ISO
International Organization for Standardization
Organisation Internationale de Normalisation

ISO/IEC JTC 1/SC 2/WG 2
Universal Multiple-Octet Coded Character Set (UCS)

ISO/IEC JTC 1/SC 2/WG 2 **N2381R**
Date: 2001-10-11

Title:	*Proposal to add eight Tamil symbols*
Source:	*V.S. Umamaheswaran (umavs@ca.ibm.com)*, Expert Contribution
Action:	*For consideration by WG2*
Distribution:	*ISO/IEC JTC 1/SC 2/WG 2 members and Liaison organizations*

Proposal:

The Tamil 99 Phonetic keyboard mandated by the Tamil Nadu State Government in India (G.O.Ms.No.3, Information Technology Department, dated 19/2/1999 - see: http://www.tamilnet99.org/gov.htm) includes three types of characters (see Annex A for details):

1. Characters that hold a one-to-one correspondence with characters in the Tamil block of the Unicode standard.
2. Two pre-composed conjuncts - KSHA and SHREE -- that can be composed of characters that exist in the Tamil block of the Unicode standard.
3. Eight Tamil Symbols that do not exist in the Unicode standard and cannot be composed of other characters. These are the eight characters shown in Table 1 below.
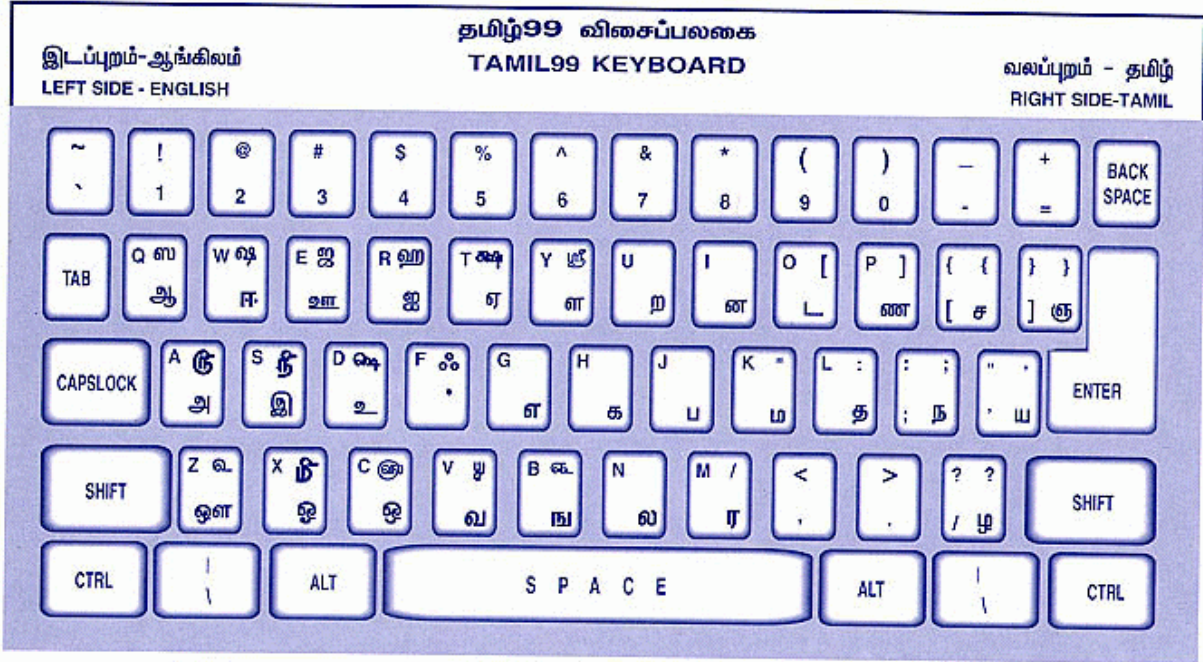
This proposal is to add the missing eight (8) Tamil symbols to the Unicode and ISO/IEC 10646 standards in the Tamil Block in the Basic Multilingual Plane. These characters cannot be composed from existing Tamil or other characters in the standard. Table 1 shows the Proposed Names, Positions and Shapes (taken from an existing font available on the web).

