

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document
Title: Response to US/UTC proposal on principles for disunification of combining diacritical marks
Source: Ireland (NSAI/ICTSCC/SC4)
Status: Member Body Contribution
Date: 2005-08-27

Summary

Document N2976, proposes text to be added to the Principles and Procedures document. While we are in broad agreement with the proposal, we disagree with some of the specifics of the text. The contribution presents a main criterion, shape, and then lists four other criteria: membership in a set, function, display behaviour and shape (again). This seems confused to us. We propose to simplify the text by merging point (d) of N2976 with the main criterion, and renumbering the points from the main criterion (a) to display behaviour (d), thus:

Criteria for disunification of combining diacritical marks

A number of criteria may be considered when deciding whether a proposed combining diacritical mark for a particular script should be unified with an existing encoded combining diacritical mark. One or more of these the criteria could favour a decision to disunify when encoding.

- a. The main criterion is that of the shape of the glyph, since that is the chief identifier of a diacritical mark. When the range of glyphic appearance of a diacritical mark may be markedly different from the range typical of the generic diacritical mark, disunification may be preferred. When the mark has been borrowed from another script, but has been significantly modified to fit with the ductus of the borrowing script, disunification may be preferred. These and other criteria have been used in the past, and may be used in the future, as deciding factors in whether to encode separate diacritical marks (or to disunify) for particular scripts.
- b. the mark forms part of a set of marks in the script (for example a set of tone marks), but only some members of the set could be considered candidates for unification with existing marks.
- c. the mark has a specific function fundamentally unrelated to the generic diacritical mark—for instance, the use of the mark as a vowel sign as opposed to the use of a similar-shaped mark as a modifying diacritic. In such case the two uses might also require explicit differences in their character properties.

- d. the display behaviour is fundamentally different and requires different support.
For example, U+A806 SYLOTI NAGRI SIGN HASANTA looks like a combining circumflex, but requires different display support.

The more of these criteria are satisfied, and the stronger the degree to which each is satisfied, the stronger the case for encoding a script-specific diacritical mark. This is not a matter of a rule that deterministically yields a “yes/no” decision; rather, it is a question of degree, which can then form a basis for a proper judgement of the encoding question.

In general, these criteria are not much different from those used for assigning script-specific punctuation.