

Software-Aids for Bharatiya Text Creation

N. S. Venkatesh
(30-8-2021)

कृत्वा मूर्ध्नि गुरोराज्ञां शास्त्रग्रन्थप्रगोपने । गुरोरस्मिन् महायत्ने विज्ञप्ये मे लवान्वयम् ॥

Bestowed with adhyayanam and kālakṣepams by Sri U. Ve. Ramanujam Svami since 1994, I have been fortunate to be assigned kaimkaryams in Svami's efforts at preserving sampradayaic texts. A gist of these is presented here.

BhashaIME

The IME is Windows software for typing Unicode text in South Bharatiya scripts. Devanagari, Grantha, Kannada, Malayalam, Tamiz and Telugu are supported currently. Additionally, Roman script with IAST and ISO schemes are also supported. All Vedic svāra-s specified in Unicode can be keyed in.

The IME is equipped with three keying schemes – Phonetic, Baraha and Inscript; Phonetic being the simplest with almost zero learning curve. The input schemes are user-configurable through simple text files.

The s/w is designed to work with any Bharatiya Unicode script. Users can easily create/modify keyboard layouts for newer/existing scripts.

BhashaTrans

Significant amount text corpus is freely available on the internet that has been created using pre-Unicode (ASCII) standards. The fonts and encoding used are multifarious. BhashaTrans transliterates more than 150 of these to Unicode.

Texts in Unicode are freely interconvertible among scripts (e.g., Devanagari to Grantha, Kannada to IAST). This helps in easy maintenance and multi-script distribution of documents.

The s/w is designed with a unique feature of preserving the layout of texts being converted. Word documents will retain all the features (e.g., header, footer, footnotes) in toto after conversion.

Another significant feature is handling MS Word/rtf documents that use multiple fonts. Texts in the document are processed with respective font-specific conversions automatically.

Transliteration is very configurable; e.g., one can choose to convert कृष्णं वन्दे जगद्गुरुम् to కృష్ణణం వందే జగద్గూరూం OR కృష్ణణం వందే జగత్తృక్రూం. Users can set these features through text files.

Transliteration schemas are defined in simple text files. These can, once again, be created/modified by users.

Fonts with Vedic svāra support

Two Kannada fonts, Goda & Vagisha, one Telugu font, Pothana2000Vedic, have

been created for preparation of Vedic texts. Free and aesthetically good fonts obtained on the internet, have been modified and enhanced for clarity and Vedic support.

A new font for Nandinagari script has been developed based on the book published by Dravidian Linguistics Association. The font uses Unicode-Devanagari-space since the operational features of Nandinagari are very similar to Devanagari. The font was deemed to supplement efforts of digitizing manuscripts written in Nandinagari.

Scripts for OCR

Google provides high quality free OCR service, but only one/small document can be converted manually. To facilitate conversion of large documents, Google scripts have been created that can handle hundreds of documents a day.

These efforts appear to be helping a significant number of shrivaishnava-s.

All software and related documentation is freely available at
<https://sites.google.com/site/bhashaime/>.

विद्यायां माम् अवैदिक्यां सक्तं शास्त्रेष्वपण्डितम् । तामेव सार्थके युङ्क्ता करोति कृतिनं हरिः ॥

श्रीकृष्णार्पणमस्तु ॥

Sri Venkatesh is BE (Electronics) from UVCE, Bangalore, 1990; ME (Comp. Science) from IISc, Bangalore, 1992. He has worked in ITI, Bangalore from 1992 till 2006 (14 years) and took voluntary retirement since then. From 1993, he has received Veda adhyayanam, Divyaprabandham, Deshika Prabandham, Rahasya kalakshepams etc. for 20 years from Dr. P. Ramanujam. An ācāra-śīla and continuing the lineage of Somāśiyāṇḍān Ācārya vaṁśa, Śri Venkatesh is ever engaged in 'Software Kainkarya', helping many Samskrta and Veda scholars/researchers/students in using Bharatiya bhashas in the Computers, through many of his free, open-source software that are highly appreciated and are very popular.

