Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

1. Introduction. Tulu is a Dravidian language spoken by nearly two million people, mainly in the southwest part of India known as Tulu Nadu. The language has a good deal of written literature, together with a rich oral literature such as the Siri Epic. The language was first written with the Tulu script, which was an adaptation of Grantha script. One Tulu inscription is ascribed to the 15th century by epigraphists; the Siri Epic can safely be ascribed to the 17th century as per the horoscope of the author recorded in the work. Today, a small number of people still write on palm leaves for astrological reasons. There is a practice of writing charms of small pieces of these leaves which are rolled and put in a small tubular metal pipe which is sealed on both ends and tied around the neck. They are considered to protect and ward of all evils from the wearer. The Jatakas (horoscopes) of the people were largely written on Palm leaves where Tulu characters and Kannada numbers were used. From the beginning of the 20th century this script was abandoned in favour of the Kannada script, but recently Tulu has begun a considerable revival. Its encoding in the UCS is a matter of some priority.

2. Structure. The Tulu script is of the Brahmic type. Consonants bear an inherent vowel *-a* which can be replaced by one or more vowel matras. Complex conjuncts as well as subscripted consonants are known; the orthographic difference between these has yet to be determined.

3. Independent vowel letters. Fourteen of the initial vowels are represented with the unique independent vowels ᰳ *a*, ᰴ *ā*, ᰍ *i*, ᰎ *ī*, ᰆ *u*, ᰇ *ū*, ᰘ *r̥*, ᰙ *r̥̄*, ᰚ *l̥*, ᰛ *l̥̄*, ᰁ *e*, ᰂ *ē*, ᰃ *o*, ᰄ *ō*. Other vowels which do not have independent forms are represented by vowel matras applied to the letters ᰳ *a*, ᰁ *e*, ᰂ *ē*, and ᰄ *ō*; ᰳ *u*, ᰳ *ū*, ᰳ *ε*, ᰳ *ε̄*, ᰁᰂ *ai*, ᰄᰍ *ov*, ᰳᰳ *aṁ*, ᰳ *aḥ*. A number of "special characters" are also represented by combnining vowel matras with ᰂ *ē*, and ᰄ *ō*: ᰂ *ayu*, ᰂ *eyi*, ᰂ *ayi*; ᰄ *ovu*, ᰄ *avu*, and ᰄ *avu*.

4. Dependent vowel signs. The dependent vowels are as follows (shown with ᰰ KA):

ᰰ *ka*   = ᰰ *ka*
ᰰ *ku*   = ᰰ *ka* + ◌ *-u*
ᰰ *kū*   = ᰰ *ka* + ◌ *-ū*
ᰰ *kā*   = ᰰ *ka* + ◌ *-ā*
ᰰ *ki*   = ᰰ *ka* + ◌ *-i*
ᰰ *kī*   = ᰰ *ka* + ◌ *-ī*

ꠀ *ku*   = ꠀ *ka* + ◌ -u
ꠀ *kū*   = ꠀ *ka* + ◌ -ū
ꠀ *kṛ*   = ꠀ *ka* + ◌ -ṛ
ꠀ *kṝ*   = ꠀ *ka* + ◌ -ṝ
ꠀ *kɛ*   = ꠀ *ka* + ◌ -ɛ
ꠀ *kɛ̄*   = ꠀ *ka* + ◌ -ɛ̄
ꠀ *ke*   = ꠀ *ka* + ◌ -e
ꠀ *kē*   = ꠀ *ka* + ◌ -ē
ꠀ *kai*   = ꠀ *ka* + ◌ -ai
ꠀ *ko*   = ꠀ *ka* + ◌ -e + ◌ -ā
ꠀ *kō*   = ꠀ *ka* + ◌ -ē + ◌ -ā
ꠀ *kov*   = ꠀ *ka* + ◌ -ē + ◌ -v

The relative order of the vowel signs as given here is the expected typing and storage order. In particular, ◌ -ā and ◌ -v complete a vowel string.

6. Final consonants and conjuncts. To be done.

7. Digits and punctuation. Script-specific digits exist and are proposed for encoding. Punctuation characters require further study, though an elaborate section-ending glyph is found in some of the manuscripts.

