L2/08-421

Comments on ISO 80000-2

Submitted by Murray Sargent III

These comments have been collected from email discussions of ISO 80000-2.

1) For Unicode discussions, it'd be worth referencing Unicode Technical Report #25 Unicode Support For Mathematics. This is the latest version (as of 9/2008). I'd think it'd be worthwhile in working on mathematical text proposals to read UTR #25 carefully. A huge amount of effort has gone into making Unicode/ISO 10646 the ideal standard for mathematical characters. Feedback on the report is welcome.

2) The character repertoire in Unicode (which is identical by design to that of 10646) has been added to over the years with the explicit aim of making certain mappings possible - in particular those for use in other ISO standards, for example public entity sets. It's critically important that such mappings "by design" don't get reinterpreted by later work - or else the identity of the coded characters are no longer stable.

3) Typography: Some of the math is beautifully typeset; some of it is pretty awful. I'd use a good looking serifed font for all the math unless sans serif characters are really the standard. One could use [La]TeX. But currently it's quite a mixture of formatting.

4) In Annex A, the word Unicode is always written with an upper-case U and doesn't have a hyphen. And ISO 10646 should be mentioned at least at the first mention of Unicode. Something like Unicode/ISO 10646 .... I've never seen the word "sedecimal" before. Since IBM introduced "hexadecimal" way back in 1964 for the IBM 360 series computers and the PC industry has used it since 1975, I'd think you could just stick with hexadecimal. Might be handy to include the Unicode values in the main text as well as in Annex A.

5) The Unicode math alphanumerics are a great way to get the mathematical variables and constants to display correctly. See Section 2.1 Mathematical Alphanumeric Symbols Block in UTR #25. They are used extensively by Word 2007 and were proposed for addition into Unicode by the STIX committee. The characters have the correct interletter spacings, whereas corresponding letters using attributes like italic or bold may not have. In addition, they allow plain text to preserve important mathematical semantic distinctions. A script H isn't the same variable as an italic H.

6) p.10, 2-7.16, last line of remarks, the minus and infinity should *not* be split
p.12, 2-9.6, remarks l.-2, "breath" ==> "breadth"
p.35, title, "Uni-codes" ==> "Unicodes"
p.35, 2-4.10 (follows 2-5.9) should be 2-5.10

7) Mathematical symbol ambiguities, such as the difference between math italic v and Greek nu are explained in Unicode Technical Report #25, which discusses ambiguities, such as the characters in the table (see Sec. 2.3 of UTR #25)

8) Item 2-8.2 Instead of U+22A5 UP TACK, the character U+27C2 PERPENDICULAR may be more appropriate. Background: UP TACK comes from APL (binary operator for "evaluate a polynomial") but is also used in mathematics and computer science for various purposes, e.g. to represent a particular, lowest, or exceptional set element, whereas PERPENDICULAR is a geometric binary operator of similar shape.

9) Item 2-9.2: Instead of U+002D (HYPHEN-MINUS), the only correct character to use here is U+2212 (MINUS SIGN). Background: A hyphen is a short line used to separate or hyphenate words, whereas glyphs for the minus sign are usually longer, thinner and higher and must be identical to the horizontal stroke in the plus sign. For economic reasons, 20th century typewriters and teletype printers did not distinguish between hyphen, minus, en dash and em dash, and therefore the ASCII character set (which was derived from teletype printer practice) did not distinguish them either and included a compromise character called HYPHEN-MINUS. In typography, however, the two characters are very different. In today's wordprocessing practice, the Unicode character U+002D HYPHEN-MINUS is used as the hyphen, while the Unicode character U+2212 MINUS SIGN is used for the minus. (There is also a Unicode character U+2010 HYPHEN, which was added to have an unambiguous code point for the hyphen, but my understanding is that it is not commonly used in practice and may be deprecated.) For keyboarding (input) it's still handy to use 002D, which should be autocorrected to U+2212.

10) Item 2-11.6 U+2223 DIVIDES belongs more to item 2-7.17

11) Item 2-18.6 Instead of U+002A ASTERISK, the character U+2217 ASTERISK OPERATOR may be more appropriate.