

## **Universal Multiple-Octet Coded Character Set International Organization for Standardization**

**Doc Type:** ISO/IEC JTC1/SC2/WG2/IRG

**Title:** Continued National Standards Development & Horizontal Extensions

**Source:** Dr. Ken Lunde (小林剣), Adobe Systems Incorporated

**Status:** Member Body Contribution

**Action:** For consideration by IRG member bodies

**Date:** 2013-10-25

The purpose of this brief paper is to describe the importance of the continued development of each member body's respective national standards, coupled with the related prudence of submitting horizontal extensions that reference characters in them.

### **The Problem**

The number of CJK Unified Ideographs in ISO/IEC 10646—when Extension E is factored in—now exceeds 80,000 characters (the exact figure is 80,379). The sheer number of characters poses the following two real-world problems to implementers of the standard, particularly font developers:

1. A single font resource cannot include more than 64K glyphs (though there is now a standard, ISO/IEC 14496-28:2012, that provides a work-around via Composite Font Representation objects, but it is not yet universally supported).
2. For reasons of practicality, most fonts—including Pan-CJK ones—are intended for use in one or more regions, locales, or languages, and thus there is no benefit in including in them the glyphs for all 80,379 CJK Unified Ideographs.

Therefore, the fundamental problem is in determining which specific CJK Unified Ideographs require glyphs in fonts that are intended for use in a particular region, locale, or language.

### **The Proposed Solution**

When a new CJK Unified Ideograph is proposed by a member body—and ultimately accepted into the standard—it includes a reference that can be used to tie the character to a particular region, locale, or language via a *kIRG\_{G,T,H,M,J,K,KP,V,U}Source* reference.

However, it is not uncommon for an existing CJK Unified Ideograph to later be deemed useful for an additional region, locale, or language. This is performed by submitting a “horizontal extension” whose effect is to “tag,” “flag,” or otherwise identify a CJK Unified Ideograph as being useful for an additional region, locale, or language.

When submitting horizontal extensions, a unique *kIRG\_{G,T,H,M,J,K,KP,V,U}Source* reference is necessary. In addition, these unique references are useful to implementers *only* if they refer to an accessible standard, such as published national standards. A published national standard also serves as a prototypical glyph reference, because the glyphs in the Code Charts do not always reflect the most current prototypical glyphs for a particular region, locale, or language. The prototypical glyph changes that took place in JIS X 0213:2004, an updated national standard from Japan, represent an example. In other words, the ISO/IEC 10646 standard does not obsolete the need to continue to develop new or update existing national standards.

It deserves to be pointed out that the Korea national body has proposed such horizontal extensions, which serves as an example to other member bodies.

### **Examples**

The first example is Hong Kong SCS-2008, all of whose hanzi—including those in its Big Five foundation—now have a *kIRG\_HSource* reference. However, any new hanzi that are deemed useful for Hong Kong must now simply be in Unicode, but because they are not being proposed in a horizontal extension with *kIRG\_HSource* references, there is nothing that ties these CJK Unified Ideographs to Hong Kong use. This suggests that Hong Kong SCS development should not have ceased.

The second example is specific to the CJK Unified Ideograph U+20BB7 (吉). Its usefulness in Japan cannot be denied by anyone who uses Japanese, yet this kanji lacks a *kIRG\_JSource* reference. This character thus deserves a *kIRG\_JSource* reference.