Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Background

The process of standardizing smaller repertoires of urgently-needed CJK Unified Ideographs is long and cumbersome, and is measured in years. This is primarily because the typical CJK Unified Ideograph "Extension" includes 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 included characters that were deferred from Extension C. Extension E is 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 was included 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 a national body to standardize smaller sets of urgently-needed CJK Unified Ideographs in a timely manner.

Proposal

In order for a formalized UNC-like process to be effective, the standardization timeline must be *short*, and the repertoires must be *small*. We recommend that a 2,048-character block, tentatively named *CJK Unified Ideographs Annual Repertoires* (AR), be established for this particular purpose. We further recommend that U+2F000 through U+2F7FF be considered as the location of this block, because it is less likely to interfere with the standardization of the parallel, and larger, CJK Unified Ideograph repertoires (aka, Extensions).

To address the timeline issue, in that it must be short, we recommend that a subrepertoire be standardized each year, then appended to this block. Each annual subrepertoire can be identified by its code point range within the block and its year of standardization.

To address the size issue of the repertoire, we recommend that each national body be limited to 25 character submissions per annual subrepertoire, 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 replaces ones that are removed for reasons such as lack of evidence or unification. Any exceptions to this 25-character limit must be approved by the IRG, and the total number of submissions per annual subrepertoire for all national bodies combined should not exceed 200. The two annual IRG meetings should be sufficient for performing two rounds of review, especially when considering the smaller size of each annual subrepertoire. With a "25 characters per national body per year" submission limit in place, this block is expected to serve this need for at least ten years, possibly longer.

We request that WG2 directs the IRG to formalize a UNC-like process per the above recommendations, and that the IRG be further tasked with documenting the detailed procedures in their Principles and Procedures document.