

**Draft Policy Document
For**

**INTERNATIONALIZED
DOMAIN
NAMES**

Language : MALAYALAM

This proposal comprises the following Parts:

- 1. ABNF Formalism**
- 2. Restriction Tables if any**
- 3. Examples**
- 4. The Code-chart for Malayalam which is compliant with Unicode 5.1 Code-charts Characters marked in yellow will NOT be considered since they do not fall within the purview of the IDN policies as laid down.**
- 5. The description of the above code-chart, in conformity with the above,**
- 6. The variant table for Malayalam with the most important homographs being identified**
- 7. Experts/Bodies Consulted**

1. ABNF FORMALISM

1.1 Naming of Variables:

Dash	→ Hyphen -
Digit	→ Indo-Arabid digits [0-9]
C	→ Consonant
V	→ Vowel
M	→ Matra
D	→ Anusvara
X	→ Visarga
H	→ Halant/Chandrakala/Virama
Y	→ Avagraha/Praslesham
L	→ Chillu

1.2 ABNF Operators

S. No.	Symbols	Functions
1	"/"	Alternative
2	"["]"	Optional
3	"**"	Variable Repetition
4	"()"	Sequence Group

In what follows the Vowel Sequence and the Consonant Sequence pertinent to Malayalam are given. To facilitate use , equivalents in Devanagari are provided.

1.3 The Vowel Sequence

A vowel sequence is made up of a single vowel. It may be followed but not necessarily (optionally) by an Anusvara (D) or a Visarga (X) The number of D or X which can follow a V in Malayalam may be restricted to one.

The possibility of a Visarga following a Chandrabindu or Anusvara is ruled out.

The vowel sequence in Malayalam is therefore V [D | X]

Examples :

V	അ	अ
V[D]	അം	अं
V[X]	അഃ	अः

1.2 Consonant Sequence

A consonant sequence admits the following shapes:

1. A single consonant (C or L)

Example:

C	ക	क
L	ണ	

2. A consonant optionally followed by dependent vowel sign[M] or Anusvara[D] or visarga[X] or Halanta [H]

C[M|D|X|H]

Example:

C[M]	കി	कि
C[D]	കം	कं
C[X]	കഃ	कः
C[H]	ക്	क् (Pure Consonant)

2.a. A CM sequence can be optionally followed by *1[D|X]
(CM) *1[D|X]

Example:

CM[D] കിം कीं
 CM[X] ചീഃ वीः

3. A sequence of consonants (up to 4) joined by Halanta *3(CH)C

Example:

CHCHCHC [॒॒॒॑॑॑॑] → ॑ + ॒ + ॑ + ॒ + ॑ + ॒ + ॑

 न + ् + त + ् + र + ् + य

Subsets

3.a. The combination may be followed by H, M, D or X

*3(CH)C[H|M|D|X]

Example:

CHC[M] ക്കി ക ൿ ക ൿ ൿ क्कि
 CHC[D] ക്കിം ക ൿ ക ൿ ൿ क्कं
 CHC[X] ക്കിഃ ക ൿ ക ൿ ൿ क्कः
 CHC[H] ക്കി ക ൿ ക ൿ क्क

3.b. *3(CH)CM may be followed by a *1 D or X

*3(CH)CM[D|X]

Example:

CHCM[D] ക്കിം ക ൿ ക ൿ ൿ क्किं
 CHCM[X] ക്കിഃ ക ൿ ക ൿ ൿ क्किः

4. A Chillu character may be followed by a Virama and consonant.

LHC

Example:

LHC **ळॠ¹**

5. A LHC can be optionally followed by a Anusvara[D], Matra[M] or Virama[H]
LHC[D|H|M]

Example:

LHC[D] **ळॠ०**

LHC[H] **ळॠँ**

LHC[M] **ळॠी**

6. A LHCM can be optionally followed by a Anusvara[D]
LHCM[D]

Example:

LHCM[D] **ळॠी०**

1. Combination only exist where L is ळ(0D7B) and the following consonant must be ळ(0D31).

ABNF Applied to MALAYALAM IDN

Consonant Sequence → *3(CH) C [H|D|X|M[D|X]] | L[HC[D|H|M[D]]]

Vowel Sequence → V[D|X]

Sequence → Consonant Sequence[Y] | Vowel Sequence[Y]

IDN-Label → (Sequence | digit)*([dash](Sequence | digit))

2. RESTRICTION RULES:

The ABN formalism is generic in nature and when applied to a specific language/script certain restriction rules apply. In other words, in a given language some of the Formalism structures do not necessarily apply. To take care of such cases restriction rules are set in place. These restrictions will help to fine-tune the ABNF formalism.

