Vedic Sanskrit; on the way of Digitization

Siji Sunny, Jui Mhatre
CDAC Mumbai, India

Introduction

Vedic Sanskrit, an ancient Indian language exhaustively used to preserve and spread Indian wisdom through oral mode. The utterances of Vedic mantras and prayers were authenticated by an exhaustive sign system with intonation marks developed when the Vedic text was written down.

Font design in context to Vedic Sanskrit language is a challenge considering its complex orthographic structure, multi-tier usages of diacritic marks of complex compositions, signs that appear above, below and at sides of the base character, integrated multi-consonant conjuncts, multiple syllabic combinations.

The text process Engine IndiX changed the concept of Simple script like Latin to complex scripting like Vedic Sanskrit. IndiX has based the grouping and reordering of characters on the Unicode standard and the minimal model of fonts. IndiX uses the ISO/IEC Technical Report 15285, which gives an recommendation for indic text processing.

The complexity of Vedic Sanskrit is implemented by IndiX followed by font processing. 5Tier Vedic Sanskrit OpenType font has been designed recently under Project IndiX using undefined Unicode values at C-DAC (Centre for Development of Advanced Computing, under Ministry of Communications and Information Technology, Govt. of India).

Indian oral tradition (Background)

Sanskrit has its own place recognized by the linguists all over the world. It is one of the most ancient languages of the world and has molded the Indian culture, rich Indian traditions and the Indian thought systems in the fields of Arts, Sciences and Philosophy. Sanskrit at the period of Vedas was exhaustively used to preserve and spread Indian wisdom through oral mode. The ancient sacred Vedas are reservoir of information, knowledge and wisdom of Indian past. This knowledge has been passed on from the teacher to the student using oral mode and its associate techniques such as recitation, repetition, memorization and oral reproductions etc. Clear and correct pronunciations of syllabic clusters by the teacher as well as attentive and meaningful listening by the pupils were integral parts of this methodology.

The utterances of Vedic mantras and prayers, were authenticated by an exhaustive sign system with intonation marks developed when the Vedic text was written down, thus providing a model of parallence in linguistic sounds and their orthographic equivalences.

Preservation of the ancient wisdom

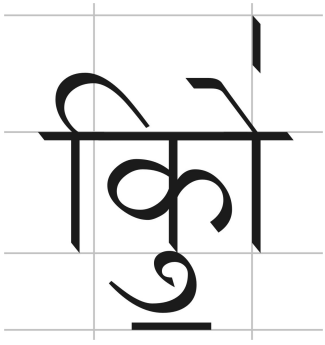
In order to preserve this ancient knowledge for the future generations there is a need to store this information which can be possible in the digital world today. Digital text generation in Vedic Sanskrit language is a means to maintain and pass this wisdom further. However Type font design and implementation of Indian scripts and languages for inputing, displaying and printing is complex as compared to Roman script. Text in Roman script generally appears in a linear way whereas Indian scripts have specific orthographic behavior.

Font design in context to Vedic Sanskrit language

Text in Roman script generally appears in a linear way whereas Indian scripts have specific orthographic behavior.

is a challenge considering its complex orthographic structure, multi-tier usages of diacritic marks of complex compositions, various signs that appear above, below and at sides of the base character, integrated multi-consonant conjuncts, multiple syllabic combinations.

Diagram – 5 tier structure of Vedic Sanskrit



The range of characters

Given the complex 5 tier structure of Vedic Sanskrit text, the font demands adequate range of characters as well as exhaustive rendering rules to achieve the advanced typographic quality. The range includes vowel signs, vowel modifying signs, consonant signs, half forms of consonant signs, consonant modifying signs, integrated multiple consonant syllables (conjuncts) ranging from integrated 2, 3, 4 and 5 consonant integrated conjuncts, various other signs to show nuances of spoken language like time duration, stress, vibration as well as specific intonation signs.

