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

Doc Type: Working Group Document
Title: Further proposed additions to ISO/IEC 10646 and comments to other proposals
Source: German NB
Status: National Body Contribution
Action: For consideration by JTC1/SC2/WG2 at the Helsinki meeting, June 6-10, 2011
Date: 2011-05-25

This document addresses proposals and other issues which are not submitted by the German NB.

Requesting or proposing additions to ISO/IEC 10646 is to be read as a request to open a new PDAM in Helsinki and either adding the characters to it.

Regarding any proposals neither submitted by the German NB nor addressed here, Germany has no special comments, and welcomes the addition of all proposed characters which get consensus of the experts.

1. Proposal to encode Teuthonista characters *(successor document to N4031, still to be submitted at the time of compiling this list)*

This proposal is worked upon by the members of the German delegation which will attend the Helsinki meeting. The German NB proposes the addition of the characters proposed there to ISO/IEC 10646, as they are needed urgently for ongoing scientific projects.

2. N3897 = ISO/IEC JTC1/SC35 N1579: Proposal to incorporate symbols of ISO/IEC 9995-7:2009 and Amendment 1 into the UCS

This proposal was presented at the previous meeting in Busan (Oct., 2010). There, it was decided to delay this proposal to let all NBs to inspect it and to raise comments. Since then, no issues regarding any of the proposed characters have been raised.
The encoding was proposed to maintain the interoperability by ISO/IEC 10646 and the standard for keyboard function symbols (ISO/IEC 9995-7), to enable the use of all keyboard symbols in plain text (e.g. operating instructions), whether they are established for a long time or newly agreed in SC35 to keep the standards up with the technological state of art. Such was requested by the German NB related to SC35/WG1. The German NB related to SC2/WG2 supports this position.

Thus, the German NB requests the addition of the characters proposed in N3897 to ISO/IEC 10646.

3. N4068: Proposal to encode missing Latin small capital and modifier letters

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.

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

4. N4070: Second revised proposal to add characters used in Lithuanian dialectology

The German NB requests the addition of all 18 characters proposed in N4070 to ISO/IEC 10646, adhering to the rationale expressed in N4070 especially on dealing with N4062 "Comments on N3914 - Characters for Lithuanian Dialectology" (N3914 is the predecessor document of N4070).

However, it is noted that the proposed code points of some of the characters collide with N3897. For the affected characters (the "intonation marks"), Germany proposes the following code points:

U+2B7C SLANTED NORTH ARROW
U+2B7D BACKSLANTED SOUTH ARROW
U+2B7E SHORT SLANTED NORTH ARROW
U+2B7F SHORT BACKSLANTED SOUTH ARROW
U+2B80 SLANTED NORTH ARROW WITH HOOKED HEAD
U+2B81 BACKSLANTED SOUTH ARROW WITH HOOKED TAIL
U+2B82 SLANTED NORTH ARROW WITH HORIZONTAL TAIL
U+2B83 BACKSLANTED SOUTH ARROW WITH HORIZONTAL TAIL
U+2B84 SLANTED ARROW POINTING DOWNWARDS THEN NORTH EAST
U+2B85 SHORT SLANTED ARROW POINTING DOWNWARDS THEN NORTH EAST

5. N4030: Proposal for the addition of six Latin characters to the UCS

The German NB requests the addition of the characters proposed in N4030 to ISO/IEC 10646.

However, it is noted that the proposed code points of the characters collide with other proposals. Therefore, Germany proposes the following code points:

U+A7CB LATIN CAPITAL LETTER B WITH FLOURISH
U+A7CD LATIN SMALL LETTER B WITH FLOURISH
U+A7CE LATIN CAPITAL LETTER REVERSED OPEN E
U+A7CF LATIN CAPITAL LETTER SCRIPT G
U+A7D0 LATIN CAPITAL LETTER SIDEWAYS I
U+A7D1 LATIN SMALL LETTER SIDEWAYS I

6. N4013: Proposal to encode additional Runic characters in the UCS

The German NB requests the addition of the characters proposed in N4013 to ISO/IEC 10646.

