

Dated: February 1, 2004

Title:	Proposal to add Tamil Digit Zero
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
0	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 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 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 (ASCII zero).

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) in ASCII range.

However usage patterns of Tamil numerals and ASCII numbers are different. Tamil numerals are mostly used for enumerations, and are not used in calculations in the current practice.

Parsers using unicode property data for digit-folding operations need all digits of a decimal system to be present in any given set of digit collection.

2.3. Compatibility between other Indic scripts

Compatibility between indic scripts for transliteration is touted as one of the main features of Unicode. Digit zero is available in all other indic scripts encoded in Unicode. Not having a basic character like a digic common between the scripts show a serious defect in this claim. Having Tamil Zero in Unicode is imporant to maintain this assertion.

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.

குழலினிது யாழினிது என்பதம் மக்கள் மழலைச் சொல் கேளாதவர் - குறள்					
திருவள்ளுவராண்டு 2032			சிலை (மார்கழி)		கி.பி. திசம்பர்
ஞாயிறு	திங்கள்	செவ்வாய்	அறிவன்	வியாழன்	வெள்ள
க திசம்பர் 16	உ 17	ந 18	ச 19	ரு 20	கா
அ ⁸ 23	க ⁹ 24	க ⁰ 25	கக 26	கஉ [*] 27	கந
கரு ⁰ 30	ககா 31	கள சனவரி 1	கஅ 2	கக 3	உ ⁰
உஉ 6	உந 7	உச 8	உரு 9	உகா 10	உள
உக 13	ந ⁰ 14	கூந்தல் (ரேமம்)	ஓசை (சப்தம்)	திருக்கோறு (பிரசாதம்)	தி (அக்தி)

“தமிழினத்தின்”		சிவலை	திசம்பர்
சாதிகளைத் தூண்டிவிட்டுக் கட்சிகளாய்	1	மயிலை சீனி வேங்கடசாமி தோற்றம்	16
	9	சுந்தர செழியார் பிணைவு	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*.

வளீவார் பெரும்புலி தெரிக்கவவ் வணநிற்சின் மாண்புடைக் கலயனாரை - மற்
றெம் கியாஞ்சென்று தேடுவே மாதலால் வளமிக்க கழகந்தொறுத், தெளிவார்
குழாங்குழுமி யோங்குமா வடுதுறைச் செல்வன்முத் தந்தருகவே - சின்மய
னருட்பெருங் குரவனம் பலவாண தேவன்முத் தந்தருகவே. (அ)

886. முன்னேகா ணின்னூல் வெலப்பட்ட கரிவன்றி முதுமீன முதுவுமு
த்து - முண்டக மலர்த்தாளி லரைபட்ட வெண்மதிய முத்துநஞ் சணவுதந்த,
புன்னேவாய் நெய்தற் கடற்றேன்று மிப்பிவளை பொறையுயிர்த் திட்டமுத்தும்-
பொருதிக லழிந்தகொக் குதவுமுக் துங்கொண்டு டுணுத னினக்கேகட, னென்
னையோ துந்திற மெனிற்சின்ன மேற்றதா ரிந்காளி வில்லையென்னி - லிலகுகண்
மணிமர்லை யிடையிடைக் கோத்திடுக வெண்ணில பழுத்தபாக, நென்னேமா வா
திபல குழுமா வடுதுறைச் செல்வன்முத் தந்தருகவே - சின்மய நருட்பெருங்
குரவனம் பலவாண தேவன்முத் தந்தருகவே. (க)

887. பங்காலு நின்னைத் துதித்துநிள் பதமுட் பதித்துவாழ் வேங்களுக்
குப் - பயனாக வொருதர மெனூந்தத் திடாயெனிற் பல்வளப் பூவணத்தி, லல்காத
மாமையிறை பொன்னையெண் முன்னேகா ளள்ளிமுத் தங்கொள்ளுமா - மளவ
ளாய் மேயதை வெளிப்பட வுரைத்திடுவ மதனிண மிங்காயெனி, னெல்காத
மற்றவ னுகிர்க்குறி நபோலத்தி லுற்றது முரைத்துவிடுவோ - மொண்கதிர்ச்
செல்வன்மதி யோனெனத் தண்ணென் றுடம்புகூ னிக்கடந்து, செல்காறும்
வளைபொழி லுயர்ந்தவா வடுதுறைச் செல்வன்முத் தந்தருகவே-சின்மய நருட்
பெருங் குரவனம் பலவாண தேவன்முத் தந்தருகவே. (க)

ரு.—முத்தப்பருவம் முற்றிற்று.

