
ISO/IEC JTC1/SC2/WG2 N 2569

Date: 2002-12-12

ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)

Doc. Type: Draft disposition of comments

Title: Proposed disposition of comments on SC2 N 3624 (FPDAM text for Amendment 2 to ISO/IEC 10646-1:2000)

Source: Michel Suignard (project editor)

Project: JTC1 02.18.01

Status: For review by WG2

Date: 2002-12-02

Distribution: WG2

Reference: SC2 N3624, N3641

Medium: Paper

Comments were received from Canada, Iran, Ireland, ROK, and USA. The following document is the draft disposition of those comments. The disposition is organized per country.

As a result of this disposition, Iran, Ireland and the USA that had voted negatively have reversed their votes to Yes.

Note - The full content of the ballot comments (minus some character glyphs) have been included in this document to facilitate the reading. The dispositions are inserted in between these comments and are marked in **Underlined Bold Serif text**, *with explanatory text in italicized italic serif*.

Canada: Yes with comments:

Editorial Comments:

Comment a:

Page 1: Italicized 'change xxx by xxx' under Page 13 items.. does not read right..

Please change to 'replace xxx with xxx' etc.

Accepted in principle

The verb 'change' will be replaced by 'replace' and will read 'replace xxx by xxx' to be consistent with other occurrences.

Comment b:

Page 3, new entries for collections:

104 and 105 have both LTR ALPHABETIC PRESENTATION FORMS. One of these (we think it is 105) must be RTL ALPHABETIC PRESENTATION FORMS.

=====

Accepted

Also requested by US comment T.5 a)

Iran: Negative

(In the following text, the P+ABCD notation is used to refer to character proposed to be encoded at the position ABCD.)

Technical comments:

T1. Pages 18 and 19: Iran requests the names (or the glyphs) for the proposed characters P+060D and P+060E be interchanged, as the current names are clearly wrong. Iran also requests a review of other characters names and shapes which are new to the Arabic block be done by the editors, to make sure there is no other such error.

Accepted

The glyphs for the proposed characters at 060D and 060E will be reversed. From document WG2 N2482 and N2483 it is clear that the intent was as described by the Iranian member body. The editor has done further study of the many involved documents (WG2 N2413, 2481, 2482, 2483, 2487 and 2489) and, to the best of his knowledge, has not found other mistakes.

T2. Pages 18 and 19: Iran requests that the character P+0659 be moved to the position 0615, since it's a combining character that, unlike the other characters in the range U+064B to U+0655, but like the characters P+0610 to P+0614, applies to the whole word (or a word ending), and not to a single letter. This will help consistency in the new characters proposed to be encoded in the Arabic block.

Accepted

T3. Pages 18 and 19: Iran requests the character P+0603 to be moved to the position 060B, or some other position in the range 061C to 061E, to help group the characters with similar functionality near each other.

Withdrawn

T4. Page 75: Iran requests the character P+FDFF to be renamed to have a name more consistent with the other characters in the range U+FDFF to U+FDFFB. As current transliterations done for the names of those characters recommend the letter E for Arabic Kasra, and since the vowels connecting the words are also Kasra (and not Fatha), the name for the new character should be changed to 'ARABIC LIGATURE BESMELLAH ERRAHMAN ERRAHIM' or 'ARABIC LIGATURE BESMELLAHERRAHMANERRAHIM'.

Accepted in principle

After further discussion with the Irish member body and the original contributor, the name is changed as followed: ARABIC LIGATURE BISMILLAH AR-RAHMAN AR-RAHEEM

Editorial comments:

E1. Pages 18 and 20: Iran requests the glyphs for the following proposed characters be revised to be consistent with the rest of the Arabic block:

P+0656 The representative glyph should be similar in shape and size to U+0670 ARABIC LETTER SUPERScript ALEF. The Alef-shaped part of the glyph is currently much larger.

P+0657 The representative glyph should be similar in shape and size to U+064F ARABIC DAMMA. It is currently smaller and comma-shaped instead of being Damma-shaped.

P+0659 The representative glyph should change to Naskhi style which is the style of all other characters in the Arabic block. The current glyph is in Nastaliq style. See also T2.

P+06EE, P+06EF, and P+06FF The representative glyphs should become larger in size (compare P+06EE with U+062F, for example) and have their "Inverted V" part consistent with the glyph for U+06C9.

Accepted in principle

The project editor suggests to the interested parties to provide a font suitable for publication.

Following the disposition, Iran changes its vote to YES.

Ireland: Negative

Technical comments

T.1. Change the character name for U+0658 to ARABIC MARK NOON GHUNNA -- note that it should not be "NOON-GHUNNA" as was seen in some earlier documentation.