In the case of Malayalam the following rules apply:

1. H | D|X |M can not occur in the beginning of an IDN domain name

Example:

ഃക	ക്
ിക	ിക
ക	ക
ഃക	ക

As can be seen they will result automatically in a “golu” marking an invalid character. This is an intrinsic property of the Indic syllable and is quasi automatically applied wherever supported by the OS.

2. H is not permitted after V, D, X, M, digit and dash

Example:

അഃ	अः
കഃ	कः
കഃ	कः
കിഃ	किः
1ഃ	1ः
-ഃ	-ः

3. Number of D or X permitted after consonant-sequence or vowel-sequence or M is restricted to one

Example:

കംഠഠ	कं
കഃഠഃ	कः
കീഠഠഠ	कीं
കീഃഠഃ	कीः
അഠഠഠഠ	अं
അഃഠഃ	अः

4. Number of M permitted after consonant-sequence is restricted to one
Example:

കീഠീ കീी

5. M is not permitted after V
Example:

ഇഠഠഠ ങ്ഠ

6. The combination of Anusvara+Visarga and vice-versa is not permissible
Example:

കഠഠഃ	കഃ
കഃഠഠ	കഃ

7. “H” is permitted after only one “L” ie. റ്റ(0D7B) and the following consonant must be റ(0D31).

3. EXAMPLES

Following are the some examples of MALAYALAM IDN Label.

C	ക	കട
CH	ക്	അക്
CM	കി	കിണർ
CD	കം	കംഗാരു
CX	നഃ	പുനഃസംഘടന
CMD	കിം	കിംവദന്തി
CMX	ദുഃ	ദുഃഖം
CHC	കെ	നടക്കുക
CHCHC	ഞ	വഞ്ജം
V	അ	അവധി
VD	അം	അംഗീകാരം
L	ൽ	പാൽ
LHC	ൻറ	അൻറാർട്ടിക്ക
LHCD	ൻറം	ഫാൻറം
LHCM	ൻറി	ആൻറി

