1. Introduction

Currently Unicode Tamil range does not include the Tamil grantha[1] character SHA. This document proposes addition of this character in Unicode character set, within the current Tamil range U+0B80 to U+0BFF.


1.1. Character proposed in this document

<table>
<thead>
<tr>
<th>Shape of proposed character</th>
<th>Proposed name</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>U0D</td>
<td>TAMIL LETTER SHA</td>
<td>0BB6</td>
</tr>
</tbody>
</table>

1.2. About INFITT and INFITT WG02

International Forum for Information Technology in Tamil (INFITT) is a non-governmental organization, with cooperation from various state and national governments in the Tamil speaking countries. It also has participation from industry representatives and individuals with expertise in Information Technology and Tamil language.

Working Group 2 (WG02) of INFITT deals with definition and implementation of Unicode in Tamil. WG02 also has a Liaison relationship with Unicode Consortium. Members of this working group are derived from various nations around the world, and represent interests of IT industry, governments, Tamil scholars, educators and other Tamil IT enthusiasts. This proposal has been prepared with the consensus of members of this working group.

1.3. Demographics

Tamil is the state language of the state of Tamil Nadu in India. It is also a national language in Sri Lanka and Singapore (Source: CIA World Factbook - http://www.cia.gov/cia/publications/factbook/index.html).

Total population of Tamil speakers around the world is estimated to be around 74 million (1999 estimate) as per the Ethnologue index prepared by Summer Institute of Linguistics. See the Ethnologue page for Tamil http://www.ethnologue.com/show_language.asp?code=TCV which provides details about the population and distribution of Tamil diaspora.

2. TAMIL LETTER SHA

This document proposes to encode Tamil grantha character SHA in unicode Tamil range at U+0BB6.
2.1. Usage of Tamil letter SHA

Devanagari and other Indian scripts have a character named SHA in their Unicode block. However in Tamil the corresponding slot is left vacant. While not in popular use, there does exist a character for letter SHA in Tamil. This document proposes to add Tamil letter SHA in Unicode Tamil range at U+0BB6.

American Library Association-Library of Congress (ALA-LC) Romanization scheme includes this character in their list. (http://lcweb.loc.gov/catdir/cps0/romanization/tamil.pdf)

Digital South Asia Library of Univ of Chicago and other American Universities also use the same ALA-LC scheme that recognizes this “sa” as one of the grantha characters repertoire. (http://www.lib.uchicago.edu/e/su/southasia/transliteration-tamil.html)

See also "Transliteration of Indic scripts; How to use ISO 15919", (http://homepage.nllworld.com/stone-catend/trind.htm) Table 3, shown as example 5 under section 3.

2.2. Data Processing for SHA

SHA behaves similar to other Tamil grantha letters, when combined with vowel signs. Specifically, U and UU vowels signs for SHA are similar to other grantha letters. In terms of other Unicode properties, the proposed character has properties same as U+0BB8 (TAMIL LETTER SA).

SHA is usually sorted after the Tamil letter NNNA and before all the grantha characters. However, minor differences might exist in this behaviour between various sort orders.

Inclusion of character SHA in Tamil range will also help in transliteration from other indic scripts into Tamil.

2.3. Ligature SRI[2]

While the base character SHA is not widely used in modern literature and communications, its ligature form SRI is used widely. The character sequence of SRI is as below:

Proposed TAMIL LETTER SHA +
TAMIL LETTER RA (U+0BB0) +
TAMIL VOWEL SIGN II (U+0BC0)

This usage corresponds with the sequence of SRI in other indic scripts.

Current implementations of unicode define SRI in Tamil as follows:

TAMIL LETTER SA (U+0BB8) +
TAMIL LETTER RA (U+0BB0) +
TAMIL VOWEL SIGN II (U+0BC0)

This is incorrect combination of SRI. This error is due to lack of TAMIL LETTER SHA in current unicode repertoire.

Proposed character SHA may also form various other ligatures in combination with MA, YA, RA, and VA. However, these ligatures are archaic and are not widely recognized. Contemporary publications only use disjoint forms.
3. Published Usage Examples

Example 1: Pronunciation of SHA

This example explains the pronunciation of the proposed character SHA.
Example 2: Transliteration of SHA

This example gives the transliteration of proposed character SHA in Devanagari.
Example 3: Shankarar

Name of Shankarar is spelt with this proposed SHA. The trasnliteration of the verse also uses SHA.