Accepted

Also requested by US comment T.3

T.2. In the note to be added at clause 25.3, it should say "When the normalization forms are applied" not "When the normalizations forms are applied"

Accepted

Similar typo in clause 24, note 2: 'Some normalizations forms', it should read 'Some normalization forms'.

T.3. Change the character name for U+FDFF to ARABIC LIGATURE BISMILLAH AL-RAHMAN AL-RAHIM. This transliteration better shows the underlying text, where the transcription ARRAHMAN ARRAHIM hides it due to the assimilation of the article as written.

Accepted in principal

See disposition for Iranian comment T.4. New name is ARABIC LIGATURE BISMILLAH AR-RAHMAN AR-RAHEEM

Editorial comments

E.1 The glyph for the hot beverage at U+2615 should be changed to one which is more conventional.

Accepted in principle

The project editor suggests to the interested parties to provide a font suitable for publication.

Following the disposition, Ireland changes its vote to YES.

Korea, Republic of (ROK), Negative

Technical Comments

T1. reason 1.

We used ISO TS 11941:1996 for naming Korean Hangul letters and syllables.

However, the transliteration used for naming the following characters in PDAM2 is different from TS 11941.
[p. 63, Table 104]

U+0321D OJEON

U+0321E OHU

U+0327C CHAMKO

U+0327D JUEUI

Considering that, in the future, we may add more Korean Hangul letters in addition to the above characters, we suggest that we set up a transliteration rule for naming Korean Hangul letters and syllables and adhere to that rule.

Noted

This is almost verbatim the same comment that was made by ROK concerning the PDAM2 ballot and could not be accommodated by WG2 (see document WG2 N2489)

T2. reason 2.

Normalization forms for Hangeul letters/syllables are not appropriate for handling Hangeul properly.

[on page 1] Page 2, clause 3 Normative references Add the following reference:

Unicode Standard Annex, UAX#15, Unicode Normalization Forms, Version 3.2.0, 2002-03-27. ...

[on page 2]

Page 13, after clause 23 Order of characters

Insert a new clause 24 between the current clause 23 and 24 (the following clauses are incremented by one):

24 Normalization forms

Normalization forms are the mechanisms allowing the selection of a unique coded representation among alternative, but equivalent coded text representations of the same text. Normalization forms for use with ISO/IEC 10646 are specified in the Unicode Standard UAX#15. ...

Not accepted

The comment was asking for complete removal of the reference to the Unicode normalization. This was not acceptable to other member bodies. However to help clarification about the goal of normalization, a new note will be added:

Note 3 – The goal of normalization is to provide a unique normalized result for any given text sequence to facilitate, among other things, identity matching. A normalized form does not necessarily represent the optimal sequence from a linguistic point of view.

USA: No with comments:

Technical comments:

T.1 Normative References

Update the reference to the Unicode Bidirectional Algorithm to version 3.2

Accepted

T.2 Greek

Add the following characters:

GREEK CAPITAL LUNATE SIGMA (proposed code position: 03F9) as documented by INCITS L2 02/314R,

GREEK CAPITAL LETTER SAN (proposed code position: 03FA) as documented by INCITS L2/313R,

GREEK SMALL LETTER SAN (proposed code position: 03FB) as documented by INCITS L2/313R

Accepted

The characters were also discussed and accepted for inclusion when covering documents WG2 N2512 and N2522. The first character name is GREEK CAPITAL LUNATE SIGMA SYMBOL (note the addition of SYMBOL).

T.3 Arabic

Change the character name for U+0658 to ARABIC MARK NOON GHUNNA

Accepted

Also requested by Irish comment T.1

T.4 Symbols

Add the character VERTICAL LINE EXTENSION (proposed code position: 23D0) as documented by WG2 N2508. The character addition would complement the existing HORIZONTAL LINE EXTENSION at code position 23AF.

Accepted

The character was also discussed and accepted for inclusion when covering document WG2 N2508.

T.5 Annex A.1 Collections of coded graphic characters

a) Change the entry 105 to read:

105 RTL ALPHABETIC PRESENTATION FORMS

Accepted

Also requested by Canadian comment a)

b) The code range for collection 110 should read 2B00-2B7F.

Accepted in principle

The exact range is 2B00-2BFF

c) To complete the change for collection 63, the first entry for collection 63 in annex A.1 (between 62 and 64) should read:

63 (Collection specified as union of other collections)

Accepted

d) Collection 251 SCRIPT-SPECIFIC FORMAT CHARACTERS should not contain the collection 207

Accepted

e) After the collection 381, a new collection should be added as follows:

382 CJK COMPATIBILITY IDEOGRAPHS-2003 * described as Collection 381 with the addition of U+FA70-FAE9. This is only required if the addition of the range FA70-FAE9 is maintained in this proposed amendment. Note that the US is asking for its removal in comment T8.

