

Critique of Proposed Changes to Inscript Layout

Swathanthra Malayalam Computing*

January 19, 2011

1 Introduction

Recently CDAC published a second draft of changes to inscript layout¹. It was published in Kerala IT Mission website² for comments. This document present the views and comments of Swathanthra Malayalam Computing on the same. This is available on Swathanthra Malayalam Computing wiki too³.

2 General Comments

Input methods are very important component of any digital system. Any change has long reaching consequence in regards to compatibility with existing data. Blindly including all new characters of Unicode 5.1+ without a clear solution in regards to the issues brought by new versions is premature.

Foreword says "These new features had marked repercussions on storage as well as inputting and an urgent need was felt for a revision whereby each and very new character introduced in Unicode would be accommodated on the keyboard and a uniform manner of entering data as well as storing data would be devised." This is contradictory in itself, because, new characters introduced by Unicode version 5.1 and above has introduced multiple ways of entering and storing the same data. For example chillu characters can be entered two different ways, and 'nta'

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

¹http://malayalam.kerala.gov.in/images/8/80/Qwerty_enhancedinscriptkeyboardlayout.pdf

²<http://malayalam.kerala.gov.in>

³<http://wiki.smc.org.in/CDAC-Inscript-Critique>

conjunct (named sequence) adds to the confusion of two existing ways of encoding nta (correct way based on language conjunct formation i.e. na + chandrakala + rra and incorrect way followed by Microsoft ie, na + chandrakala + ZWJ + rra). With the existing inscript keyboard layout, chillu characters are entered and stored in a uniform manner, these changes will allow entering and storing chillu characters in two different ways incompatible with each other. There is no mention about how to deal with the existing encoding of chillu character in Unicode 5.1 or how to deal with two different encoding of chillu characters.

On page 2 case b is incorrect conclusion of the underlying problem. The problem described in case a and b is examples of ZWJ and ZWNJ use. This has absolutely no relation with what input method you use. The problem is when the fonts used for display uses different sequences for displaying the same conjunct. Encoding of Malayalam conjunct 'nta' is a classic example of this. Irrespective of the input methods used, be it a key board layout like inscript/lalitha or transliteration method like varamozhi or swanalekha the encoding ultimately depends on the fonts used by the user. Users of Microsoft's karthika font enters/encodes 'nta' as na + chandrakala + ZWJ + rra irrespective of the input methods used and users of other fonts following the rules of the language enters/encodes it as na + chandrakala + rra.

This is approach to solve the problem of multiple encoding by standardizing keyboard layout is destined to fail. Dual encoding is introduced by Unicode for characters (basic and named sequences) included in its table, and by font creators for conjuncts in the private area (nta for example). Even if the proposed inscript layout changes are accepted, the users will ultimately decide which encoding to use based on their choice of Unicode version and Malayalam font. If at all there is any chance of success, that would be allowing only one way of entering a particular character - but that would mean breaking compatibility with old version of inscript layout as well as different versions of Unicode. Since the problem originates at Unicode and in fonts, the solution also has to happen at the same place.

Some problems are enumerated correctly but the solutions does not address those or in many cases adds to the problem. Take the case of search, when one more method of encoding chillu character is added, the user will have to try searching all possible encodings to get desired results. What does "porting and transferring a document input by one mechanism across OSs or even within the same OS." even mean? Does it mean exchanging documents encoded in variants of the standard? Even in that case the deciding factor is font, once the document is created, the role of input method ended. Exchange of data depends only on the underlying encoding - in this case Unicode. The only question is about the variations like different nta or chillu (atomic or joiner formed), but that is processed by font.

