

## **Atomic chillu's are Unacceptable**

Swathanthra Malayalam Computing

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 relation of a chillu with its base character.
2. The examples cited to justify semantic difference between words only separated by ZWJ are non-existent in dictionary
  - a) വന്യവനിക/വന്യവനിക (vanYavanika/vanyavanika) in L2/06-207 is not existste s in Most widely used & comprehensive dictionaries like Sabdatharavali
  - b) കൺവലയം/കണലയം (kanvalayam/kanualayam) which is mensioned in various other proposals of pro-atomic chillu group & in Indic unicode list is not included in any dictionary
3. Here , the fundamental problem lies in Unicode's way of treating only representational forms without looking at its linguistic correctness. Unlike Latin in most of the indic languages collations are based on linguistic rules. If you are not considering it, it will become a play yard of people with vested interests.
4. Atomic codepoint for chillu is coming up in the name of convinience. At the same time it introduces the security issue of duplicate encoding. All words with chillu characters can be spoofed **2<sup>n</sup>** times, where n is the no of chillaksharams in that word. In computing world security is always considered important than convenience
5. 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 above 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.
6. If atomic chillu is introduced it will create dual encoding and makes URL spoofing very easy. This has already been illustrated with the following examples .

Do main	Punic ode Eq uivalent
<b>http:// റാൽമിനോവ്.blogspot.com</b> (using chillu joiner sequence)	<a href="http://xn--uwclier4cj1hof.blogspot.com">xn--uwclier4cj1hof.blogspot.com</a>
<b>http://രാ® മിനോവ്.blogspot.com</b> (using atomic chillu)	<a href="http://xn--uwclis6bh9fra04b.blogspot.com">xn--uwclis6bh9fra04b.blogspot.com</a>

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.

8. 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.
9. As per the [unicode stability policy](#), session 'Named Character Sequence Stability', the existing chillu sequence has to be indefinitely supported. Enormous amount of data encoded in Unicode 5.0 standard with chillu letters in consonant +virama+zwj continue to exist in digital form. If atomic chillus are introduced, inorder to process the text with atomic chillu and exiting 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.
10. If the incorrect words projected as “words with different meaning differed by zwj” was the reason for introducing atomic chillu in unicode 5.1 draft document , various words exists in which zwnj will make the meaning change. There exist 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.

കൊയിരാള (koirala)	0D15 0D4A 0D2F 0D4D 200C 0D30 0D3E 0D33
സദ്വാരം	0D38 0D26 0D4D 200C 0D35 0D3E 0D30 0D02

See word pair below which have difference in meaning only with the difference of zwnj

സദ്വാരം (good week)	0D38 0D26 0D4D 200C 0D35 0D3E 0D30 0D02
സദ്വാരം (with hole)	0D38 0D26 0D4D 0D35 0D3E 0D30 0D02

11. We believe all consonants have a property to form chillu. chillu is a vowel less consonant. It is either marked by Cons+virama+zwnj if representational forms doesnt exists in script or as Cons+virama+zwj if representational forms exists in script
12. Chillu's never form conjuncts. All proposals for such definitions are linguistically incorrect (function of virama is to create vowel-less and you can't use it with a chillu because these are already vowel-less forms of the underlying consonants)
13. In Table 1 of [http://www.unicode.org/versions/Unicode5.1.0/#Significant\\_Character\\_Additions](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

14. Dot repha must be as follows

If the font supports Repha	<RA><VIRAMA><CONSONANT   CONJUNCT> => <REPH (CONSONANT   CONJUNCT)>
fallback (if the font not supports repha)	<RA><VIRAMA><CONSONANT   CONJUNCT> => <RA><VIRAMA><ZWJ><CONSONANT   CONJUNCT> => <CHILLU RA><CONSONANT   CONJUNCT>

15. 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 has 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 .
16. The document <http://www.rachanamalayalam.org/docs/ChilluEncodingsIsWrong.pdf> already questioned the new code points and there was no satisfactory reply from the people who proposed atomic chillu.
17. 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.

Praveen A, Anivar Aravind, Santhosh Thottingal, Baiju.M

**Swathanthra Malayalam Computing**

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

<http://savannah.nongnu.org/projects/smc>