

New Malayalam code points in Unicode 5.1

Draft version of Unicode 5.1.0 suggests new code points for Chillu characters. We, Swathanthra Malayalam Computing, a developer community working on localization, development and standardization of Malayalam Computing softwares has the following comments on the draft version of Unicode 5.1.0 .

1. The atomic chillu's are unacceptable because it destroys the link of a chillu with its base character.
2. The examples used to justify semantic difference between words only separated by ZWJ are non-existent in dictionary , not in are grammatically wrong or meaningless without proper context.

a) വൻയവനിക/വന്യവനിക (vanYavanika/vanyavanika), കൺവലയം/കണ്ഡലയം (kanvalayam/kanualayam) -- contrived examples not found in dictionary

Here , the fundamental problem lies in Unicode's way of treating only representational forms without checking linguistic correctness. Most of the Indic languages are unlike Latin and collations are based on linguistics. If you are not considering it, it will become a play yard of people with vested interests.

3. All these arguments were once considered and rejected by UTC and the only new argument in support of atomic chillus is the issue of missing domain names in IDN. The examples given in 1) can't be considered real as these are contrived just to make a case for atomic chillus. Even if were real it is similar to case folding in Latin (You can't register two sites PenIsland.com and PenisLand.com). How can already rejected proposal be accepted when the new arguments in supports is not only proved to be real, but creates a lot of new chaos and security problems.

4. Introducing atomic chillus will create dual encoding and makes URL spoofing very easy. This has already been illustrated with the following examples .

<http://ഓലൂനോവ്.blogspot.com> (using chillu joiner sequence) <http://ഓമിനോവ്.blogspot.com> (using atomic chillu)

because both of these have different punicode. The existing chillu encoding with joiners is best solution because all of the combinations of joiners and non-joiners give exactly same punicode.

5. Since the joiners has to be supported for backward compatibility it creates unnecessary complexity to all text processing application (sorting, searching) and it

makes atomic chillus redundant and useless.

6. As per the unicode stability policy, session 'Named Character Sequence Stability', the existing chillu sequence has to be supported. In that case, in order to process the text with atomic chillu and existing chillu, there should be a canonical equivalence to old sequence. It is not provided and not mentioned in Unicode 5.1 and that breaks the existing applications and data and violates unicode stability policy.

7. Even after atomic chillus are made part of the standard many words cannot be written without joiners and it would be increasing the chaos. Thereby the atomic chillu does not solve the issue of ignorability of ZWJ or ZWNJ as mentioned in the proposal to encode chillu and atomic chillu is a partial incomplete solution. Eg: കൊയ്തള (koirala), സദ്വാരം (sadvaram)

8. Using virama with chillus is linguistically incorrect (function of virama is to create vowel-less and you can't use it with a chillu or pure consonant because these are already vowel-less forms of the underlying consonants)

9. In Table 1 of

http://www.unicode.org/versions/Unicode5.1.0/#Significant_Character_Additions the chillu-n sequence is <na, virama, zwj>. So for consistency, in Table 2 whenever chillu sequences occur, a zwj insertion showing chillu = <cons, virama, zwj> is essential

10. Everybody knows that there was no consensus reached in the discussion in indic@unicode.org mailing list and still the problem is controversial. Another thing is even though the new changes will have a major impact on the language technology, the linguistics and language experts in Malayalam is not at all aware of the facts. We doubt that language experts/authorities accepted by the public were given an explanation of what Unicode is and what the atomic chillu proposal is about. Only some ivory-tower discussions among some academicians were carried out and even those have reached the conclusions that there is no particular necessity for atomic chillus. Even among the IT literate Malayalees (people who use Malayalam on a regular basis) only a handful know the Unicode representation of Malayalam and issues surrounding it. We would like to start a process which explains the pros and cons to the language experts and getting their opinion in this matter. So any hurry on adding new code points will, in our opinion, be ill-informed and will have a bad impact on the future of the language.

11. The document <http://www.rachanamalayalam.org/docs/ChilluEncodingIsWrong.pdf> already questioned the new code points and there was no satisfactory reply from the people who proposed atomic chillu.

Malayalam has already gone through a round of mutilation during the typewriter reform era. We know that the means provided by current computing platforms can resurrect the language and its script and restore its former beauty . If such hurried and ill informed moves are taken instead of careful and well thought out ones , we will be murdering the language instead of resurrecting it.

We strongly oppose the inclusion of these characters for encoding chillus in the Unicode standard as it not only fails to solve all the problems with joiner , but it also creates new problems and introduces the need for providing backward compatibility leading to more chaos.

Swathanthra Malayalam Computing

<http://www.smc.org.in>