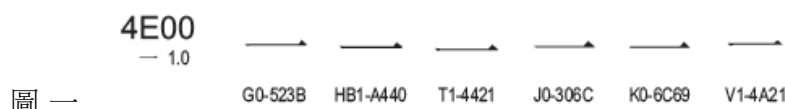


討論在 ISO/IEC 10646 國際編碼標準字集  
以橫向擴展(Horizontal Extension)方式  
加入香港字源資料及相關字形的建議

背景

ISO/IEC 10646 國際編碼標準字集收錄不同來源的表意文字（漢字）並加以編碼。字集的編碼表除載列字符編碼外，還顯示其字源資料及相關字形，例如圖一的“一”字(U+4E00)顯示六項字源資料：G0-523B、HB1-A440、T1-4421、J0-306C、K0-6C69、V1-4A21，當中 G、H、T、J、K、V 是來源標準類別代表字母。



2. 《香港增補字符集》字符的字源資料以“H”註明，例如圖二的“埗”字(U+57D7)，來源標準類別代表字母為“H”，即該字符的來源標準是《香港增補字符集》。



3. 國際標準化組織轄下表意文字小組在 2014 年 11 月舉行的第 43 次會議上，提出並討論橫向擴展(Horizontal Extension)的建議，讓表意文字小組成員為已納入 ISO /IEC 10646 國際編碼標準字集的字符，補充字源資料和字形。有關討論文件見附件一。

4. 在 2008 年 3 月 31 日至 2015 年 2 月 28 日期間，中文界面諮詢委員會（中諮會）共審核七百多個字符增收申請，除七個字符缺乏充分的佐證資料外，其餘字符均已收錄於 ISO/IEC 10646 國際編碼標準字集，因此不獲納入《香港增補字符集》，

亦因此這些字及其字源資料和字形未能沿用“已收納在《香港增補字符集》”的依據申請加入 ISO/IEC 10646 國際編碼標準。討論文件以此為例，說明橫向擴展有利把這些香港字源資料和字形加入 ISO/IEC 10646 國際編碼標準字集。

## 建議

5. 橫向擴展的建議讓香港能夠因應需要，為 ISO/IEC 10646 國際編碼標準字集內適用於香港的漢字，加入香港字形，利便軟件開發商開發支援本地要求的字型產品。為確保橫向擴展反映本地的用字需要，我們建議訂立機制，進一步審核不獲納入《香港增補字符集》的字符，確保有關字形符合香港慣常寫法。附件二列出考慮要點，供工作小組成員參考。

6. 秘書處會收集工作小組成員對上述建議的意見，整理後提交中諮會審議。

## 徵詢意見

7. 請工作小組成員發表意見。

2015 年 3 月

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

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

**Title:** Horizontal Extension Proposal

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

**Status:** Individual Contribution

**Action:** For consideration by the IRG

**Date:** 2014-11-24

The purpose of this brief document is to propose a method for submitting horizontal extensions in a way that does not involve the publishing of new national standards and assigning new character codes, which are processes that can be associated with bureaucratic hurdles, logistical overhead, and other obstacles.

Background documents include [IRG N1964](#) (aka [L2/13-192](#)) from IRG #41 (*Continued National Standards Development & Horizontal Extensions*) and [IRG N2050](#) from IRG #43 (*Recommendations of IRG Meeting #43*), specifically referring to Recommendation IRG M43.2 (*Horizontal Extension of H Source for existing characters*).

### The Problem

The number of CJK Unified Ideographs in Unicode and 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 glyphs for all 80,379 CJK Unified Ideographs.

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

### Horizontal Extensions

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

However, it is not uncommon for an existing CJK Unified Ideograph to later be deemed useful for additional regions, locales, or languages. 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. These unique source references correspond to code points in published national standards.

## The Proposed Solution

As a way to eliminate the overhead involved in publishing new national standards for the purpose of horizontal extensions, I propose the use of the “U” identifier immediately following the source prefix, and followed by a zero-padded five-digit hexadecimal character code of the UCS code point to which it corresponds.

This solution has the following benefits:

- Explicitly indicates that no corresponding national standard exists, so font developers need not bother searching for one.
- Eliminates the need to publish new national standards to accommodate horizontal extensions.
- Explicitly indicates that the character has been deemed useful for a particular region, locale, or language.
- Provides the representative glyph for the region, locale, or language.
- Promotes the code charts as the most up-to-date and authoritative reference for representative glyphs across all regions, locales, and languages.

