L2/20-080

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

Author: Ken Lunde

Date: 2020-03-16 (replaces L2/18-063)

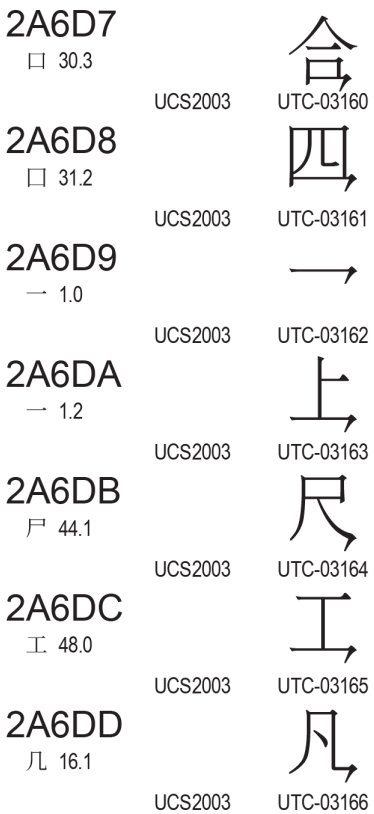
This is a two-part proposal that seeks to remove the UCS2003 representative glyphs from the CJK Unified Ideographs Extension B code charts.

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: *1) Remove the UCS2003 representative glyphs from the Extension B code charts starting from Unicode Version 14.0, because they have simply outlived their usefulness; and 2) generate a separate archival Extension B code chart that includes only the UCS2003 representative glyphs.*

The following six points serve as justifications for accepting this proposal:

1. **Static**: Unicode Version 13.0 introduced seven new CJK Unified Ideographs that were appended to Extension B, U+2A6D7 through U+2A6DD, which will forever lack UCS2003 representative glyphs, and may therefore appear to be an error:



2. **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.
3. **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 nor developers.

4. **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 𠯿 and U+25426 𥐦 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 ⿰口己/⿰口已 and ⿰石己/⿰石已, respectively, per IRG N2240:

5. **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.
6. **Preserving History**: Even if the static UCS2003 representative glyphs are removed from the Extension B code charts for Unicode Version 14.0, they can still be easily referenced in the Extension B code charts for Versions 6.1 through 13.0, and in the third through sixth editions of ISO/IEC 10646. As proposed in this document, a separate Extension B code chart is produced 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:

   This archival Extension B code chart includes only the UCS2003 representative glyphs that are considered historical from when its code chart included a single representative glyph for each of its 42,711 code points, which spanned Unicode Versions 3.1 through 5.1. Its content is static and includes errors, known and unknown. For more accurate and up-to-date representative glyphs, please reference the current Extension B code chart.

That is all.