

**Title:** Proposal to remove the **kRSKangXi** property from the Unihan database

**Author:** Ken Lunde

**Date:** 2022-08-29

This is a proposal to remove the *provisional* **kRSKangXi** property from the Unihan database, which offers no value beyond what the *informative* **kRSUnicode** property already provides.

This property currently covers **63,696** ideographs in the following five blocks (the number of records is indicated in parentheses):

- **CJK Unified Ideographs Extension A** (9): 3687, 3A16, 3ADA, 3B39, 4091, 4347, 49F9, 4CED, 4D60
- **CJK Unified Ideographs** (20,717): 4E00..4E1F, 4E21..4E28, 4E2A..4E45, 4E47..4EB0, 4EB2..4EB6, 4EB8..4EBD, 4EBF..4F28, 4F2A..5033, 5035..5078, 507A..51AD, 51AF..51E7, 51E9..5233, 5235..5258, 525A..5285, 5287..528D, 528F..52FC, 52FE..534F, 5351..53C3, 53C5..543E, 5440..5518, 551A..5527, 5529..556A, 556C..5570, 5573..55B9, 55BB..55ED, 55EF, 55F1..5621, 5623..5627, 5629..5691, 5693..56B7, 56BA, 56BC..56C3, 56C5..570D, 570F, 5711..5716, 5718, 571A..572A, 572C..576B, 576D..57B0, 57B2..583B, 583D..5842, 5844, 5846..585B, 585D..5883, 5885..5894, 5896..58CA, 58CC..594B, 594D..59E1, 59E3..5BBA, 5BBC..5BD4, 5BD6..5BDB, 5BDD..5C6A, 5C6C..5C71, 5C73..5C84, 5C86..5C9D, 5C9F..5CE1, 5CE3..5CE6, 5CE8..5D84, 5D86..5DA4, 5DA6..5DF9, 5DFB..5E59, 5E5B..5EBA, 5EBC..5EBF, 5EC1..5EF8, 5EFA..5EFC, 5EFE..5F41, 5F43..5F59, 5F5B..618E, 6190..61D2, 61D4..6437, 6439..6448, 644A..6599, 659B..6728, 672A..6C1D, 6C1F..6D5A, 6D5C..6DAC, 6DAE..6E99, 6E9B..6F47, 6F49..6F4B, 6F4D..6F55, 6F57..6F74, 6F76..70C8, 70CA..7150, 7152..7192, 7194..722D, 722F..7293, 7295..7436, 7438..746B, 746D..759D, 759F..75B6, 75B8..762B, 762D..764E, 7650, 7652..7672, 7674..7701, 7703..770E, 7710..777A, 777C..7807, 7809..784A, 784C..7887, 7889..7902, 7904..7931, 7934..7985, 7987..79F3, 79F5..7AC1, 7AC3..7C14, 7C16..7C17, 7C19, 7C1B..7C31, 7C33..7C43, 7C45..7CE3, 7CE5..7E01, 7E03..7EEB, 7EED..7F3F, 7F42..802E, 8030..806A, 806C..806D, 806F..807B, 807D..8103, 8105..824B, 824D..8264, 8266..8312, 8314..8404, 8406..84DD, 84DF..84E1, 84E3..8607, 8609..8627, 8629..8648, 864A..8691, 8693..8715, 8717..87A4, 87A6..8865, 8867..8886, 8888..88B2, 88B4..88E9, 88EB..8907, 8909..8964, 8966..898B, 898D, 898F..89C2, 89C4..8A4A, 8A4C..8B90, 8B92..8B9A, 8B9C..8C43, 8C45..8C51, 8C53..8CD3, 8CD5..8E76, 8E78..8F30, 8F32..8FA9, 8FAB..9024, 9026..902F, 9031..9039, 903B..905E, 9060..910B, 910D..9112, 9114..916A, 916C..91A8, 91AA..9272, 9274..9344, 9346..95D6, 95D8..96B4, 96B6..97D1, 97D3..983E, 9840..9900, 9902..9903, 9905..9AA8, 9AAA..9AE7, 9AE9..9B13, 9B15..9B2F, 9B31..9B3F, 9B41..9B4F, 9B51..9B68, 9B6A..9B9B, 9B9D..9BCA, 9BCC..9C97, 9C99..9CED, 9CF0..9D38, 9D3A..9D6D, 9D6F..9E4F, 9E51..9EA1, 9EA3..9FA5
- **CJK Compatibility Ideographs** (301): F900..FA1F, FA21..FA2D
- **CJK Unified Ideographs Extension B** (42,669): 20000..20861, 20863..2230A, 2230C..22B4E, 22B50..22C80, 22C82..23562, 23564..2420A, 2420E..243B8, 243BA..249B9, 249BB..24B56, 24B58..253C2, 253C4..26224, 26226..2622D, 2622F..26232, 26234..2623C, 26241..26242, 26247..26249, 2624E..26251, 26255, 26259..2625A, 2625D..26260, 26262..26263, 26265, 26267, 26269..27440, 27442..2780B, 2780D..2883F, 28841..2934B, 2934D..2A6D6
- **CJK Compatibility Ideographs Supplement** (1): 2FA0E

