC-OVL D ₁ D ₂		
15.	DIAGONAL FROM RIGHT TO LEFT	2	IDC-DRL D ₁ D ₂		
16.	OVERLAID	n	IDC-OVL D _n	