Dated: February 1, 2004

Title:	Proposal to add Tamil Digit Zero (DRAFT)
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

Unicode Tamil range encodes Tamil numerals. However it doesn't include the number zero. This document proposes addition of this character in Unicode character set, within the current Tamil range U+0B80 to U+0BFF.

ISO/IEC JTC 1/SC 2/WG 2 Proposal summary form N2352-F accompanies this document.

1.1. Character proposed in this document

Shapes of proposed characters	Proposed name	Position
O	TAMIL DIGIT ZERO	0BE6

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 implemenation 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 DIGIT ZERO

Tamil digit zero is not currently encoded in Unicode range. This document proposes to encode this character in Unicode Tamil range at position U+0BE6.

2.1. Tamil number system

Ancient Tamil number system did not include the digit zero. It had separate glyphs for numbers 10, 100, and 1000. However in modern practice, Tamil numerals are used in the same way as decimal numbers, with a zero that looks similar to the digit zero in international form of indian numerals.

2.2. Tamil numerals in Unicode

Currently unicode encodes Tamil numerals 1 to 9, 10, 100, and 1000. The proposed position in the Tamil Unicode range (U+0BE6) is reserved and has the comment as equalent to digit zero (U+0030). It is proposed that Tamil zero be encoded at this position.

2.3. Compatibility between other Indic scripts

Digit zero is available in all other indic scripts encoded in Unicode. Having the digit in Tamil block would help during transliteration between the indian languages.

2.4. Usage of Tamil Digit Zero

Even though the Tamil Digit Zero is not widely used, it is often used in some traditional publications, especially in Tamil calendars. Examples in section 4 give various areas this digit is used.

3. Data processing for Tamil Digit Zero

Data processing for digit zero is same as other Tamil digits.

Property values are:

```
0BE6;TAMIL DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
```

4. Published Usage Examples

Example 1: Tamil Calendar

Page from calendar for Tamil month *markazi* (December 2001- January 2002). Gregorian dates are shown in the bottom right corner, while the Tamil dates are shown in the center in Tamil numerals, with international form of indian numerals in the inset.

"தமிழினத்தின்"	சிலை		டிசம்பர்
சாதிகளைத் தூண்டிவிட்டுக் கட்சிகளாய்	1	மயிலை சீனி வேங்கடசாமி தோற்றம்	16
	9	ரங்க பெரியார் நிளைவு	24

Example 2: Book: pirapantat tirattu

Tamil numerals including Tamil number zero are used in the page numbers, and verse numbers in the following example from the book *pirapantat tirattu* by *makAvitvAn mInATcicuntaram piLLai*, compiled and edited by *U.Ve. cAminAta Aiyar*.

வளிவார் பெரும்புலி தெரிக்கவவ் வணரிந்தின் மாண்புடைக் கலயதுரை - மற்
றெம் இயாஞ்சென்று தேடிவே மாதலால் வளயிக்க கழகந்தொருத், தெளிவார்
குழாக்குழுமி யோக்குமா வடிதுறைச் செல்வன்முத் தத்தருகவே - சின்மய
னரூட்பெருங் குரவனம் பலவான தேவன்முத் தத்தருகவே. (௭)

886. முன்கா னினருல் வெலப்பட்ட கரிவன்றி முதுமீன முதவுமு
த்து - மண்டக மலர்த்தாளி லரைபட்ட வெண்மதிய முத்துஞ் சுணவுதத்த,
புன்னைவாய் செய்தற் கடற்றேன்று மிப்பிவனீ பொரையுயிர்த் திட்டமுத்தம்-
பொருதிக லழிந்தகொக் குதவுமுத் துங்கொண்டு பூணத நினக்கேட, னென்
கூயோ துந்திற மெனிந்சின்ன மேற்றதா ரிக்காளி லில்லையென்னி - விலகுகண்
மணிமார்கல யிடையிடைக் கோக்திடக வெண்ணில பழுத்தபாக, தென்னைமா வா
திபல குழுமா வடிதுறைச் செல்வன்முத் தத்தருகவே - சின்மய னரூட்பெருங்
குரவனம் பலவான தேவன்முழ தத்தருகவே. (௬)

887. பல்காவு நின்கைக் துதித்துமின் பதமுட் பதித்துவாழ் வேங்களுக்
குப் - பயனுக வொருதர மெனுந்தக் கிடாயெனித் பல்வளாப் பூவனத்தி, லல்காத
மாமைநிறை பொன்னனையன் முன்கா எள்ளவிமுத் தங்கொள்ளுமா - ளாவ
ளாய் மேயதை வெளிப்பட வுரைத்திடவ மதனிய மிஃகாயெனி, ஜெல்காத
மற்றவ ஞதிர்க்குறி கபோலத்தி யுற்றது முரைத்துவிடேயோ - மொன்கதிர்ச்
செல்வன்மதி யோனெனத் தண்ணென் றுடம்புக னிக்கடந்து, செல்காரம்
வளபொழி யுயர்த்தவா வடிதுறைச் செல்வன்முத் தத்தருகவே-சின்மய னரூட்
பெருங் குரவனம் பலவான தேவன்முத் தத்தருகவே. (௬௦)

ரு.—முத்தப்பருவம் முற்றிற்று.