Table 1: Proposed Character Names, Glyphs and Positions in the BMP:

Shapes Using Maduram Font[1]	Name	Position	Name in Tamil Nadu Font Standard[2]	Reference Key (Shifted)[3]
ஐ	TAMIL DAY SIGN (Naal)	0BF3	TAMIL SYMBOL DAY	Z
ம	TAMIL MONTH SIGN (Maatham)	0BF4	TAMIL SYMBOL MONTH	X
வரு	TAMIL YEAR SIGN (Varudam)	0BF5	TAMIL SYMBOL YEAR	C
பு	TAMIL DEBIT SIGN (Patru)	0BF6	TAMIL LETTER DEBIT	V
வ	TAMIL CREDIT SIGN (Varavu)	0BF7	TAMIL LETTER CREDIT	B
வழ	TAMIL DITTO SIGN (Merpadi)	0BF8	TAMIL SYMBOL DITTO	D
ந	TAMIL RUPEE SIGN (Rupai)	0BF9	TAMIL SYMBOL RUPEE	A
ந	TAMIL NUMBER SIGN (Akkam)	0BFA	TAMIL SYMBOL NUMBER	S

Notes:
1. Using: http://kamban.com.au/fonts/TAMMADUR.TTF
2. Tamil monolingual font encoding standard and character names (see http://www.tamilnet99.org/tabtam.htm)
3. From Tamil 99 (Phonetic) Keyboard shown below (http://www.tamilnet99.org/phonetichtm):

Note: The Tamil Letter Shree (at Shifted position on the Y key) is composed of the following characters (identified by their Unicode values): U+0B9A + U+0BCD + U+0BB0 + U+0BC0.   Similarly the Tamil Letter Ksha (at Shifted position on the T key) is composed of the following Unicode characters: U0B95+U0BCD+U0BB7.

They are in wide use in Tamil Nadu state in India, and including them in the Tamil 99 Encoding and Keyboard makes them available to the larger Tamil community all over the world. (http://www.tamilnet99.org).

Usage of the proposed symbols:

In addition to being mandated on the TNG Keyboard, these symbols are in current wide use.

The following examples were provided by *The International Forum for Information Technology in Tamil (INFITT)* Working Group on Tamil in Unicode (see http://www.infitt.org/wgfull.html) (Messrs. Muthu Nedumaran, A. Elangovan and Michael Kaplan).

Example 1 – Manuscript form of an accounting sheet

Example 2 – A Newspaper Heading Line

Example 3 – Page from an Almanac

Number Sign

Year Sign

Day Sign

Ditto Sign

Transcoding with ISCII Tamil G-set

ISCII does not encode any of these Tamil symbols in it.

One of the current characteristics of Unicode Indic charts is its alignment from repertoire point of view with ISCII, and the associated phonetic transliteration advantages thereby obtained among the major languages using the Indic scripts. If we add these additional symbols, some considerations will need to be given on how to transcode these symbols into equivalents in ISCII.

The iLeap word processor from CDAC (see http://www.cdac.org.in), which has implemented the ISCII standard encoding for the different Indic scripts, has the following information in its Tamil tutorial page. Note, however, the actual implementation has not been verified by the authors. Also, the glyph shown in iLeap help file for the TAMIL DAY SYMBOL looks closer to the TAMIL RUPEE SYMBOL than what is shown below. The Tamil numerals are not (apparently) supported in it.)

"Nukta:

Nukta character (NUK) is used for deriving four special signs நீ , மீ , வரு and உ . It is provided on the "]" key. These special signs can be typed as follows:

நீ	+ NUK =	நீ	(Day)
மீ	+ NUK =	மீ	(Month)
வ	+ NUK =	வரு	(Year)
உ	+ NUK =	உ	(Number)"

Note that the Nukta character is alien to the Tamil script, and is NOT used with Tamil in any other context.

Annex A

This annex shows the mapping of the characters from the TNG Keyboard to UCS, identifying the missing ones. The handling of the Dependent and Independent vowels generated from this keyboard are specific to this keyboard. See http://www.tamilnet99.org/annex2.htm . The input method editor will be responsible to output the correct UCS code positions in the data stream sent to the application.

Tamil99 keys (with Tamil Chars only)	Character Names (Tamil99)	UCS Values as input to Input Method Editor	Remarks
Q ஸ ஆ	TAMIL LETTER SA	U+0BB8	
	TAMIL LETTER AA	U+0B86	
W ஷ ஈ	TAMIL LETTER SHA	U+0BB7 *(Tamil Letter SSA)*	
	TAMIL LETTER II	U+0B88	
E ஜ உா	TAMIL LETTER JA	U+0B9C	
	TAMIL LETTER UU	U+0B8A	
R ஹ ஐ	TAMIL LETTER HA	U+0BB9	
	TAMIL LETTER AI	U+0B90	
T க்ஷ ஏ	TAMIL LETTER KSHA	U+0B95, U+0BCD, U+0BB7	
	TAMIL LETTER EE	U+0B8F	
Y ஸ்ரீ ஏ	TAMIL LETTER SHRI	U+0B9A, U+0BCD, U+0BB0, U+0BC0	
	TAMIL LETTER LLA	U+0BB3	
U ற	TAMIL LETTER RRA	U+0BB1	
I ண	TAMIL LETTER NNA	U+0BA9 *(Tamil Letter NNNA)*	
O [ட	Left Square Bracket	U++005B	
	TAMIL LETTER TTA	U+0B9F	
P] ண	Right Square Bracket	U++005D	
	TAMIL LETTER NNA	U+0BA3	