3 Issues

1. u+0D4C ഞ - is present in the current map, in the position of normal q, and which is in frequent use. But in the proposal, it is moved to the position of ext q. This change is unacceptable, as it will break compatibility with existing layout. In the new layout also its at Ext q and interestingly named backward compatibility. How does a change in layout maintain backward compatibility, especially when there is no canonical equivalence with the new character in its place?
2. u+0D57 ഞ - is introduced to the existing position (normal q) of the sign u+0D4C ഞ, which is a duplicate encoding of the sign ഞ. This change is unacceptable because, this will break the compatibility between the data entered using the existing keyboard layout and that entered with the proposed keyboard layout. This continues in the second draft. Since there is no canonical equivalence between u+0D4C and u+0D57 introducing one at the others place is not a good idea. For keeping the data clean, its desirable to have only one and most probably to keep backward compatibility, u+0D4C.
3. u+0D79 - Malayalam date mark - mapped on Ext-shifted v. It should be there in the second layer, as it is frequent in usage, especially while drafting letters. It can be positioned on shifted v, as there are no letters are mapped in shift v position, in existing Inscript layout. It should at least be changed to Ext v which is empty. Not getting the logic of using 4th layer when third is empty. That too for a frequent character.
4. Positions proposed for all the chillus in the new proposal, are unacceptable. Chillaksharams are very frequent in usage, and are already mapped on existing inscript keyboard layout. Proposed positions differ largely from the existing layout. Starting from primary school students, employees in government offices, people in the DTP industry, all are familiar with the existing logical layout, regarding the mapping of chillaksharams. The new proposal breaks the backward compatibility it offer. Besides,
 - (a) The position of normal is proposed for u+0D7C (Atomic chillu RR), where the ZWNJ is mapped in the existing inscript keyboard layout. This change make no sense and thus, unacceptable.
 - (b) In the proposal document, u+0D7D (Atomic chillu L) is proposed to be in the position of normal shifted fullstop. This position is occupied by the symbol > which is frequently used in mathematics. This change

will make unnecessary difficulty for the school children, who are already familiar with the existing layout. Again, it breaks the backward compatibility.

- (c) Similarly, u+0D7E (Atomic chillu LL) is proposed in the position of normal shifted 8 where * lies. Again, it is a frequently used symbol in mathematics, as well as in programming. This change will make unnecessary difficulty especially for the school children.
 - (d) No comments are done, why ഏ excluded. It is already mapped in existing inscript keyboard layout. In latest draft, its simply written EXCLUDED. May be CDAC should remember with out chillu K its difficult to write ക്ക.
 - (e) The comment column for atomic chillus in the inscript keyboard layout, are filled with the sentence "Placed as per Kerala Govt. Gazette (Vol.46 Thiruvananthapuram Dated. 18th Dec. 2001 No.2023) G.O. (Rt) No. 93/2001/ITD. dated 2-6-2001" We all know that atomic chillus are introduced in Unicode 5.1, without considering the factual defences by the people who really know Malayalam. This version of Unicode was released in April 2008. Then how the government order which was released much earlier (in June 2001) proposed for mapping atomic chillus on keyboard? We don't think that the Honourable Secretary to the Government foreseen the future inclusion of these code points, and issued an order. We couldn't understand its logic. This needs explanation. It would be useful in this discussion to get a copy of this GO. Government of Kerala has a Malayalam Computing project right now. These comments on Malayalam inscript, are based on the keyboard layout released in the website of this project⁴.
5. Proposed mapping of ZWJ (shift 9) and ZWNJ (shift 0) are also questionable, as they are already mapped in the positions of \ and] respectively. This change breaks the backward compatibility.
6. Regarding the Caps lock key. The document says: "After due discussions with OS developers it was decided that the Toggling key between layer 1 and layer 3 will be different for different Operating Systems. Though toggling between layer 1 and layer 2 (which is English layer) will remain through the Caps-Lock key as mentioned in Annexure-D of "Bureau of Indian Standard document for ISCII-91" which is as follows: "The Inscript overlay gets selected when Caps-Lock is active, otherwise normal lower case English overlay gets selected." This logic is not easy to understand,

⁴<http://malayalam.kerala.gov.in/images/7/78/Inscript.jpg>

because, Bureau of Indian Standard document for ISCII-91 is an age old document. It was prepared in a period when the user circle was much small, and facilities available in computers were less comparing today. It should be a nice option to let the user to customise the toggling keys. GNU/Linux already has this facility.

4 Contributors

Jaisen Nedumpala, Jinesh K J, Praveen Arimbrathodiyil, Swathanthra Malayalam Computing