

ISO/IEC JTC1/SC2/WG2 N4230 IRG N1843 L2/11-442R2

Universal Multiple-Octet Coded Character Set International Organization for Standardization

Doc Type: Working Group Document

Title: Proposal to establish a CJK Unified Ideographs “Urgently Needed Characters” process

Authors: Dr. Ken Lunde (小林劍) & John Jenkins (井作恆)

Source: The Unicode Consortium & US National Body

Status: National Body Contribution

Action: For consideration by JTC1/SC2/WG2

Date: 2012-02-10 (originally submitted to the UTC on 2011-11-17, and revised on 2012-02-07)

Background

The process of standardizing repertoires of CJK Unified Ideographs is long and cumbersome, and is almost always measured in years. This is primarily because the typical CJK Unified Ideograph repertoire includes thousands or tens of thousands of characters, and thus requires several rounds of review and discussion before it can be standardized. Extension E, for example, began as Extension D, whose national body submissions were accepted in early 2007, and incorporated characters that were deferred from Extension C. Extension E is currently at the final stages of standardization.

To address this particular process shortcoming, the IRG established a one-time UNC (*Urgently Needed Characters*) repertoire as one of the IRG 29 Resolutions (see IRG N1377, specifically Resolution IRG M29.5), which eventually became CJK Unified Ideographs Extension D, with 222 CJK Unified Ideographs (U+2B740 through U+2B81D), and which was subsequently included in ISO/IEC 10646:2003 Amendment 8, and in Unicode Version 6.0.

Without a formalized UNC-like process in place, which would serve as a parallel pipeline for smaller repertoires of urgently-needed CJK Unified Ideographs, it is extraordinarily difficult for national bodies to standardize smaller sets of urgently-needed CJK Unified Ideographs in a timely manner.

Proposal

We hereby propose that a formalized CJK Unified Ideographs “Urgently Needed Characters” process be established. In order for such a process to function effectively, the standardization timeline must be *short*, and the repertoires must be *small*.

To address the timeline issue, in that it must be short, we recommend that each repertoire be standardized on a regular basis, such as every one or two years, and subsequently included in an amendment or new edition of the standard.

To address the size issue of the repertoire, in that it must be small, we recommend that each national body be limited to 25 character submissions per repertoire, in terms of what is originally submitted, not what actually becomes standardized. In other words, national bodies should not be allowed to submit additional characters to replace ones that are removed for reasons such as lack of sufficient evidence or unification. Any exceptions to this 25-character limit must be approved by the IRG, and the total number of submissions per repertoire for all national bodies combined should not exceed 200. With a “25 characters per national body per repertoire” submission limit in place, the effort needed to bring each repertoire to a state in which it can be standardized becomes much more manageable than for the larger repertoires.

We request that WG2 direct the IRG to formalize an “Urgently Needed Characters” process—as a parallel pipeline to that intended for standard CJK Unified Ideograph submissions—per the above recommendations, and that the IRG be further tasked with documenting the detailed procedures in their Principles and Procedures document.

Title: Comments on "Urgently Needed Characters" Process

Re: WG 2 N4230, IRG N1843

Date: 2012-6-10

Source: Japan

Status: National Body Contribution

Requested Action: To discuss on IRG #38

Abstract

In this contribution, Japan states its position against the proposed UNC process in WG 2 N4230, IRG N1843, and proposes an alternative solution: *isolation* process. The isolation process will speed the IRG process up in general, and it will especially work well for very small set of proposed additions such as those the UNC process was intended to facilitate.

1. Japanese position on the original proposal (WG 2 N4230, IRG N1843)

Japan agrees on the points that the current IRG process is long and cumbersome, that it is sometimes difficult to publish an additional set of unified ideographs in timely manner under the current process, and that we may sometimes have some urgent requirements to standardize some additional unified ideographs very quickly. So, Japan supports the idea to have some solution to the said issue.

However, Japan doesn't think the proposed UNC process works well for two reasons: (1) the 25 characters limit proposed in N1843 appears arbitrary and will make some trouble to use the process, and (2) it is complicated to define what case will be considered urgent as a part of the rules and it will make another trouble when a member wants to use the process..

As explained in N1843 itself, IRG and its parent WG 2 have some experiences on handling such urgent requirements (not just CJK-D but HKSCS additions for example can be considered as a similar case.) Japan believes that, if a member has an urgent need in a future, the best way will be to raise the issue to IRG or WG 2 so that we can handle it in a case-by-case manner as we did in the past. Japan doesn't think IRG needs

to establish a new rule.

2. Analysis of the current IRG process

As previously said, Japan agrees on the point that the current IRG process takes too long time. N1843 appears attributing it to the large number of candidate characters, but it is not so simple.

Past experiences show that members' proposed sets varied greatly. While one proposal contained tens of thousands of characters, another proposal contained just tens. While one proposal was stable enough and very few critical comments such as unification errors were made, another proposal contained a lot of critical problems, for example, many unification errors or changes of glyph shapes in a middle of review processes. If all the proposed candidate ideographs were very stable, just a few review cycles with some small number of corrections would convince member editors the proposed set under review is just fine for inclusion in the International Standard.

The current IRG process collects proposed ideographs from all members, making a single set of candidates for review. The set is divided into several subsets, i.e., M set, D set, ..., but those sets contain ideographs from all proposed sets. If someone submitted a set of very instable candidates, most of the review efforts and editorial discussion would be consumed on those instable ones, and other stable candidates would be just there undiscussed (because they were stable, very few comments were made against) along with the instable candidates.

In other words, some problematic candidate ideographs delay the review process of whole other candidates. Japan considered that is what happens in IRG.

3. An alternative proposed process: isolation process

Based on the above analysis, Japan proposes an alternative process. We call it an *isolation* process in this contribution.

Under the isolated process, each member body submits its own set of proposed additions to IRG as before. However, the chief editor doesn't merge them into a single set. Instead, the chief editor keeps those proposals as a set of separate sets of candidate ideographs.

(An isolation.)

The set of sets will go out for members' internal reviews and the submitted comments are consolidated as before. The IRG will run several review cycles as before. The only difference is that we don't merge proposed candidates from other member bodies. We will keep them isolated throughout the review process.

At some point, it will be apparent for member editors that some proposed sets are good, meaning no many critical problems were found, and some other sets are bad, meaning a lot of reasonable comments were made against. During an editorial meeting, the editors discuss which sets are good and which sets are bad, and then the chief editor takes only the sets considered good and merges them into a single ordered list for submission to WG 2.

Isolation process doesn't limit the size (number of candidate characters) of a proposal from a member. However, a submitter is free to do so. If a member body limits its own proposal to a set of less than 25 candidates and reviews very carefully before submitting to IRG, the proposal will be stable enough when it first comes to IRG. Under the isolation process, such a submission will go through only a few review cycles and will go to WG 2 very quickly, even if during the same time frame someone else submitted a large set of instable candidates.

The current IRG process is synchronous and serial, i.e., all member bodies submit their own proposal at a same time, as we are about to do for CJK-F, and no further full submission will be accepted unless the whole CJK-F is done. The isolation process should be asynchronous, i.e., as soon as a proposed set from a member body is considered good enough to be submitted to WG 2, the member body should be allowed to submit the next proposed set, even if proposals from other member bodies were not finalized yet. Otherwise years would be just wasted for that particular next proposal as before. Moreover, it seems better that the isolation process allows parallel submission, i.e., a member body is allowed to submit its own next proposal before its own previous proposal is finalized. It is to facilitate some urgent needs just raised after its previous submission.

4. Discussions on the isolation process

The core difficulty of the isolation process is the cross-unification of candidate ideographs in sets from different member bodies. Of course member editors can make a comment on such cross unification whenever they find ones, we want to make sure all such unifiable ideographs are detected before a proposal is sent to WG 2. A special review cycle, just before WG 2 submission, dedicated for cross-unification, may be needed to solve the issue.

When some good sets are submitted to WG 2, they are considered to be a part of "unified ideographs already in the international standard". The member editors need to check duplicates against them as always.

In the UNC process proposed in N1843, the number of candidates included in an urgent proposal is limited to 25 characters. If a member had an urgent needs to add 26 ideographs, it would not be covered by the UNC process. However, the UNC process requires some limits anyway, because it doesn't work if we allow a member to propose some thousands of ideographs as an urgent submission.

In the isolation process, we don't need any set limit. A smaller proposal has a better chance of going out to WG 2 in a shorter review periods, but it is up to the member body's decision how many candidate ideographs to be submitted at once, considering their own urgency.

Although it seems better to allow parallel submissions from a member body, we may need some limits on the number of *sets* to be submitted at a same time. Or, we could limit the member editors review process, e.g., only one set from a member body goes for a review in a single review period and other sets from the same body are suspended. We need more discussion on this point.

5. Conclusion

Japan proposed isolation process in this contribution. Japan considers substituting the current process with the isolation process is a better alternative than adding the UNC process to supplement the current process, because Japan believes the isolation process speeds the whole IRG process up, solving "urgent needs" requirements at the same time.