

കമ്പ്യൂട്ടറിലെങ്ങനെ

മലയാളം കാണുന്നു?

The text rendering stack in free software world

Santhosh Thottingal

You need

data

data

data

d	64	10000000
a	61	111101
t	74	1001010
a	61	111101

data

d	64	01000100
a	61	0011101
t	74	1001010
a	61	111101

ASCII

1000000 111101 1001010 111101

64

61

74

61

d

a

t

a

1000000 111101 1001010 111101

64

61

74

61

d

a

t

a

data

1000000 111101 1001010 111101

64

61

d

a

t

data



Font

1000000 111101 1001010 111101

64

61

Font

d a t a

data

മലയാളം

മലയാളം

മ	?	?
ല	?	?
യാ	?	?
ളം	?	?

മലയാളം

മ	?	?
ല	?	?
യ	?	?
ാ	?	?
ള	?	?
ം	?	?

ASCII is 8 bit

256 characters

Unicode

മലയാളം

മ	3374	0D2E
ല	3378	0D32
യ	3375	0D2F
ഓ	3390	0D3E
ള	3379	0D33
ഌ	3330	0D02

മലയാളം

എഴുതാനറിയാമോ?

മലയാളം




~~എഴുതാനറിയാമോ?~~

ടെക്സ്റ്റ് ചെയ്യാനറിയാമോ?



	3349	0D15
	3405	0D4D
	3349	0D15



	3349	0D15
	3405	0D4D
	3349	0D15





Complex script rendering

റെക്ക

ക	3349	0D15
െ	3398	0D46



	3349	0D15
	3498	0D46



Complex script rendering

ക + േം → കേം

Complex script rendering



Complex script rendering

Ⓔ+Ⓕ+Ⓔ → Ⓔ

Complex script rendering

ഖ് + ന → ഖ്ന

Complex script rendering

ക + ് + ഷ → ക്ഷ

Complex script rendering

ന്+ത്+ര+് +യ → (നതൃ)

Complex script rendering

സ + ് + ത + ്
+ ര + ീ

→ (സ്മി)

