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

L2/05-295

ISO/IEC JTC 1/SC 2/WG 2 *N 2970*
2005-08-23

Title:	**Proposal to add 3 Malayalam Numbers 10, 100, 1000 and 3 Fraction symbols 1/4, 1/2 and 3/4**
Source:	**UTC, USA, Kerala State IT Mission & other experts (Compiled by: V.S. Umamaheswaran – umavs@ca.ibm.com)**
References:	
Action:	**For WG2 consideration and adoption in the next Amendment to the standard**
Distribution:	**ISO/IEC JTC 1/SC 2/WG 2**

The Unicode Technical Committee had received proposals for Malayalam numerals Ten, One Hundred and One Thousand, and for Malayalam symbols used for fractions in old Malayalam writings. The proposal for the numerals was part of the Public Review Issue #71 seeking feedback from interested parties (see http://www.unicode.org/review/pr-71.html). The feedback was supportive of the proposal for the numerals. The feedback also included request for inclusion of several fraction symbols used in old Malayalam material. Of these there was sufficient evidence to proceed with three of the fraction symbols 1/4, 1/2 and 3/4 at this time.

Even though these numbers and fraction symbols are not in current use they are required for creating digital versions of old Malayalam manuscripts and mathematical texts. Figure 1 shows a palm leaf manuscript and Figure 2 shows a table from a reference mathematical text. See the attached documents from different sources for more details.

The UTC has reviewed these and has found these to be acceptable as candidates for encoding in Unicode. This document is a proposal to add the following characters to ISO/IEC 10646, preferably in the current Amendment 2.

Mr. Michael Everson has kindly volunteered to provide the necessary font additions.

Location	Character Name	Representative Glyph
U+0D70	MALAYALAM NUMBER TEN	See figure 2
U+0D71	MALAYALAM NUMBER ONE HUNDRED	See figure 2
U+0D72	MALAYALAM NUMBER ONE THOUSAND	See figure 2
U+0D73	MALAYALAM FRACTION ONE QUARTER	See figure 2
U+0D74	MALAYALAM FRACTION ONE HALF	See figure 2
U+0D75	MALAYALAM FRACTION THREE QUARTERS	See figure 2

