

To : UTC
From: Peter Scharf, Brown University
Date: 2 February 2007
RE: Vedic Unicode Workshop Report

Vedic Unicode Workshop Report

The International Digital Sanskrit Library Integration project at Brown University sponsored a workshop 14-18 January 2007 to draft recommendations to the Unicode Consortium to revise the Unicode Standard to include characters necessary for encoding Vedic Sanskrit in the major Indic scripts. Vedic is of paramount importance because it is the principal language for the religious heritage of India and because it represents the earliest attested phase of the Indic language family. The meeting included the following participants: Peter Scharf (PI) (chair), Malcolm Hyman (co-PI), Susan Rosenfield (post-doctoral research associate), R. Chandrashekar (research associate), Michael Everson (Unicode technical committee member; typographer; expert in preparing non-Roman script recommendations for Unicode), Michael Witzel (Harvard University; recognized authority in Vedic), Michel Angot (L'École des hautes études en sciences sociales; authority in Vedic), B. V. Venkatakrisna Sastry (Hindu University of America; authority in Vedic).

The meeting began with a brief review of encoding principles and evaluation of previous Vedic encoding standards and proposals, including principally the Indian Standard Code for Information Interchange Annex G (1988), the draft proposal of Omkarananda Ashram (www.omkarananda-ashram.org/Sanskrit/Vedicaccents.htm), and the proposal of the Centre for Development of Advanced Computing (C-DAC) in Pune.

Unicode has not accepted previous proposals to extend the Devanāgarī block to include Vedic characters owing to two fundamental shortcomings in the proposals. First, the principles guiding the proposed sets of characters were inconsistent with Unicode's principle of encoding contrastive graphic signs. Second, the proposals were insufficiently comprehensive. They did not present thorough evidence from printed and manuscript sources, nor did they propose corresponding extensions to other Indic scripts besides Devanāgarī.

The workshop examined the usage of Vedic characters in Devanāgarī, Bengali, Kannada, Oriya, and Telugu, as well as Sharada and Grantha scripts, in printed texts and manuscripts. The group concluded with specific recommendations in a draft proposal. Most Indic scripts utilize a standard set of Vedic characters, usually copying the glyphs used in Devanāgarī, but often either modifying them to distinctive glyphs or inventing characters by combining existing glyphs. Sharada and Grantha scripts, not yet included in the Unicode standard, provide independent sources for Vedic characters. The workshop concluded by drafting a Working Group Document, entitled "Proposal to encode characters for Vedic Sanskrit in the BMP of the UCS," for consideration by the Technical Committee 1 (JTC1), the International Standards Organization (SC2), the Unicode Consortium's Working Group 2 (WG2) and the Unicode Technical Committee (UTC) at meetings early in 2007. The proposal recommends additions to the Devanāgarī code block and the allocation of a code block for Vedic Extensions. The proposal will be circulated for comment in the Vedic discussion list at Evertime.com. Vedic scholars worldwide will be invited to participate.