Doc Type: **Working Group Document**
Title: **Preliminary Proposal to add the Leke Script in the SMP of the UCS**
Source: **Erich Fickle, Martin Hosken**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2**
Date: **2012-09-20**

1. Introduction: This document is a preliminary proposal on the Leke script used by the Karen in Thailand and neighbouring countries.

The Karen people live mostly in mountainous regions on the border between Thailand and Myanmar, speaking a range of dialects. There is a continuous spectrum of dialects that are frequently grouped into two dialect clusters, the Pwo Karen and the S'gaw Karen. The Pwo Karen write with several different writing systems, some based on already encoded scripts Latin, Thai, Ancient Mon and other scripts. Leke and several monastic scripts are not encoded. Some monastic scripts might be inappropriate for encoding since they are considered secret by the users and there is a lack of publicly available samples. However, the Leke script is used publicly and in different contexts by Buddhists monks, Christians and in fully secular contexts.

The Leke script was developed before the 1860s (Stern) with some assumed borrowings from Ancient Mon and other scripts from the area.

This preliminary proposal lacks information which is outlined in the section labelled "Issues" below. The authors **seek any additional information or text samples.**

2. Structure:

Texts in this script are written left to right. The writing is structured by syllables like in other scripts in the region. The vowels and syllable coda are structured around the initial letter. The initial is a consonant. For reference, we are calling the base character a consonant, regardless of the actual linguistic value. Consonant finals on closed syllables are represented by a consonant with a LEKE KILLER.

Tones are phonemic, but phonetically they may be a combination of vowel breathiness (2008 Phillips) and tone.

There are no contextual forms of the letters. Consonants without a LEKE KILLER have a default inherent vowel sound of [a] after them. Leke has an irregular system of vowel plus final consonant sequences which are read as open vowels with no final consonant. Some consonant sounds can only occur in word medial positions, and this probably aids in the reader's parsing of text strings when spacing characters are not used.

Numbers can be surrounded by a special punctuation marker LEKE NUMBER MARKER. There is a unique Leke set of digits.

A line break should not occur inside a syllable. The difficulty occurs because some vowel ligatures appear before consonants visually. This difficulty is solved by using a logical storage encoding order, rather than a

visual storage order. Line breaking may also occur outisde a LEKE NUMBER MARKER, after a space or after other punctuation.

There are a few marks in texts whose function is not understood at the time of this proposal. It is assumed they are decorative, but this may not be true. They are not widely spread across different texts.

Some characters always function as medial consonants, and this is reflected in the names.

There is a special character for LEKE END OF SECTION. There is a large amount of variation in the glyphs.

Four of the vowels have the same shape as consonant letters, but these glyphs are marks which will attach below consonant base characters. They are written smaller in size than the consonants.

2.1 Issues

2.1.1. Script Coverage

This is a Preliminary Proposal because the analysis of the function of the script has not been community tested. Many sources claim that there may be letters to provide transcription of Pali (ISO 639-3:pli) texts representing sounds which are not normally used for Karen languages. We do not have examples of these transcribed Pali texts at this time. There are over 30 vowel phonemes in Pwo Karen (Phillips), but the text examples we have do not show more than a handful. Some of the older sources (Stern 1968), claim that the Leke script over-differentiates phonemically, but the evidence we have shows an extreme under differentiation. Further research is planned to confirm the script analysis in this Preliminary Proposal.

We have verbal reports that the script is used by the S'gaw Karen (ISO 639-3:ksw) around Kanchanaburi province in Thailand. This use by S'gaw Karen speakers has not been confirmed, and it is unknown if they have extended the script. By far, the largest community of Leke users is among the Pwo Karen (ISO 639-3:pwo).

The large number of empty slots in the charts are to accommodate Pali extension letters (Brown, 1985).

