

Doc Type: Working Group Document

Title: Proposal for Encoding of Vaidika Sanskrit Characters & Symbols in the BMP of UCS

Source: Ministry of Communications and Information Technology
Government of India

Status: Institutional Member's Contribution

Action: For consideration by UTC

Consolidates: L2/07-395, L2/07-396, L2/07-397, L2/08-042 & L2/08-043, L2/08-110

Date: 5 May 2008

This document consolidates the summary of the views of the Vaidika experts in India. It also reconfirms the encoding of various characters and symbols proposed through various proposals submitted to Unicode Consortium and recommendations of the South Asia Sub committee & UTC.

The combined proposal dated 23 January 2008 (L2/08-042 & L2/08-043) was submitted to Unicode Consortium by Government of India for consideration of Unicode Technical Committee (UTC). The proposal was discussed by the South Asia Sub-Committee on February 4th, 2008 prior to the UTC#114 meeting. The South Asia Sub-Committee recommended most of the proposed characters for inclusion in the future version of the standard (L2/08-110).

Shri Bharat Bhushan, JS&FA, GC, LCG, Ministry of Communications and Information Technology (MC&IT), Govt. of India participated in the UTC#114. Prof. R K Joshi was also deputed to attend the UTC meeting to discuss and present Government of India's view point on the encoding of the Vaidika characters & symbols but due to untimely demise of Prof. Joshi in USA, he could not participate in the meeting.

The various recommendations of the South Asia Sub Committee were further discussed with various scholars, technologists and representatives of Institutes from India. This document is the consolidation of various recommendations of Vaidika experts in India and Government of India officials.

Opening Remarks

In the earlier proposal document L2/07-395, 396 and 397, 79 character codes were identified under the block entitled Vaidika Extensions and requested for its inclusion in the suitable slot in the BMP of UCS.

The UTC has recommended most of the proposed characters for encoding in the future version of the Unicode Standard. UTC recommended to encode the various characters and symbols required for representation of Vaidika text in three different blocks named as:

- | | | |
|----|---------------------------------------|---------------|
| 1. | Devanagari (U+900 – U+97F) | 08 Characters |
| 2. | Vaidika Extensions (U+1CD0 – U+1CF1) | 32 Characters |
| 3. | Devanagari Extended (U+A8E0 – U+A8F9) | 24 Characters |

In the Vaidika Extensions block (1CD0 – 1CFF), 32 characters have been recommended for encoding and 24 characters have been recommended for encoding in the Devanagari Extended block (A8E0 – A8FF). The two codes A8FE and A8FF namely Vaidika Prarambha Chihna and Vaidika Anta Chihna were proposed for addressing syllabic tonal / accent marks in Jaiminiya branch. Eight characters have been recommended to be added in Devanagari block.

Since Dr. Joshi couldn't present the case in the meeting, some of the explanations in support of the proposal couldn't be made available to Script Sub-Committee/ UTC. Therefore, we revisited the characters agreed by UTC and the experts in India have re-confirmed the encoding of most of the characters/ symbols as already recommended. However, the following needs to be reconsidered:

They have suggested relocating the characters exclusively used in Vaidika text to "Vaidika Extension block" from proposed "Devanagari block". Also experts have proposed to encode some more characters in the Devanagari Extended block. The summary of the views of the experts is presented in this document to further ease / focus discussion on the unresolved issues.

The characters as documented in the tables below are reconfirmed for encoding in the various blocks:

VAIDIKA EXTENSIONS

Tone marks for the Samaveda

UTC Code	Glyph	Name of character/ symbol	Reference from L2/08-042 08
1CD0		VEDIC TONE KARSHANA = <i>vaidika saamasvara karshanna</i>	A8F5
1CD1		VEDIC TONE SHARA = <i>vaidika svarita uurdhva shara</i>	A8F6
1CD2		VEDIC TONE PRENKHA = <i>vaidika saamasvara prenkha</i> • <i>indicates vibrato</i>	A8F7
1CD3		VEDIC SIGN NIHSVASA = <i>vaidika saamagaana yogakaala</i>	A8F8

