L2/10-165

Preliminary Proposal to Encode Sinhala Digits and Numerals

From: ICT Agency of Sri Lanka
Date: 2010-05-03

Background:
In L2/07-002R (ISO/IEC JTC1/SC2/WG2 N3195R) dated 2007-02-08, Mr. Michael Everson proposed to encode twenty archaic numbers for Sinhala to the BMP of the UCS.

In L2/08-068 dated 2008-01-28, ICTA Sri Lanka requested additional time to study the issue, and in L2/08-105 dated February 2008, we presented our preliminary findings on Sinhala numeration.

Thereafter, at the request of ICTA, Mr. Harsha Wijayawardhana of the University of Colombo School of Computing (UCSC), has conducted a thorough study of the subject.

Outcome of the research:
The research has revealed that five different types of numerations were used in the Sinhala language:

i. Sinhala Illakam - does not have a zero and has separate symbols for 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000.
ii. Lith Illakam - digits 0..9; widely used for calculations especially in astrology.
iii. Katapayadia - numerals based on Sinhala letters.
iv. Page numbering of Ola leaves using Sinhala "Swara"
v. Bhootha Anka or Butha Sankya - method of depicting numbers by words; used primarily by poets.

Of these, the first two had been widely used. Sinhala Illakkam was in use up to the early part of the nineteenth century, and Lith Illakkam was in use well into the 20th century. A large amount of evidence for both of these forms exists, and will be included in the final proposal.

A Synopsis of the research was presented at the National Archeological Symposium held in July 2009, conducted by the Department of Archeology, and published Volume II of the Symposium's proceedings.

Shapes of the Sinhala and Lith Illakkam

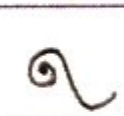
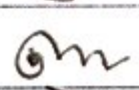
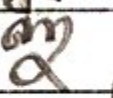
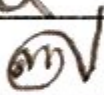
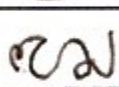
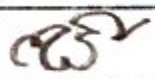
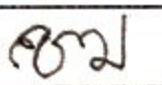
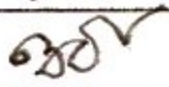
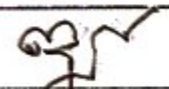
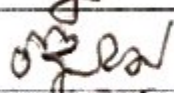
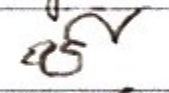
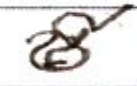
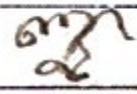
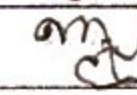
i. Sinhala Illakkam

Shapes of Sinhala Illakkam were given by Mendis Gunsekera in his famous book 'A comprehensive Grammar of Sinhala Language', and also in W.A De Silva's 'Catalogue of Palm leaf manuscripts in the library of Colombo Museum'. Sinhala Illakkam were also used in the Kandyan Convention signed by the British and Kandyan Chieftains in 1815.

Accordingly the following shapes are proposed for the Sinhala Illakkam:

(more clear glyphs will be provided)

Shapes of Sinhala Illakkum

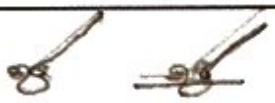
	Sinhala Illakkum Digit one
	Sinhala Illakkum Digit two
	Sinhala Illakkum Digit Three
	Sinhala Illakkum Digit Four
	Sinhala Illakkum Digit Five
	Sinhala Illakkum Digit Six
	Sinhala Illakkum Digit Seven
	Sinhala Illakkum Digit Eight
	Sinhala Illakkum Digit Nine
	Sinhala Illakkum Digit Ten
	Sinhala Illakkum Digit Twenty
	Sinhala Illakkum Digit Thirty
	Sinhala Illakkum Digit Fourty
	Sinhala Illakkum Digit Fifty
	Sinhala Illakkum Digit Sixty
	Sinhala Illakkum Digit Seventy
	Sinhala Illakkum Digit Eighty
	Sinhala Illakkum Digit Ninety
	Sinhala Illakkum Digit Hundred
	Sinhala Illakkum Digit Thousand

ii. Lith Illakkam

Three versions of Lith Illakkam have been discovered and some form of Halant had been used for zero in all three versions. Mendis Gunesekera has stated that symbol for zero in Arabic numerals can be used in place of the Sinhala Virama in Lith Illakkam. Ancient scribes of Sinhala had used a line across Brahmi numerals to separate numerals from characters.

