Open Font Format Specification for Scripts of India

Elaborate, Illustrated font development guideline document for scripts of India by font designers, developers, language experts.

Author
Santhosh Thottingal

June 7, 2013

Dedicated to all font designers

Contents

1 Introduction **1**
Need for document 2
Objectives of document 3
Target Audience for this document 4
Scope 5
How to use this document 6
Notes on Collaboration 7

2 General Concepts **9**
Complex script 10
Glyph 11
Ligatures 12
Cluster/Syllable 13
Akhand 14
Ra Forms 15
Split Matra 16
Reordering 17
Zero Widh Joiner 18
Zero Width Non Joiner 19
Stacking 20

3 Open font format **21**
Introduction 22
GPOS 23
GSUB 24
GDEF 25
Shaping Engine 26
Shape Glyph sequence 27
Position Glyph sequence 28

Reference fonts 29
Reference Rendering Engine 30

4 Devanagari **31**

5 Tamil **33**

6 Kannada **35**

7 Panjabi **37**

8 Telugu **39**

9 Gujarati **41**

10 Odiya **43**

11 Malayalam **45**
Introduction 46
Orthography variation 47
11.0.1 Quasi Vowel Consonants 47
Reference fonts 48
11.0.2 Meera Font 48
11.0.3 Lohit Malayalam Font 48

12 Glossary **49**

13 References **51**

14 Appendices **53**

1

Introduction

Need for document

In present era if we simply do the comparison between availability of Latin and Indian fonts one can easily observe so much different. only handful fonts available for Indian language compare to thousands or even more fonts and style available for Latin script.

Objectives of document

Target Audience for this document

Scope

- This document covers only languages and scripts recognized in India.
- This document covers all complex OFF feature required for Indian scripts.
- This document is based on the current expertise of community members working/worked on this area.
- This document is from the typography perspective for each script and many not be linguistically correct.
- This document does not cover design or calligraphy style aspects but covers only technical aspects.
- This document is based on Unicode 6.2 and ISO/IEC 14496-22:2009 (Second Edition) "Open Font Format" standard.
- This document is not a tutorial on font design or development and does not teach typography.

How to use this document

Elaborate, Illustrated font design guideline document for Indic fonts by font designers, developers, language experts. - designer freedom to adapt

Notes on Collaboration

2

General Concepts

Complex script

[What is a complex script? What makes it complex, some examples, screenshots] 2-3 paragraph + images How it differs from simple scripts like Latin Character

Glyph

Ligatures

What is a Ligature, example, image How to identify a ligature Virama AKA - Pulli, Chandarakkala, Vattu, Halant

Cluster/Syllable

Akhand

Ra Forms

Explain Reph Explain Rakar

Matra

[explain prebase, postbase matra with examples and images]

Split Matra

Reordering

Zero Widh Joiner

Zero Width Non Joiner

Stacking

3

Open font format

Introduction

GPOS

GSUB

GDEF

Shaping Engine

[Some conent from http://behdad.org/text/ can be used]

Shape Glyph sequence

Position Glyph sequence

Reference fonts

To illustrate the concepts in this document, we will be using a set of reference fonts for each script, some times more than one per script. List of fonts table

Reference Rendering Engine

We are using Harfbuzz as a reference rendering engine. Examples given are working perfectly with Harfbuzz but any rendering engine conforms to the Open font specification will give same result.

4

Devanagari

5

Tamil

6

Kannada

7

Panjabi

8

Telugu

9

Gujarati

10

Odiya

11

Malayalam

Introduction

Like many other Indic scripts, it is an abugida, or a writing system that is partially "alphabetic" and partially syllable-based. The modern Malayalam alphabet has 13 vowel letters, 36 consonant letters, and a few other symbols. The Malayalam script is a Vattezhuttu script, which had been extended with Grantha script symbols to represent Indo-Aryan loanwords. The script is also used to write several minority languages such as Paniya, Betta Kurumba, and Ravula. The Malayalam language itself was historically written in several different scripts.

As is the case for many other Brahmi-derived scripts in the Unicode Standard, Malayalam uses a virama character to form consonant conjuncts. The virama sign itself is known as candrakala(ചന്ദ്രക്കല) in Malayalam.

When the candrakala sign is visibly shown in Malayalam, it usually indicates the suppression of the inherent vowel, but it sometimes indicates instead a reduced schwa sound, often called "half-u" or samvruthokaram. In the later case, there can also be a -u vowel sign, and the base character can be a vowel letter. In all cases, the candrakala sign is represented by the character U+0D4D malayalam sign virama, which follows any vowel sign that may be present and precedes any anusvara that may be present.

FIXME samvruthokaram needs more explanation, refer `http://thottingal.in/documents/Malayalam_Unicode_Report_of_Workshop_Kerala_University.pdf` and `http://smc.org.in/doc/rachana-malayalam-collation.pdf`

Orthography variation

Malayalam has two orthography variations, both actively used. Generally they are known as Old orthography and Modern orthography. Old orthography is also known as traditional orthography. Modern orthography is also known as reformed orthography.

Old orthography is generally identified by large number of ligature glyphs, often formed by more than one consonants or vowel combination. For example, in old orthography, consonant ക + vowel ൂ will form as a single ligature. similarly a conjunct ക + ് + ത will form ക്ത as a single ligature. Font developers of traditional stye fonts design and draw this large set of glyphs. For a reference font Meera, the number of glyphs is approximately over a thousad.

In the 1970s and 1980s, Malayalam underwent orthographic reform due to printing difficulties. The treatment of the combining vowel signs u and uu was simplified at this time. These vowel signs had previously been represented using special cluster graphemes where the vowel signs were fused beneath their consonants, but in the reformed orthography they are represented by spacing characters following their consonants.

// FIXME: insert image here to illustrate both orthography variation. Refer chapter 9 of Unicode standard page 307 for a sample image

11.0.1 Quasi Vowel Consonants

In Malayalam some consonants like YA(യ), RA(ര) are quasi-vowels and they have post(pre)/below base forms. As per OT spec the pstf(pref) and blwf features are always applied and this leads to weird shaping when some consonants come as base. Eg. യ, യ്യ, ര, ല. This is more prominent in reformed scripts because one can do away with these forms in traditional scripts.

Reference fonts

Since we need to illustrate both orthographies we will be using 2 fonts. For old orthography, we will use Meera font and for modern orthography we will use Lohit Malayalam.

11.0.2 Meera Font

മീര മാതൃക // FIXME: Short introduction, designers, maintainers, usage infor, popularity of the font

11.0.3 Lohit Malayalam Font

// FIXME: Short introduction, designers, maintainers, usage infor, popularity of the font

12

Glossary

13

References

State of Text Rendering By Behdad Esfahbod <behdad behdad com>
http://behdad.org/text/

14

Appendices

[Editor notes]