7. Hungarian Runic (also called Szekely-Hungarian Rovas):

There are two proposals for the same script:
N3697: Proposal for encoding the Hungarian Runic script
N4007: Revised proposal for encoding the Szekely-Hungarian Rovas script in the SMP

N4064: "Comparison of Hungarian Runic and Szekely-Hungarian Rovas proposals"
shows that there is a large common subset, and that N3697 contains exactly this subset plus one single character.

Thus, the German NB requests the addition of the characters proposed in N3967 to ISO/IEC 10646, with the exception of the character "NIKOLSBURG OE" which is not in the common subset.
(Germany will agree to add this character also if the experts present at Helsinki will decide so; this also will apply for any character in N4007 not contained in N3697, or for any naming or sorting issue.)

Also, N3640: "Old Hungarian/Szekler-Hungarian Rovas Ad hoc report"
states "The numbers ... 500 can be added if the supporting evidence can be provided".
The German NB considers the evidence now shown in N4007 as sufficient and requests the addition of this character as "HUNGARIAN RUNIC NUMBER FIVE HUNDRED" to ISO/IEC 10646.
Germany proposes the code point U+10CFE, implying that the code points of the characters proposed at 10CFA...10CFE in N3697 will be decreased by one.

In addition, the German NB requests the addition of the characters proposed in N3664 "Proposal for encoding generic punctuation used with the Hungarian Runic script" to ISO/IEC 10646.

8. N4022: Proposal to add Wingdings and Webdings Symbols

The German NB strongly opposes to an acceptance of this proposal in its present form.

The major concerns are:

- The proposal shows no benefit for the user for any character.

    Indicators for such benefit for the user can be:

    – Evidence of actual use.
    – Evidence of prevention of an otherwise probable actual use due to the lack of encoding.
    – Conformance with or compliance to another standard.

    (Note that we list as "indicator of benefit for the user" not only "evidence of actual use".)

- The proposal violates the character/glyph model by proposing glyph variants as different characters without any explanation.

    For instance. on p.31 is shown:

ID	glyph	UCS	glyph	Name
2093	‽	203D	‽	INTERROBANG
2094	‽	1F678		MEDIUM HEAVY INTERROBANG
2095	‽	1F679		SANS-SERIF INTERROBANG
2096	‽	1F67A		MEDIUM HEAVY SANS-SERIF INTERROBANG

    Presumably, the glyph variants of the interrobang went into the Wingdings font at a time when the interrobang was not encoded (and thus was not available as a character), to enable the

user to include the interrobang in their texts. In doing so, the user was enabled to select the symbol which fits best to the font and style used.

Now, as the interrobang is available as a character, there is no use for doing such: The user simply inserts U+2093, and the glyph will be selected automatically by the selected font.

Thus, encoding glyph variants of the interrobang as separate characters U+1F678 ... U+1F67A, as N4022 in fact proposes, violates the character/glyph model without providing any advantage in exchange.

However, a part of the characters contained in N4022 in fact may deserve encoding.

Therefore, the German NB requests:

- The list of the symbols contained in N4022 shall be the starting point for a work like it is done on Emoji, with the objective to compile an acceptable proposal for the next WG2 meeting of March, 2012.

- In doing this work, other symbols commonly used on Internet sites, or being related otherwise, shall be included.

- Also, in doing this work, other fonts widespread on the computers of leading manufacturers (e.g. Apple) shall be included, thus avoiding the impression that Unicode or SC2/WG2 favor a single manufacturer.

- For any symbols which can be regarded as user interface symbols, the work shall be coordinated with ISO/IEC JTC1/SC35/WG1, which is working on a new amendment 2 for the keyboard symbol standard ISO/IEC 9995-7.

However, the German NB proposes three of the characters proposed in N4022 to be added to ISO/IEC 10646 now:

U+1F441 EYE
   (ID 0078)

o   This symbol was already proposed by the German and Irish NB by the same name and code point in N3607.

U+1F10B DINGBAT CIRCLED SANS-SERIF DIGIT ZERO
   (ID 1128; named in N4022: CIRCLED SANS-SERIF DIGIT ZERO)

U+1F10C DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT ZERO
   (ID 1139; named in N4022: NEGATIVE CIRCLED SANS-SERIF DIGIT ZERO)

o   These are straightforward extensions of the existing dingbat sans-serif digit sets, adding a zero to them which is already done to similar sets. To denote the characters being part of this set, the name shall be selected according, including "DINGBAT".