

Proposal to Deprecate the Kangxi Field in UAX #45

John H. Jenkins

2 July 2020

Summary

The data in UAX #45 currently contains a field for Kangxi position information. This field contains only redundant information and considerable effort is required to fill it in. It is recommended that it be formally deprecated and the data removed.

Proposal

The Kangxi data in UAX #45 consists of data in the same format as the `kKangXi` field of the Unihan database. Because of the nature of the data in UAX #45, all these Kangxi values are all pointers to virtual positions within the dictionary, meaning that in practice they are merely a more clumsy way of indicating the radical-stroke counts. The data is therefore redundant. The erratic mapping between the Kangxi values and radical-strokes means that generating this redundant data requires a not insignificant amount of work.

(Technically, the mapping from radical-stroke values to virtual Kangxi positions is not injective, meaning that multiple radical-stroke values can map to the same virtual Kangxi position, making the Kangxi data worse than redundant.)

The field itself cannot be removed without altering the format of the data file, and so we recommend only that it be obsoleted and its data removed. At a future time, it might be possible to repurpose the field to include useful information (such as total strokes), but we feel that it is premature to consider this possibility.