Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation

Doc Type:	**Working Group Document**
Title:	**Proposal to Add Tamil Nukta Character**
Source:	**Martin Hosken**
Status:	**Individual contribution**
Action:	**For consideration by UTC and WG2**
Date:	**2015-10-26**

Executive Summary This proposal is to add a character U+0BBC TAMIL SIGN NUKTA. To the Tamil block of Unicode. The Unicode character data for it will be:

```
0BBC;TAMIL SIGN NUKTA;Mn;7;NSM;;;;;N;;;;;
```

The representative glyph will be a lower single dot following Devanagari. It's sorting and shaping behaviour will follow Devanagari.

Introduction The Tamil language does not make use of a Nukta. But there are a few minority languages that use the Tamil script that need a Nukta. The following table shows examples of their use:

Language	Example	Unicode
Chetti (ctt)	ப்ரீயிக்ஞ	`0BAA 0BBC 0BC0 0BAF 0BBF 0B95 0BCD 0B9E`
Betta Kurumba (xub)	ஒண்டெட	`0B92 0BA3 0BCD 0B9F 0BBC 0BC6 0BBC`
Jennu Kurumba (xuj)	நட்த	`0BA8 0B9F 0BBC 0BA4 0BBC`
Irula (iru)	பெல	`0BAA 0BBC 0BC6 0BB2`

There are two key points to notice about the Tamil Nukta from these examples:

• There are different styles of nukta: single dot, as used in Chetti; single ring as used in Betta and Jennu Kurumba and double dot as used in Irula.

• The nukta may occur under a vowel, as in the Betta Kurumba example, where the vowel U+0BC6 TAMIL VOWEL SIGN E is reordered in front of the U+0B9F TAMIL LETTER TTA and both take a nukta.

Rationale While there are nukta letters encoded in other blocks, there is a need for a nukta character in the Tamil block to ensure script consistency across a run containing this diacritic.

Representative Glyph One of the styles of the Tamil nukta needs to be chosen for the representative glyph. Since Chetti is a more dominant language of the region, the relevant communities have decided that the default glyph should be the single dot as used for Chetti. This form is also the most prevalent nukta, being used in Devanagari, Bengali, Gurmukhi, Gujarati and Oriya, with only Kannada having a different form.

Examples

ஸ்வர அக்ஷர

அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ

வஞ்ஞன அக்ஷர

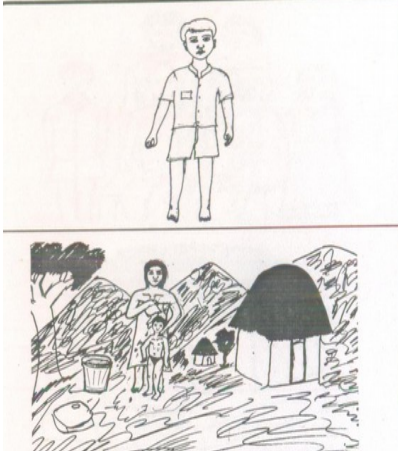| க	க	ங	
ச	ஜ	ஞ	
ட	ட	ண	
த	த	ந	ன
ப	ப	ம	
ய	ர	ற	வ
ல	ள	ஷ	ஸ
ஹ	க்ஷ	ழ	

Illustration 1: Chetti alphabet from spelling book

நங்கள சமுதாயத பற்றிட்டுள்ள பிவற

நங்கள சமுதாயா சுமாராயிட்டு மொத்த ஆறாயிர ஜனத்தொகெ இத்தீ

விகிதாச்சார ஏகதேச சம ஆப்புது நங்க இப்பத்திநாக்கு தரவாடு உ

குலதெய்வங்ஙளும், உட்டு. இதனொளெ ஐது காராய்மெ காரும் உட்டு

Illustration 2: Sample from chetti newsletter where in desperation they used a double nukta, despite their style preference

பேரு கடக்க குருவ
ரோனுங்ச குருவனு
பேரு எலிசெபெத்து

Illustration 3: Irula sample text with nukta use marked

Illustration 4: Betta Kurumba story book

Illustration 5: Irula story showing alternative style of nukta

	0B8
0	ற
1	ற
2	ல
3	ள
4	ழ
5	வ
6	ஶ
7	ஷ
8	ஸ
9	ஹ
A	
B	
C	ஂ
D	
E	ா
F	ி

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:**	*Tamil Nukta*
2. Requester's name:	*Martin Hosken*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual contribution*
4. Submission date:	*10/Sep/2015*
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	*X*
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:
    a. This proposal is for a new script (set of characters):
        Proposed name of script:
    b. The proposal is for addition of character(s) to an existing block: *X*
        Name of the existing block: *Tamil*

2. Number of characters in proposal: *1*

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary *X*	B.1-Specialized (small collection)	B.2-Specialized (large collection)
C-Major extinct	D-Attested extinct	E-Minor extinct
F-Archaic Hieroglyphic or Ideographic		G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided?   *yes*
    a. If YES, are the names in accordance with the "character naming guidelines"
        in Annex L of P&P document?   *yes*
    b. Are the character shapes attached in a legible form suitable for review?   *yes*

5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
    publishing the standard?   *NLCI*
    If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
    used:

6. References:
    a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?   *no*
    b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
    of proposed characters attached?   *yes*

7. Special encoding issues:
    Does the proposal address other aspects of character data processing (if applicable) such as input,
    presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?   *no*

---

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?   *no*
   If YES explain
2. Has contact been made to members of the user community (for example: National Body,
   user groups of the script or characters, other experts, etc.)?   *yes*
   If YES, with whom?   *Chetti community: NIRMAAN*
   If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example:
   size, demographics, information technology use, or publishing use) is included?   *no*
   Reference:
4. The context of use for the proposed characters (type of use; common or rare)   *common*
   Reference:
5. Are the proposed characters in current use by the user community?   yes
   If YES, where? Reference:
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP?   *yes*
   If YES, is a rationale provided?   *yes*
   If YES, reference: *addition to existing BMP block*
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?   *no*
8. Can any of the proposed characters be considered a presentation form of an existing
   character or character sequence?   *no*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either
   existing characters or other proposed characters?   *no*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function)
    to an existing character?   *yes*
    If YES, is a rationale for its inclusion provided?   *yes*
    If YES, reference: *This document*
11. Does the proposal include use of combining characters and/or use of composite sequences?   *yes*
    If YES, is a rationale for such use provided?   *yes*
    If YES, reference: *This document*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?   *no*
    If YES, reference:
12. Does the proposal contain characters with any special properties such as
    control function or similar semantics?   *no*
    If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility character(s)?   *no*
    If YES, is the equivalent corresponding unified ideographic character(s) identified?
    If YES, reference: