

Universal Multiple-Octet Coded Character Set International Organization for Standardization

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

Title: Proposal to clarify radical assignments in IRG working sets and other submissions

Source: Ken Lunde, Andrew West, Lee Collins, Toshiya SUZUKI, John Jenkins, Michel Suignard

References: [IRG N955](#), [IRG N2310](#)

Date: 2019-10-09 (original dated 2019-08-06)

The IRG Principles & Procedures, [IRG N2310](#) (Version 11), does not provide clear guidance about radical assignment for ideographs in IRG working sets and other submissions.

Radical Assignment Issues

The technically “correct” (aka semantic indicator) radical for the vast majority of ideographs in IRG working sets and other submissions also happens to be the most intuitive one, which can be described as a radical assignment made by an ideograph expert who is neither a specialist in the history of the Han script nor its etymology. However, a non-trivial number of ideographs have a technically correct indexing radical that is non-intuitive, which exposes the following real-world issues:

- **Discoverability**—If an ideograph is assigned a technically correct but non-intuitive radical, it is virtually guaranteed to be overlooked, which can also lead to the possibility of the same ideograph appearing in a future IRG working set. In other words, non-intuitive radical assignments waste time, in terms of trying to locate an ideograph among the nearly 100K CJK Unified Ideographs in multiple blocks, and possibly going through the effort of needlessly proposing it again for encoding. The technically correct but non-intuitive suggested radical changes for IRG Working Set 2017 [#01037](#) and [#01206](#) serve as examples.
- **Region-dependent**—The technically correct radical for an ideograph is sometimes region-dependent. Assigning the most intuitive radical can be applied to such cases. SAT’s 2018-09-03 comment against IRG Working Set 2017 [#01166](#) demonstrates this.
- **ISO/IEC 10646 is a character set encoding, not a dictionary**—One of the reasons why source references, some of which correspond to actual dictionaries, are associated with CJK Unified Ideographs is precisely so that the correct (according to the source) radical and other attributes, such as readings and meanings, can be discovered.

Proposal

Append to Section 2.2.1.d.(5).c), *Kangxi Radical Code*, in IRG Principles & Procedures:

If the technically correct (aka semantic) radical for an ideograph hampers its discoverability, or is region-dependent, the primary radical shall be assigned as though made by an ideograph expert who is neither a specialist in the history of the Han script nor familiar with ideograph etymology. The technically correct radical can be assigned as a second radical. Both are shown in the code charts, though the primary one serves as the basis for ordering within a CJK Unified Ideographs block.

That is all.