

ISO/IEC JTC1/SC2/WG2/IRG

Ideographic Rapporteur Group

(IRG)

Source/Contribution Identifier :	Ideograph Radical Ad Hoc
Meeting :	20 th IRG Meeting in Hanoi, Vietnam
Title :	IRG Radical Classification
Status :	For use by the IRG

As work on Extension C1 has proceeded, it has become clear that the effort of checking for already encoded characters and duplicate submissions has become prohibitive. Characters in SuperCJK and Extension C1 are grouped by their KangXi radical, but the radical is sometimes ambiguous, and different editors may expect a character to be classified under different radicals. This means that a careful editor will look for already encoded characters or duplicate submissions under multiple radicals, which can be slow and error-prone.

The Ideograph Radical Ad Hoc group has developed an algorithm which can be used to assign a character to a KangXi radical. This “IRG editorial radical” is something which can be determined by an algorithm without regard for the character’s actual meaning. As such, it should not be used for producing the final ordering of Extension C1. The IRG should continue to use an actual radical to which the character is assigned as if in a dictionary.

The Algorithm

1. Write the character using an Ideographic Description Sequence (IDS). This should be done in such a fashion as to minimize the nesting of the decomposition so long as a KangXi radical (in whatever form) is found, without unnaturally breaking up components.
2. If one or more KangXi radical is found in the IDS, then the *first* KangXi radical in the sequence should be used as the character’s IRG Editorial Radical.
3. If there are no KangXi radicals in the IDS, then the character’s first stroke is used, with the five stroke types mapped to the first five KangXi radicals.

Example One

Consider the hypothetical submission 井蛙. We first write it using an IDS as <井蛙>. There are, however, no radicals in this IDS, so we further write it as <井虫圭>. This IDS now has a radical and does not break up the 井 into smaller pieces, which is unnatural for this component even though it is possible and is the basis for its KangXi radical classification.

Since 虫 is the first KangXi radical in this IDS, it (radical 142) is the IRG Editorial Radical for the character.

Example Two

Consider the character 乂 from the Unicode submission to Extension C1. In this case, there is no natural way to break the character down using an IDS. We therefore use the character’s first stroke (a dot) to assign it radical 3 as its IRG Editorial Radical.