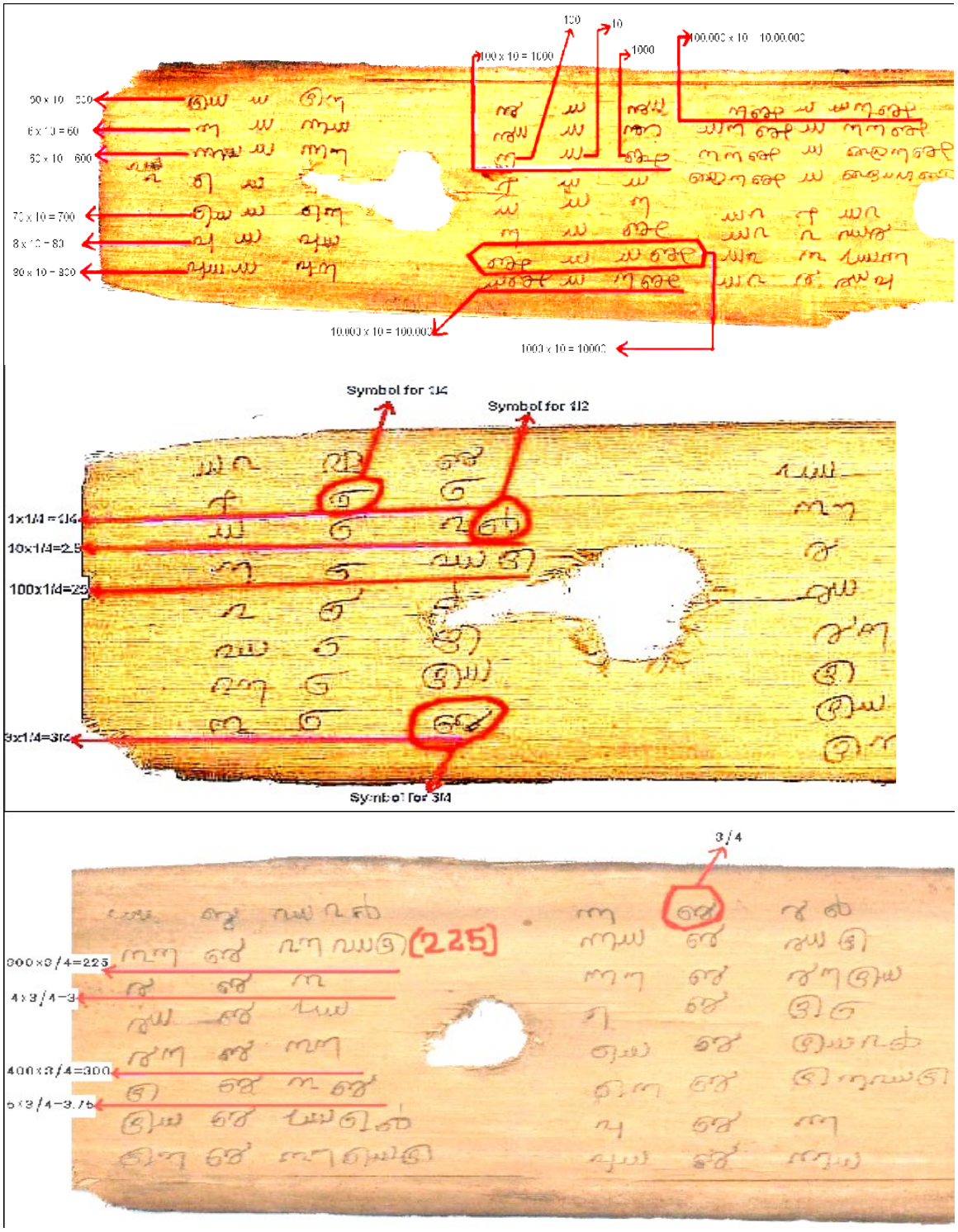


Figure 1: Palm leaf manuscripts (Extracted from L2/05-164 - K.G.Sulochana, CDAC Trivandrum, Kerala; "Material as evidence for the symbols used for 10,100,1000 and fractions in Malayalam" http://www.malayalamresourcecentre.org/Mrc/symbols/malsymbols.html)
(Note: the above images have been adjusted for brightness etc. to make the numerals more legible on paper medium).