Signs for Yajurvedic

1CD5		VEDIC TONE YAJURVEDIC AGGRAVATED INDEPENDENT SVARITA = <i>vaidika svarita adho nyubja</i>	1CEA
1CD6		VEDIC TONE YAJURVEDIC INDEPENDENT SVARITA = <i>vaidika svarita adho konna</i>	1CEB
1CD7		VEDIC TONE YAJURVEDIC KATHAKA INDEPENDENT SVARITA = <i>vaidika svarita adho vakra rekha</i>	1CE8
1CD8		VEDIC TONE CANDRA BELOW = <i>vaidika svarita adho ardha vakra</i>	1CE9
1CD9		VEDIC TONE YAJURVEDIC KATHAKA INDEPENDENT SVARITA SCHROEDER = <i>vaidika svarita adho samyukta rekhaa</i>	1CEE
1CDA		VEDIC TONE DOUBLE SVARITA = <i>vaidika svarita uurdhva dvi rekha</i>	1CE5
1CDB		VEDIC TONE TRIPLE SVARITA = <i>vaidika svarita uurdhva tri rekha</i>	1CE6
1CDC		VEDIC TONE KATHAKA ANUDATTA = <i>vaidika svarita adho rekhaa</i>	1CEC
1CDD		VEDIC TONE DOT BELOW = <i>vaidika svarita adho bindu</i>	1CED

Tone marks for the Satapathabrahmana

1CDE		VEDIC TONE TWO DOTS BELOW = <i>vaidika svarita adho dvi bindu</i>	1CFA
1CDF		VEDIC TONE THREE DOTS BELOW = <i>vaidika svarita adho tri bindu</i>	1CFB

Tone mark for the Rigveda

1CE0		VEDIC TONE RIGVEDIC KASHMIRI INDEPENDENT SVARITA = <i>vaidika uurdhva vakra rekha</i>	1CE7
------	---	--	------

Tone mark for the Atharvaveda

1CE1		VEDIC TONE ATHARVAVEDIC INDEPENDENT SVARITA = <i>vaidika svarita dvi vakra khaṇḍa</i>	1CE4
------	---	--	------

Diacritics for Visarga

1CE2		VEDIC SIGN VISARGA SVARITA = <i>vaidika madhyarekhaa</i>	1CDE
1CE3		VEDIC SIGN VISARGA UDATTA = <i>vaidika visarga dakshinnatah uurdhvaga</i>	1CDF
1CE4		VEDIC SIGN REVERSED VISARGA UDATTA = <i>vaidika visarga vaamatah uurdhvaga</i>	1CE1
1CE5		VEDIC SIGN VISARGA ANUDATTA = <i>vaidika visarga vaamatah adhoga</i>	1CE0
1CE6		VEDIC SIGN REVERSED VISARGA ANUDATTA = <i>vaidika visarga dakshinnatah adhoga</i>	1CE2
1CE7		VEDIC SIGN VISARGA UDATTA WITH TAIL = <i>vaidika visarga dakshinnatah uurdhva vakra</i>	1CF8
1CE8		VEDIC SIGN VISARGA ANUDATTA WITH TAIL = <i>vaidika visarga vaamatah adho vakra</i>	1CF9

Ardhavisarga

1CF1		VEDIC SIGN ARDHAVISARGA = <i>vaidika jihvaamuliiya upadhamaaniya</i>	1CE3
------	---	---	------

Experts informed that the usage of Ardhavisarga is before the consonant to which it is applied so the glyph should be modified to as shown above.