In contrast, the *informative* **kRSUnicode** property has 100% coverage of the **98,060** ideographs in the nine CJK Unified Ideographs blocks (**97,058**) and the two CJK Compatibility Ideographs blocks (**1,002**) in Unicode Version 15.0.

The following bullets provide justification for removing this Unihan database property:

- The *kRSKangXi* property does not indicate whether a radical is simplified, and given the nature of simplified forms and their relationship with the *Kāngxī Dictionary* (康熙字典 *kāngxī zìdiǎn*), one could easily argue that this property should not even include such entries. There are approximately 1,300 such *kRSKangXi* property records. The *kRSUnicode* property uses a single straight quote to indicate whether the specified radical is the simplified form. For example, the *kRSUnicode* property value of U+89C8 览 is 147'.5, but its *kRSKangXi* property value is 147.5.
- The *kRSKangXi* property, in theory, supports multiple property values, but in practice it does not. There are approximately 80 ideographs that include multiple *kRSUnicode* property values, and that figure tends to increase in each subsequent version of the Unicode Standard. Multiple *kRSUnicode* property values aid in discoverability.
- There are many erroneous *kRSKangXi* property values. For example, the *kRSKangXi* property values for U+27A20 𠮩言𠮩穴又 and U+2959A 𠮩風勺 are 19.7 and 185.3, respectively, but their *kRSUnicode* property values are 149.7 and 182.5, respectively. In other words, the digit “4” is missing from the former *kRSKangXi* property value, and 185 (首) should have been 182 (風) in the latter *kRSKangXi* property value.
- Ideographs that include a simplified component that is not treated as the radical tend to have *kRSKangXi* property values with bogus residual stroke count values. For example, the *kRSKangXi* property value of U+841A 𢦏 is 140.16, but its *kRSUnicode* property value is 140.8.
- The *kRSKangXi* property includes a large number of ideographs that are not actually present in the *Kāngxī Dictionary*.
- The *kRSUnicode* property has coverage of over 34,000 more ideographs than the *kRSKangXi* property, and serves as the basis for producing the [RSIndex.pdf](#), [IICoreRSIndex.pdf](#), and [UnihanCore2020RSIndex.pdf](#) Radical-Stroke indexes.

After factoring out any problematic cases as described above, what remained, in terms of differences between the *kRSKangXi* and *kRSUnicode* properties, were less than 250 cases of property values with either one-off residual stroke counts or different radicals. I took a fine-tooth comb through the data to come up with 123 suggested *kRSUnicode* property value changes, which, if accepted, would effectively preserve in the *kRSUnicode* property any useful bits from the *kRSKangXi* property.

## Data file

The *kRSUnicode-data-1.txt* data file, which is a PDF attachment, provides the proposed property value changes for the Unihan database.

That is all.