Diagram - Range of characters

The Rules

An exhaustive number of compositional rules and tables are applied to the font in order to generate text and the complex

syllabic structure of Vedic Sanskrit. Various rules for joineries, combining characters (conjuncts), variant glyph forms, substitution of glyphs (characters), positioning of vowel signs and intonation signs above or below the vowel signs are written to obtain the orthographic and linguistic needs of Vedic Sanskrit language.

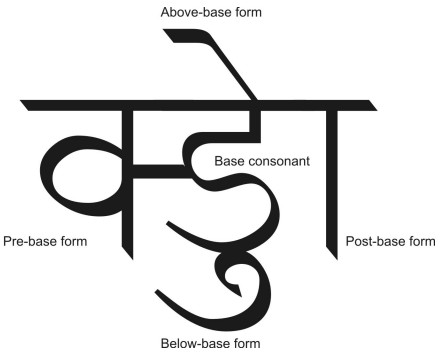
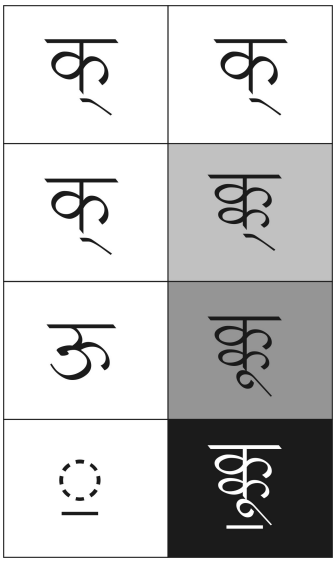
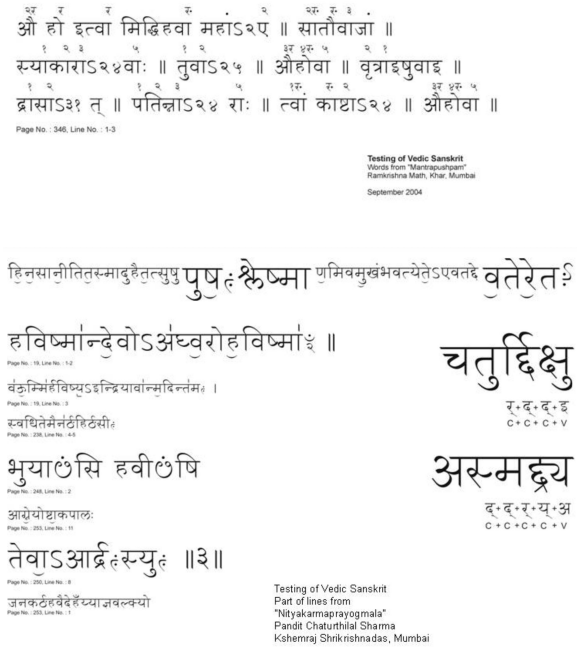


Diagram - Output through Substitution and Positioning tables



Vedic Sanskrit font contains 1174 characters (glyphs) in total with tables that include 10 Features, 36 Substitution lookups, 6 Positioning lookups, 82 ligature rules, 68 glyph groups, 253 consonant conjunct ligatures.

Diagram – Testing of Vedic Sanskrit font



This font with the help of exhaustive rules and font processing can generate about 24,702 multiple syllabic combinations. The information, the knowledge and the wisdom of Indian culture/tradition will be now exposed globally.