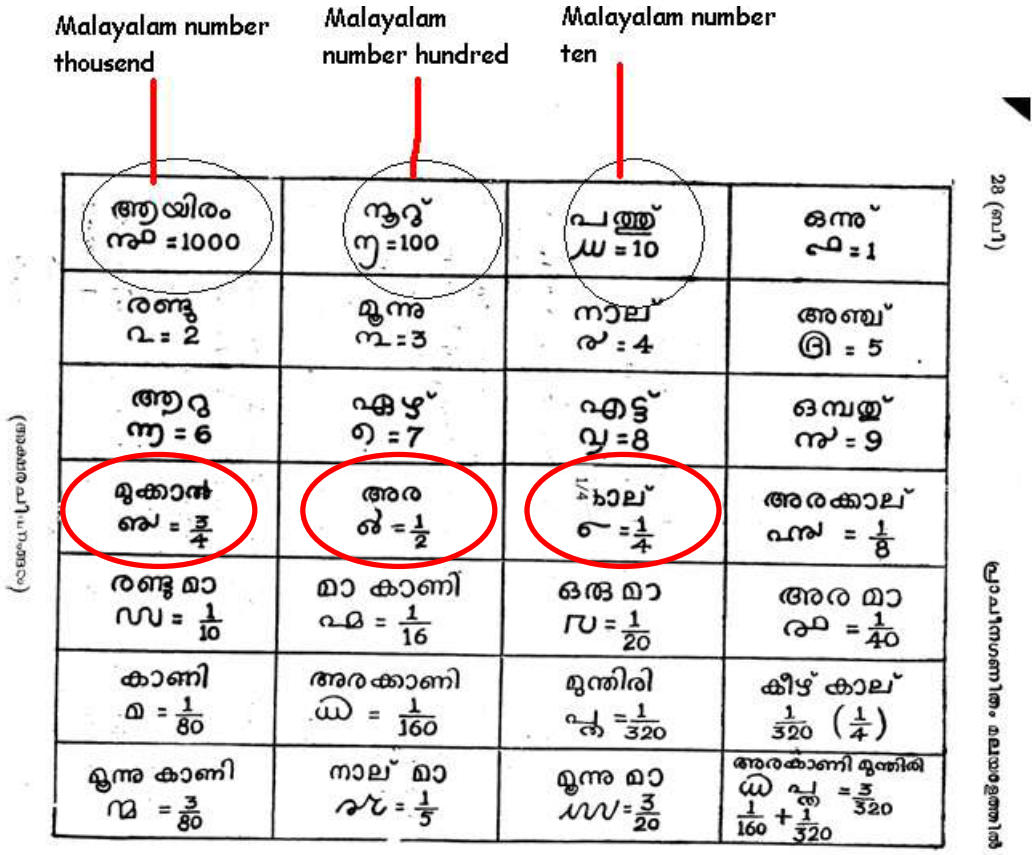


Figure 2: Table from *Kanakkadhikaaram* by Manavan Mapila
(See Attachment 3)

Attachments:
(Note: some of the attachments contain information about Malayalam digit zero also).

1. Proposal Summary Form
2. Document L2/05-87 - proposal from Dr. N. Ganesan
3. Document L2/05-175 - proposal from Kerala State IT Mission, Center for Linguistic Computing Keralam,
4. Document L2/05-173 - feedback on Unicode PRI 71 from Cibu Johny.

(Attachment 1 to document N2970)

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** **Malayalam Numbers 10, 100, 1000 and Fractions 1/4, 1/2 and 3/4**
2. Requester's name: **UTC, USA, Kerala State IT Mission and other experts (contact: umavs@ca.ibm.com)**
3. Requester type (Member body/Liaison/Individual contribution): **Joint Member Body, Liaison and Individual**
4. Submission date: **2005-08-23**
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: **Yes**

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: **MALAYALAM**
2. Number of characters in proposal: **SIX**
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary _____ B.1-Specialized (small collection) _____ B.2-Specialized (large collection) _____
   C-Major extinct _____ D-Attested extinct _____ E-Minor extinct **XX**
   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): **3**
   Is a rationale provided for the choice? _____
   If Yes, reference: **(Same as MALAYALAM script)**
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. Michael Everson has kindly volunteered.**
   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)?
   Similar to other Number / Fraction Symbols in the standard.
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.

[1] Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

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? **Indic@unicode.org list; Kerala State IT mission.**

   If YES, available relevant documents: **See attachments**

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? **(See attached Info)**

   Reference: **These symbols are not in contemporary use. Needed for Digital Archiving of Manuscripts**

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

   Reference: **Ancient**

5. Are the proposed characters in current use by the user community? **No**

   If YES, where? Reference: _____

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? _____

   If YES, reference: **Part of Malayalam Script - used with Malayalam Digits**

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? **If possible**

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? **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? _____

    If YES, describe in detail (include attachment if necessary) _____

13. Does the proposal contain any Ideographic compatibility character(s)? _____

    If YES, is the equivalent corresponding unified ideographic character(s) identified? _____

    If YES, reference: _____

Attachment 2 to document N2970

Extract of Relevant parts from: L2/05-087

Title:	**…. and adding Malayalam numerics (U+0D70, U+0D71, U+0D72)**
Author:	**Dr. Naga Ganesan, Houston, Texas**

(Note: Only section 2 dealing with Malayalam numerics is included here)

Author: Nagamanickam Ganesan, Ph.D.
16923 Sky Harbor Ct.
Friendswood, TX 77546, USA
281-648-8636 (Phone)
naga_ganesan@hotmail.com

2.0 Malayalam Numerics 10 (U+0D70), 100 (U+0D71), 1000 (U+0D72):

In both Malayalam and Tamil scripts, digit zero is a modern innovation introduced in the nineteenth century. Like its closely allied Tamil script, Malayalam also employed numeric signs for 10, 100 and 1000 to write numbers larger than nine. Tamil and Roman zero-less numerical systems are compared by Michael Kaplan: http://blogs.msdn.com/michkap/archive/2005/01/24/359347.aspx.

