ISO/IEC JTC 1/SC 2 N 3652/WG2 N2527

DATE: 2002-11-25

L2/02-429

ISO/IEC JTC 1/SC 2
Coded Character Sets
Secretariat: Japan (JISC)

DOC. TYPE	National Body Contribution
TITLE	Canadian Comments on SC 2 N 3638: WG 2 Proposal for Merging ISO/IEC 10646 Parts 1 & 2 into One Part [WG 2 N 2506, N 2408] and N 3639: Working Draft for Merged ISO/IEC 10646 [WG 2 N 2499 R]
SOURCE	National Body of Canada
PROJECT	
STATUS	This document is forwarded to WG 2 for consideration. It will also be considered at the 12th SC 2 Plenary meeting to be held in Tokyo, Japan.
ACTION ID	ACT
DUE DATE	
DISTRIBUTION	P, O and L Members of ISO/IEC JTC 1/SC 2 ; ISO/IEC JTC 1 Secretariat; ISO/IEC ITTF
ACCESS LEVEL	Def
ISSUE NO.	152
FILE	**NAME** 02n3652.pdf **SIZE (KB)** **PAGES** 5

Canada accepts the project sub division proposal and the rationale provided in document SC2/N3638, and will participate in the project.

Canada also accepts the document SC2/N3639 as a suitable draft text for the combined third edition for 10646.

We have two small editorial comments.

   1) N2408 (attachment to N3638) in its last paragraph on page 1 has « At the October 2001 meeting in Singapore, WG2 initiated work on two project sub-divisions -- Amendment 2 to Part 2, and Amendment 1 to Part 1, both involving character additions ». It should be corrected to « ... Amendment 2 to Part 1, and Amendment 1 to Part 2, ... » (as mentioned in last paragraph of page 2).

   2) In N3639, page 7, figure 2 : it is not clear why plane E0 is still mentioned and has a solid border rather than a dotted one. Suggest removing labels on planes 80 and E0, re-label 0F as 10 and show all the planes beyond the new 10 with dotted lines all around.

Canada
Additional Comments on WG2 N2999R (SC2 N3639)
(Expert contribution from: V.S. Umamaheswaran, IBM Canada, 2002-11-08)

---

1.   Editorial -- Throughout the document:

   a.   Use of Plane 0, Plane 00, Plane 1, Plane 01, Plane 2, Plane 02, .. Plane E, Plane 0E, Plane F, Plane 0F etc. is inconsistent in many places in the document. Suggest rewrites to Plane 00, Plane 01, Plane 0E, Plane 0F ..consistently. Also, BMP, SMP, SIP and SSP are used instead of Plane 00 etc. in some places. In other places you see Plane 0 (BMP) etc. By using only BMP and SIP, instead of Plane 00, Plane 02 etc. everywhere it makes sense, may make life of editor easier and improve the consistency.
   Use of Group 0, Group 00 .. is inconsistent in many places in the document. Suggest rewrites to use Group 00 wherever Group 0 occurs.
   The above inconsistency in notations also occurs in the Collections detailed in Annex A.

   b.   References to clauses and annexes occur in many styles .. the following variants are some of them: "see clause xx", "see sub-clause xx', "see xx", "(see clause xx)", "(see xx)", "(xx)" , "described in xx", "listed in Annex xx", Annex is sometimes capitalized, mostly not, etc. These references should be checked and cleaned up to use a consistent style -- for example "see clause xx", "(specified in Annex xx)", etc. throughout the document. There may be an ISO Style guide that specifies how this should be done.
   Whichever form is chosen, please apply it consistently throughout the document.

===========
The following editorial comments to the combined parts draft text, may be candidates for Editorial Corrigenda to be processed along with FDAM-2 to Part 1, or FDAM-1 to Part 2, as appropriate, if it will help in progressing the combined part without any further delay.
===========
2.   Individual Editorial
   a.   Page 1, Left column, paragraph above Note 1
            This reads clumsy .. At first reading, it meant that a character assigned to the BMP will not be duplicated in the Supplementary Planes. Also no duplicates were allowed within the set of Supplementary Planes. But the reverse was sort of left OPEN--i.e. A character already coded in the SMP (for example) could be duplicated in the BMP in the future??!!
            Suggest combing the two sentences should be combined into one. Something like:

"A graphic character will be assigned only one code position in the standard, located either in the BMP or in one of the supplementary planes."

b. Page 3, Clause 4.23
     change the first 'a' to 'A' -- to be consistent with other definitions in 4.
c. Page 4, Clause 4.33
     Missing '.' at the end of the sentence -- to be consistent with other definitions in 4.
d. Page 4, Clause 4.38
     Add to the definition: A plane other than Plane 00 of Group 00; a plane that accommodates ...
e. Page 4, Clause 4.42, second bullet
     Rewrite to read "... preceded by an RC-element from .. " -- consistent with text in the first bullet.
f. Page 5, Third paragraph in the left column
     Suggest reword "... from 16 supplementary planes of group 00 (Planes 01 to 10), in addition to the BMP (Plane 00), in a form ...".
g. Page 5, Last paragraph in the left column
     Suggest reword "... according to ISO/IEC 4873 that have .. " or "... according to ISO/IEC 4873, which have ... ". Grammar checkers consistently insist that a comma should be present before a 'which' or change to use 'that'.
h. Page 6 Figure 1 - this could be improved:
     Move the arrow and text Plane 00 of Group 00 .. such that the text is outside the boxes, and the arrow points to the first plane.
     Move the arrowhead associated with 'Plane FF of Group 00' a little bit away from the edge of the previous plane.
i. Page 7, Figure 2 - needs some fixes
     Remove the labels 80 and E0 from the planes. Change the edges of plane E0 to dotted lines. The private use planes and groups beyond Planes 0F and 10 have been removed from the standard.
j. Page 8, Clause 6.4-c
     It would be clearer if reworded as .. "follows the source-reference rule given in Clause 27 ...". Also, suggest adding a forward reference to the construction of names for Ideographs in clause 28.1 and Hangul Syllables in clause 28.2, after the last paragraph in this clause.
k. Page 8 - last line of the right column:
     Suggest splitting and rewording as ... "to F, or a to f). For example:"
l. Page 11 - Title of clause 13.1
     The title should be: "Two-octet BMP form (UCS-2)" -- There is a forward reference to clause 13.1 from Clause 8 First Paragraph on page 9 -- which hints that 13.1 is defining UCS-2; also to be consistent with the title of 13.2 for UCS-4.
m. Page 12 - right column, first line after the last ESC ..
     Suggest rewording .. "or from the lists in C.5 for UTF-16 and D.6 for UTF-8 forms.
n. Page 13, clause 16.4, after the last ESC.. The sentence beginning with
     "For a subset of C0 or C1 sets, the final octet F ... " should be changed to "For other C0 or C1 sets, ...".
     This change is needed to be fully in line with the intent of the first paragraph of Clause 11 -- which allows for 'sets or subsets' not only from 6429 but from other sources as well.
o. page 13, Clause 17
     Reword to start with: "Clause 33 ... "
p. Page 14, left column, before the last paragraph.
     Suggest a forward reference to Annex E via a note .. such as:
     NOTE - A list of characters that may be rendered as mirror images in a bidirectional context is provided in Annex E.
q. Page 14, Clause 19, First line
     Suggest reword to : "A class of left and right handed .."
r. Page 14, clause 19.1
     It will be helpful to give a reference to the normative references here .. by rewording:
     "The Unicode Bidirectional Algorithm (see clause 3) ... ".
s. Page 15, Clause 20.4, Right Column, First paragraph
     I suspect this paragraph should really be a NOTE.
t. Page 17, Right Column, NOTE 4

There is a reference to the Unicode data base here. It will be useful to give the URL for this database at this spot.

There are mentions of Unicode sprinkled at various places in the document -- besides the BiDi and Normalization TRs, some way of adding a non-normative reference to Unicode standard and the Unicode database should be found.

The standard has only Normative Reference clause 3, and Annex S for Sources of Characters, as two places where references to external sources can be made. Not sure if references to the Unicode database and to the Unicode standard V3.2, V4.0 etc. can be included in Annex S -- under a separate sub-item called "Other related standards".. or something like that.

u. Page 18, Clause 22, Fourth paragraph

Suggest rewording .. "... (characters that are part of the CJK ..)" .. occurs at two place.

v. Page 18, right column, last line of first paragraph:

It will be helpful by rewording ".. UAX#15 (see clause 3)."

w. Page 19, Clause 25.3, Last sentence

Add a reference .. "Normalizing (see clause 24) these coded ... ".

x. Page 20, Clause 26.2 Title

Suggest rewording to: "Features of Indic and similar alphabetic scripts"
or "Features of scripts used in India and some other South Asian countries"
Myanamar, in some peoples' minds, is NOT Indic.

y. Page 20, Clause 27.1, First paragraph, last line

Reword to: "... in the code tables in Clause 33, are described in annex S. .. "

z. Page 20, Clause 27.1, Paragraph after Note 1

Reword to: "The following ... in both the BMP and the SIP. The set of .. ". If Plane 0, Plane 2 etc. is retained, it has to be changed to Plane 00 in Group 00 ad Plane 02 in Group 02 etc.

aa. Page 21, Right Column - bullet for columns 27--33

K4-dddd is the only Decimal digits only use in this entire table. If this is not finalized, and if DPRK can be requested to provide the resource in K4-hhhh format, it will make it consistent with the rest of the source references in this file, and will make the life of any parser a little easier -- especially if these dddd-s have to be converted to hhhh by the reading software of the file.

bb. Page 22, Note under the Figure in the rectangle.

At first glance, reading the text of the note alone, raised the question -- how come the decimal numbers in the last line are not equivalent to the hex numbers in the second last line. The explanation is in the middle of the third paragraph following this Note.
Suggest either -- delete this NOTE, since the explanation is in the text;
or copy the second sentence starting with "The first line below the graphic symbol ... for position number." as the text to replace the current text of the NOTE.

cc. Page 22, First paragraph after the Note

Suggest rewording .. "both in decimal (in row / cell format) and in hexadecimal form.

dd. Page 22, clause 27.3, First paragraph

Reword to: "The following ... in both the BMP and the SIP. The set of .. ".
If Plane 0, Plane 2 etc. is retained, it has to be changed to Plane 00 in Group 00 and Plane 02 in Group 02 etc.

ee. Page 22, Right Column, Bullet with 07-12 octet:

Suggest reword to "octet: Code position of corresponding CJK ... "

ff. Page 23, clause 28.2, Second paragraph, second line:

Suggest reword to ".. representation is their two-octet value expressed as four hexadecimal digits."

gg. Page 25, Figure 3

FDD0 to FDEF -- range should also be shown in black (not graphic characters).

hh. Page 26, Figure 4

Suggest removing the Black zones completely. Replace the Black Zones for 00-1F and 7F-9F with white zones with 'Controls' in each of these ranges.

ii. Page 27, clause 30

Suggest adding:
The Plane 01 of Group 00 is the Supplementary Multilingual Plane (SMP). (Similar to the first paragraph of clause 31, on page 28).

jj. Page 27, clause 30, first paragraph

Suggest rewording to:

> "The SMP is used for encoding graphic characters used in other scripts of the world that are not encoded in the BMP. Most, but not all, of the scripts encoded to date in the SMP are not in use as living scripts by modern user communities."

(This rewording makes it similar to what is in clause 31, and avoids any hints at scripts scheduled for the BMP or SMP etc. "... not already scheduled for encoding in the BMP" and ".. scheduled for encoding in the SMP ...", etc. parts of the current text hint at some other plan / document that has already decided in some fixed manner where the different scripts will be encoded. While WG2 / UTC use the Roadmap as a tool to guide in the categorization of scripts their possible planes etc. it is not appropriate to include this in clause 30. The word 'scheduled' can also imply some timetable behind it.)

kk. Page 1004 - Right Column

The list of collection names under Note 3, is part of 3, and would be better to be using the same point size as the text in the note. It is smaller than the point sizes of the text in the main clauses, but not small enough to match the point size of the text in the Note. This will ensure that the list is not considered to be normative and is part of the Note.

ll. Page 1009, clause A.5

The collections in this clause start from Unicode 3.1. It will be useful to add a NOTE to indicate the equivalent collection numbers corresponding to (even though not labeled as such) Unicode 2.0, 2.1 and Unicode 3.0 -- which were all entirely fixed BMP collections with one or two additional characters. Unicode 2.0 was = BMP Amd.7 collection 301, I think Unicode 2.1 was 301 + 20AC (the Euro Sign), Unicode 3.0 was BMP 2nd Ed 302 + One ??Greek?? character.

mm.   Page 1030 Right Column, Last line

There is a reference to Unicode V3.2 here. My earlier comment about reference for such mentions of Unicode in the text applies here too.

3.   More than just editorial:

Page 1027 Alternate format characters - Missing 2060 WJ (probably should be added to FPDAM2 of Part 1)

The character 2060 WORD JOINER should be added to the list .. it was targeted to be taking over the role of ZWNBSP (FEFF) and freeing up FEFF to be exclusively useable for BOM in the long run. I suspect its description will be identical to what is there for ZWNBSP in the current text. (Has to be verified with the text in Unicode, and the submitter of the proposal to add this character.)

Not sure if there are other Alternate format characters that have been added to the characters list, for which no explanation is added in this Annex (in Amd 1 to Part 1, Amd 2 to Part 1). There should not be any Alt Fmt Chars in Part 2.

4.   The following comment .. if accepted may be processed as a Technical Corrigendum to the standard.

Page 10 Clause 11
Not sure if the standard should have a specific SC identified in it as the one to maintain / update etc. If SC2 disappears, or JTC1 disappears, which is possible under normal reorganizations of ISO/IEC, the standard has to Revised / Amended to make a change to this Clause alone, in the future. I have not seen too many ISO standards -- other than Registry type of documents which even has a similar clause in it. Under normal ISO process the revision and updating of any ISO standard is undertaken by a TC/SC which is assigned to the task -- this is business as usual, and need not be entered inside the standard itself. I know this clause has been there from the beginning of this standard.

WG2 should consider deleting this clause entirely, or do not hardcode JTC1/SC2 in this clause.

========================= End of Comments =========================