8. Collating order. At present it is expected that a Devanagari-style alphabetic sorting order will apply to Tulu.

9. Character names. The character names used follow the usual Brahmic conventions for the UCS.

10. Linebreaking. Opportunities for line-break occur after any full orthographic syllable, as in other scripts of India.

11. Unicode Character Properties.
To be done.

12. Bibliography.
To be supplied

	11B5	11B6	11B7	11B8	11B9	11BA
0	◌o 11B50	ⓐ 11B60	⾅ 11B70	ⓝ 11B80	ℙ◌ 11B90	
1	◌: 11B51	ⓥ 11B61	Ɜ 11B71	₂ 11B81	ℙℙ◌ 11B91	
2	℥ 11B52	ⓛ 11B62	ω 11B72	☻ 11B82	ℙ◌ 11B92	
3	℥ 11B53	ⓦ 11B63	⋒ 11B73	◌̈ 11B83	ℙℙ◌ 11B93	
4	ⴳ 11B54	ⴳ 11B64	⌣ 11B74	◌̔ 11B84	◌⌐ 11B94	
5	ⴳ 11B55	ⴲ 11B65	⌣ 11B75	◌⌐ 11B85	⌐ 11B95	
6	ℓ 11B56	Ɋ 11B66	⋒ 11B76	◌⌐ 11B86	ⵙ◌ 11B96	o 11BA6
7	ℓ 11B57	ⴳ 11B67	⋒ 11B77	◌⌐ 11B87	ⴲⴲ◌ 11B97	⌐ 11BA7
8	ⴳ 11B58	ⴲ 11B68	Ƨ 11B78	◌̥ 11B88	◌ɾ 11B98	⌐ 11BA8
9	ⴳ 11B59	ⴲ 11B69	⌣ 11B79	◌̥ 11B89	Ƨ 11B99	ⴽ 11BA9
A	ⴲ 11B5A	Ƨ 11B6A	⌒ 11B7A	◌̥ 11B8A	◌̊ 11B9A	ⴼ 11BAA
B	ⴲ 11B5B	⊙ 11B6B	ⴲ 11B7B	◌̥ 11B8B	◌́ 11B9B	Ɋ 11BAB
C	℥ 11B5C	ⴲ 11B6C	⌒ 11B7C	◌̥ 11B8C	◌̊ 11B9C	ε 11BAC
D	℥ 11B5D	ⴲ 11B6D	⋒ 11B7D	◌̥ 11B8D		℥ 11BAD
E	Ɜ 11B5E	ⴽ 11B6E	ⴲ 11B7E	◌̔ 11B8E		℮ 11BAE
F	Ɜ 11B5F	ⴲ 11B6F	ⴽ 11B7F	◌̈ 11B8F		ℇ 11BAF

Independent vowels

11B50	○ం	TULU LETTER ANUSVARA
11B51	○ః	TULU LETTER VISARGA
11B52	ꭰ	TULU LETTER A
11B53	ꭰ	TULU LETTER AA
11B54	ꭰ	TULU LETTER I
11B55	ꭰ	TULU LETTER II
11B56	ꭰ	TULU LETTER U
11B57	ꭰ	TULU LETTER UU
11B58	ꭰ	TULU LETTER VOCALIC R
11B59	ꭰ	TULU LETTER VOCALIC RR
11B5A	ꭰ	TULU LETTER VOCALIC L
11B5B	ꭰ	TULU LETTER VOCALIC LL
11B5C	ꭰ	TULU LETTER E
11B5D	ꭰ	TULU LETTER EE
11B5E	ꭰ	TULU LETTER O
11B5F	ꭰ	TULU LETTER OO

Consonants

11B60	ꭰ	TULU LETTER KA
11B61	ꭰ	TULU LETTER KHA
11B62	ꭰ	TULU LETTER GA
11B63	ꭰ	TULU LETTER GHA
11B64	ꭰ	TULU LETTER NGA
11B65	ꭰ	TULU LETTER CA
11B66	ꭰ	TULU LETTER CHA
11B67	ꭰ	TULU LETTER JA
11B68	ꭰ	TULU LETTER JHA
11B69	ꭰ	TULU LETTER NYA
11B6A	ꭰ	TULU LETTER TTA
11B6B	ꭰ	TULU LETTER TTHA
11B6C	ꭰ	TULU LETTER DDA
11B6D	ꭰ	TULU LETTER DDHA
11B6E	ꭰ	TULU LETTER NNA
11B6F	ꭰ	TULU LETTER TA
11B70	ꭰ	TULU LETTER THA
11B71	ꭰ	TULU LETTER DA
11B72	ꭰ	TULU LETTER DHA
11B73	ꭰ	TULU LETTER NA
11B74	ꭰ	TULU LETTER PA
11B75	ꭰ	TULU LETTER PHA
11B76	ꭰ	TULU LETTER BA
11B77	ꭰ	TULU LETTER BHA
11B78	ꭰ	TULU LETTER MA
11B79	ꭰ	TULU LETTER YA
11B7A	ꭰ	TULU LETTER RA
11B7B	ꭰ	TULU LETTER LA
11B7C	ꭰ	TULU LETTER VA
11B7D	ꭰ	TULU LETTER SHA
11B7E	ꭰ	TULU LETTER SSA
11B7F	ꭰ	TULU LETTER SA
11B80	ꭰ	TULU LETTER HA
11B81	ꭰ	TULU LETTER LLA
11B82	ꭰ	TULU LETTER LLLA