Georges Ifrah, The universal history of Numbers from prehistory to the invention of the computer, John Wiley, 2000. Page 373 (the scanned page attached at the end),

> " Malayalam figures
> These figures are used by the Dravidian people of Kerala state, on the ancient coast of Malabar, in the southwest of India. They have the same name as the form of writing used in the area. Like the Tamils, the people of Kerala did not use zero in their notation system for many centuries: Malayalam figures are not based on the place-value system, and there are specific figures for 10, 100 and 1,000. It was only since the middle of the nineteenth century, under the influence of Europe, that zero was introduced and combined with the symbols for the nine units according to the positional principle. Thus the Tamil and Malayalam figures were the only ones in India that did not include zero and were not based on the positional principle relatively recently."

While Malayalam digit zero was introduced in 1850 CE or later, the Tamil Nadu government (then called Madras Presidency under the British colonial rule) introduced Tamil digit zero (U+0BE6) prior to 1820 CE. A citation from a 1825 CE school textbook is given at: http://www.geocities.com/thamizh@sbcglobal.net/tamil_zero.PDF. Like the Tamil script having numerics for 10, 100 and 1000 at U+0BF0, U+0BF1, U+0BF2 respectively, Malayalam historically also has numeric signs. Only with these numeric signs, large numbers greater than nine can be written in the traditional Malayalam script. Hence it is recommended that Unicode allocates separate code points for the Malayalam numeric signs also.

> Malayalam numerics
> U+0D70 MALAYALAM NUMBER TEN
> U+0D71 MALAYALAM NUMBER ONE HUNDRED
> U+0D72 MALAYALAM NUMBER ONE THOUAND

The code points are assigned in a manner parallel to the Tamil numerics code points. The shapes of the Malayalam numerics glyphs for U+0D70, U+0D71, U+0D72 must be as shown in page 335, Figure 23.21, Georges Ifrah, The universal history of Numbers from prehistory to the invention of the computer, John Wiley, 2000 which is enclosed in this proposal.

First appearance: c.600 – 900 CE

Type: B5 (hybrid number-system of the fifth type: Fig.23.37). Base 10

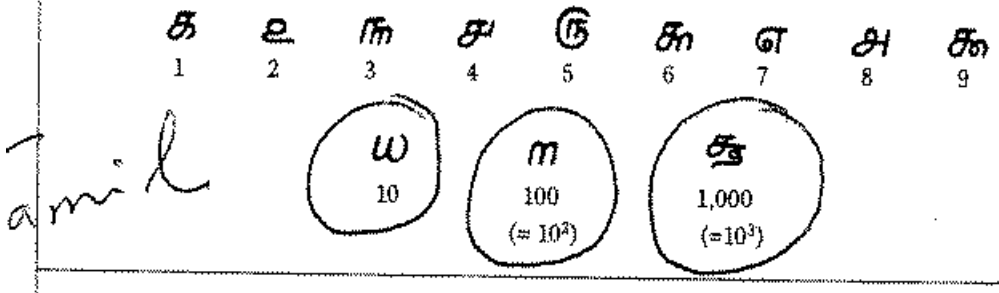
Need for zero sign: No, when the hybrid principle is rigorously applied. Yes, when the simplified rule below is applied.

Existence of zero sign: Not before the modern era

Capacity for representation: Limited in the case of the unsimplified system (see Chapter 24, p.372)

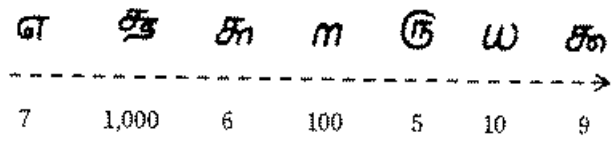
System used among the Tamils (southern India)

Base numbers (modern script)



Example: 7,659

Normal script



Representation based entirely on hybrid principle, broken down thus:

$$7 \times 1{,}000 + 6 \times 100 + 5 \times 10 + 9$$

Abridged script in use since modern times

The above representation was sometimes produced in the simplified form below, thus tending towards an application of the positional principle with base 10:

