

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document**Title: Preliminary proposal for encoding the Varang Kshiti script in the UCS****Source: Michael Everson****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Replaces: N1958 (1999-01-29)****Date: 2008-04-08**

This is a preliminary proposal to encode the Varang Kshiti script in the BMP of the UCS.

1. Introduction. The Varang Kshiti script is used to write the Ho language. Ho is a North Munda language, which family, together with the Mon-Khmer languages, makes up Austro-Asiatic. Varang Kshiti was devised by charismatic community leader Lako Bodra as part of a comprehensive cultural program, and was offered as an improvement over scripts used by Christian missionary linguists. Varang Kshiti is used in primary and adult education, and is published in a variety of printed materials.

2. Structure. Varang Kshiti is written from left to right, with the vowel characters in the order they are spoken; complex vowel placement and variant forms found in Brahmi-derived scripts does not obtain in Varang Kshiti. Letters have an inherent vowel, typically pronounced [a], but sometimes [o] or [e]. The letter 𑄢 HA is used after a vowel to indicate lengthening: 𑄢𑄢 *ah*. Long *i* and *u* are described as “ligatures” and evidently 𑄢 *ī* and 𑄢 *ū* are written, not *𑄢𑄢 and *𑄢𑄢.

3. Digits and numbers. No zero is given in Zide 1996; the tens are used only in stand-alone position; 11 is coded 𑄢 1+1, 23 is coded 𑄢𑄢 2+3, 57 is coded 𑄢𑄢 5+7. The tens could (possibly) be considered ligatures of the digits 1-9 plus a notional zero (so 𑄢 = 𑄢+𑄢, 𑄢 = 𑄢+𑄢, 𑄢 = 𑄢+𑄢, 𑄢 = 𑄢+𑄢, 𑄢 = 𑄢+𑄢, 𑄢 = 𑄢+𑄢, 𑄢 = 𑄢+𑄢, 𑄢 = 𑄢+𑄢, 𑄢 = 𑄢+𑄢). Information on numbers greater than 99 are cannot be derived from Zide’s article.

4. Collating order. Collation order is as in the code chart.

5. Character names. Consonant letter names are derived by adding the inherent vowel [a] to the consonant.

6. Punctuation. I have seen no text in Varang Kshiti. It can be surmised that punctuation common to Latin may be used. Zide notes that there is a special symbol for the syllable *om*, but he does not show it. I have presented an *o* with *anusvara* as a placeholder for the time being.