Vowel signs

11B83	○	TULU VOWEL SIGN UE
11B84	○	TULU VOWEL SIGN UEE
11B85	○	TULU VOWEL SIGN AA
11B86	○	TULU VOWEL SIGN I
11B87	○	TULU VOWEL SIGN II
11B88	○	TULU VOWEL SIGN U
11B89	○	TULU VOWEL SIGN UU
11B8A	○	TULU VOWEL SIGN VOCALIC R
11B8B	○	TULU VOWEL SIGN VOCALIC RR
11B8C	○	TULU VOWEL SIGN VOCALIC L
11B8D	○	TULU VOWEL SIGN VOCALIC LL
11B8E	○	TULU VOWEL SIGN SHORT E
11B8F	○	TULU VOWEL SIGN SHORT EE
11B90	○	TULU VOWEL SIGN AVU
11B91	○	TULU VOWEL SIGN AYU
11B92	○	TULU VOWEL SIGN EE
11B93	○	TULU VOWEL SIGN AI
11B94	○	TULU VOWEL SIGN OV
11B95	○	TULU VOWEL SIGN OVU
11B96	○	TULU VOWEL SIGN EYI
11B97	○	TULU VOWEL SIGN AYI

Virama

11B98	○	TULU VIRAMA
11B99	ꭰ	TULU AVAGRAHA
11B9A	○	TULU CANDRABINDU
11B9B	○	TULU ARKA
11B9C	○	TULU ARKA HALANT

Digits

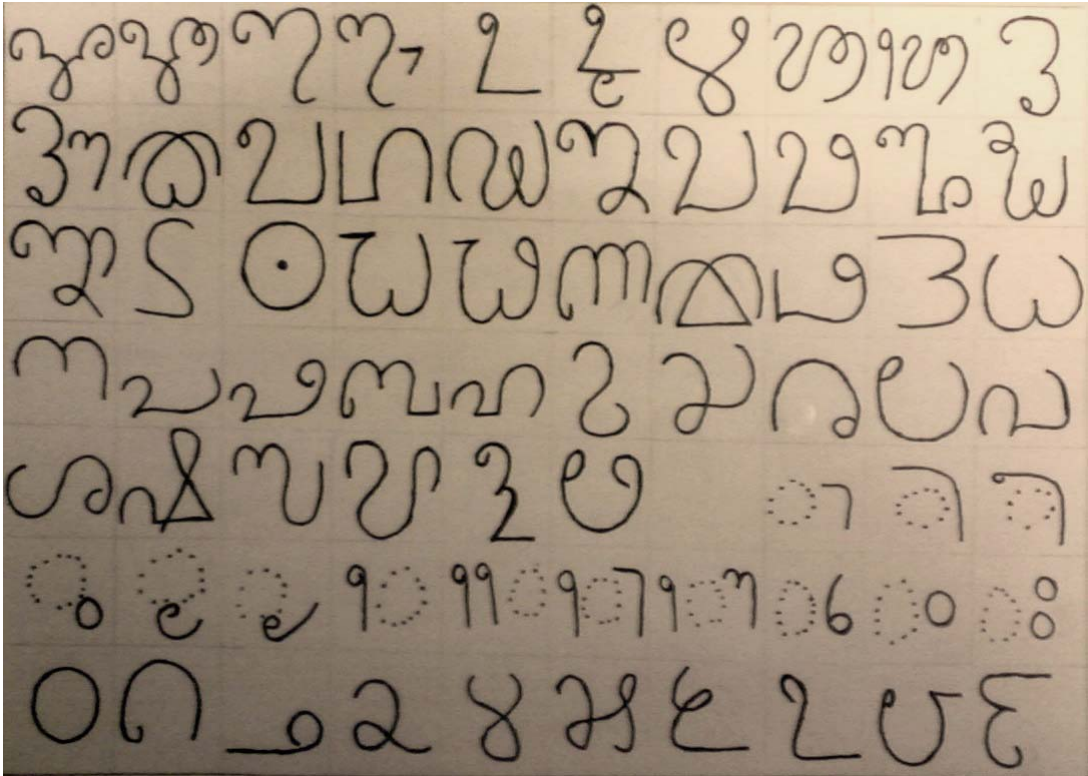|		
11BA6	ꭰ	TULU DIGIT ZERO
11BA7	ꭰ	TULU DIGIT ONE
11BA8	ꭰ	TULU DIGIT TWO
11BA9	ꭰ	TULU DIGIT THREE
11BAA	ꭰ	TULU DIGIT FOUR
11BAB	ꭰ	TULU DIGIT FIVE
11BAC	ꭰ	TULU DIGIT SIX
11BAD	ꭰ	TULU DIGIT SEVEN
11BAE	ꭰ	TULU DIGIT EIGHT
11BAF	ꭰ	TULU DIGIT NINE