4. UNICODE CHART SHOWING ALLOWED / NOT ALLOWED SUBSET

	0D00	0D01	0D02	0D03	0D04	0D05	0D06	0D07
0	☐	എ <small>0D00</small>	ഓ <small>0D01</small>	ര <small>0D02</small>	ീ <small>0D03</small>	☐	ഈ <small>0D06</small>	യ <small>0D07</small>
1	☐	☐	ഡ <small>0D21</small>	റ <small>0D31</small>	ൂ <small>0D41</small>	☐	ത്ര <small>0D61</small>	ത <small>0D71</small>
2	ഃ <small>0D02</small>	ഐ <small>0D12</small>	ഡ <small>0D22</small>	ല <small>0D32</small>	ൂ <small>0D42</small>	☐	ര <small>0D62</small>	ന <small>0D72</small>
3	ഃ <small>0D03</small>	ഐ <small>0D13</small>	ണ <small>0D23</small>	ള <small>0D33</small>	്യ <small>0D43</small>	☐	ര <small>0D63</small>	റ <small>0D73</small>
4	☐	ൗ <small>0D14</small>	ത <small>0D24</small>	ഴ <small>0D34</small>	്യ <small>0D44</small>	☐	☐	ർ <small>0D74</small>
5	അ <small>0D05</small>	ക <small>0D15</small>	ഥ <small>0D25</small>	വ <small>0D35</small>	☐	☐	☐	ൻ <small>0D75</small>
6	ആ <small>0D06</small>	ഖ <small>0D16</small>	ഭ <small>0D26</small>	ശ <small>0D36</small>	ഓ <small>0D46</small>	☐	ഓ <small>0D66</small>	☐
7	ഇ <small>0D07</small>	ഗ <small>0D17</small>	ധ <small>0D27</small>	ഷ <small>0D37</small>	ഓ <small>0D47</small>	ഊ <small>0D57</small>	ഏ <small>0D67</small>	☐
8	ഈ <small>0D08</small>	ഘ <small>0D18</small>	ന <small>0D28</small>	സ <small>0D38</small>	ഓ <small>0D48</small>	☐	ഏ <small>0D68</small>	☐
9	ഉ <small>0D09</small>	ഘ <small>0D19</small>	☐	ഹ <small>0D39</small>	☐	☐	ഏ <small>0D69</small>	ന <small>0D79</small>
A	ഊ <small>0D0A</small>	ഘ <small>0D1A</small>	വ <small>0D2A</small>	☐	ഓ <small>0D4A</small>	☐	ർ <small>0D6A</small>	ൻ <small>0D7A</small>
B	ഈ <small>0D0B</small>	ഘ <small>0D1B</small>	ഫ <small>0D2B</small>	☐	ഓ <small>0D4B</small>	☐	ർ <small>0D6B</small>	ൻ <small>0D7B</small>
C	ഈ <small>0D0C</small>	ഘ <small>0D1C</small>	ബ <small>0D2C</small>	☐	ഊ <small>0D4C</small>	☐	ന <small>0D6C</small>	ർ <small>0D7C</small>
D	☐	ഡ <small>0D1D</small>	ഭ <small>0D2D</small>	ഴ <small>0D3D</small>	ൂ <small>0D4D</small>	☐	ഊ <small>0D6D</small>	ൽ <small>0D7D</small>
E	എ <small>0D0E</small>	ഐ <small>0D1E</small>	മ <small>0D2E</small>	ഓ <small>0D3E</small>	☐	☐	ഊ <small>0D6E</small>	ൾ <small>0D7E</small>
F	എ <small>0D0F</small>	ട <small>0D1F</small>	യ <small>0D2F</small>	ീ <small>0D3F</small>	☐	☐	ൻ <small>0D6F</small>	ൽ <small>0D7F</small>

5. DESCRIPTION : MALAYALAM CODE CHART

Unicode Value	Character	Character Name
ANUSVARA (D)		
0D02	◌ം	MALAYALAM SIGN ANUSVARA
VISARGA (X)		
0D03	◌ഃ	MALAYALAM SIGN VISARGA
VOWEL LETTERS (V)		
0D05	അ	MALAYALAM LETTER A
0D06	ആ	MALAYALAM LETTER AA
0D07	ഇ	MALAYALAM LETTER I
0D08	ഈ	MALAYALAM LETTER II
0D09	ഉ	MALAYALAM LETTER U
0D0A	ഊ	MALAYALAM LETTER UU
0D0B	ഋ	MALAYALAM LETTER VOCALIC R
0D0C	ൠ	MALAYALAM LETTER VOCALIC L
0D0E	എ	MALAYALAM LETTER E
0D0F	ഈ	MALAYALAM LETTER EE
0D10	ഐ	MALAYALAM LETTER AI
0D12	ഒ	MALAYALAM LETTER O

0D13	ഓ	MALAYALAM LETTER OO
0D14	ഔ	MALAYALAM LETTER AU
CONSONANTS (C)		
0D15	ക	MALAYALAM LETTER KA
0D16	ഖ	MALAYALAM LETTER KHA
0D17	ഗ	MALAYALAM LETTER GA
0D18	ഘ	MALAYALAM LETTER GHA
0D19	ങ	MALAYALAM LETTER NGA
0D1A	ച	MALAYALAM LETTER CA
0D1B	ഛ	MALAYALAM LETTER CHA
0D1C	ജ	MALAYALAM LETTER JA
0D1D	ഝ	MALAYALAM LETTER JHA
0D1E	ഞ	MALAYALAM LETTER NYA
0D1F	ട	MALAYALAM LETTER TTA
0D20	ഠ	MALAYALAM LETTER TTHA
0D21	ഡ	MALAYALAM LETTER DDA
0D22	ഢ	MALAYALAM LETTER DDHA
0D23	ണ	MALAYALAM LETTER NNA