சூ.—வாராணப்பருவம்.

888. உறத்தகதிர் மண்டிலத் துறைவது தெரிப்பதென வொண்மேனி
சேயொளிசெய - வோங்குமதி மண்டிலத் துறைவது தெரிப்பதென வொளிர்
நீறு வெள்ளொளிசெய, வறக்தல பெடிப்பவொளி ரிவ்வொளிக னேக்குதொறு
மடியார்மல பத்தருமுழுது - மடலற வழிழ்த்தழி விலாததே ரின்பநில யருளிற் கொ
டுத்திடலாற், பிறந்தபிற தாபமும் புகழுமே யென்றநிருர் பேசிமகிழ் ஞக்கவ
எவை-பேசிய துணர்த்துதகை செய்தென முகத்தப் பிறக்குபுன் மூராேன்னச்,
சிறந்ததிரு வாவடி துறைக்கண்வளர் திருவருட் செல்வப்பிரான் வருகவே-செம்
மைமதிகழ் சிவஞான வடிவனம் பலவான தேசிகோத் தமன்வருகவே. (க)

889. உற்றபெரு மாதவ முருந்துறகர்க் கல்லா திருஞ்றுதவ மொன்றுமில்
லா - வுவர்வருக வென்றமுழத் திடா மடைந்திடத வரிமையன் றென்றுநினே
யல், பொற்றமதி நிலவுகக ரன்றிவெக காட்டிலும் புணரிவா யுண்டெழுந்த -
புயன்மழை செழும்பூண நிலத்தன்றி யுவர்பற்று புயியகத் தம்பொழியுமே,
தற்றமலர் தவிர்மெத் தெனப்பொர்ப் பினமுளத் தாயர்கை கொடுக்கின்றிர்-தோ
லாத விருதகளரு மெழிரெடித் தேகின் முணத்தாள் பெயர்த்துவளமே, செற்ற
திரு வாவடி துறைக்கண்வளர் திருவருட் செல்வப் பிரான்வருகவே - செம்மை
திகழ் சிவஞான வடிவனம் பலவான தேசிகோத் தமன்வருகவே. (௨)

Example 3: List of Tamil numbers and how to read them

This page lists Tamil numbers in both traditional and decimal notations.

இந்தத் தமிழ் இலக்கங்களைப் படிக்கவும்:—

க, உ, ங, ச, ரு, சூ, எ, அ, கூ, ம (௰ 10) கக, கஉ, கங, கச, க௫, கசூ, கஎ, கஅ, ககூ, உம, (௨0) ௱ (௱00), உ௱, ங௱, ௱௫சூ (156), ௲, ௲க, ௲௰ (1040), அ௲ (8000), எ௰௲ (70,000), கூம௲ (90,000), ௱௲ (க,00,000 இலட்சம்); அ௱௲ (8,00,000), ம௱௲ (பத்து இலட்சம்,) கூம௱௲ (90 இலட்சம்), ௱௱௲ (கோடி, 100 இலட்சம்).

தமிழ் இலக்கங்கள் க (ஒன்று), ம (பத்து), ௱ (நூறு), ௲ (ஆயிரம்), ம௲ (பதினாயிரம்), ௱௲ (இலட்சம்) ம௱௲ (பத்து இலட்சம்), ௱௱௲ (கோடி) — ம௱௱௲, ௱௱௱௲, ௲௱௱௲, ம௲௱௱௲, ௱௲௱௱௲, ௱௲ம௱௱௲ (கோடி கோடி) என்று செல்லும்.

<table>
<tr><td align="center">

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.)
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html for latest Form.
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for latest Principles and Procedures document.
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest roadmaps.

(Form number: N2352-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09)

</td></tr>
</table>

A. Administrative

```
1. Title:  __Addition of Tamil Digit Zero _ _____

2. Requester's name: _International Forum for Information Technology_____
                   _in Tamil (INFITT)_____
3. Requester type (Member body/Liaison/Individual contribution): __Liaison___
4. Submission date:                                        __2004-02-01___
5. Requester's reference (if applicable): _____WG02 Zero 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
     guidelines in Annex L of ISO/IEC 10646-1: 2000?              __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)?
     _____NO_____
```

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 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_____
      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 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 2.3 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?       _____YES_____
    If YES, is a rationale for its inclusion provided?      _____
     If YES, reference: _____See attached proposal_____
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)?                              _____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: _____
```

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 <u>distinct</u> 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 <u>as characters</u> 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.