Figures

Figure 1. Handwritten sample of part of the Tulu alphabet and digits.
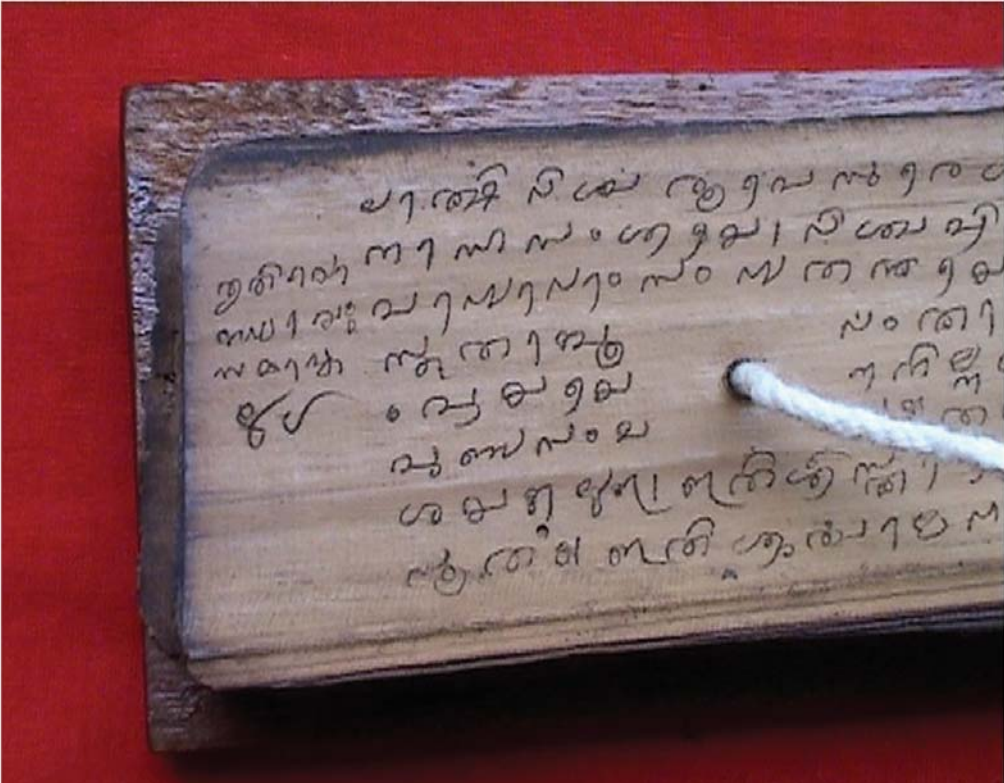
Figure 2. Part of a Tulu palm-leaf manuscript.

Figure 3. More samples of handwritten Tulu.

A. Administrative

1. Title

Preliminary proposal for encoding the Tulu script in the SMP of the UCS

2. Requester's name

UC Berkeley Script Encoding Initiative (Universal Scripts Project)

3. Requester type (Member body/Liaison/Individual contribution)

Liaison contribution.

4. Submission date

2011-04-22

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

No.

6b. More information will be provided later

Yes.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Tulu.

1c. The proposal is for addition of character(s) to an existing block

No.

1d. Name of the existing block

2. Number of characters in proposal

87.

3. Proposed category (A-Contemporary; 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)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, FontLab.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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.

8. 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/Public/UNIDATA/ UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. 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.

2b. If YES, with whom?

Peter Bilak.

2c. 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?

See above.

4a. The context of use for the proposed characters (type of use; common or rare)

To write the Tulu language.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

In general publications.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

No.

6b. If YES, is a rationale provided?

6c. If YES, reference

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?