

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/IRGN2472

Date: 2021-3-12

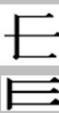






Doc. Type:	Member body contribution
Title:	Considerations of IRGN2225 and decomposition Strategies for IDS
Source:	TCA
Status:	Input to IRG meeting 56
Action:	For consideration by JTC1/SC2/WG2/IRG
Distribution:	IRG Members and Ideographic Experts
No. of pages:	2
Appendixes	None
Medium:	Electronic

TCA is currently reviewing the CNS11643 Website (全字庫) for IDS, based on the IDS Code rule provided by IRGN2225 , but we find that the specification is not very clear at this stage, so TCA suggests that the IRG experts can focus on the IDS rule for discussion.

According to the results of the review, three phenomena were found, and if there is no consistent principle, TCA are afraid there will be difficulties for the next submission of new sets.

1. The confusion of CJK Supplementary Component representation

Such as “IDS code vs U code”, is it not necessary to use IDS code to represent the glyphs that already have U code? Should the IDS code be removed? be given to its use?

	FS Code	IDS code	Description	Glyph	Example	IDS.TX T NO.	U code
1	H	H3-01	CJK RADICAL PAW FOUR2		Example: bottom of 虐 / 疴, bottom of 虐 / 台 = CDP-8DC1 / UTC-00969 (variant glyph = USAT-105803; may be confusable with P3-01 below)	{60}	30002 - 1.2 UTC-00969 E 30004 - 1.2 SAT-05803 E
2	H	H3-02	CJK COMPONENT C-SIMPLIFIED TRIPLE EARTH		Example: 尧 = CDP-89D9		2D80D 手 64.0 JMJ-059643 支
3	H	H3-05					TCA WS2017 30001 - 1.2 UK-01671 丰
4	Z	Z4-02	CJK COMPONENT RAN		Example: left of 那 = CDP-89FA		2D0C2 刀 18.2 GZ-0602101 刃
5	S	S5-01	CJK COMPONENT C-SIMPLIFIED OVERSIGHT		Example: top of 監 / 覽 = CDPO-81F0 / UTC-02965	{44}	30022 2.4 UK-02965 𠄎
6	S	S4-01	CJK COMPONENT C-SIMPLIFIED LEAD		Example: top of 堅 / 豎 / 緊 / 腎 / 賢	{43}	30021 2.3 UK-02964 収
7	P	P2-03					TCA WS2017 3002A J 4.1 SAT-06022 𠄎

2. Inconsistent disassembly methods

As shown in the table below.

Unicode	Glyph	IDS-1	IDS-2
U+35EF	𪗇	𠄎者叱	^𠄎者𠄎口匕\$(K)
U+3675	壙	𠄎土臺(U+365C)	^𠄎土臺(U+213B2)\$(GT)
U+36AC	姁	𠄎女勻	^𠄎女勻\$(GHT)
U+38F4	𨾏	𠄎彳黃	^𠄎彳黃\$(K)
U+5E42	冪	𠄎冥巾	^𠄎一冥巾\$(GHKT)
U+6811	𣎵	𠄎杈寸	^𠄎木对\$(GT)
U+7638	癩	𠄎疒疒	^𠄎痂肉\$(GHJKT)
U+20AEF	𠄎	𠄎𠄎人	^𠄎𠄎^\$(G)
U+20B1E	𠄎	𠄎𠄎又	^𠄎𠄎又\$(GT)
U+237BF	𣎵	𠄎榘糸	^𠄎木𠄎糸\$(T)
U+2393F	鰾	𠄎魚𠄎	^𠄎鰾欠\$(GT)
U+272F2	𧈧	𠄎一𧈧	^𠄎一北虫\$(G)

3. How to maintain the database?

If, according to IRGN2225, each submitter can add an IDS code of his own, how will all the data be maintained? How to publish the IDS code after adding it? How to make sure the IDS code is the later version?

In view of this, TCA recommend that IRG could set up an IDS group and those who are interested in this topic could join and discuss it together. Because of the inconsistency of each expert's idea and the difference of ideographs of each member body, there will be differences in the disassembly process.

(End of document)

Subject: A Feedback to "IRG N2472, Considerations of IRG N2225 and decomposition Strategies for IDS", TCA

Date: 2021.03.13.

Author: KIM, Kyongsok (Republic of Korea)

Status: Individual Contribution

Relevant documents: IRG N2225 (Henry CHAN and Qin LU), IRG N2472 (TCA)

This is a feedback to IRG N2472, "Considerations of IRG N2225 and decomposition Strategies for IDS" by TCA, 2021-03-12.

Three issues pointed out by TCA will be discussed one by one.

1. What if Supp. Comps. in N2225 are encoded in UCS?

- Although I am not competent enough to evaluate the details regarding this issue, TCA comment seems reasonable.

- If the 7 Supp. Comps. mentioned by TCA are really encoded in UCS (it looks like they are), it seems natural to use UCS cp's instead of Supp. Comps. and to remove those 7 Supp. Comps. from the list of Supp. Comps.

