

## Introduction

Issues with the approach and process

Criticism on the policy

General Comments

1. The variant table is defined based on random glyphs taken from a 900+ possible glyphs for Malayalam. No explanation on how two entries in variant table becomes homo morphs. One entry in variant table is just because of the fact that one is mirror image of other. Since b, d are not excluded from English, no need to exclude mirror imaged glyphs in variant table

Ans: The IDN system devised for Malayalam is based only on the modern script. It doesn't address the old script or the fonts based on old script. Also, a detailed study was done before proposing homographs in each of the languages. The study included observing the visual form of the conjunct in the point size of the Address bars of major browsers. The mirror imaged nature of the glyphs was not the criterion for the two glyphs to be qualified as variants.

2. Visually identical glyphs are not only the entries to be considered for variant table. Unicode chart itself has ambiguous dual representations for the same code point with out canonical equivalence. An example for this is au signs in Tamil and Malayalam. ീ- ു and ൂ - ൃ . The document does not consider this special cases.

Ans: The IDN policy does not permit the entry of syllables having structure CMM or MCM, where M stands for Matra or vowel sign. The ABNF rules takes care of this.

3. There are different orthographic forms for many glyphs in Malayalam. The variant table does not address different scenario arising while considering the visual similarity. For example in traditional orthography TTA is written stacked form. While in modern orthography it can be written in non-stacked form and this non-stacked form is visually identical to two RA sequence.

Ans : Only the stacked form is considered to be the conjunct TTA in modern orthography.

## ABNF rules

1. Section 2 says ീ as pure consonant of ു. Chillu of ീ is considered as pure consonant of ka.

Ans: The policy document doesn't address the obsolete characters in the script, although those characters might have been included in Unicode code chart.

2. Section 2.a says CM can be followed by only D (anuswara) or X (visarga). This excludes the Samvruthokarams of Malayalam. All consonant can have cons + u vowel sign + virama and forming samvruthokaram form of that consonant. Examples: ീ, ു, ൂ, ൃ

Ans: The use of samvruthokarams is considered to be the part of traditional orthography which the policy doesn't permit.

3. Section 3.a restrict the count of consonant in syllable as 4. But ശ്ല(ല) has 5 consonants

Ans : Complex conjuncts like ശ്ല(ല) have been simplified in modern orthography.

4. Section 3.b excludes syllables with samvruthokram like ഞ്ഞ.

5. Section 4 states a chillu can be followed by a vowel sign. Since chillu is dead consonant, there is no possibility of having virama after chillu.

Ans : The document doesn't state that a chillu can be followed by a vowel sign. The observation that a virama can appear after a chillu is based on the recommendation of Unicode 5.1 official document on rendering the conjunct 'nTa'.

6. The example used for LHC - ഞ്ഞ does not exist in printing or digital format. None of the input methods or Malayalam writers ഞ്ഞ in this way. The sequence for nta is ഞ + ഞ് + ഞ . ie there is no LHC sequence in Malayalam.

Ans: The ഞ്ഞ as displayed in the document is the wrongly rendered form of the conjunct 'nTa'. This happened because none of the rendering engines available today does support the rendering of 'nTa' in the Unicode 5.1 way.

7. Since LHC is invalid for Malayalam, including L = ഞ്ഞ , section 5 of the document cancels itself.

8. Because of argument #6, section 6 also cancels itself.

9. Because of arguments #1 to #8 the IDN rule "Consonant Sequence → \*3(CH) C [H|D|X|M[D|X]] | L[HC[D|H|M[D]]]" is completely wrong and need to be reformulated.

#### Restriction Rules

1. Section 2 says "H is not permitted after V, D, X, M, digit and dash" This is wrong since samvruthokaram requires H after V

Ans : See the explanation for section 2 under ABNF Rules

2. Section 7 says H can follow L if it is followed by ഞ , This is wrong as explained above. L can never followed by H. It can only followed by C

Ans : See the explanation for section 5 under ABNF Rules

#### nta criticism

This document does not address the case of stacked and non stacked forms of nta, which are

interchangeably used. For example ഏൻ്റെ can be spoofed with എൻ്റെ. Severity of this issue is increased by having one more sequence to represent the same conjunct (ൻ + ് + റ) is introduced in Unicode 5.1

Ans : Modern orthography treats റൻ as 'nRa' and റൻ as 'nTa'. The interchangeable usage of stacked and non-stacked forms for the conjunct 'nTa' is wrong by convention.

#### Chart of allowed characters

1. Malayalam chillus - the 5.1 version ങ് is removed from the tables. which is having same characteristics and use cases of other chillus. So excluding it from the allowed code points does not make any sense. Moreover the existing chillu representation - non-atomic - is not mentioned in the document at all.

Ans: As mentioned earlier, the IDN policy document doesn't permit the obsolete characters in the script to be part of domain names. And the policy document is based on Unicode 5.1 which provides an atomic representation for chillu characters.

2. Malayalam au sign - റൗ is not allowed. Instead the au length mark റൗ is provided. The inscript standard does not allow one to type റൗ and allows only റൗ. Other input methods allows to type both. But the document does not say anything on the equivalence of both. Allowing both vowel signs is also a spoofing issue. And hence this should be handled in variant table.

Ans: The inscript standard being revised. The new standard allows both the characters to be inputted. For restricting spoofing and phishing, only one form i.e. റൗ (used in modern text) has been allowed by IDN policy.

#### Variant Table and Visual Spoofing

Variant table is not logical. Only ഉഉ and ഉഉ makes sense. None of the other entries should be considered as spoofing. റൻ and റൻ is not even close. Mirror images are already used in Latin, eg. b and d. Hence റൻ and റൻ cannot be blocked. Moreover it is not clear why the same logic does not apply for റൻ and റൻ. It did not consider the case of റൻ and non stacked form of റൻ common in new lipi.

Ans : The variant table is based on the observations how Malayalam characters and conjuncts are rendered in the address bars of standard browsers like IE, Mozilla and Safari. While റൻ and റൻ are

perfectly rendered in Mozilla and Safari, they are not legibly rendered in various versions of IE. The mirror imaged nature of the glyphs was not the criterion for the two glyphs to be qualified as variants. Also note that the variant table is not a full-proof mechanism which can prevent spoofing.

Even though similarity is considered, dual encoding is not mentioned. In case of dual encoding of chillus, both forms (atomic chillu and consonant chandrakkala ZWJ) of chillus will look SAME.

Ans: IDN policy doesn't allow control characters such as ZWJ and ZWNJ to be part of domain names.