Dated: April 28, 2006

Title: Proposal to add TAMIL OM

Source: International Forum for Information Technology in Tamil (INFITT)

Action: For consideration by UTC and ISO/IEC JTC 1/SC 2/WG 2

Distribution: ISO/IEC JTC 1/SC 2/WG 2 members and Liaison organizations

1. Introduction

Currently Unicode Tamil range does not include the sign, TAMIL OM. This document proposes addition of this character in Unicode, within the current Tamil range U+0B80 to U+0BFF.

ISO/IEC JTC 1/SC 2/WG 2 Proposal summary form N2652-F accompanies this document.

1.1. Character proposed in this document

Shape of proposed character	Proposed name	Proposed position
ௐ	TAMIL OM	0BD0

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). Tamil is one of the widely spoken languages found in almost all countries of the world, and its non-conjunct script makes it easy for use in the net and learning even by Diaspora children.

So, Tamil in Unicode is one of the widely employed scripts originating in India. There are approximately 1200 bloggers in Tamil (Source: http://www.thamizmanam.com), and 'anbudan' Unicode Tamil e-list (http://groups.google.com/group/anbudan/about) had 28,000 messages in its first year of operation.

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 Ethonologue 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 OM

This document proposes to encode the sign named TAMIL OM in Unicode Tamil range at U+0BD0. TAMIL OM property values for data processing are:
0BD0;TAMIL OM;Lo;0;L;;;;;N;;;;;

2.1. Usage of Tamil OM

Devanagari and Gujarati scripts have a sign named OM in their Unicode ranges. However in Tamil the corresponding slot is left vacant. Gurmukhi script also has an OM sign. Tamil OM sign is widely used in Hindu religious texts, temple publications, and as neon lamps of sign boards in shops etc. OM is a religious concept found in all three major religions born in India viz. Hinduism, Jainism and Buddhism. This document proposes to add the character TAMIL OM in Unicode Tamil range at U+0BD0.

2.2 Transliterating Tamil OM

In Indian scripts, Om has an anusvara following long O. There are linguistic differences between a full "m" and an anusvara "m". Tamil Om (Section 1.1) has an anusvara M integrated inside just like other Indic scripts treat the anusvara m of Om in unique ways.

Linguists have considered the question of OM origins in Indian languages and religions (Hinduism, Buddhism and Jainism) and concluded that OM is native to Dravidian. OM usually called Ongaaram in literature and is considered a single letter. This is because OM in Indian scripts (eg., Hindi, Gujarati, Tamil) is generally a separate sign because the M in OM is a nasalized m (m with a bindu/candrabindu) and not full m. A linguistics reference for OM sign originating from Dravidian, and subsequently borrowed into Indo-Aryan: Asko Parpola "On the Primary Meaning and Etymology of the sacred Syllable OM" in *Studia Orientalia* [Finnish Oriental Society] vol 50, 1980, pp. 195-213.

Until Unicode 4.1 there was no Tamil Zero in Unicode. Tamil Zero was added to facilitate interoperability scenarios between Indic scripts. Similarly Tamil Om sign in Unicode will help immensely for transliterating between different Indic scripts. See the examples of Tamil and Devanagari scripts in Section 4. The following Table 1. provides the standard transliteration for OM into Roman script and a sample of OM signs in various Indic scripts including Tamil script are given .

Table 1. OM sign in some Indic scripts
and its Roman transliteration

Script	Om sign	Transliteration (ISO 15919)	Unicode code-point
Gujarati	ૐ	ōm̐	0AD0
Devanagari	ॐ	ōm̐	0950
Tamil	ௐ	ōm̐	0BD0 (proposed)
Gurmukhi	ੴ	ōm̐	0A74

All the encoded OM signs in Indic scripts in Unicode do not involve the use of joiners like ZWJ or ZWNJ. It is recommended that this procedure of nonuse of ZWJ be followed for encoding Tamil OM also. Use of ZWJ has been avoided for other unique letters of Indic scripts, for example, in the encoding of Bengali Khanda Ta [Reference 1]. For the Malayalam chillu pre-pausal consonant letters, the existence of about 25+ chillus have been found. There are proposals to avoid their implementation using ZWJ [an example, Reference 2]. For several Dravidian script unique letters such as Telugu alveolar markers on letters CA and JA, it is proposed that joiners be not used as already done for Bengali

Khanda TA. On the non-use of ZWJ for Malayalam Chillu letters, also refer to the letters supporting that position from Drs. Gautam Sengupta, a linguist and a native expert on Bengali and K. G. Sulochana, a Malayalam expert from Kerala in the Indic (Unicode) list during April, 2006 discussions.

