

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

Doc Type: Working Group Document
Title: Myanmar ad-hoc report
Source: Erkki Kolehmainen (chair), Deborah Anderson, Michael Everson, Asmus Freytag, Lisa Moore, Thein Oo, Tseng Shih-Shyeng (Bear), Wei Lin-Mei, Ken Whistler, Zaw Htut
Status: Ad-hoc report
Action: For consideration by JTC1/SC2/WG2
Date: 2006-04-26

The ad-hoc group on Myanmar character additions notes that the changes proposed to the encoding model for the Myanmar script needed to be taken seriously, and recognizes that such changes cannot be made lightly, and must be made only if there are clear technical needs to do so.

A statement of the fundamental requirements was made: that the Myanmar encoding must be adequate to represent all of the languages of the Union of Myanmar which use the script, in a plain-text monofont implementation without recourse to language tagging or language-specific fonts. The core character set must be completed with language-specific characters, and the encoding model must be able to support multilingual documents correctly. Another requirement is that it be possible to handle alternate spellings of the same word in Burmese.

Discussion of specific problems showed that the current model is in fact inadequate to represent two features of S'gaw Karen alongside Burmese text (AA/TALL AA and MEDIAL YA/LA as described in N3043R). It is also determined that some variant spellings in Burmese itself involving AA/TALL AA as different spellings for the same word exist, and cannot be handled by the current encoding. For example, *pwā*: 'duplicate' can be written both ဝုး and ဝုဲ. As proposed, addressing this inability to represent required distinctions by disunifying TALL AA from AA and by disunifying MEDIAL YA from VIRAMA + YA has other implications. In particular, the ad-hoc recommends the disunification of the other three medials and the disunification of ASAT from VIRAMA. This permits the correct representation of S'gaw Karen and Mon text; it also regularizes the representation of Burmese, and permits the representation of some rare Pali and Sanskrit clusters as well. In balance, the fact that this approach addresses the requirements for the representation of minority languages in Myanmar script significantly outweighs the transition problems that the ad-hoc sees in adopting these characters. At present, the amount of data available in Myanmar script is still low, but it is expected to increase very soon, due to new e-government initiatives on the part of the authorities in the Union of Myanmar.

The ad-hoc agreed that the simplification for Burmese alone is not sufficient reason to make these changes. The source of many of the technical problems that drive the assessment regarding the need for encoding medials as stated above is that the Myanmar script is not just for Burmese but must also represent all of the languages that use it. The proposed additions in N3043R address this problem. While this is the fundamental reason for making the changes implied by the addition of the characters, these changes also have the benefit of simplifying the overall representation of Burmese itself and making possible the representation of distinctions that could not be covered by the current encoding within Burmese.

The ad-hoc recommends to WG2 that the seven characters proposed in N3043R be added to PDAM3, in order for them to be available as soon as possible for Myanmar I.T. implementations. However, noting the significant technical implications of these additions, and the concerns of some national bodies that sufficient review be allowed for such changes, the ad-hoc also recommends that PDAM3 should be progressed as a second PDAM, to give the greatest opportunity for National Body review and feedback during balloting.