7. Linebreaking. Letters and digits behave as in Latin and other alphabetic scripts.

8. Unicode Character Properties.

AB00 VARANG KSHITI LETTER NGA;Lo;0;L;;;;N;;;;;
AB01 VARANG KSHITI LETTER GA;Lo;0;L;;;;N;;;;;
AB02 VARANG KSHITI LETTER KA;Lo;0;L;;;;N;;;;;
AB03 VARANG KSHITI LETTER NYA;Lo;0;L;;;;N;;;;;
AB04 VARANG KSHITI LETTER JA;Lo;0;L;;;;N;;;;;
AB05 VARANG KSHITI LETTER CA;Lo;0;L;;;;N;;;;;
AB06 VARANG KSHITI LETTER NNA;Lo;0;L;;;;N;;;;;
AB07 VARANG KSHITI LETTER DDA;Lo;0;L;;;;N;;;;;
AB08 VARANG KSHITI LETTER TTA;Lo;0;L;;;;N;;;;;
AB09 VARANG KSHITI LETTER NA;Lo;0;L;;;;N;;;;;
AB0A VARANG KSHITI LETTER DA;Lo;0;L;;;;N;;;;;
AB0B VARANG KSHITI LETTER TA;Lo;0;L;;;;N;;;;;
AB0C VARANG KSHITI LETTER MA;Lo;0;L;;;;N;;;;;
AB0D VARANG KSHITI LETTER BA;Lo;0;L;;;;N;;;;;
AB0E VARANG KSHITI LETTER PA;Lo;0;L;;;;N;;;;;
AB0F VARANG KSHITI LETTER HA;Lo;0;L;;;;N;;;;;
AB10 VARANG KSHITI LETTER LA;Lo;0;L;;;;N;;;;;
AB11 VARANG KSHITI LETTER RRA;Lo;0;L;;;;N;;;;;
AB12 VARANG KSHITI LETTER RA;Lo;0;L;;;;N;;;;;
AB13 VARANG KSHITI LETTER SSA;Lo;0;L;;;;N;;;;;
AB14 VARANG KSHITI LETTER SA;Lo;0;L;;;;N;;;;;
AB15 VARANG KSHITI LETTER ANUSVARA;Lo;0;L;;;;N;;;;;
AB16 VARANG KSHITI LETTER A;Lo;0;L;;;;N;;;;;
AB17 VARANG KSHITI LETTER I;Lo;0;L;;;;N;;;;;
AB18 VARANG KSHITI LETTER U;Lo;0;L;;;;N;;;;;
AB19 VARANG KSHITI LETTER E;Lo;0;L;;;;N;;;;;
AB1A VARANG KSHITI LETTER O;Lo;0;L;;;;N;;;;;
AB1B VARANG KSHITI LETTER GLOTTAL;Lo;0;L;;;;N;;;;;
AB1C VARANG KSHITI LETTER Y;Lo;0;L;;;;N;;;;;
AB1D VARANG KSHITI LETTER II;Lo;0;L;;;;N;;;;;
AB1E VARANG KSHITI LETTER UU;Lo;0;L;;;;N;;;;;
AB27;VARANG KSHITI DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
AB28;VARANG KSHITI DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
AB29;VARANG KSHITI DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
AB2A;VARANG KSHITI DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
AB2B;VARANG KSHITI DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
AB2C;VARANG KSHITI DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
AB2D;VARANG KSHITI DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
AB2E;VARANG KSHITI DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
AB2F;VARANG KSHITI DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
AB30 VARANG KSHITI OM;Lo;0;L;;;;N;;;;;
AB37 VARANG KSHITI NUMBER TEN;No;0;L;;;10;N;;;;;
AB38 VARANG KSHITI NUMBER TWENTY;No;0;L;;;20;N;;;;;
AB39 VARANG KSHITI NUMBER THIRTY;No;0;L;;;30;N;;;;;
AB3A VARANG KSHITI NUMBER FORTY;No;0;L;;;40;N;;;;;
AB3B VARANG KSHITI NUMBER FIFTY;No;0;L;;;50;N;;;;;
AB3C VARANG KSHITI NUMBER SIXTY;No;0;L;;;60;N;;;;;
AB3D VARANG KSHITI NUMBER SEVENTY;No;0;L;;;70;N;;;;;
AB3E VARANG KSHITI NUMBER EIGHTY;No;0;L;;;80;N;;;;;
AB3F VARANG KSHITI NUMBER NINETY;No;0;L;;;90;N;;;;;

9. Bibliography.

Zide, Norman. 1996. "Scripts for Munda languages", in Peter T. Daniels and William Bright, eds. *The world's writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0

10. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Varang Kshiti encoding.

A. Administrative

1. Title

Proposal for encoding the Varang Kshiti script in the BMP of the UCS

2. Requester's name

Michael Everson

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

Individual contribution.

4. Submission date

2008-04-08

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

No.

6b. More information will be provided later

Yes.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Varang Kshiti.

1c. The proposal is for addition of character(s) to an existing block

No.

1d. Name of the existing block

2. Number of characters in proposal

50.

3. Proposed category (A-Contemporary; 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)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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)?

Yes.

8. 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.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Yes, in N1958.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

No.

2b. If YES, with whom?

2c. 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?

Speakers of the Ho language, whose population is 1,026,000, with 25%-50 % literacy, according to the SIL Ethnologue. The Ethnologue mentions only the Oriya and Devanagari script areas.