Different stylistic glyph variants of the Tamil OM symbol are known to exist, and so the proposed U+0BD0 code-point may contain slightly different glyphs for the Tamil OM symbol in the future. In the absence of the Tamil OM symbol in a font or a typeface, the sequence O + Ma + PuLLi/Virama may be used for rendering purposes.

When one searches for the Tamil OM sign, the search engine should get hit only on that sign. It should not produce the O+Ma+Virama. This is because of the fact that there is a difference between O followed by a whole consonant m and O followed by an anuswara M. In India, OM consists of the anuswara M as shown in Table 1, and this is found in the formal academic transliteration of OM into Roman script with a candrabindu (U+0310) placed over the letter, m.

3. References

1. Review of Bengali Khanda Ta and PRI-30 Feedback, P. Constable.
L2/04-252, June 16, 2004.
2. Dravidian script "markers" code-points in Unicode, Dr. N. Ganesan,
L2/06-134, April 17, 2006.

4. Usage Examples

Example 1:

Taittiriya Upanishad is an ancient Vedic text. It is given in the Tamil script at,
http://www.srivaishnavam.com/stotras/shiixavalli_tamil.pdf
Tamil OM sign from page 7 is shown below:

ப்₃ரூஹஸ்பதி: | ஶங்கோ₃ விஷ்ணுருருக்ரம: | நமோ ப்₃ரஹ்மணே | நமஸ்தே

வாயோ | த்வமேவ ப்ரத்யக்ஷம் ப்₃ரஹ்மாஸி | த்வாமேவ ப்ரத்யக்ஷம்

ப்₃ரஹ்மாவாதி₃ஷம் | ருதம வாதி₃ஷம் | ஸத்யமவாதி₃ஷம் | தன் மாமாவீத் |

தத்₃வக்தார மாவீத் | ஆவீந்மாம் | ஆவீத்₃வக்தாரம் | ௐ ஶாந்தி: ஶாந்தி:

ஶாந்தி: ||

ஹரி: ௐ(3) → Tamil OM

The corresponding text of the Taittiriya Upanishad in Devanagari text is at
http://www.srivaishnavam.com/stotras/shiixavalli.pdf   The Devanagari OM from
page 6 is shown below:

शन्नो मित्रः शं वरुणः । शन्नो भवत्वर्यमा । शन्न इन्द्रो बृहस्पतिः । शन्नो विष्णुरुरुक्रमः ।
नमो ब्रह्मणे । नमस्ते वायो । त्वमेव प्रत्यक्षं ब्रह्मासि । त्वामेव प्रत्यक्षं ब्रह्मावादिषं ।
ऋतमवादिषं । सत्यमवादिषं । तन् मामावीत् । तद्वक्तारमावीत् । आवीन्माम् ।
आवीद्वक्तारम् । ॐ शान्तिः शान्तिः शान्तिः ॥

हरिः ॐ (३) → Devanagari OM corresponding to Tamil OM.

Example 2:

A popular religious magazine published in Tamil Nadu is Om Sakthi. Its January 2006 cover page is given below. Note the Tamil Om followed by the word, Sakthi.

Example 3:

Mahakavi Bharatiyar is the national poet of Tamil Nadu, and his famous poem extolling the great Goddess is from page 45 of March 2006 issue.

"மறுபிறவி உண்டு"

"திருக்குரானில் மறுபிறவியா? என்ற கேள்வி எழுப்பியுள்ள கட்டுரையாளர் ஹஜ்ஜா சித்தை செளதா அவர்கள், திருக்குரான் வசனங்களுள் "மறுபிறவி" பற்றி ஏதேனும் குறிப்பு உள்ளனவா என்று கேட்கிறார்!

