Email and web in Malayalam

L2/03-309

An unofficial patched mozilla can render Malayalam for email and web. More details here

Unicode Malayalam rendering page

Handling Of Chillakasharam

Chillu are a unique feature of Malayalam. They are consonants with out vowels, thus can be said as dead consonants. These rules can be used to render Chillu in Malayalam.

1. When a Chillu consonant combines with a virama, you get the dead form of the consonant, which is consonant followed by chandrakala.
2. When Chillu consonant is followed by a chandrakala and zwj the Chillu form is rendered.
3. When they combine with other consonants they produce the corresponding Conjunct, if they exist or two consonants joined by chandrakala if they do not exist
4. When chandrakala is followed by Zero Width Joiner, the Chillu form is explicitly shown even if a conjunct is present for the consonant cluster
5. When chandrakala is followed by Zero Width Non-Joiner, the virama has to be shown explicitly.

Thus all the four cases of Chillu can be handled properly in Unicode. The figure below gives a full visual representation of the rules above. The Unicode text used to produce the example is here

Latest Unicode standard has a small paragraph on handling Chillu letters. I am reproducing a small relevant portion of it as image for easy reference.

Five sonorant consonants merge with the virama when they appear in syllable-final position with no inherent vowel. A consonant when so merged is called *cillaks.aram*:

ൺ	*ṇ*
ൻ	*n*
ർ	*r̠*
ൽ	*l*
ൾ	*ḷ*

It is important to note the use of the ZERO-WIDTH JOINER and ZERO-WIDTH NON-JOINER in these environments.

$$ന + \ddot{ } + മ \rightarrow ന്മ \ (nma)$$

$$ന + \ddot{ } + \boxed{\text{ZWNJ}} + മ \rightarrow ന്‌മ \ (nma)$$

$$ന + \ddot{ } + \boxed{\text{ZWJ}} + മ \rightarrow ൻമ \ (nma)$$

Chill at non syllable-final position

Unicode defines Chillu as Five sonorant consonants merge with the virama when they appear in syllable-final position with no inherent vowel. A consonant when so merged is called cillaks.aram: But this is not always true as shown from the counter example below.

അവൾക്ക് – means for her. The ൾക്ക് is a single cluster.

Ambiguity in language rules

Though strictly not a Unicode issue, lack of knowledge of Malayalam rules can cause problems in the Unicode storage of Malayalam text. sorting searching spell check etc relays on a common spelling for words. 2 such pairs are highlited here.

- RA and RRA
- LA and LLA

Please see the figures (provided by Ajailal) for examples.

1.a. മൈക്രോസോഫ്റ്റ് (Microsoft)

മ+ൈ+ക+്+ര+ോ+സ+ോ+ഫ+്+റ+്+റ+് ->മൈക്രോസോഫ്റ്റ്

മ+ൈ+ക+്+ര+ോ+സ+ോ+ഫ+്+റ+്+റ+് ->മൈക്രോസോഫ്റ്റ്

1.b. മൈക്രോപ്രോസസ്സർ (Microprocessor)

മ+ൈ+ക+്+ര+ോ+പ+്+ര+ോ+സ+സ+്+സ+ര+്+ZWJ -> മൈക്രോപ്രോസസ്സർ

മ+ൈ+ക+്+ര+ോ+പ+്+ര+ോ+സ+സ+്+സ+ര+്+ZWJ -> മൈക്രോപ്രോസസ്സർ

മ+ൈ+ക+്+ര+ോ+പ+്+ര+ോ+സ+സ+്+സ+ര+്+ZWJ -> മൈക്രോപ്രോസസ്സർ

മ+ൈ+ക+്+ര++ോപ+്+ര+ോ+സ+സ+്+സ+ര+്+ZWJ -> മൈക്രോപ്രോസസ്സർ

2. ക്ലോക്ക് (Clock)

ക+്+ല+ോ+ക+്+ക+് ->ക്ലോക്ക്

ക+്+ല+ോ+ക+്+ക+് ->ക്ലോക്ക്

The wrong representations are in red circles

3. ഴ+്+മ -> ഴ്മ   ഴ്മ

    ഴ+്+സ -> ഴ്സ   ഴ്സ

Comments to raj at linuxense dot com