

18/03/2020

Proposal to Encode Telugu Letter Nakaara Pollu

Vinodh Rajan vinodh@virtualvinodh.com
 Shriramana Sharma jamadagni@gmail.com
 Suresh Kolichala suresh.kolichala@gmail.com

Introduction

L2/11-049 initially proposed the special vowel-less form of Telugu Letter NA U+0C28 నా, *Nakaara Pollu* ె be handled using a complicated mixture of joiners and font-level delegation (see below section). However, there haven't been much progress in term of vendor support as there are not any fonts that support this historical character. This is perhaps, partially, due to its complicated rendering requirement.

Since the current trend of the Unicode process is to remove cumbersome dependence on joiners and simplify the overall rendering & representation of a text, it is a good time to revisit the Telugu *Nakaara Pollu* in TUS.

Current Status

Table 12-31. Rendering of Telugu *na + virama*

Font	Sequence	Rendering
Old Style	<i>na + virama</i>	ె
	<i>na + virama</i> + ZW NJ + <i>da</i>	ెద
New Style	<i>na + virama</i>	న
	<i>na + virama</i> + ZW NJ + <i>da</i>	నద
All Fonts	<i>na + virama + da</i>	ద

Unicode standard 13.0, Chapter 12, Section 12.7

As noted earlier, currently, the representation of *Nakaara Pollu* is font dependent along with the need to use ZWNJ to avoid conjunct formation. Furthermore, the standard is silent about rendering the regular NA +Virama న in *Old Style* fonts and *Nakaara Pollu* ె in *New Style* fonts.

One can only assume that TUS recommends mixing fonts to achieve simultaneous rendering of న and ె. This would mean those two characters can never be unambiguously represented in plain text. Hypothetically one may recommend further sequences with ZWJ/ZWNJ for their simultaneous rendering. But we will end up with complicated joiner-based sequences that isn't ideal for both the users that would need to input them and rendering engines that need to support them.

18/03/2020

Cognate dead consonantal forms in other Indic scripts such as the *Chillu* letters in Malayalam and *Khanda Ta* in Bengali currently exist as atomic characters. Even though, previously they were also rendered using complicated and confusing joiner sequences. Their current atomic forms have simplified their rendering requirements and eased how users input them, without ever resorting to joiners. A similar solution can be adopted for *Nakaara Pollu* as well.

Atomic Encoding of Telugu Nakaara Pollu

An atomic encoding of *Nakaara Pollu* will simplify its rendering in TUS and bring its encoding on par with similar vowel-less characters in the UCS. Also, as a minority and historic character it would achieve better vendor support as an atomic letter.

We propose that *Telugu Letter Nakaara Pollu* be encoded in the Telugu block of the UCS with the associated character properties and recommended annotation.

0C5B  Telugu Letter Nakaara Pollu

- Special vowel-less form of NA

→ Malayalam Letter Chillu N

0C5B;TELUGU LETTER NAKAARA POLLU;Lo;0;L;;;;N;;;;;


Indic Syllabic Category

The following addition should be made to the IndicSyllabicCategory.txt file under:

```
# Indic_Syllabic_Category=Consonant_Dead
```

```
0C5B                   ; Consonant_Dead # Lo                   TELUGU LETTER NAKAARA POLLU
```

Sorting and Searching

Nakaara Pollu should be treated as an equivalent to the Unicode sequence  U+0C28 U+0C4D for sorting and searching purposes.

Telugu Chapter of TUS

The Telugu Chapter of TUS must be edited to reflect the atomic handling of *Nakaara Pollu*.

18/03/2020

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Proposal to Encode Telugu Letter Nakaara Pollu		
2. Requester's name:	Vinodh Rajan, Shriramana Sharma, Suresh Kolichala		
3. Requester type (Member body/Liaison/Individual contribution):	Individual		
4. Submission date:	18/03/2020		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:			Yes
(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:		Yes	
Name of the existing block:		Telugu	
2. Number of characters in proposal:		1	
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary	B.1-Specialized (small collection)	Y	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?			
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. Fonts related:			
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?		Vinodh Rajan	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):		Vinodh Rajan, vinodh@virtualvinodh.com	
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?		Yes	
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)?		Yes	
		Sorting & Searching	

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 Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

18/03/2020

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? Suresh Kolichala & Vinodh Rajan	
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?	Yes
Reference: See Proposal	
4. The context of use for the proposed characters (type of use; common or rare)	Rare
Reference: See Proposal	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference: See Proposal	
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: Telugu is in BMP	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
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?	Yes
If YES, is a rationale for its inclusion provided? Yes	
If YES, reference: See Proposal	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	No
If YES, is a rationale for such use provided?	
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
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 characters?	No
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	