

ISO/IEC JTC 1/SC2/WG2/IRG Ideographic Rapporteur Group

Source / Contribution Identifier: Japan

Title: Japan's Opinion Regarding U Set

Status: Member body contribution

Short Description:

Japan requests to remove U set characters from CJK C1.

Proposed Conclusion / Requested Action:

To be accepted by IRG member bodies and remove U characters from the next C1 draft.

Arguments / Text of Contribution: See below.

Japan's Opinion Regarding U Set

1. What IRG Was Doing

IRG started its effort to add more ideographs than CJK UNIFIED IDEOGRAPHS EXTENSION B in the past century. The ways to establish the next extension set, usually referred to by CJK C1, was so tough that IRG spent more years than expected.

During its meeting #25 on Berkeley, IRG agreed to reduce the size of the C1 candidate to ensure quality. IRG also agreed that "Characters with no question after review will be submitted to WG2 as CJK_C1." (See IRG N1179.) Japan understands this sentence as follows: IRG drops all questioned characters from its C1 candidate to get the refined set of never-be-questioned characters.

After the meeting #25, IRG editors reviewed v60 and submitted comments against several characters. Then v61 came, and the editors reviewed it again. IRG also employed IDS-based machine-supported method to detect suspicious characters easily. The purpose of this particular review cycle after the meeting #25 was, in Japan's understanding, to filter all suspicious characters out from C1 candidate. The idea was as follows: By keep removing all suspicious characters, we will get absolutely stable, error-free set of characters.

Based on this understanding, Japan did not spend much time for reviewing G, D, nor U sets, since Japan considered it was not urgent. Japan concentrated on reviewing the M set. There must be very few, if any, IRG editors who actually reviewed U set before the Hue meeting.

2. What IRG Is About To Do

After the Hue meeting, IRG Chief Editor stated his intension to put the U set back in the C1 submission to WG2. He wrote in a mail:

Since the U set was NOT discussed at IRG#26 in Hue, the characters in U set are ACQUIESCENTLY added to the CJK_C1 final version.

Japan is so confused by this statement. It suggests that IRG should put all characters that was once questioned and never be discussed back to the candidate. It sounds very unreasonable. Japan cannot accept such idea.

U set contains a lot of mechanically detected *false* duplicates, due to the limitation of the algorithm. However, Japan also believes that at least some of them are real duplicates that should never be in UCS.

Just for an example, please look at a case regarding C1 number 06027 listed in IRGN1194_CJK_C1_V61U.pdf:

垺	垺	垺
06027	05829	U-00021377

It is apparent that C1 06027 and existing U+5829 is same. It is true that IRG has never discussed on this particular character. It is in the U set. Japan is uncomfortable with the IRG Chief Editor's judgement.

3. What IRG Should Do

Japan still believes that the most important thing at this moment regarding C1 is to make C1 stable and error-free. If WG2 considers C1 has no enough quality, it will not be a part of the Amendment 3. IRG needs to do the following two: (1) to *make* the C1 submission eligible for the International Standard, and (2) to prove it.

IRG should keep removing all questioned characters from C1, postponing the discussion about the questioned characters to C2 timeframe. IRG should exclude all U characters from C1.

4. What Japan Will Do

Japan will object against any idea to make an unstable repertoire a part of the 10646. Japan is sure that the current M set is stable and is eligible for the International Standard. However, U set is not. If IRG submitted C1 proposal including U set, Most of the WG2 members including Japan will not support it.

Japan will keep reviewing latest C1 candidates. Japan wants to know other member bodies' opinion regarding this issue.

(END OF DOCUMENT)