

Universal Multiple-Octet Coded Character Set
International Organization for Standardization

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

Title: Proposal to remove the UCS2003 representative glyphs from the Extension B code charts

Source: Ken Lunde (Adobe) & Henry Chan

Status: Individual Contribution

Action: For consideration by IRG & UTC

Date: 2018-02-22

Speaking in terms of Unicode versions, the CJK Unified Ideographs “Extension B” code charts originally included a single representative glyph for each of its 42,711 code points, meaning from [Version 3.1](#) (2001-03), and continued to use the same one-per-code-point representative glyphs through [Version 5.1](#) (2008-03). The multiple-column Extension B code charts were introduced in [Version 5.2](#) (2009-10), and the so-called “UCS2003” representative glyphs—intended to preserve the original one-per-code-point representative glyphs—were introduced in [Version 6.1](#) (2012-01), and continue to be used in the Unicode and ISO/IEC 10646 Extension B code charts.

Proposal: *Remove the UCS2003 representative glyphs from the Extension B code charts starting from the next version of both standards, Unicode and ISO/IEC 10646, because they have simply outlived their usefulness.*

The following five points are justifications for taking this prudent action:

1. **Static & Unmaintained:** There are many errors among the UCS2003 representative glyphs, and more continue to be found, but because they are intended to preserve history, they cannot be corrected.
2. **No Practical Use:** The purpose of the multiple-column code charts is to provide a representative glyph for regions that use a particular ideograph. The UCS2003 representative glyphs are associated with no particular region, and therefore have no practical use for experts or developers.
3. **Misleading:** The very first representative glyph that is shown in the Extension B code charts is the UCS2003 one, which can lead experts and developers to believe that it is somehow more “representative” than the one or more region-specific representative glyphs that follow. This misunderstanding may have led to the propagation of UCS2003 errors, such as in online dictionaries and in shipping fonts, and there are even traces of UCS2003 errors in the GB 18030-2005 standard. U+20BC7 𠄗 and U+25426 𠄗 that are shown below serve as excellent examples of the confusion that is caused by the UCS2003 representative glyphs, because both are ripe for disunification as 𠄗𠄗己/𠄗𠄗巳 and 𠄗石己/𠄗石巳, respectively, per [IRG N2240](#):

20BC7	𠄗	𠄗	25426	𠄗	𠄗
𠄗 30.3			石 112.3		
UCS2003		V0-3069	UCS2003		V0-3F27

4. **Unnecessary Reporting of Errors:** Known UCS2003 errors are recorded in ISO/IEC 10646 Annex P, *Additional information on CJK Unified ideographs* (informative). When an expert or developer discovers a UCS2003 error, Annex P may not be consulted, which may result in the same error being reported over and over again.
5. **Preserving History:** Even if the static UCS2003 representative glyphs are removed from the Extension B code charts for Unicode Version 11.0, they can still be easily referenced in the Extension B code charts for Versions 6.1, 6.2, 6.3, 7.0, 8.0, 9.0, and 10.0, and in the third, fourth, and fifth editions of ISO/IEC 10646. Alternatively, a separate—and more compact—Extension B code chart that includes only the UCS2003 representative glyphs—attached to which is very strong language that clearly states that its content is static, and includes errors, only some of which are known—could be produced.

That is all.