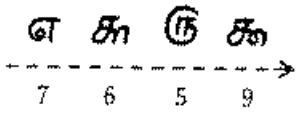


FIG. 23.20. *Tamil number-system*

(From: George Ifrah - Universal History of Numbers, John Wiley, 2000; page 334)

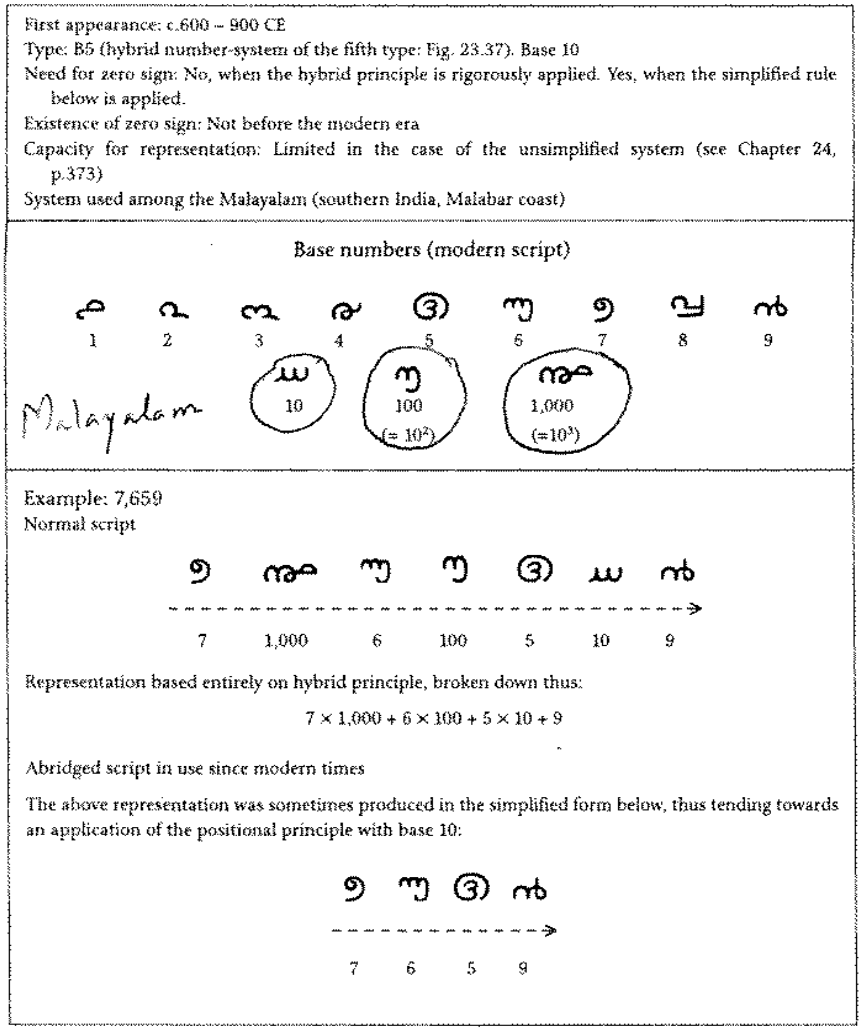
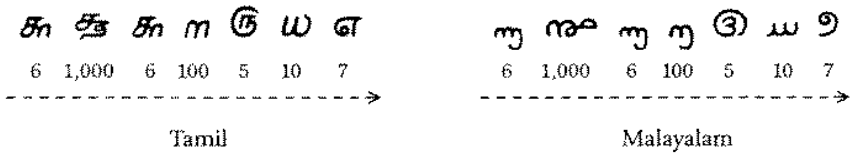


FIG. 23.21. *Malayalam number-system*

In this way, the number 6,657, for example, would usually be written as follows:



which corresponded to the decomposition

$$6 \times 1{,}000 + 6 \times 100 + 5 \times 10 + 7.$$

(From: George Ifrah - Universal History of Numbers, John Wiley, 2000; page 335)
(Note page 373 from George Ifrah deals with Malayalam numerals already encoded in the standard and is not included in this attachment..)