1. The confusion of CJK Supplementary Component representation

Such as "IDS code vs U code", is it not necessary to use IDS code to represent the glyphs that already have U code? Should the IDS code be removed? be given to its use?

FS Code	IDS code	Description	Glyph	Example	IDS.TXT NO.	U code
1	H H3-01	CJK RADICAL PAW FOUR2		Example: bottom of 虐 / 疝, bottom of 虐 / 合 = CDP-8DC1 / UTC-00969 (variant glyph = USAT-05803; may be confusable with P3-01 below)	{60}	30002 - 12 UTC-00969 30004 - 12 SAT-05803
2	H H3-02	CJK COMPONENT C-SIMPLIFIED TRIPLE EARTH		Example: 堯 = CDP-89D9		2D80D 手 840 JMJ-009643
3	H H3-05					TCA WS2017 30001 - 12 UN-01871
4	Z Z4-02	CJK COMPONENT RAN		Example: left of 那 = CDP-89FA		2D0C2 刀 18.2 QZ-0602101
5	S S5-01	CJK COMPONENT C-SIMPLIFIED OVERSIGHT		Example: top of 監 / 覽 = CDPO-81F0 / UTC-02965	{44}	30022 I 24 UK-02965
6	S S4-01	CJK COMPONENT C-SIMPLIFIED LEAD		Example: top of 堅 / 豎 / 緊 / 腎 / 賢	{43}	30021 I 23 UK-02964
7	P P2-03					TCA WS2017 3002A J 41 SAT-08022

2. Different decomposition (disassembly) ?

- It seems that the 12 examples in N2472 can be classified into the following two categories.

2.1 Different but equivalent IDSeS

- If we further decompose 叱 in IDS-1 for U+35EF, IDS-1 will eventually become IDS-2. Therefore, I would say that IDS-1 and IDS-2 are not identical, but equivalent.

- I assume that IDS check program should be able to figure out equivalent IDSeS.

Unicode	Glyph	IDS-1	IDS-2
U+35EF	𪗗	𠄎者叱	^𠄎者𠄎口匕\$(K)

- The same can be said of the following 6 examples too.

U+5E42	冪	𠄎冥巾	^𠄎一冥巾\$(GHKT)
U+6811	𣎵	𠄎杈寸	^𠄎木对\$(GT)
U+7638	瘡	𠄎疒𠄎	^𠄎痂肉\$(GHJKT)
U+237BF	櫛	𠄎櫛糸	^𠄎木𠄎糸\$(T)
U+2393F	鰾	𠄎魚𠄎	^𠄎鰾欠\$(GT)
U+272F2	𧈧	𠄎一𧈧	^𠄎一北虫\$(G)

2.2 Decomposition (disassembly) into similar but different components

- Let's consider U+3675. U+365C and U+213B2 are different but quite similar.

- I assume that IDS check program should be able to figure out similar components.

- I don't know if IRG maintains a list of similar components.

U+3675	壘	𠄎土臺(U+365C)	^𠄎土臺(U+213B2)\$(GT)
U+36AC	姁	𠄎女匀	^𠄎女匀\$(GHT)
U+38F4	𧈧	𠄎彳𧈧	^𠄎彳𧈧\$(K)

U+20AEF	𠄎	㊦ 𠄎人	^㊦ 𠄎^\$(G)
U+20B1E	𠄎	㊦ 𠄎又	^㊦ 𠄎又\$(GT)

2.3 Overall: This issue could be solved by an IDS check program and should not be a problem.

3. How to maintain the list of Supp. Comps.?

3. How to maintain the database?

If, according to IRGN2225, each submitter can add an IDS code of his own, how will all the data be maintained? How to publish the IDS code after adding it? How to make sure the IDS code is the later version?

- TCA mentions that, according to IRG N2225, each submitter can add an IDS code of his own. However, I could not find a statement to that effect in IRG N2225. (If I am wrong, please correct me by indicating the statement(s) to that effect.)

- I already pointed out similar problems in 2.2 and 2.3 of IRG N2464 (part 1). The Subsection titles of 2.2 and 2.3 are shown below. For detailed information, see IRG N2464 (part 1).

2.2 Criteria and procedure to review and accept newly suggested Supp. Comps.

2.3 Supp. Comp. Naming procedure

4. Setting up an IDS group for maintaining IDSes ?

In view of this, TCA recommend that IRG could set up an IDS group and those who are interested in this topic could join and discuss it together. Because of the inconsistency of each expert's idea and the difference of ideographs of each member body, there will be differences in the disassembly process.

- TCA suggests to form an "IDS group" which will maintain IDS-related issues such as IDSes, a list of Supp. Comps., a list of similar characters, etc.

- Personally, I totally agree with TCA suggestion. The current situation regarding IDS is somewhat confusing as I mentioned in IRG N2464 (part 1) and N2462_2 (part 2) and, therefore, it is suggested that IRG specify all IDS-related issues in PnP.

* * *