2.1.2 Incomplete analysis

The names of the characters are not satisfactory in this proposal. There are characters where exact pronunciation was not known, especially the vowels and a few consonants. In some cases there is not even a reasonable guess that could be tentatively made. This orthography is in the process of being resurveyed so that a Latin script English language transcription of the names can be made more accurately for the Unicode/ISO 10646 names list.

There are characters which are not in the normal consonant sequence and appear to function as vowels, but are not joining diacritics. We have not yet found any information on what these characters are named.

There is one source which claims that some consonants have different vowel phonetic values. If true, this would leave Leke spelling with an indeterminate pronunciation.

3 Characters:

	0	1	2	3	4	5
0	౧	౸	▓	◌ᩡ	▓	○
1	/	౦	▓	◌ᩢ	▓	∨
2	౭	⊙	▓	◌ᩣ	▓	४
3	✓	ౙ	▓	◌ᩤ	▓	ᨿ
4	౨	⋈	▓	◌ᩥ	▓	ᨿ
5	౬	✖	▓	◌ᩦ	▓	४
6	౮	⊛	▓	◌ᩧ	▓	४
7	౧	౨	▓	◌ᩨ	▓	ᨿ
8	౧	౬	▓	◌᩠=	▓	४
9	౧	▓	▓	◌ᩩ	▓	౨
A	౩	▓	▓	◌ᩪ	▓	∗
B	//	▓	—	▓	◌ᩫ	•
C	�	▓	÷	▓	◌̣	ᨶ
D	ౝ	▓	÷	▓	◌:	▓
E	౧౧	▓	⊙	▓	◌ᩬ	▓
F	౧౧	▓	౮	▓	◌̇	▓

Consonants

00	ɔ	LEKE LETTER KA
01	⁄	LEKE LETTER KHA
02	૨	LEKE LETTER NGA
03	˅	LEKE LETTER CA
04	੧	LEKE LETTER CHA
05	୯	LEKE LETTER NJA
06	σ	LEKE LETTER NYA
07	∩	LEKE LETTER THA
08	૧	LEKE LETTER SA
09	୨	LEKE LETTER SHA
0A	੩	LEKE LETTER PA
0B	∥	LEKE LETTER PHA
0C	₹	LEKE LETTER MA
0D	↳	LEKE LETTER YA
0E	୨୨	LEKE LETTER RA
0F	၉	LEKE LETTER ZA
10	�	LEKE LETTER LA
11	೮	LEKE LETTER JHA
12	∩	LEKE LETTER HA
13	੮	LEKE LETTER AA
14	⋈	LEKE LETTER JZA
15	×	LEKE LETTER GLOTAL
16	⦿	LEKE LETTER BA
17	ၐ	LEKE LETTER DA
18	ഗ	LEKE LETTER NA

Non-diacritic Vowels or poorly understood

2B	–	LEKE LETTER unknown
2C	ϛ	LEKE LETTER unknown
2D	÷	LEKE LETTER unknown
2E	⊙	LEKE LETTER unknown
2F	൰	LEKE MEDIAL RA

Vowels

30	໐ᕉ	LEKE VOWEL AA
31	̍	LEKE VOWEL I
32	̌	LEKE VOWEL AE
33	́	LEKE VOWEL AI
34	̂	LEKE VOWEL U
35	̤	LEKE VOWEL AU
36	̩	LEKE VOWEL E
37	̫	LEKE VOWEL O
38	̈	LEKE VOWEL OJ
39	̬	LEKE VOWEL UU
3A	̯	LEKE VOWEL unknown

Tones

4B	̥	LEKE KILLER
4C	̣	LEKE TONE HIGH
4D	ː	LEKE TONE RISING
4E	̔	LEKE TONE FALLING
4F	̇	LEKE TONE BREATHY

Digits

50	○	LEKE DIGIT ZERO
51	∨	LEKE DIGIT ONE
52	૪	LEKE DIGIT TWO
53	ʊ	LEKE DIGIT THREE
54	ʊ	LEKE DIGIT FOUR
55	૪	LEKE DIGIT FIVE
56	૪	LEKE DIGIT SIX
57	⑂	LEKE DIGIT SEVEN
58	�	LEKE DIGIT EIGHT
59	ſ	LEKE DIGIT NINE

