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

Doc Type: Working Group Document
Title: Request to add the characters proposed in WG2 N4068 to PDAM 1.3
Source: German NB
Action: For consideration by JTC1/SC2/WG2 at the discussion of PDAM 1.2
Date: 2011-10-18

WG2 N4068 was submitted to SC2/WG2 as a liaison proposal by SC35, requesting the encoding of missing Latin small capital and modifier letters.
This proposal was on the agenda of the JTC1/SC2/WG2 June 6-10, 2011, in Helsinki.
In fact, at that meeting, it was not discussed due to lack of time.

Germany already had requested the addition in WG2 N4085, section 3.
As this German request persists, the text of that section is repeated here, with added text shown in italics.

The German NB welcomes this proposal, especially (besides the reasons listed in the proposal) as it will save the time to discuss and decide any single character whenever new evidence is found.

In discussions on the Unicode mailing list and forum, concerns were issued that having the sets of modifier letters derived from the basic Latin letters (A...Z) complete will invite users to "misuse" them, by using them to obtain a different style instead denoting of a different semantic (like using the small capital letters to get smallcaps). However, this already can be done with the current almost complete sets, and has to be taken into account by search engines etc. anyway (and imposes no real problem as all these characters have compatibility equivalences to the respective base letters). It also should be noted that quasistandards like the Adobe Glyph List 2.0 already have encoded such complete sets (using the PUA), thus indicating style by character selection rather by markup is already an existing practice. (Such practice may be considered "bad", but standards are there for serving their users rather than for educating them.)

In addition, concerns were issued that encoding the modifier forms of the basic Latin letters (A...Z) will encourage proposals for more letters. The addition of the complete set of basic Latin squared letters A...Z in Unicode 6.0 (like the encircled letters long ago) has proven that such is not an automatic consequence.

Moreover, the German NB is concerned about the fact that SC35/WG2, which identifies characters in keyboard standards by Unicode values only recently, has partially to deviate from this in a recent standard draft (ISO/IEC 9995-9 CD), by the plain fact that specifying a method to enter raised, lowered, and small capital forms of the basic Latin letters A...Z/a...z (as part of language orthographies, not as style) results in a mapping to a character set which cannot completely enumerated by Unicode. Even if the PUA positions used in that draft "are only for internal reference", there is the danger that they will in fact be used by referencing to an international standard, thus weakening the position of Unicode for a larger audience. This can be avoided when SC2/WG2 is prepared to cooperate with SC25/WG2 by accepting the characters proposed in N4068 in time.

Thus, the German NB requests the addition of the characters proposed in N4068 now to PDAM 1.3 of ISO/IEC 10646.

Technical note: The code range of the block "Phonetic Extensions Supplement B" referred to in N4068 is now part of the block "Latin Extended-E". However, the proposed code points for the characters there are still appropriate.