

To: LSA and UC Berkeley Communities
From: Deborah Anderson, UCB representative and LSA liaison
Date: 24 March 2005 (revised 4 May 2005)
Topic: Unicode Technical Meeting #102, 7 -10 February 2005, Mountain View, California

[The comments below are written for the LSA and UC Berkeley communities and others who may find them of interest. They are not to be considered official or exhaustive; for the full minutes, readers are directed to the UTC minutes page on the Unicode Consortium website where the minutes will be posted soon at <http://www.unicode.org/consortium/utc-minutes.html>.]

1. Report on ISO/IEC JTC1/SC2/WG2 (Universal Character Set) ("Working Group 2", or WG2) meeting in Xiamen, China, 24-28 Jan. 2005

The Final Preliminary Draft Amendment 1 (FPDAM1) to ISO 10646:2003 passed. Scripts contained in the amendment include Buginese, Coptic, Glagolitic, New Tai Lue, Nuskhuri (to extend Georgian), Syloti Nagri, Tifinagh, Kharoshthi, and Old Persian Cuneiform, as well as a host of new characters. Unicode 4.1, was released 31 March 2005 on the Unicode Consortium website (www.unicode.org), will include these new characters and scripts.

The Preliminary Draft Amendment 2 (PDAM2) to ISO 10646:2003 also passed, though a few national bodies voted "no." Amendment 2 includes the following scripts: Phags-pa, Sumero-Akkadian Cuneiform, Phoenician, N'Ko, and Balinese. Another round of technical balloting (which will enable national bodies to send in comments) will take place over the course of the next few months. The comments on this amendment will be discussed at the next WG2 meeting, which takes place in Sophia-Antipolis, France during the week of 12 September. Note that attendees must pay a fee of 50 EUR per day to help offset hosting costs. The scripts and newly proposed characters in Amendment 2 are slated to appear in Unicode 5.0. (The Amendment can be viewed at: <http://std.dkuug.dk/jtc1/sc2/wg2/docs/N2936.pdf>.)

2. UAXes and UTRs

Updates and corrections to several Standard Annexes (UAX), Technical Standards (UTS), and Technical Reports (UTR) were discussed and many were approved for posting on the Unicode.org website.¹

¹ UAX #9 Bidirectional Algorithm, UAX #14 Line-breaking Properties, UAX #15 Normalization Forms, UAX #29 Text Boundaries, Draft UAX #34 Unicode Named Character Sequences, UTR #10 Unicode Collation Algorithm, UTR #33 Unicode Conformance Model, UTR #36 Security Consideration, UTS #22 Character Mapping Mark-up Language, Preliminary Draft UTS #37 Variation Selection Registration

3. *Line-Breaking*

Details on line-breaking were requested for specific characters in several scripts, including Kharoshthi, Coptic, Greek, and Tibetan. It was also highly recommended that all future script proposals include line-breaking properties. Prose descriptions are suitable, as long as they describe whether a line-break is allowed before (or after) a particular character, and if a space is necessary before or after a character which occurs at a line-break.

4. *Script Proposals*

a. Greek Epigraphic Characters

The proposals by Nick Nicholas for Greek heta and Greek epigraphical characters (archaic sampi, “tsan”/Pamphylian digamma, and Corinthian EI) were discussed. Clarifications regarding the requested characters were needed; comments from the Committee will be sent to the author.

b. Uighur Latin characters

Lorna Priest of SIL International proposed 6 characters for the Uighur Latin alphabet (also used for the Judeo-Tat orthography from the 1920s and 1930s). These were approved. The characters include (with the following codepoints):
U+2C65 LATIN CAPITAL LETTER H WITH DESCENDER
U+2C66 LATIN SMALL LETTER H WITH DESCENDER
U+2C67 LATIN CAPITAL LETTER K WITH DESCENDER
U+2C68 LATIN SMALL LETTER K WITH DESCENDER
U+2C69 LATIN CAPITAL LETTER Z WITH DESCENDER
U+2C6A LATIN SMALL LETTER Z WITH DESCENDER

The proposal (with glyphs of the characters) can be viewed at <http://std.dkuug.dk/jtc1/sc2/wg2/docs/N2931.pdf>.

c. Balinese

The Balinese script proposal, which had been approved for Amendment 2 of 10646, was discussed. The UTC had no objections to the characters, their names, or code positions. Line-breaking properties and combining classes are needed before the script can be approved. (The Balinese script proposal is available at: <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2908.pdf>.)

d. Lepcha

A proposal from the Government of India for Lepcha was discussed, and a recommendation was made for the authors to work closely with others--

including Michael Everson, who also has a proposal for this script--so a single, final proposal can be submitted to the UTC.

e. Vai

This script will be proposed in the near future. Work is ongoing and a preliminary codechart with nameslist is available, both for the modern and archaic characters.

f. Indic script issues

Some of the characters requested at this meeting by India may be eligible for encoding, but some are not (without additional evidence). In order to improve communication between the UTC and the Government of India, specific steps were outlined in an ad hoc meeting.

5. Language Tags

The UTC endorsed the development of a successor to RFC 3066.

6. Other Misc.

* 1034A GOTHIC LETTER NINE HUNDRED was given a numeric value in UnicodeData.txt. Other characters marked as numbers should likewise be checked to verify they have numeric values assigned to them.

* An update on Common Locale Data Repository (CLDR) was provided. (Note: Input on CLDR is encouraged, especially reports on missing or incorrect data.)

7. Public Review Issues and Comments

Various Public Review Issues (PRI) were discussed, along with comments made on the online input form. Discussion on the disunification of the dandas question (PRI #59) will be extended as a Public Review Issue until May 3.

8. Upcoming meetings:

May 10-13, 2005 (SF Bay Area)

August 16-19, 2005, Redmond, Washington

November 1-4, 2005, San Jose, CA

Feb. 6-9, 2006 Bay Area