கூ.—வாராணைப்பருவம்.



888. உறத்தகதிர் மண்டிலத் துறைவது தெரிப்பதென வொண்மேனி
செயொளிசெய - வோங்குமதி மண்டிலத் துறைவது தெரிப்பதென வொளிர்
நீறு வெள்ளொளிசெய, வறந்தலை பெரிப்பவொளி ரிவ்வொளிக ணைக்குதொறு
மடியர்மல பத்தமுமுது - மடலற வழித்தழி விலாதபே ரின்பரிலை யாருளிற் கொ
டுத்திடுதலாற், பிறந்தபிர தாபமும் புகழுமே யென்றறிஞர் பேசுமகிழ் தூங்கவ
ளவை-பேசிய துணர்ந்துகை செய்தென முகத்துப் பிறங்குபுன் மூரறேன்றச்,
சிறந்ததிரு வாவடு துறைக்கண்வளர் திருவருட் செல்வப் பிரான் வருகவே-செம்
மைதிகழ் சிவஞான வடிவனம் பலவாண தேசிகோத் தமன்வருகவே. (க)

889. உற்றபெரு மாதவ முஞ்ந்துநர்க் கல்லா துஞ்ந்துதவ மொன்றுமில்
லா - வுவர்வருக வென்றழைத் திடநா மடைத்திடுத லுரிமையன் நென்றுநினை
யல், பொற்றமதி நிலவுக ரன்றிவெங் காட்டிலும் புணரிவா யுண்டெழுந்த -
புயன்மழை செழும்பணை நிலத்தன்றி யுவர்பற்று புலியகத் தும்பொழியுமே,
துற்றமலர் தளிர்மெத் தெனப்பரப் பினமுளத் தூயர்கை கொடுக்கரின்றார்-தொ
லாத விருதுகளு மெதிரெடுத் தேநின் றுணைத்தாள் பெயர்த்துவளமே, செற்ற
சிறு வாவடு துறைக்கண்வளர் திருவருட் செல்வப் பிரான்வருகவே - செம்மை
திகழ் சிவஞான வடிவனம் பலவாண தேசிகோத் தமன்வருகவே. (உ)

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

This page lists Tamil numbers in both traditional and decimal notations. (Source: *iniya tamiz ilakkam* by Yogisri Cuddhanantha Bharathiyar, page 201.

இத்தத் தமிழ் இலக்கங்களைப் படிக்கவும்:—

க, உ, ன, ச, ங, கூ, எ, அ, க, ய (கௌ) கக, கஉ,
கந, கச, கஞ, ககூ, கஎ, கஅ, ககூ, உய, (கௌ) ன (கௌ),
உந, னந, னகூ (156), து, துக, துசய (1040),
அது (8000), எய்து (70,000), கூய்து (90,000), னது
(௧,00,000 இலட்சம்); அநது (8,00,000), யநது (பத்து
இலட்சம்,) கூயநது (90 இலட்சம்), னநது (கோடி, 100
இலட்சம்).

தமிழ் இலக்கங்கள் க (ஒன்று), ய (பத்து), ன
(நூறு), து (ஆயிரம்), ய்து (பதினாயிரம்), னது (இலட்சம்)
யநது (பத்து இலட்சம்), னநது (கோடி)—
யநநது, னநநது, துநநது, ய்துநநது, னதுநநது,
நநதுயநது (கோடி கோடி) என்று செல்லும்.

(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 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.