

18/03/2020

Proposal to Encode Telugu Reph

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

Introduction

The Telugu *reph* (Telugu: *Valapalagilaka*) is currently handled using a mixture of font-level delegation and joiners (as recommended by L2/12-017). However, the vendor font support for the character has been non-existent up to this point.

Given that, now-a-days UTC recommends the atomic encoding *reph* for scripts that show alternate conjunct behavior with the consonant RA to simplify rendering and text encoding, we consider it is time that Telugu also benefitted from a similar approach and moved away from joiners.

Reph in Telugu

However, in older texts, U+0C30 TELUGU LETTER RA takes the reduced (or *reph*) form ఱ when it appears first in a consonant cluster, and the following consonant maintains its nominal form:

$$U+0C30 \text{ ఱ } ra + U+0C4D \text{ ఱ } virama + U+0C2E \text{ మ } ma \rightarrow \text{ఱమ} (rma)$$

U+200D ZERO WIDTH JOINER is placed immediately after the *virama* to render the *reph* explicitly in modern texts:

$$U+0C30 \text{ ఱ } ra + U+0C4D \text{ ఱ } virama + U+200D \text{ ZWJ } + U+0C2E \text{ మ } ma \rightarrow \text{ఱమ}$$

To prevent display of a *reph*, U+200D ZERO WIDTH JOINER is placed after the *ra*, but preceding the *virama*:

$$U+0C30 \text{ ఱ } ra + U+200D \text{ ZWJ } + U+0C4D \text{ ఱ } virama + U+0C2E \text{ మ } ma \rightarrow \text{ఱమ}$$

Unicode Standard 13.0, Chapter 12, Section 12.7

Reph is not currently used in Modern Telugu and is only used in historical and archaic orthographies. Therefore, it is necessary to preserve the identity of the character in historical documents, which the current handling of Repha does not allow. For instance, the syllable /rka/ is written as ఱక in the modern orthography and ఱక in the archaic orthography. But switching fonts might change the appearance. One may possibly over-ride using joiners and specify, which form of the conjuncts one desires. But the identity of *reph* is at the mercy of fonts, vendors and rendering engines, who are very patchy when it comes to supporting magical joiner sequences.

18/03/2020

Atomic Character for Reph

As a historical character, it would benefit from atomic encoding to gain widespread, simplified and consistent vendor support in terms of fonts and rendering engines, all the while also having its own identity in digitized documents.

Similar to the *reph* characters in Malayalam, Masaram Gondi & Dives Akuru, we recommend that the Telugu *reph* be placed before the character. However, the character should be re-ordered to the right of the consonant (like Masaram Gondi).

Archaic	Modern
$\boxed{\text{ఱ}} + \text{క} \rightarrow \text{కఱ}$	$\bar{\delta} + \text{క} \rightarrow \bar{\delta}\text{క}$
$\boxed{\text{ఱ}} + \text{క} + \text{ఱ}^6 \rightarrow \text{ఱ}^6\text{క}$	$\bar{\delta} + \text{క} + \text{ఱ}^6 \rightarrow \bar{\delta}\text{క}^6$
$\boxed{\text{ఱ}} + \text{క} + \text{ఱ}^8 \rightarrow \text{కఱ}^8$	$\bar{\delta} + \text{క} + \text{ఱ}^8 \rightarrow \bar{\delta}\text{క}^8$

(The placement of the Telugu *reph* is pretty much the same as the *reph* in Kannada)

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

0C5C $\boxed{\text{ఱ}}$ Telugu Reph

- *Valapalagilaka*
- *Must be placed before a consonant*

0C5C;TELUGU REPH;Lo;0;L;;;;;N;;;;;

Indic Syllabic Category

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

```
# Indic_Syllabic_Category=Consonant_Preceding_Repha
0C5C                    ; Consonant_Preceding_Repha # Lo                    TELUGU REPH
```

Indic Positional Category

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

```
# Indic_Positional_Category=Right
0C5C                    ; Right # Lo                    TELUGU REPH
```

18/03/2020

Sorting and Searching

The *reph* should be treated as an equivalent to the Unicode sequence $\bar{\text{ఠ}}$ U+0C30 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 the *reph*.

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 Reph
2. Requester's name:	<i>Vinodh Rajan, Shriramana Sharma, Suresh Kolichala</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual</i>
4. Submission date:	<i>18/03/2020</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/> Yes
(or) More information will be provided later:	<input type="checkbox"/>

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:	<input checked="" type="checkbox"/> Yes
Name of the existing block:	<i>Telugu</i>
2. Number of characters in proposal:	<i>1</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input type="checkbox"/> B.1-Specialized (small collection) <input checked="" type="checkbox"/> B.2-Specialized (large collection) <input type="checkbox"/>	
C-Major extinct <input type="checkbox"/> D-Attested extinct <input type="checkbox"/> E-Minor extinct <input type="checkbox"/>	
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/> G-Obscure or questionable usage symbols <input type="checkbox"/>	
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?	<input checked="" type="checkbox"/> Yes
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/> Yes
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Vinodh Rajan</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Vinodh Rajan, vinodh@virtualvinodh.com</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> Yes
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> 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)?	<input checked="" type="checkbox"/> Yes
	<i>Sorting & Searching</i>
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? If YES explain	No
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.)? If YES, with whom? If YES, available relevant documents:	Yes Suresh Kolichala & Vinodh Rajan -
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Yes See Proposal
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Rare See Proposal
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes See Proposal
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	Yes Yes 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? If YES, is a rationale for its inclusion provided? If YES, reference:	No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	Yes Yes 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? If YES, is a rationale for its inclusion provided? If YES, reference:	No
11. Does the proposal include use of combining characters and/or use of composite sequences? 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:	Yes Yes See Proposal Yes See Proposal
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No