Kerala State IT Mission
Center for Linguistic Computing Keralam

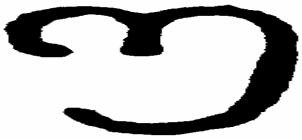
L2/05-175

Title: Proposal to includes Malayalam Numbers as distinct characters
Source: Centre for Linguistic Computing Keralam @ C-DIT
Action: For consideration by UTC and ISO/IEC JTC 1/SC 2/WG 2
Distribution: ISO/IEC JTC 1/SC 2, ISO/IEC JTC 1/SC 2
Contact: linguafranka@yahoo.com

1. Introduction

Unicode Malayalam range encodes Malayalam numerals. The numbers of TEN, HUNDRED and THOUSEND had not been included as separate characters even though Malayalam facilitates separate characters for those.

1.1. Characters Proposed in this Document

No	Shape of the character	Proposed Name	Proposed Character Position
1		MALAYALAM NUMBER TEN	U+0D70
2		MALAYALAM NUMBER HUNDRED	U+0D71
3		MALAYALAM NUMBER THOUSAND	U+0D72

1.2. About Centre for Linguistic Computing Keralam@ C-DIT

Centre for Linguistic Computing Keralam, known as CLiCK @ C-DIT is a join venture of Kerala State IT Mission, Government of Kerala; Department of Linguistics, University of Kerala and Computational Linguistic Team @ C-DIT for the development research, facilitation and dissemination of Malayalam Language Computing in Keralam and the Malayalee speech communities around the world. www.clickeralam.org

1.3. Demographics

Malayalam is the language of the state of Kerala in India. Total population of Malayalam language speakers around the world is estimated to be around three cores (1999 Census)

2. Malayalam Numeric

With the representation of ten hundred, thousand by one, and zero Malayalam numeric system, possess separate symbols for the above

2.1. Malayalam number system

Ancient Malayalam number system did not include the digit zero. It had distinctive glyphs for numbers 10, 100, 1000. However, in modern practice, Malayalam numerals are used in the same way as decimal number, with a zero that looks similar to the digit zero in international form of Indian numerals but not as represented in the Unicode Malayalam character map

2.2. Compatibility between other Indian scripts

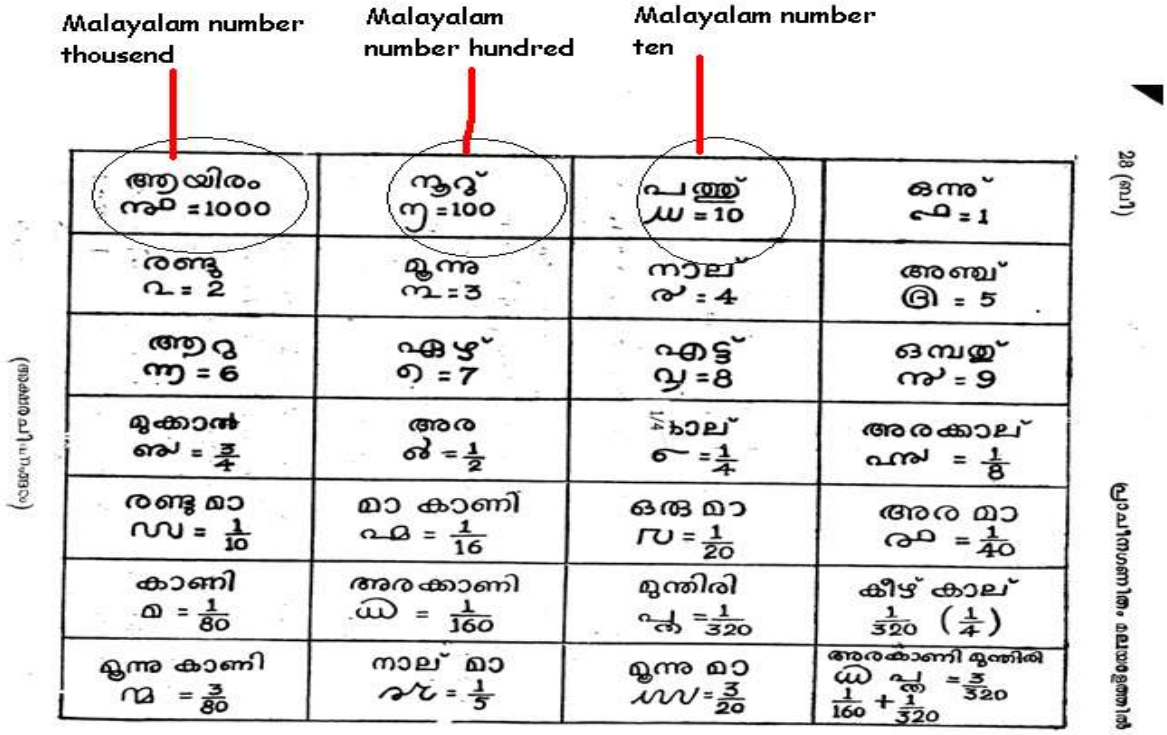
Malayalam numbers in the proposed form is available in all Indic script encoded in Unicode explicitly in most related language Tamil also, see Tamil Unicode chart