It is suggested that a horizontal line to be drawn across Lith Illakkam to disambiguate digits and letters. (more clear glyphs will be provided)

Lith Illakkum

	Lith Illakkum digit zero
	Lith Illakkum digit one
	Lith Illakkum digit two
	Lith Illakkum digit three
	Lith Illakkum digit Four
	Lith Illakkum digit Five
	Lith Illakkum digit Six
	Lith Illakkum digit Seven
	Lith Illakkum digit Eight
	Lith Illakkum digit Nine

Conclusion:

ICTA's Local Language Working Group, after reviewing the research and after extensive discussions and consultations with other experts and stakeholders agreed that:

1. Lith Illakkam, (digits 0-9) should be included in the Basic Multilingual Plane (Sinhala block 0Dxx) of the Unicode standard.
2. Sinhala Illakkam (numerals 1-0, 10-90, 100 and 1000) be included in the supplementary area of the Unicode standard.

These encodings were also submitted to the Sri Lanka Standards Institute to be incorporated into SLS 1134: Sinhala Character Code for Information Interchange. The SLSI Sectoral Committee on IT has approved the commencement of the standardisation process for these encodings.

Proposal:
Accordingly, we propose the following 10 digits be encoded as follows.

Code	**Name**
0DE6	SINHALA LITH DIGIT ZERO
0DE7	SINHALA LITH DIGIT ONE
0DE8	SINHALA LITH DIGIT TWO
0DE9	SINHALA LITH DIGIT THREE
0DEA	SINHALA LITH DIGIT FOUR
0DEB	SINHALA LITH DIGIT FIVE
0DEC	SINHALA LITH DIGIT SIX
0DED	SINHALA LITH DIGIT SEVEN
0DEE	SINHALA LITH DIGIT EIGHT
0DEF	SINHALA LITH DIGIT NINE

We also propose that the following 20 Sinhala numerals be encoded in a suitable block. We also recommend that space be reserved in that block for the numerals 10000, 100000 ... 10^{12} and Sinhala fractions, to be encoded if evidence is found for them.

It is not possible to encode these in the current Sinhala block as it does not contain sufficient free space.

Code	**Name**
	ARCHAIC SINHALA NUMBER ONE
	ARCHAIC SINHALA NUMBER TWO
	ARCHAIC SINHALA NUMBER THREE
	ARCHAIC SINHALA NUMBER FOUR
	ARCHAIC SINHALA NUMBER FIVE
	ARCHAIC SINHALA NUMBER SIX
	ARCHAIC SINHALA NUMBER SEVEN
	ARCHAIC SINHALA NUMBER EIGHT
	ARCHAIC SINHALA NUMBER NINE
	ARCHAIC SINHALA NUMBER TEN

Code	Name
	ARCHAIC SINHALA NUMBER TWENTY
	ARCHAIC SINHALA NUMBER THIRTY
	ARCHAIC SINHALA NUMBER FORTY
	ARCHAIC SINHALA NUMBER FIFTY
	ARCHAIC SINHALA NUMBER SIXTY
	ARCHAIC SINHALA NUMBER SEVENTY
	ARCHAIC SINHALA NUMBER EIGHTY
	ARCHAIC SINHALA NUMBER NINETY
	ARCHAIC SINHALA NUMBER ONE HUNDRED
	ARCHAIC SINHALA NUMBER ONE THOUSAND