Punctuation

5A	∗	LEKE NUMBER MARKER
5B	●	LEKE BULLET
5C	ᔕ	LEKE END OF SECTION

Unicode Properties

```
00;LEKE LETTER KA;Lo;0;L;;;;;N;;;;;
01;LEKE LETTER KHA;Lo;0;L;;;;;N;;;;;
02;LEKE LETTER NGA;Lo;0;L;;;;;N;;;;;
03;LEKE LETTER CA;Lo;0;L;;;;;N;;;;;
04;LEKE LETTER CHA;Lo;0;L;;;;;N;;;;;
05;LEKE LETTER NJA;Lo;0;L;;;;;N;;;;;
06;LEKE LETTER NYA;Lo;0;L;;;;;N;;;;;
07;LEKE LETTER TA;Lo;0;L;;;;;N;;;;;
08;LEKE LETTER SA;Lo;0;L;;;;;N;;;;;
09;LEKE LETTER SHA;Lo;0;L;;;;;N;;;;;
0A;LEKE LETTER PA;Lo;0;L;;;;;N;;;;;
0B;LEKE LETTER PHA;Lo;0;L;;;;;N;;;;;
0C;LEKE LETTER MA;Lo;0;L;;;;;N;;;;;
0D;LEKE LETTER YA;Lo;0;L;;;;;N;;;;;
0E;LEKE LETTER RA;Lo;0;L;;;;;N;;;;;
0F;LEKE LETTER ZA;Lo;0;L;;;;;N;;;;;
10;LEKE LETTER LA;Lo;0;L;;;;;N;;;;;
11;LEKE LETTER JHA;Lo;0;L;;;;;N;;;;;
12;LEKE LETTER HA;Lo;0;L;;;;;N;;;;;
13;LEKE LETTER AA;Lo;0;L;;;;;N;;;;;
14;LEKE LETTER ZJA;Lo;0;L;;;;;N;;;;;
15;LEKE LETTER GLOTTAL;Lo;0;L;;;;;N;;;;;
16;LEKE LETTER BA;Lo;0;L;;;;;N;;;;;
17;LEKE LETTER DAA;Lo;0;L;;;;;N;;;;;
18;LEKE LETTER NA;Lo;0;L;;;;;N;;;;;

2A;LEKE LETTER unknown;Lo;0;L;;;;;N;;;;;
2B;LEKE LETTER unknown;Lo;0;L;;;;;N;;;;;
2C;LEKE LETTER unknown;Lo;0;L;;;;;N;;;;;
2D;LEKE LETTER unknown;Lo;0;L;;;;;N;;;;;
2E;LEKE LETTER unknown;Lo;0;L;;;;;N;;;;;
2F;LEKE MEDIAL RA;Mn;0;L;;;;;N;;;;;

30;LEKE VOWEL AA;Mc;0;NSM;;;;;N;;;;;
31;LEKE VOWEL I;Mn;0;NSM;;;;;N;;;;;
32;LEKE VOWEL AE;Mn;0;NSM;;;;;N;;;;;
33;LEKE VOWEL AI;Mn;0;NSM;;;;;N;;;;;
34;LEKE VOWEL U;Mn;0;NSM;;;;;N;;;;;
35;LEKE VOWEL AU;Mn;0;NSM;;;;;N;;;;;
36;LEKE VOWEL E;Mn;0;NSM;;;;;N;;;;;
37;LEKE VOWEL O;Mn;0;NSM;;;;;N;;;;;
38;LEKE VOWEL OJ;Mn;0;NSM;;;;;N;;;;;
39;LEKE VOWEL UU;Mn;0;NSM;;;;;N;;;;;
3A;LEKE VOWEL unknown;Mn;0;NSM;;;;;N;;;;;

4B;LEKE KILLER;Mn;0;NSM;;;;;N;;;;;
4C;LEKE TONE HIGH;Mn;0;NSM;;;;;N;;;;;
4D;LEKE TONE RISING;Mn;0;NSM;;;;;N;;;;;
4E;LEKE TONE FALLING;Mn;0;NSM;;;;;N;;;;;
4F;LEKE TONE BREATHY;Mn;0;NSM;;;;;N;;;;;

50;LEKE DIGIT 0;Nd;0;L;;0;0;0;N;;;;;
51;LEKE DIGIT 1;Nd;0;L;;1;1;1;N;;;;;
52;LEKE DIGIT 2;Nd;0;L;;2;2;2;N;;;;;
53;LEKE DIGIT 3;Nd;0;L;;3;3;3;N;;;;;
54;LEKE DIGIT 4;Nd;0;L;;4;4;4;N;;;;;
55;LEKE DIGIT 5;Nd;0;L;;5;5;5;N;;;;;
56;LEKE DIGIT 6;Nd;0;L;;6;6;6;N;;;;;
57;LEKE DIGIT 7;Nd;0;L;;7;7;7;N;;;;;
58;LEKE DIGIT 8;Nd;0;L;;8;8;8;N;;;;;
59;LEKE DIGIT 9;Nd;0;L;;9;9;9;N;;;;;
5A;LEKE NUMBER MARKER;Po;0;ON;;;;;N;;;;;
5B;LEKE BULLET;Po;0;ON;;;;;N;;;;;
5c;LEKE END OF SECTION;Po;0;ON;;;;;N;;;;;
```
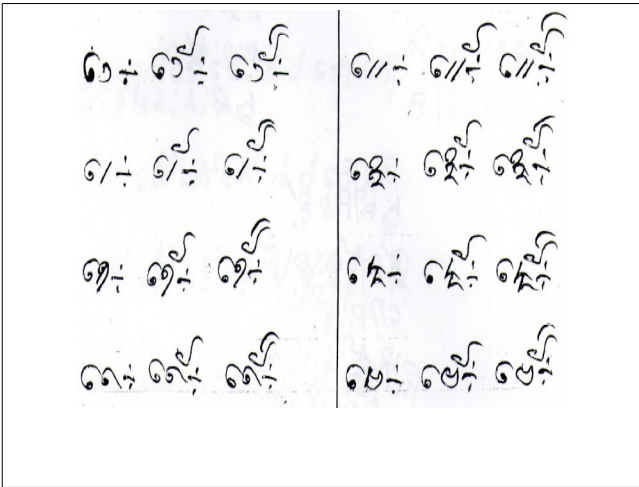
Examples