2.3 Usage of proposed Malayalam Numbers

Malayalam old numerals is not being widely used because it belongs to the minor extinct category of character, but the adapted Indian glyph form of zero is widely used with modern Arabic numerals. However, it necessary for the digitization of old Malayalam texts.(see cdac Link)

3. Published Usage Example

3.1 The existing Malayalam Numerals: The following document named കണക്കധികാരം *(kanakkadhikaaram*) is the mathematics text in Malayalam, which describes this.



ആയിരം ൟ =1000	നൂറ് ൱ =100	പത്ത് ൰ = 10	ഒന്ന് ൯ = 1
രണ്ട് ൨ = 2	മൂന്ന് ൩ = 3	നാല് ൪ = 4	അഞ്ച് ൫ = 5
ആറ് ൬ = 6	ഏഴ് ൭ = 7	എട്ട് ൮ = 8	ഒമ്പത് ൯ = 9
മുക്കാൽ ൾ = $\frac{3}{4}$	അര ൽ = $\frac{1}{2}$	കാൽ ¼ ർ = $\frac{1}{4}$	അരക്കാല് ൻ = $\frac{1}{8}$
രണ്ടു മാ ൿ = $\frac{1}{10}$	മാ കാണി ൌ = $\frac{1}{16}$	ഒരു മാ ൿ = $\frac{1}{20}$	അര മാ ൌ = $\frac{1}{40}$
കാണി ൎ = $\frac{1}{80}$	അരക്കാണി ൏ = $\frac{1}{160}$	മുന്തിരി ൐ = $\frac{1}{320}$	കീഴ് കാല് $\frac{1}{320}$ ($\frac{1}{4}$)
മൂന്നു കാണി ൗ = $\frac{3}{80}$	നാല് മാ ൾ = $\frac{1}{5}$	മൂന്നു മാ ൿ = $\frac{3}{20}$	അരക്കാണി മുന്തിരി ൏ ൐ = $\frac{3}{320}$ $\frac{1}{160}+\frac{1}{320}$

4. References

Manavan Mapila ( ) കണക്കധികാരം

L.A Ramavarma (1971) പ്രാചീനകേരള ലിപികള്: Kerla Sahithy Academy

K.C Musaad (1980) പ്രാചീന ഗണിതം മലയാളത്തിൽ: State Institute of Language

P.K Narayana Pillai (1951) *Pracheena malayaala grandhaa mathrakakal*

Malayalam Related Topics

L2/05-173

2005-07-21

On Unicode Public Review issue #71: Malayalam Digits

Author: Cibu C Johny
Date: July 21, 2005
Email: yahoo.com id 'cibu'

Please read the details of the issue at Unicode Public Review listing.

On the right, you can see the scanned image of the number symbols which I jotted down in my old note book.

Following are my suggestions for inclusion in Unicode chart:

1. Shortcuts for numbers 10, 20, .., 90, 100 and 1000. These symbols are very archaic.
2. The fraction symbols 1/8, 1/4, 1/2 and 3/4. My notes miss 1/8. I know there is a symbol for 1/8; but don't have a document to show how it looks. These are also very archaic.
3. The symbol used in '17-൰' to mean '17th'. In contrast to previous ones, this number related symbol is commonly used. Many existing fonts have this in their glyph set. If at all, some of these number related symbols make it to Unicode chart, this should be the first.