INSPIRATIONS OF ACHARYA

I know not Sabda, I know not Artha,
I know not: Padyam, I know not Gadyam.
But there shines in My Hrudadam six faced luminous light,
Brilliant words flow out from my mouth extempore
Example 4: Vishnu sahasranaamam

First verse of Vishnu sahasranaamam uses SHA in several places.
Example 5: Transliteration of Indic Scripts: How to use ISO 15919

Transliteration of proposed character SHA is shown as "s<acute>a" in the table below.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pa</td>
<td>प</td>
<td>ป</td>
<td>འ</td>
<td>প</td>
<td>ೆ</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>pha</td>
<td>फ</td>
<td>ผ</td>
<td>ཡ</td>
<td>ফ</td>
<td>ೇ</td>
<td>th</td>
<td>th</td>
<td>th</td>
<td>th</td>
<td>th</td>
<td></td>
</tr>
<tr>
<td>ba</td>
<td>ब</td>
<td>བ</td>
<td>ལ</td>
<td>ব</td>
<td>ೋ</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td></td>
</tr>
<tr>
<td>bha</td>
<td>भ</td>
<td>བ</td>
<td>ས</td>
<td>ভ</td>
<td>ೊ</td>
<td>th</td>
<td>th</td>
<td>th</td>
<td>th</td>
<td>th</td>
<td></td>
</tr>
<tr>
<td>ma</td>
<td>म</td>
<td>ད</td>
<td>ཕ</td>
<td>ম</td>
<td>ೋ</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>ſba</td>
<td>ſ</td>
<td>ſ</td>
<td>ſ</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td></td>
</tr>
<tr>
<td>ra</td>
<td>(र)</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>9.</td>
</tr>
<tr>
<td>ta</td>
<td>(ट)</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>10.</td>
</tr>
<tr>
<td>na</td>
<td>(न)</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>la</td>
<td>(ल)</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>ya</td>
<td>य</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>ša</td>
<td>श</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>11.</td>
</tr>
<tr>
<td>ša</td>
<td>ष</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td></td>
</tr>
<tr>
<td>sa</td>
<td>स</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>ha</td>
<td>ह</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>〈</td>
<td>〈</td>
<td>〈</td>
<td>〈</td>
<td>〈</td>
<td>〈</td>
<td>〈</td>
<td>〈</td>
<td>〈</td>
<td>〈</td>
<td>〈</td>
<td>2</td>
</tr>
<tr>
<td>qa</td>
<td>क</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td></td>
</tr>
<tr>
<td>kha</td>
<td>क़</td>
<td>kḥ</td>
<td>kḥ</td>
<td>kḥ</td>
<td>kḥ</td>
<td>kḥ</td>
<td>kḥ</td>
<td>kḥ</td>
<td>kḥ</td>
<td>kḥ</td>
<td></td>
</tr>
<tr>
<td>ġa</td>
<td>ġ</td>
<td>ġ</td>
<td>ġ</td>
<td>ġ</td>
<td>ġ</td>
<td>ġ</td>
<td>ġ</td>
<td>ġ</td>
<td>ġ</td>
<td>ġ</td>
<td></td>
</tr>
<tr>
<td>za</td>
<td>ज़</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td>z</td>
<td></td>
</tr>
<tr>
<td>fa</td>
<td>फ</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>wa</td>
<td>व</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td></td>
</tr>
</tbody>
</table>
Example 6: Ligature SRI

This wedding invitation contains the letter SRI (also transliterated in English as SHRI or SHREE).

Notes:

[1] Grantha
   Grantha characters are part of Tamil script. Historically these characters are not considered to be part of proper Tamil Language, but are used to write non-Tamil words and names. These were first introduced by rulers of Pallava dynasty during 6-7th centuries A.D. These characters are widely used by Tamils in Indian subcontinent and the Tamil diaspora in other countries.