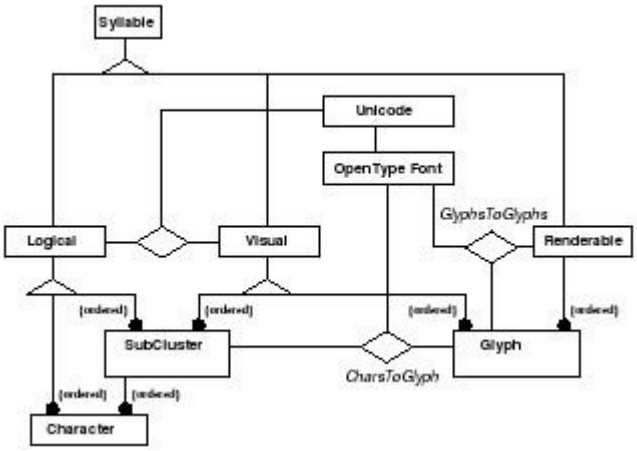
Rendering Vedic Sanskrit IndiX

Rendering of Latin (Internationalized) text is often assumed to be simply matter of one to one mapping. All languages are thought to be like English where it is simply a matter picking the right symbols from the font and displaying by the help of rendering engine in the order they occur in the memory representation this view internationalized rendering is very simple. But in indic script specially in Vedic Sanskrit the unit of orthography is based on syllable, which consists of one or more characters that can be rendered by a sequence of glyphs.

For Getting Vedic Sanskrit support in GNU/LINUX platform we used Indix library as for rendering the Vedic Text followed by font processing.

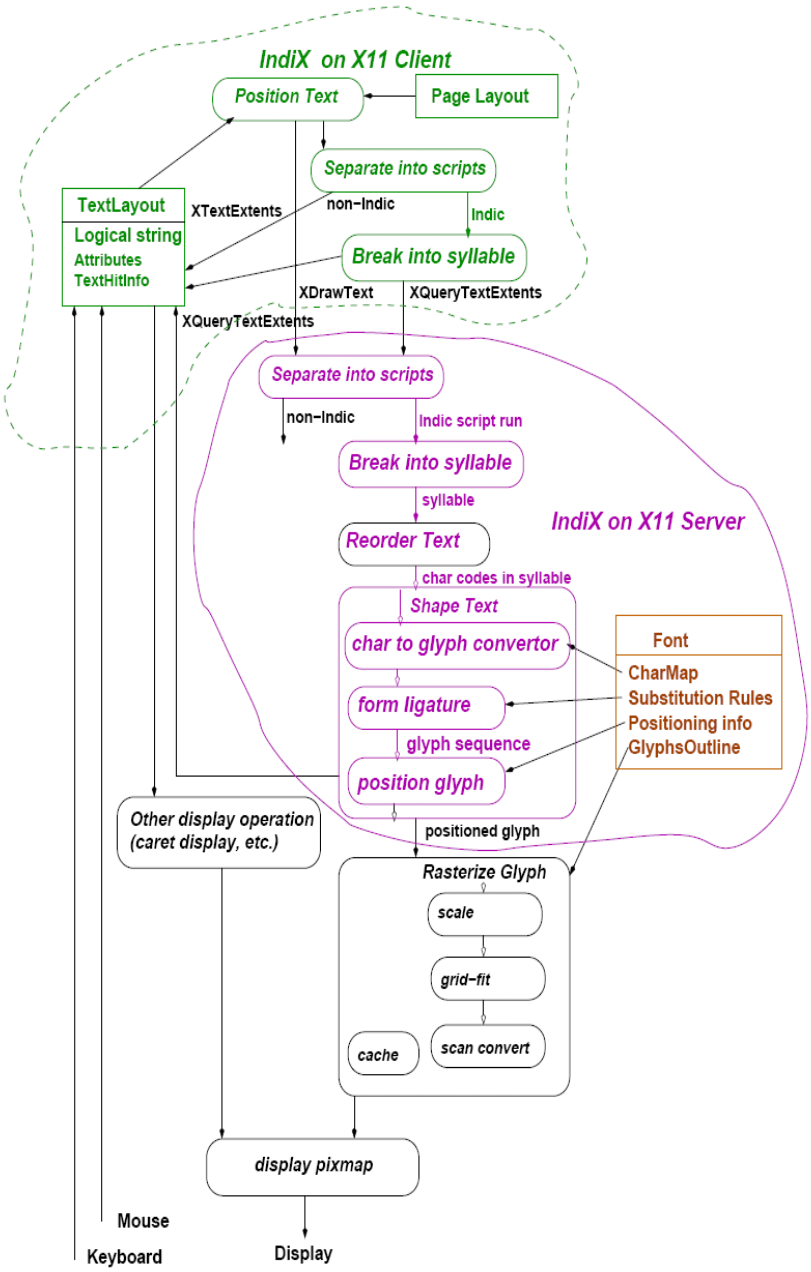
Indix rendering engine, takes text content and formatting information and displays the formatted content on the screen. It "paints" on the content area of a window, which is displayed on a monitor or a printer. Currently This rendering engine is used for Mozilla Browser and Moziila composer for displaying (and editing) of Vedic contents.