Figure 1: Examples from literacy copybook of unknown value characters which are apparently vowels.
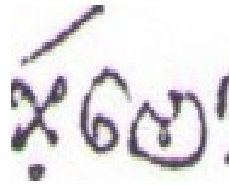

Figure 2 : Example of unknown value character, dotted circle with dot inside.

Figure 3: Example of Leke Leaf Book with End of Section or Paragraph marker in a different, highly ornate style.
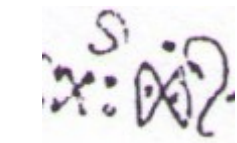

Figure 4: Example of dots inside, but with a dot outside to suggest this is not just ornate placement of vowel marks.
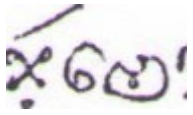
Figure 5: Example of unknown glyph with notch in the top of the dotted circle.
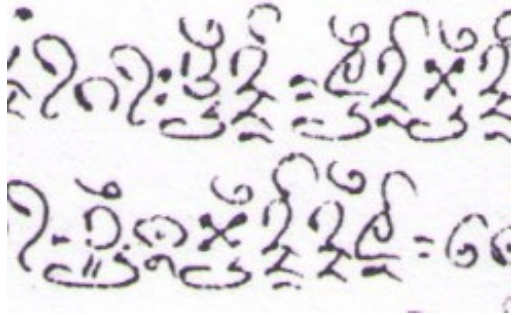
Figure 6: Examples of how the vowel diacritics and tone marks can stack.
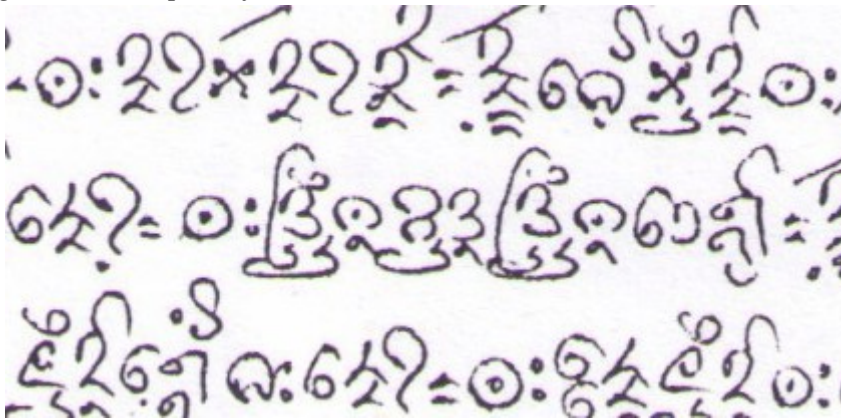
Figure 7: Example of dotted circle character with unknown value or function.
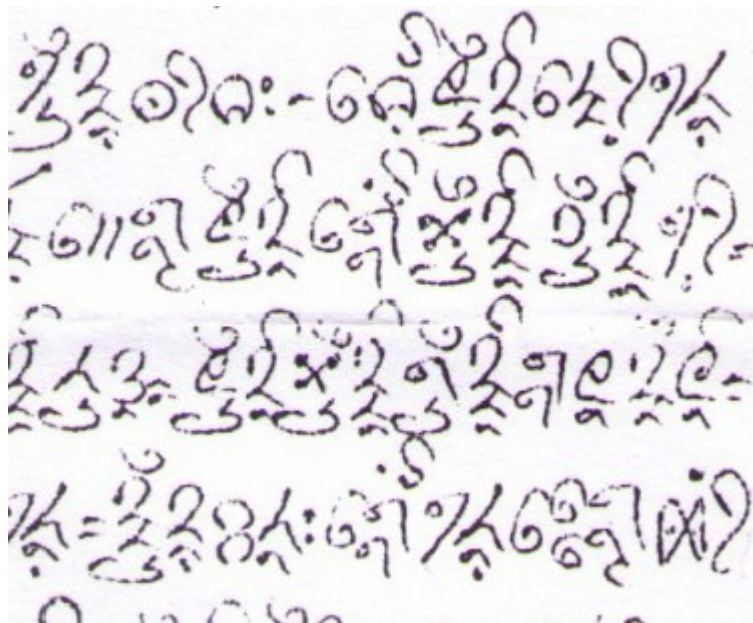
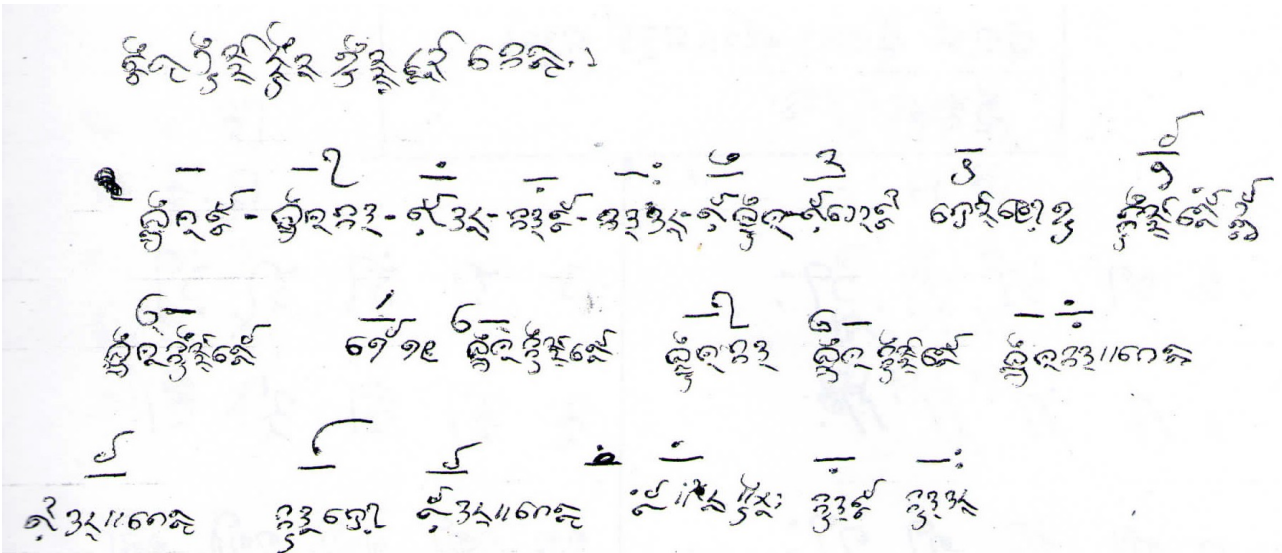*Figure 8: Example showing staff-like glyphs of unknown value or function.*