[2] Sri
   Also written as SHRI or SHREE. Sri in Sanskrit means wealth. Hindu goddess of wealth is called "SRIDEVI". The prefix SRI is often added to many gods and goddesses and also while formally addressing elders.
ISO/IEC JTC 1/SC 2/WG 2

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

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

(Please read Principles and Procedures Document for guidelines and details before filling this form.)


A. Administrative

1. Title: _______Addition of Tamil character SHA___________
2. Requester’s name: _International Forum for Information Technology______
   ___in Tamil (INFITT)_____________________________________
3. Requester type (Member body/Liaison/Individual contribution): __Liaison___
4. Submission date:                                           __2003-03-10___
5. Requester’s reference (if applicable): _____WG02 SHA Proposal_____________
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):     _____NO______
      Proposed name of script: _______________________________________________
   b. The proposal is for addition of character(s) to an existing block: _YES_
      Name of the existing block: ___Tamil__Range: 0B80-0BFF____________________
2. Number of characters in proposal:                           ______1_______
3. Proposed category (see section II, Character Categories):   ______A_______
4. Proposed Level of Implementation (1, 2 or 3)
   (see clause 14, ISO/IEC 10646-1: 2000):   _______1_______
   Is a rationale provided for the choice?                      ______YES_____
   If Yes, reference:__See attached proposal_______________________________
5. Is a repertoire including character names provided?         ______YES_____
   a. If Yes, are the names in accordance with the ’character naming
   b. Are the character shapes attached in a legible form suitable for review? ___YES__
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
   ________________________________________________________________
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
   ___________________________________________________________________
   ___________________________________________________________________
   ___________________________________________________________________
7. 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___
8. 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. See attached proposal.

9. 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, 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.

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? __INFITT WG02_____________________________

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 section 1.3 of attached proposal________________________

4. The context of use for the proposed characters (type of use; common or rare) __Common____
   Reference: __See attached proposal______________________________________

5. Are the proposed characters in current use by the user community? __YES___
   Reference: __See attached proposal________________________

6. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP? __YES____
   If YES, is a rationale provided? __YES____
   If YES, reference: __See section 2.1 of attached proposal________________

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? _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? _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 (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)? __YES__
    If YES, is a rationale for such use provided? __YES__
    If YES, reference: __See attached proposal section 2.3__________________

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

A.1 Submitter’s Responsibilities

The national body or liaison organization (or any other organization or an individual) proposing new character(s) or a new script shall provide:

1. Proposed category for the script or character(s), character name(s), and description of usage.
2. Justification for the category and name(s).
3. A representative glyph(s) image on paper:
   If the proposed glyph image is similar to a glyph image of a previously encoded ISO/IEC 10646 character, then additional justification for encoding the new character shall be provided.
   **Note:** Any proposal that suggests that one or more of such variant forms is actually a distinct character requiring separate encoding, should provide detailed, printed evidence that there is actual, contrastive use of the variant form(s). It is insufficient for a proposal to claim a requirement to encode as characters in the Standard, glyphic forms which happen to occur in another character encoding that did not follow the Character-Glyph Model that guides the choice of appropriate characters for encoding in ISO/IEC 10646.
   **Note:** WG 2 has resolved in Resolution M38.12 not to add any more Arabic presentation forms to the standard and suggests users to employ appropriate input methods, rendering and font technologies to meet the user requirements.
4. Mappings to accepted sources, for example, other standards, dictionaries, accessible published materials.
5. Computerized/camera-ready font:
   Prior to the preparation of the final text of the next amendment or version of the standard a suitable computerized font (camera-ready font) will be needed. Camera-ready copy is mandatory for final text of any pDAMs before the next revision. Ordered preference of the fonts is True Type or PostScript format. The minimum design resolution for the font is 96 by 96 dots matrix, for presentation at or near 22 points in print size.
6. List of all the parties consulted.
7. Equivalent glyph images:
   If the submission intends using composite sequences of proposed or existing combining and non-combining characters, a list consisting of each composite sequence and its corresponding glyph image shall be provided to better understand the intended use.
8. Compatibility equivalents:
   If the submission includes compatibility ideographic characters, identify the equivalent unified CJK Ideograph character(s).
9. Any additional information that will assist in correct understanding of the different characteristics and linguistic processing of the proposed character(s) or script.