Developing a local language capability at system level is better than developing it in application level, Indix followed was to make changes to the software at the core level modified the lowermost layer of X window system, the X-server to support Vedic Sanskrit.



Object Model For Characters Indix

Indix analyased and Enabled the same way how the X11 server handles the fonts and converts Latin text characters to glyphs on the window screen,IndiX bases its characters to glyphs processing on the ISO\IEC 10646 model with small modification. So, for example, all the characters of an Indic syllable are treated as an indivisible unit like a single Latin character and presented for conversion to glyph at the lowermost level of the font. It is the font that has to form conjuncts and other shapes and place them appropriately. In IndiX these operations are not carried out at the higher levels as a text layout operation.



Indix Architectural Diagram

In the X server the text is separated into Indic and non Indic runs based on the Unicode range alloted to the Indic script. But in Vedic Sanskrit yet Unicode is not allocated any values so we used 0880- serious values for the Vedic symbols and defined these values in Indix Client library.

The indic sequence broken into syllable,the characters are recored to facilitate proper display of the forms those characters. After these transformation ,the text is shaped. Using the CharMap in the font,the char codes are converted to glyph indices.Using substitution rules,groups of glpyh indices are converted to indices of their ligature or alternate forms.Then using information from the font,the final glyph are positioned.This has been done by OpenTypefont,that include the substitution and postilioning rules appropriate to the type faces with in the font

Most display system have an efficient glyph rastering machinery that can take a sequence of glyph, scale the glyph outline according to the point size and scaling of the display window,grid fit the points on the outline and scan convert the outline to bitmap relative positioning of the glyphs is generally decided by the glyph metrics and so the machinery can rasterize a glyph sequence fast. This rasterised pixmap can be displayed by most devices in hardware .

Input Methodology for Vedic Sanskrit

There are two ways to map Devanagari keyboard on X window, using X Keyboard Extension (XKB) and using xmodmap. We had created a keyboard map file for xmodmap. By using the utility xmodmap to map Vedic keyboard.Normally xmodmap is used to load a keyboard configured file.For most Linux distributions, when you start X window with startx, X server will find Xmodmap in /etc/X11/ first. And by using a single line command can start Vedic typing.

Acknowledgment

We acknowledge Mr Zia Saquib, Executive Director, C-DAC Mumbai and Dr Alka Irani, Language Computing group, C-DAC Mumbai,Professor.R.K.Joshi and Vinod Kumar .

References

1 IBM and others. International Components for Unicode(ICU 2.8 Documentation, 2003)

2 IndiX homepage http://www.cdacmumbai.in/projects/indix

3 Sandeep Rao, Vinod Kumar, Standards for Visual sequence of characters for Indic Scripts, C-DAC Mumbai, Mar 2005

4 Edwin Hart, Alan Griffee. An operational model for characters and glyphs. Technical report ISO/IEC TR 15285, 1998

5. R.K.Joshi ,Dr.T.N.Dharmadhikari and Dr.Vijay Vasudev Bedeker,29-31 October 2007 “The Phonemic approach for Indian Text with Special Reference To Sanskrit -The first International Sanskrit Computational Linguistic Symposium and Workshop organised by INRIA at Rocquencourt,Versalles,France