Complex script rendering

ശ + റ് + ശ + ൾ → ശ്രീ

Complex script rendering

Font

Font

Glyph drawings

Rendering Rules

Font

Glyph drawings **#art**

Rendering Rules **#program**

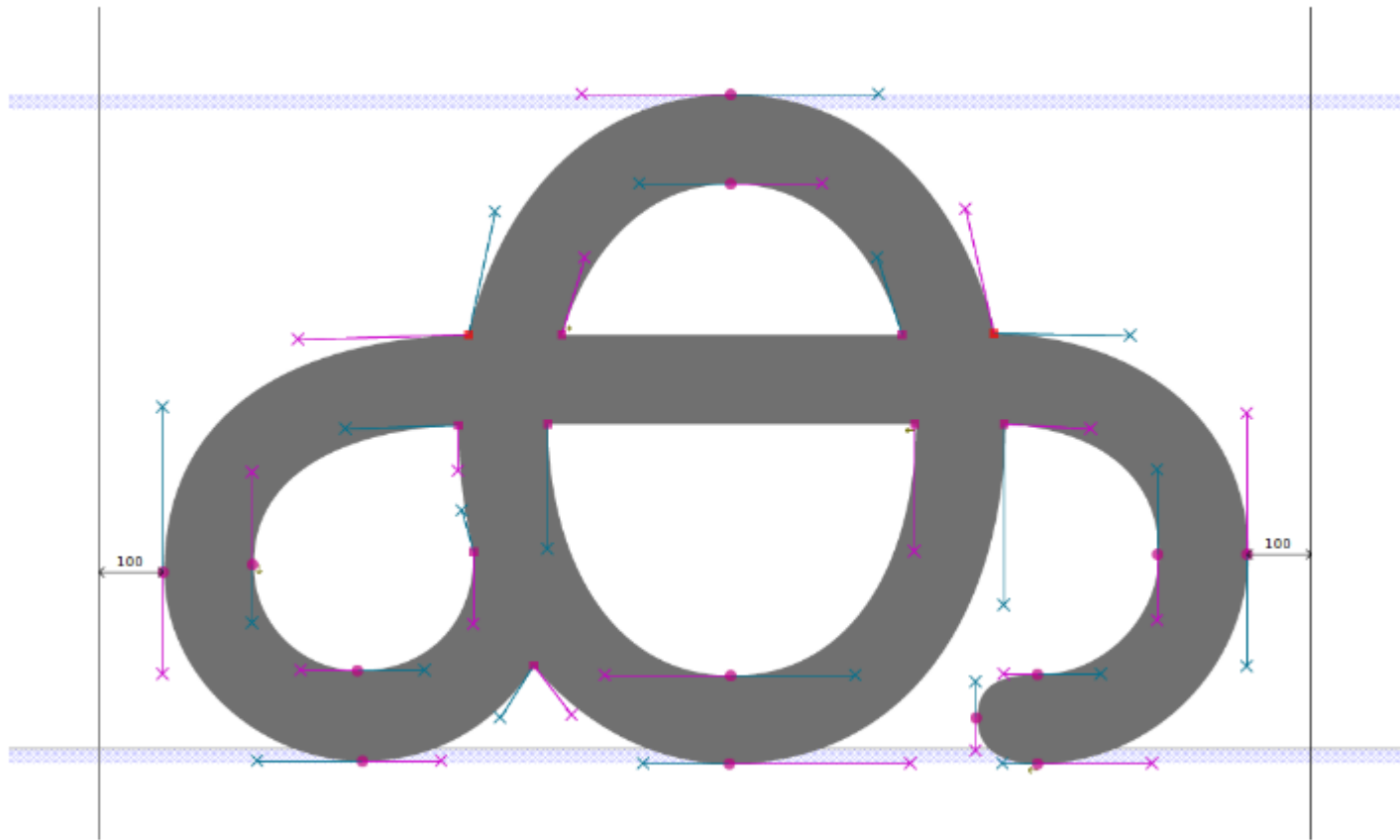
Font

Glyph drawings

Rendering Rules

Opentype

Glyph



Rendering Rules

```
feature akhn {  
    script mlm2;  
        language dflt ;  
        lookup chillu;  
        lookup akhn_conjuncts;  
        lookup akhn_conditional_stacking;  
    script mlym;  
        language dflt ;  
        lookup chillu;  
        lookup akhn_conjuncts;  
        lookup akhn_conditional_stacking;  
} akhn;  
  
lookup akhn_conjuncts {  
    lookupflag 0;  
    sub k1 l1 by k1l1;  
    sub k1 xx k1 by k1k1;
```

Shaping Engine

Shaping Engine

HarfBuzz

Converting Unicode
text to glyph indices
and positions

Behdad Esfahbod



HarfBuzz

Converting Unicode text to glyph indices and positions

```
$ hb-shape Manjari.ttf ெ  
[k1=0+1896]
```

```
$ hb-shape Manjari.ttf ெ  
[k1k1=0+2529]
```

Harfbuzz

Converting Unicode text to glyph indices and positions

```
$ hb-shape Manjari.ttf ே൧ശ  
[e2=0+1022|m1=0+1398|z1=2+1909]
```

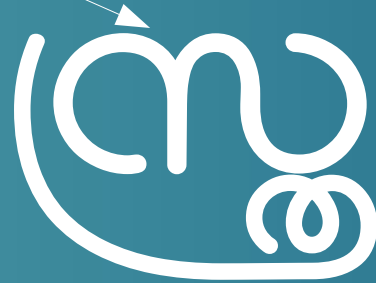


Three Malayalam characters are displayed: ே, ே, ശ. Arrows point from the code above to each character. Below the characters is a horizontal line with two vertical tick marks.

HarfBuzz

Converting Unicode text to glyph indices and positions

```
$ hb-shape Manjari.ttf (ꣳ  
[s1th1r3=0+2840]
```



HarfBuzz

Converting Unicode text to glyph indices and positions

```
$ hb-shape Manjari.ttf കൊച്ചി  
[e1=0+1395|k1=0+1896|a2=0+1001|  
ch1ch1=2+1525|i1=2+341]
```

കൊച്ചി

HarfBuzz

Chrome
Firefox
LibreOffice
Android
Xelatex
Gnome
QT/KDE

* Uniscribe (Windows) & Coretext (OS X) are proprietary alternatives

Fontconfig

Font matching and selection

Fontconfig

Font matching and selection

```
$ fc-match :lang=ml  
Manjari-Regular.ttf: "Manjari" "Regular"
```

```
$ fc-match :lang=ta  
Lohit-Tamil.ttf: "Lohit Tamil" "Regular"
```


Text Layout

Rendering the content in given area, line break, paragraphs,
text formatting

Text Layout

Rendering the content in given area, line break, paragraphs,
text formatting

Pango, ICU

Refer:

State of Text Rendering
behdad.org/text