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

To write the Ho language.

4b. Reference

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

Yes.

5b. If YES, where?

In India.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Contemporary use and accordance with the Roadmap.

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

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. 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.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

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

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

	AB0	AB1	AB2	AB3
0	 AB00	 AB10		 AB30
1	 AB01	 AB11		
2	 AB02	 AB12		
3	 AB03	 AB13		
4	 AB04	 AB14		
5	 AB05	 AB15		
6	 AB06	 AB16		
7	 AB07	 AB17	 AB27	 AB37
8	 AB08	 AB18	 AB28	 AB38
9	 AB09	 AB19	 AB29	 AB39
A	 AB0A	 AB1A	 AB2A	 AB3A
B	 AB0B	 AB1B	 AB2B	 AB3B
C	 AB0C	 AB1C	 AB2C	 AB3C
D	 AB0D	 AB1D	 AB2D	 AB3D
E	 AB0E	 AB1E	 AB2E	 AB3E
F	 AB0F		 AB2F	 AB3F

Consonants

AB00		VARANG KSHITI LETTER NGA
AB01		VARANG KSHITI LETTER GA
AB02		VARANG KSHITI LETTER KA
AB03		VARANG KSHITI LETTER NYA
AB04		VARANG KSHITI LETTER JA
AB05		VARANG KSHITI LETTER CA
AB06		VARANG KSHITI LETTER NNA
AB07		VARANG KSHITI LETTER DDA
AB08		VARANG KSHITI LETTER TTA
AB09		VARANG KSHITI LETTER NA
AB0A		VARANG KSHITI LETTER DA
AB0B		VARANG KSHITI LETTER TA
AB0C		VARANG KSHITI LETTER MA
AB0D		VARANG KSHITI LETTER BA
AB0E		VARANG KSHITI LETTER PA
AB0F		VARANG KSHITI LETTER HA
AB10		VARANG KSHITI LETTER LA
AB11		VARANG KSHITI LETTER RRA
AB12		VARANG KSHITI LETTER RA
AB13		VARANG KSHITI LETTER SSA
AB14		VARANG KSHITI LETTER SA

Vowels

AB15		VARANG KSHITI LETTER ANUSVARA
AB16		VARANG KSHITI LETTER A
AB17		VARANG KSHITI LETTER I
AB18		VARANG KSHITI LETTER U
AB19		VARANG KSHITI LETTER E
AB1A		VARANG KSHITI LETTER O
AB1B		VARANG KSHITI LETTER GLOTTAL
AB1C		VARANG KSHITI LETTER Y
AB1D		VARANG KSHITI LETTER II
AB1E		VARANG KSHITI LETTER UU

Digits

AB27		VARANG KSHITI DIGIT ONE
AB28		VARANG KSHITI DIGIT TWO
AB29		VARANG KSHITI DIGIT THREE
AB2A		VARANG KSHITI DIGIT FOUR
AB2B		VARANG KSHITI DIGIT FIVE
AB2C		VARANG KSHITI DIGIT SIX
AB2D		VARANG KSHITI DIGIT SEVEN
AB2E		VARANG KSHITI DIGIT EIGHT
AB2F		VARANG KSHITI DIGIT NINE

Sign

AB30		VARANG KSHITI OM
------	---	------------------

Numbers

AB37		VARANG KSHITI NUMBER TEN
AB38		VARANG KSHITI NUMBER TWENTY
AB39		VARANG KSHITI NUMBER THIRTY
AB3A		VARANG KSHITI NUMBER FORTY
AB3B		VARANG KSHITI NUMBER FIFTY
AB3C		VARANG KSHITI NUMBER SIXTY
AB3D		VARANG KSHITI NUMBER SEVENTY
AB3E		VARANG KSHITI NUMBER EIGHTY
AB3F		VARANG KSHITI NUMBER NINETY