உலகின் இறுதிக் காலத்தில் இயேசு தோன்றுவார் என்று அம் மதத்தவரால் எதிர்பார்க்கப்படு கிறார்! அது சரியோ, தவறோ.

இறுதித் தீர்ப்பு நாள் வேளையில், இறைவனால் மறைக்கப்பட்ட ஈஸா நபி (அலை) அவதரிப்பார்; காட்சி யளிப்பார் என்று இஸ்லாம் கூறு மானால், 'மறுபிறவி' உண்டு என்ற பொருளைத்தானே அது தருகிறது.

'மறுபிறவி' கூற்றுக்கு ஆதார மாகத் திருக்குர்ஆனில் இருந்து எடுக்கப்பட்ட ஆதாரங்கள் போது மானதாக இல்லை என்பதையே இங்கு கவனத்திற்குக் கொணர வேண்டியிருக்கிறது. முதலில் 39 : 5 வசனம் அவசியமே இல்லை.

7 : 16 வசனம் "இழிந்த குரங்கு களாக ஆகிவிடுங்கள்" என்பது, மறு பிறவி என்று தானே பொருளாகிறது.

இறைவன் படைப்பிலே மனி தனே மேன்மையானவன் என்பதை மனிதர்களே உணர்ந்தாலும், அவர் களில் சிறிய குணநலன் இல்லாத சிலர், தம்மை இன்னவரின் மறுபிறவி என்றோ – முன்னோருள் முன்னவ

ஓசக்தி

உயிரெனத் தோன்றி உணவு கொண்டே வளர்ந்து
ஓங்கிடும் சக்தியை ஓது கின்றோம்.
பயிரினைக் காக்கும் மழையென எங்களைப்
பாலித்து நித்தம் வளர்க்க என்றே.

சித்தத்திலே நின்று சேர்வது உணரும்
சிவசக்தி தன்புகழ் செப்பு கின்றோம்.
இத்தரை மீதினில் இன்பங்கள் யாவும்
எமக்குத் தெரிந்திடல் வேண்டு மென்றே.

மாறுதல் இன்றிப் பராசக்தி தன்புகழ்
வண்ணமுற நித்தம் பாடு கின்றோம்.
நூறு வயது புகழுடன் வாழ்ந்து உயர்
நோக்கங்கள் பெற்றிட வேண்டு மென்றே.

ஓசக்தி ஓசக்தி ஓசக்தி ஓசக்தி
ஓசக்தி யென்று உரை செய்திடுவோம்
ஓசக்தி யென்பவர் உண்மை கண்டால் சுடர்
ஒண்மை கொண்டார், உயிர் வண்மை கொண்டார்.

– மகாகவி பாரதியார்

Tamil OM (handwritten annotation)

பரம்பரையினர் என்று சொல்ல வேண்டுமேயொழிய, குலத்தவரின் மறுபிறவிகள் என்று கட்டுரையாள ரால் சொல்லப்படுவது சரியாகப் படவில்லை.

A. Administrative

```
1. Title:    ____Addition of Tamil OM_____
2. Requester's name: _International Forum for Informational
   Technology_____ _in Tamil (INFITT)_____
3. Requester type (Member body/Liaison/Individual contribution):
   __Liaison___
4. Submission date:          __2006-4-30___
5. Requester's reference (if applicable): ___WG02 OM 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 (select one from below - see section 2.2 of P&P
   document):
  A-Contemporary   _YES_  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. Proposed Level of Implementation (1, 2 or 3)
      (see Annex K in P&P document):       _____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
```

```
   guidelines" in Annex L of P&P document?     ____YES____
   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? ___Mr. Muthu Nedumaran_(muthu@murasu.com)__
   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, 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? _____INFITT WG02_____
   If YES, available relevant documents: _See section 1.2 of
   attached_proposal._____
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_    If YES, where?  Reference: __See attached 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:    __See 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?  __POSSIBLE YES__
   If YES, is a rationale for its inclusion provided?  __YES__
   If YES, reference: __See section 2.2 of attached proposal.____
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 section 2.2 of attached proposal.___
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?       _____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
    character(s)?  _NO_
    If YES, is the equivalent corresponding unified ideographic
    character(s) identified?  _____
    If YES, reference: _____
```