Figure 9: List of vowels from a Leke Copy Book. The dash appears to function like the Unicode Dotted Circle, except in some cases which look like / ÷ /.

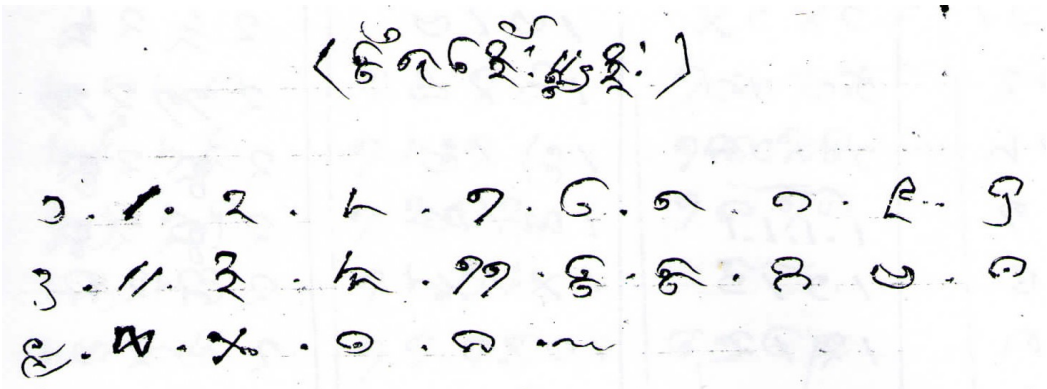*Figure 10: Leke Copy Book tone drill section which shows several unknown characters.*

Figure 11: Yellow book example showing complete consonant list in accepted order. The glyph at the end is a LEKE END OF SECTION.

Bibliography

Information specific to the Leke script

Morse, David (2012), Leke keyboard layout.
Morse, David (2012), Leke font layout.

Stern, Theodore (1968), Three Pwo Karen Scripts: A study of alphabet formation. Anthropological Linguistics, Vol. 10, No. 1.

