standards reports

Some Standardization News

by Keld Simonsen

Keld Simonsen is active in ISO standardization, particularly in internationalization, POSIX, C, C++, and making it all available on the Web.

<keld@dkuug.dk>

Here are some updates on work in ISO standardization, in the field of IT, internationalization, and character sets.

JTC 1 – the group responsible for standardization of all IT in ISO and IEC – had a meeting in Tromsø, Norway, in November. Tromsø is a little below 70° North, and the days were shortening considerably during our stay – two weeks later it would have been always pitch dark.

There will be a trial to have a number of ISO standards available for a modest fee, say $25, for binary copies. Another trial will be continued for another free set of ISO standards; today about 50 standards are available this way. See *<http://isotc.iso.ch/livelink/livelink/fetch/2489/Ittf_Home/PubliclyAvailableStandards.htm>* for this list.

A proposal that participating companies could buy a voting membership for $25,000 a year on a level with national bodies was amended to a trial in which everybody in participating groups could take part on an equal level with national experts. Personally, I think it would have been very problematical if a dozen big firms had been allowed to buy a majority for an ISO standard, and I welcome everybody's individual participation in the trial without a fee. However, you still need to pay for your own travel.

A proposal to always use MS Word 7 as the internal document format in JTC 1 was changed to allow other document formats, including PDF, .txt, HTML, MS Word 6, WordPerfect 5.1, and RTF. I am happy that JTC 1 did not choose to only allow a non-standard document format, which is not fully supported with macros, for example, on some platforms, including Linux.

This is my pet area and there have been quite a few developments in standardization since my last snitch report. In August the convener of WG20 – the internationalization working group in ISO – reported in a personal note to the parent group SC22 that he thought that WG20 should be disbanded, a few projects transferred to some other standardization groups, and the rest of the projects cancelled. One of the main points was that the technology being standardized was arcane, as it builds on the C/POSIX locale internationalization model. Another point was that this area best be left to the industry to standardize. One project – the i18n API project ISO 15435 – had not progressed to the first ballot after three years and should, the convener thought, therefore be cancelled.

SC22 had quite a discussion on this and by a small majority decided to continue the project for one year and, furthermore, to ask WG20 whether the group itself thought it should be disbanded (the discussion report was a personal report from the convener and had not been discussed officially in WG20 before submission). In the WG20 meeting in November, a majority in the group disagreed with the view of the convener. WG20 also decided to send the cultural registry standard (which registers POSIX locales, charmaps, etc.) for its first ballot, to send the enhanced POSIX locale standard TR 14652 for its final DTR ballot, and to begin an addendum for the new sorting standard ISO 14651 covering the additions of characters to the ISO 10646 UCS standard. I think that this latter discussion derives from the USA's L2 group, which is very oriented toward character-set issues and which also represents the USA in the ISO character-set group SC22. The L2 group also has close con-

Our standards report editor, David Blackwood, welcomes dialogue between this column and you, the readers. Please send your comments to *<dave@usenix.org>*

nections with the Unicode Technical Committee, as they hold all their meetings together. Some of the WG20 group work is in direct competition with the Unicode work. Other US groups are quite supportive of the WG20 efforts: the C and C++ groups have implemented the WG20 guidelines on what characters/letters can be used in extended identifiers, while the new revision of the COBOL standard will use the specification in the enhanced locale standard on how to map from uppercase to lowercase for all of 10646. I hope that the countries can come to an agreement on common standards about internationalization, and that individual countries will not try to either push their own standards through or sabotage international standardization in this area.

ISO 10646 – the huge character standard – is now being extended beyond the 16 bits. This is ISO 10646-2, which is now out for its final FDIS ballot. It contains mainly an extension with many ideographic (Chinese/Japanese/Korean) characters, and a few exotic scripts, plus some characters for language declarations. Two bytes are thus not enough any longer for UCS. The UTF-8 format of UCS is now increasingly being implemented in UNIX and Linux. Kde is doing all its messages in UTF-8, and glibc 2.2 supports conversion of all messages to UTF-8 via the gettext package and iconv. Glibc 2.2 even supports TR 14642-style transliteration if the execution character set does not contain specific characters in a message. The ISO 10646-1:2000 standard is now available for about $50 on a CD from ISO. This is not that expensive compared to the paper version that is priced according to normal (expensive) ISO rates. A number of 8-bit character sets are being finalized, including a character set supporting Romanian, and a revised 693 standard that covers most Latin letters with accents.