Withdrawn

The block was removed

f) In the alphabetically list in Note 3, add collections 98 and 99 to the 'Arrows' entry and 380, 381, 382 (as appropriate, see above) to the 'Ideographs' entry.

Accepted

As per disposition of comment T.8, the value 382 is not added

T.6 Annex A.2 Block lists

Modify the code position range for MISCELLANEOUS SYMBOLS AND ARROWS to 2B00-2B7F

Accepted in principle

The exact range is 2B00-2BFF

T.7 Annex Q and R

In the context of merging the two parts of ISO 10646, the US is asking for an overhaul of Annex Q and R of Part 1 by making the code mapping for Hangul Syllables (Annex Q) and the Names of Hangul Syllables (Annex R) available as separate text files. This will make the information much more usable.

Accepted in principle

Annex Q will be replaced; however Annex R will still contain the current name table with an additional link to a plain text file containing the full names as specified by clause 26.2 of ISO/IEC 10646-1:2000

T.8 DPRK Compatibility characters

The US asks for removal of characters from FA70 to FAE9.

Rationale:

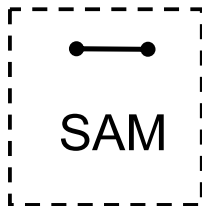
The US has found many discrepancies in the list of characters (incorrect mapping). For example, the reference mapping for characters FA9D to FAE9 is shifted by one position, leaving the proposed code position FA9D without reference mapping to a CJK Unified character. Some characters like FA8C and FA90 and many others seem to have incorrect mappings. Furthermore, it looks like not enough research has been done to look for existing variants in the CJK repertoire that could have been used for encoding the KP1 characters. Given these issues, the US does not see these proposed additions as mature enough for encoding and is therefore asking for their removal pending further study.

Accepted

Note that WG2 accepted in principle the revised repertoire (107 characters) for further consideration in future amendment (see resolution WG2 M43.3)

T.9 Syriac

Change the glyph for 070F SYRIAC ABBREVIATION MARK to include a default rendering in addition to the current picture, suggested new glyph:



Accepted in principle

The glyph will be as suggested but with an additional segment in the upper part (dot-line-dot-line-dot)

T.10 Annex F, Soft Hyphen format character

The SOFT HYPHEN entry in clause F.1.1 should be reworded as follows:

SOFT HYPHEN (00AD): SOFT HYPHEN (SHY) is a format character that indicates a preferred intra-word line-break opportunity. If the line is broken at that point, then whatever mechanism is appropriate for intra-word line-breaks should be invoked, just as if the line break had been triggered by another mechanism, such as a dictionary lookup. Depending on the language and the word, that may produce different visible results, such as:

- inserting a graphic symbol indicating the hyphenation and breaking the line after it,
- inserting a graphic symbol indicating the hyphenation, breaking the line after the symbol and changing spelling in the divided word parts,
- not showing any visible change and simply breaking the line at that point.

The inserted graphic symbol, if any, can be take a wide variety of shapes, such as HYPHEN (2010), ARMENIAN HYPHEN (058A), MONGOLIAN TODO SOFT HYPHEN (1806), as appropriate for the situation.

When encoding text that includes explicit line breaking opportunities, including actual hyphenations, characters such as HYPHEN, ARMENIAN HYPHEN, and MONGOLIAN TODO SOFT HYPHEN may be used, depending on the language.

When a SOFT HYPHEN is used to represent a possible hyphenation point, the spelling is that of the word without hyphenation (for example: "tug<00AD>gumi"). When encoding text that includes hard line breaks, including actual hyphenations, the spelling must reflect the changes due to hyphenation (for example: "tugg<2010>" / "gumi").

Accepted in principle

The last paragraph is changed as follows:

When a SOFT HYPHEN is used to represent a possible hyphenation point, the character representation is that of the text sequence without hyphenation (for example: "tug<00AD>gumi"). When encoding text that includes hard line breaks, including actual hyphenations, the character representation of the text sequence must reflect the changes due to hyphenation (for example: "tugg<2010>" / "gumi").

Editorial Comments:

E.1 Clause 25.3 Alternate coded representations

In the 6th line add a 's' at the word 'representation' (after multiple).

Accepted

E.2 Clause 27 CJK Unified Ideographs

Add a mention that the clause number is now 28 (following the insertion of clause 24).

Accepted

Following the disposition, the US changes its vote to YES.