Tamil99 keys (with Tamil Chars only)	Character Names (Tamil99)	UCS Values as input to Input Method Editor	Remarks
{ { [ச	Left Curly Bracket	U++007B	
	TAMIL LETTER CA	U+0B9A	
} }] ஞ	Right Curly Bracket	U++007D	
	TAMIL LETTER NJA	U+0B9E *(Tamil Letter NYA)*	
A ரு அ	TAMIL SYMBOL RUPEE *(Rupai)*	DOES NOT EXIST	
	TAMIL LETTER A	U+0B85	
S கி இ	TAMIL SYMBOL NUMBER *(Enn)*	DOES NOT EXIST	
	TAMIL LETTER I	U+0B87	
D மெ உ	TAMIL SYMBOL DITTO *(Merpadi)*	DOES NOT EXIST	
	TAMIL LETTER U	U+0B89	
F ஃ ·	TAMIL LETTER AKH	U+0B83 *(Tamil Sign Visarga)*	
	TAMIL SIGN VIRAMA *(Pulli)*	U+0BCD	
G எ	TAMIL LETTER E	U+0B8E	
H க	TAMIL LETTER KA	U+0B95	
J ப	TAMIL LETTER PA	U+0BAA	
K " ம	DOUBLE QUOTE	U+0022	As in PonMozhi Keyboard
	TAMIL LETTER MA	U+0BAE	
L : த	TAMIL LETTER THA	U+0BA4 *(Tamil Letter TA)*	
: ; ; ந	TAMIL LETTER NA	U+0BA8	

Tamil99 keys (with Tamil Chars only)	Character Names (Tamil99)	UCS Values as input to Input Method Editor	Remarks
	SINGLE QUOTE	U+0027	As in PonMozhi Keyboard
	TAMIL LETTER YA	U+0BAF	
	TAMIL SYMBOL DAY *(Naal)*	DOES NOT EXIST	
	TAMIL LETTER AU	U+0B94 *(Tamil Letter AU+)*	PonMozhi Keyboard uses two (O + LLA) glyphs to create AU
	TAMIL SYMBOL MONTH *(Maatham)*	DOES NOT EXIST	
	TAMIL LETTER OO	U+0B93	
	TAMIL SYMBOL YEAR *(Varudam)*	DOES NOT EXIST	
	TAMIL LETTER O	U+0B92	
	TAMIL SYMBOL DEBIT *(Patru)*	DOES NOT EXIST	
	TAMIL LETTER VA	U+0BB5	
	TAMIL SYMBOL CREDIT *(Varavu)*	DOES NOT EXIST	
	TAMIL LETTER NGA	U+0B99	
	TAMIL LETTER LA	U+0BB2	
	TAMIL LETTER RA	U+0BB0	
	TAMIL LETTER LLLA	U+0BB4	

Note: For 'PonMozhi' keyboard and related documents visit http://www.kanithamizh.org

Annex B

A. Administrative

1. **Title:**   **EIGHT Tamil Symbols from TNG Keyboard**_____

2. Requester's name: V.S. Umamaheswaran, IBM Canada, and INFITT Working Group on Tamil in Unicode

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

4. Submission date: 2001-10-10

5. Requester's reference (if applicable): _____

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 (suggest : 0BF3—0BFA in BMP)

2. Number of characters in proposal: EIGHT (8)

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: _____Attached Contribution (this document) Self Evident_____

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?   http://www.kanithamizh.org or http://kamban.com.au/fonts for example

          If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
          used: _____ http://www.tamilnet99.org/tabtam.htm (Fonts related to the Tamil Nadu Government Order)

          _____

          _____

7. References:

          a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?   NO

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

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)?

          Some. Input - see: http://www.tamilnet99.org/annex2.htm for information on use of the keyboard

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

_____

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

see 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? Tamil Nadu State Government, India; Kanithamizh Sangam, Chennai
        INFITT - International Forum for Information Technology in Tamil.

        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?   YES

    Reference: See Tamilnadu Government, India (http://www.tamilnet99.org/)_____

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

    Reference: _____(Samples attached)_____

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

    If YES, where? Reference: Tamil Nadu and the larger Tamil speaking Community

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: Characters should be part of the Tamil block

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?   NO

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?   NO

        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)?   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: _____