Huston, Peter, "Refugee/ Asia Stuff: S'gaw Karen, Mon, Shan language script and encoding", [http://peterhuston.blogspot.com/2009/06/refugee-asia-stuff-sgaw-karen-mon-shan.html]

Saw Kya Shin, et. al., "A brief outline on the traditional background of the Lehkai (Ariya) religious sect [and] its Leit-Hsan Wan", [http://www.ibiblio.org/obl/docskaren/Karen%20Heritage %20Web/pdf/Lehkai.pdf]

Background information on the Karen people and writing systems

unknown (1970). Minority groups in Thailand: Chapter 13 The Karen. Cultural Information Analysis Center, Department of the [USA] Army.

Brown, J. Marvin, "From ancient thai to modern dialects: and other writings on Historical Thai linguistis"", Bangkok, Thailand: White Lotus Co., LTD.

Hickey, Gerald C. and Wright, Jesse. 1978. The hill people of Northern Thailand: Social and Ecconomic Development. USAID.

Joseph R. Cooke, J. Edwin Hudspith & James A. Morris. 1976. "Phlong (Pwo Karen of Hot District, Chiang Mai)," Phonemes and Orthography: Language Planning in Ten Minority Languages of Thailand, ed. William Smalley, 187-220.

Iayami, Yoko. "To be Karen and to be be cool", [http://horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_4/sci_hum/39291.pdf]

Stern, Theodore. 1968. "Ariya and the Golden Book: A millenarian Buddhist sect among the Karen", The journal of asian studies, Vol.27, No. 2 (Feb., 1968), pp. 297-328, [https://www.eastwestcenter.org/fileadmin/resources/education/asdp_pdfs/asdp_pdfs_2011/NehRead _Mat_Stern.pdf]

Phillips, Audra. Omkoi Pwo Karen Phonology and Orthography, [http://ic.payap.ac.th/graduate/linguistics/papers/Omkoi_Pwo_Karen_Phonology_and_Orthography. pdf]

Manson, Ken. 2009. "Prolegomena to reconstructing Proto-Karen." La Trobe Papers in Linguistics 12: 26. [http://www.latrobe.edu.au/linguistics/LaTrobePapersinLinguistics/Vol%2012/Contents.htm]

Anderson, Kirsten Ewers. 1978. Elements of Pwo Karen Buddhism. The Scandinavian institute of asian studies: Copenhagen. [http://www.ibiblio.org/obl/docs/KirstenEA-Karen%20Buddhism.htm]

Acknowledgements

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:**	*Leke*
2. Requester's name:	*Erich Fickle, Martin Hosken*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual contribution*
4. Submission date:	*20012/09/20*
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	
(or) More information will be provided later:	*X*

B. Technical – General

1. Choose one of the following:
   - a. This proposal is for a new script (set of characters):   *X*
     - Proposed name of script: *Leke*
   - b. The proposal is for addition of character(s) to an existing block:
     - Name of the existing block:

2. Number of characters in proposal: *57*

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary *X*	B.1-Specialized (small collection)	B.2-Specialized (large collection)
C-Major extinct	D-Attested extinct	E-Minor extinct
F-Archaic Hieroglyphic or Ideographic		G-Obscure or questionable usage symbols

4. 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?   *partially*
   - b. Are the character shapes attached in a legible form suitable for review?   *yes*

5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
   publishing the standard?   *Marten Hoskin or David Morse*
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
   used:

6. 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*

7. 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*

8. 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.

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.)? *partially*
   If YES, with whom?
   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: *this document*
4. The context of use for the proposed characters (type of use; common or rare) *common*
   Reference:
5. Are the proposed characters in current use by the user community? yes
   If YES, where? Reference: *Thailand and others*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP? *no*
   If YES, is a rationale provided?
   If YES, reference:
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? *yes*
    If YES, reference: *this document*
11. Does the proposal include use of combining characters and/or use of composite sequences? *yes*
    If YES, is a rationale for such use provided? *yes*
    If YES, reference: *this document*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? *no*
    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: