Source: Japan
Title: Some issues in ISO/IEC 10646
Date: 2012-10-17

In Japan, a JIS committee is working on translation of ISO/IEC 10646:2012 into Japanese to publish it as a revision to JIS X0221. During the translation, the committee found some possible errors in ISO/IEC 10646:2012 texts. Japanese NB received and reviewed the findings, and decided to send the list to WG 2 for consideration.

The following list is organized in appearance order. Most of the items are simple editorial mistakes and minor technical errors, but some items may have major technical impact (item f, for example.)

a) In 4.1, NOTE 2, "They can be" appears a mistake and should read as "There can be". It further appears to be better to be rewritten as "There are" to avoid "can".

b) In 4.5, NOTE, "A character" should be "A graphic symbol".

c) In 4.44, the number "65536" should be written as "65 536" to follow the directives.

d) In 4.49, the definition for *row* is now incomplete, because we removed detailed specification formerly placed in clause 6. The suggested definition is: "subdivision of a plane consisting of contiguous 256 code points beginning at a multiple of 256."

e) In 4.59, the definition contains an undefined term *surrogate code unit*. The current 10646 has definitions of *high-surrogate code unit* and *low-surrogate code unit*, but it has no definition of (unqualified) *surrogate code unit*. In this particular context, just say "code unit" in place of "surrogate code unit" seems sufficient. Alternatively, we could add a definition of *surrogate code unit* as "either high-surrogate code unit or low-surrogate code unit".

f) In 6.3.8, the sentence "Future editions of this International Standard will not allocate any characters to these reserved code points." is wrong, because this subclause is for code points reserved for future standardization. The sentence should read as "Future editions of this International Standard may allocate characters to some of these reserved code points."

g) In 9.1, there is a word "CC-sequences", which should be "code unit sequences."

h) In 23.2, the 5th item, "5^h field" should be "5th field" ("t" is missing).

i) In 23.2, the 5th item, the format for JH source is specified as "`(JH-xxxxxx)`" but it is wrong. JH source reference is either 6 digits or 7 digits, so the 7 digits version "`(JH-xxxxxxx)`" should be inserted after the 6 digits version, e.g., "`(JA-hhhh), (JH-xxxxxx), (JH-xxxxxxx), (JK-ddddd)`". (Note that the last *digit* for the 7 digits version of JH source reference is always "S", so we can write the format as "`(JH-xxxxxxS)`" if it is preferred.)

   * This issue has been found by Prof. Kim Kyongsok.

j) In 31.2, EXAMPLE, the character name for 01C9 is written as "LATIN SMALL LETTER IJ", but it should be "LATIN SMALL LETTER LJ" ("LJ" for "IJ").

k) In A.1, the code point range for the collection 105 RTL ALPHABETIC PRESENTATION FORMS uses 2013 EN DASH to indicate a range, whereas other ranges use 002D HYPHEN-MINUS.

l) In A.1, definition for the collection 401 is inappropriate; it says "G=00" to mean "Group 00", but the notion of the *group* has been removed from 10646.   We can simply erase it.   The notation "P=" looks strange, because specification that explained such notation (in clause 6) has been removed.   It's better to say "Planes 0F and 10".

m) In A.1, NOTE 3 (the list of keywords appearing in the collection names), there is an entry "Counting Rod numerals" (with C and R in uppercase), but it was "Counting Rod Numerals" (C, R, and N in uppercase) in the previous edition.

n) In A.1, NOTE 3, the collection numbers for the entry "Game Tiles" is written as "1028, 1029" (with comma), but it should be "1028 1029" (without comma.)

o) In A.1, NOTE 3, the entry "Nko" should be spelled as "NKo" (with uppercase K.)

p) In A.4.2, NOTE 2, the file name "JIEx.txt" should be "JIExt.txt" (with "t" before a dot.)

q) In A.5.7, it says "Planes 00-10", but it should be "Plane 00", because all code points of BASIC JAPANESE are on the Plane 00.

r) In Annex I, there are I.1.1 and I.1.2 but no I.1.   They should be renumbered as I.1 and I.2, respectively.

s) Throughout A.6.x, there are many occurrences of a phrase "a collection from A.1" (or its plural form).   Repeating "from A.1" many times doesn't make sense.   It's better to remove the "from" part and simply to say "a collection" or "collections".   Alternatively, each "from" part can refer to the exact subclause that gives the

definition, as in past editions.

t) In Annex P, NOTE, 1st line, a closing parenthesis should be inserted after "amendments 1 to 5".

u) In many places such as 17, 22.2, 26, or A.3, there are texts that assume each row has an identification number in range 00-FF. It was true in the previous editions, but the current edition dropped it. Re-introducing the notion of "row number" seems a better idea than rewriting all related texts.