

ISO/IEC JTC1/SC2/WG2/IRG N1736

Universal Multiple-Octet Coded Character Set
International Organization for Standardization

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

Title: Add Big Five mappings/glyphs to the “H” Source for the Unihan Database and Code Charts

Source: Dr. Ken Lunde, Adobe Systems Incorporated (lunde@adobe.com)

Status: Individual Contribution

Action: For consideration by the IRG

Date: 2010-10-29

Background

Hong Kong SCS, whose current version is Hong Kong SCS-2008, is a character set standard that is based on Big Five, and defines an extension to Big Five that is specifically designed to satisfy the needs of Hong Kong.* The current Unihan Database and Code Charts do not reflect the Big Five portions of Hong Kong SCS-2008, meaning that the inclusion of Big Five proper is implicit, not explicit.

This proposal recommends that the Big Five proper mappings be added to the “kIRG_HSource” field of the Unihan Database, and that the representative glyphs for Big Five proper be added to the appropriate Code Charts.

Unihan Database Additions

In order to distinguish sources that correspond to Big Five proper and those that correspond to Hong Kong SCS-2008, the “H” prefix for the “kIRG_HSource” mappings needs to be modified to reflect this. I propose the following four prefixes:

H0 (10 ideographs among Big Five Symbols)

H1 (Big Five Level 1; 5,401 ideographs)

H2 (Big Five Level 2; 7,652 ideographs)

H3 (Hong Kong SCS-2008; 4,579 ideographs)

Note that all current instances of the “H” prefix simply need to be changed to “H3” to effect this change. The “H0” through “H2” entries need to be added. The Unihan Database changes (for “H3”) and additions (for “H0” through “H2”) are attached to this document: *hkscs2008-big5-h0-2.txt* and *hkscs2008-h3.txt*.

In addition to updating the Unihan Database itself, UAX #38 (*Unicode Han Database*)[†] also needs to be updated, specifically the “Syntax” and “Description” fields for *kIRG_HSource*. Below are my suggested changes:

Syntax

Current: `^H-[0-9A-F]{4}$`

Proposed Change: `^H[0-3]-[0-9A-F]{4}$`

* http://www.ogcio.gov.hk/ccli/eng/hkscs/hkscs_2008_summary.html

† <http://www.unicode.org/reports/tr38/>

Description

Current: *The IRG “H” source mapping for this character in hex. The IRG “H” source consists of data from the Hong Kong Supplementary Character Set – 2008.*

Proposed Change: *The IRG “H” source mapping for this character in hex. The IRG “H” source consists of data from the Hong Kong Supplementary Character Set – 2008.*

H0 Big Five Symbols

H1 Big Five Level 1

H2 Big Five Level 2

H3 Hong Kong SCS-2008

Code Chart Additions

This proposal also includes the recommendation that the representative glyphs for Big Five proper be added to the “H” column for the Code Charts, which affect only those for the URO (13,061 new glyphs) and the BMP’s CJK Compatibility Ideographs (2 new glyphs).

In addition, Adobe Systems grants permission to use its “Adobe Ming Std L” font, whose PostScript name is AdobeMingStd-Light, for the purpose of adding representative glyphs for Big Five in the “H” column of the Code Charts. Note that the same “H0” through “H3” prefixes should be used for the sources associated with the “H” column representative glyphs.

That is all.