0D24	ത	MALAYALAM LETTER TA
0D25	ഥ	MALAYALAM LETTER THA
0D26	ദ	MALAYALAM LETTER DA
0D27	ധ	MALAYALAM LETTER DHA
0D28	ന	MALAYALAM LETTER NA
0D2A	പ	MALAYALAM LETTER PA
0D2B	ഫ	MALAYALAM LETTER PHA
0D2C	ബ	MALAYALAM LETTER BA
0D2D	ഭ	MALAYALAM LETTER BHA
0D2E	മ	MALAYALAM LETTER MA
0D2F	യ	MALAYALAM LETTER YA
0D30	ര	MALAYALAM LETTER RA
0D31	റ	MALAYALAM LETTER RRA
0D32	ല	MALAYALAM LETTER LA
0D33	ള	MALAYALAM LETTER LLA
0D34	ഴ	MALAYALAM LETTER LLLA
0D35	വ	MALAYALAM LETTER VA
0D36	ശ	MALAYALAM LETTER SHA

0D37	ഷ	MALAYALAM LETTER SSA
0D38	സ	MALAYALAM LETTER SA
0D39	ഹ	MALAYALAM LETTER HA
VOWEL SIGNS (MATRAS) (M)		
0D3E	ാ	MALAYALAM VOWEL SIGN AA
0D3F	ി	MALAYALAM VOWEL SIGN I
0D40	ീ	MALAYALAM VOWEL SIGN II
0D41	ു	MALAYALAM VOWEL SIGN U
0D42	ൂ	MALAYALAM VOWEL SIGN UU
0D43	ൃ	MALAYALAM VOWEL SIGN VOCALIC R
0D46	െ	MALAYALAM VOWEL SIGN E
0D47	േ	MALAYALAM VOWEL SIGN EE
0D48	ൈ	MALAYALAM VOWEL SIGN AI
0D4A	ൊ	MALAYALAM VOWEL SIGN O
0D4B	ോ	MALAYALAM VOWEL SIGN OO
0D57	ൗ	MALAYALAM AU LENGTH MARK
0D62	ൌ	MALAYALAM VOWEL SIGN VOCALIC L
VIRAMA (H)		
0D4D	്	MALAYALAM SIGN VIRAMA

AVAGRAHA (Y)		
0D3D	/	MALAYALAM SIGN AVAGRAHA
CHILLU LETTERS (L)		
0D7A	ണ	MALAYALAM LETTER CHILLU NN
0D7B	ൻ	MALAYALAM LETTER CHILLU N
0D7C	ർ	MALAYALAM LETTER CHILLU RR
0D7D	ൽ	MALAYALAM LETTER CHILLU L
0D7E	ൾ	MALAYALAM LETTER CHILLU LL

6. VARIANT TABLE FOR MALAYALAM

VARIANTS	
<p>ഉള</p> <p>0D33 + 0D33</p>	<p>ഉള</p> <p>0D33 + 0D4D + 0D33</p>
<p>സ</p> <p>0D24 + 0D4D + 0D38</p>	<p>സ</p> <p>0D1D</p>
<p>ഗ</p> <p>0D17 + 0D4D + 0D17</p>	<p>ഗ</p> <p>0D17 + 0D4D + 0D32</p>
<p>ന</p> <p>0D28 + 0D4D + 0D24</p>	<p>ന</p> <p>0D28 + 0D4D + 0D28</p>
<p>ശ</p> <p>0D36 + 0D4D + 0D32</p>	<p>ശ</p> <p>0D36 + 0D4D + 0D36</p>
<p>സ</p> <p>0D38 + 0D4D + 0D38</p>	<p>സ</p> <p>0D21 + 0D4D + 0D21</p>
<p>ബ</p> <p>0D2C + 0D4D + 0D27</p>	<p>ബ</p> <p>0D2C + 0D4D + 0D26</p>

7. EXPERTS/BODIES CONSULTED

Expertise provided by C-DAC Thiruvananthapuram.

Note : You can send your feedbacks to ids-feedback@cdac.in