Others have suggested putting this information into an appendix or other document, but doing so is less useful, and easily overlooked. Having the information closely tied to the character, such as in the multicolumn code charts and as the *kIRG\_{G,T,H,M,J,K,KP,V,U}Source* reference, is more convenient for developers and less likely to be overlooked.

Of course, horizontal extensions that use this method would still require that representative glyphs be provided for use in the corresponding column of the code charts.

## Specific Use Cases

Hong Kong SAR is the primary beneficiary of this proposal because it has thus far identified several hundred CJK Unified Ideographs that have been deemed useful for Hong Kong, yet they lack a *kIRG\_HSource* reference. Font developers have no convenient way of knowing what these additional characters are.

ROK (South Korea) submits horizontal extensions on a somewhat regular basis, which are tied to obscure KS standards that seem to exist for the sole purpose of submitting horizontal extensions.

Japan would be in a position to replace the existing “JA-XXXX” source references—whose corresponding standard cannot be found—with “JU-HHHHH” ones, and could also propose “JU-20BB7” as the *kIRG-JSource* reference for U+20BB7 (吉).

有關橫向擴展(Horizontal Extension)的考慮要點

香港是否應該進行橫向擴展取決於是否有一些已在 ISO/IEC 10646 編碼的非 H 列漢字在香港有較廣泛的使用，而需要造字商配合支持。

1. 可否不採用表意文字小組的橫向擴展建議，而沿用現行機制，待《香港增補字符集》更新後，才向國際標準化組織申請加入香港特區的字符來源和字形？

現行《香港增補字符集》字符增收原則訂明，凡在 ISO/IEC 10646 國際編碼標準字集及《香港增補字符集》內已有可視為等同之字符者，概不增收。也就是說香港目前只做 ISO/IEC 10646 的縱向增收工作。目前提出增收申請的字符，絕大部分在 ISO /IEC 10646 國際編碼標準字集內已有可視為等同的字符，因此現行機制不會批准把這些字符納入《香港增補字符集》，這些字符亦未能沿用“已收納在《香港增補字符集》”的依據申請加入”ISO/IEC 10646 國際編碼標準字集內。

2. 香港特區以橫向擴展方式建立的字集與《香港增補字符集》的功能有何異同？

《香港增補字符集》和日後建立的香港特區橫向擴展字集都是香港特區在中文電子通訊上使用的字符。《香港增補字符集》旨在向國際標準化組織提交 ISO /IEC 10646 國際編碼標準字集尚未收錄的本地用字；香港特區橫向擴展字集則是為已納入 ISO /IEC 10646 國際編碼標準字集的字符，補充香港特區字源資料和字形。兩者在功能上互相補足。

3. 在 ISO/IEC 10646 國際編碼標準字集中，香港特區以橫向擴展方式加入的字形和《香港增補字符集》字符會否分別註明？

《香港增補字符集》字符是以“H” + “4 個 16 進制數字”註明。根據表意文字小組有關建議，以橫向擴展方式加入的香港特區字形，可採用以下字源資料：“U” + “H” + “5 個 16 進制的 ISO/IEC 10646 碼位”。以字符(U+4F06)為例，橫向擴展加入香港特區字形後，該字的字源資料將會加入“UH-04F06”。

4. 為何不直接把所有不獲納入《香港增補字符集》的字符，以橫向擴展方式加入 ISO /IEC 10646 國際編碼標準字集？

不獲納入《香港增補字符集》的字符，部分缺乏充分的佐證資料，部分沒有中文輸入法支援，部分則未必符合香港慣常寫法，所以不宜未經進一步審核便以橫向擴展方式加入 ISO /IEC 10646 國際編碼標準字集。

5. 如何甄別合適的香港用字，以橫向擴展方式加入 ISO /IEC 10646 國際編碼標準字集？是否需要全數審核七百多個不獲納入《香港增補字符集》的字符？

這有待確立香港特區的橫向擴展字符審議機制後才能準確評估。現把 2008 年 3 月 31 日至 2015 年 2 月 28 日不獲納入《香港增補字符集》字符的分類資料表列如下，以供參考。

字符分類	數目
已收錄於 ISO/IEC 10646 國際編碼標準字集但不屬大五碼或《香港增補字符集》的字符	242
已收錄於 ISO/IEC 10646 國際編碼標準字集並屬大五碼的字符	424
已收錄於 ISO/IEC 10646 國際編碼標準字集並屬《香港增補字符集》的字符	114
缺乏充分佐證資料的字符	7
總數：	787