Marks of nasalization

Table -1

1CE9	ॐ	VEDIC SIGN ANUSVARA ANTARGOMUKHA = <i>vaidika anusvaara antarmukha</i>	
1CEA	ॐ	VEDIC SIGN ANUSVARA BAHIRGOMUKHA = <i>vaidika anusvaara naagaphanna</i>	
1CEB	ॐ	VEDIC SIGN ANUSVARA VAMAGOMUKHA = <i>vaidika anusvaara vaamamukha</i>	
1CEC	ॐ	VEDIC SIGN ANUSVARA VAMARGOMUKHA WITH TAIL = <i>vaidika anusvaara vaamamukha sa-vakra</i>	

Table -2

	ॐ	VAIDIKA ANUSVAARA ANTARMUKHA	1CD6
	ॐ	VAIDIKA ANUSVAARA VAAMAMUKHA	1CD7
	ॐ	VAIDIKA ANUSVAARA VAAMAMUKHA SA-VAKRA	1CD8
	ॐ	VAIDIKA ANUSVAARA BAHIRMUKHA	1CD9
	ॐ	VAIDIKA ANUSVAARA VAAMAMUKHASHCANDRA	1CDA
	ॐ	VAIDIKA ANUSVAARA BAHIRMUKHASHCANDRA	1CDB

There are multiple characters which use the few basic shapes (listed in Table 1) in conjunction with Bindu, Chandrabindu and Anudatta (such characters in current useage as listed in Table 2). Keeping in mind the fact that more such characters could be in usage (it may not have come to notice right now), it is felt desirable to encode the basic shapes and compose the characters in Table 2 through combination of characters which are already present in the code chart. This would keep the provision for extensibility of the standard and will also preserve code space.

DEVANAGARI EXTENDED

Cantillation marks for the Samaveda

A8E0	ॐ	COMBINING DEVANAGARI DIGIT ZERO = <i>vaidika saamasvara anka shuunya</i>	A8E0
A8E1	॑	COMBINING DEVANAGARI DIGIT ONE = <i>vaidika saamasvara anka eka udaatta</i>	A8E1
A8E2	॒	COMBINING DEVANAGARI DIGIT TWO = <i>vaidika saamasvara anka dvi svarita</i>	A8E2
A8E3	॒॑	COMBINING DEVANAGARI DIGIT THREE = <i>vaidika saamasvara anka tri anudaatta</i>	A8E3
A8E4	॒॑॑	COMBINING DEVANAGARI DIGIT FOUR = <i>vaidika saamasvara anka chatur</i>	A8E4
A8E5	॒॑॑॑	COMBINING DEVANAGARI DIGIT FIVE = <i>vaidika saamasvara anka pancha</i>	A8E5
A8E6	॒॑॑॑॑	COMBINING DEVANAGARI DIGIT SIX = <i>vaidika saamasvara anka shat</i>	A8E6
A8E7	॒॑॑॑॑॑	COMBINING DEVANAGARI DIGIT SEVEN = <i>vaidika saamasvara anka sapta</i>	A8E7
A8E8	॒॑॑॑॑॑॑	COMBINING DEVANAGARI DIGIT EIGHT = <i>vaidika saamasvara anka ashta</i>	A8E8
A8E9	॒॑॑॑॑॑॑॑	COMBINING DEVANAGARI DIGIT NINE = <i>vaidika saamasvara anka nava</i>	A8E9
A8EA	अ॑	COMBINING DEVANAGARI LETTER A = <i>vaidika saamasvara abhinihita</i>	A8EA
A8EB	उ॑	COMBINING DEVANAGARI LETTER U = <i>vaidika saamasvara u</i>	A8EE
A8EC	क॑	COMBINING DEVANAGARI LETTER KA = <i>vaidika saamasvara ka</i>	A8ED
A8ED	न॑	COMBINING DEVANAGARI LETTER NA = <i>vaidika saamasvara namana</i>	A8EB
A8EE	प॑	COMBINING DEVANAGARI LETTER PA = <i>vaidika saamasvara prannatam</i>	A8F2
A8EF	र॑	COMBINING DEVANAGARI LETTER RA = <i>vaidika saamasvara ra</i>	A8EC

A8F0		COMBINING DEVANAGARI LETTER VI = <i>vaidika saamasvara vinata</i>	A8F3
A8F1		COMBINING DEVANAGARI SIGN AVAGRAHA = <i>vaidika saamasvara diirghiibhaava</i>	A8F4

Marks of Nasalization

A8F2		DEVANAGARI SIGN SPACING CANDRABINDU = <i>vaidika candrabindu</i>	1CD0
A8F3		DEVANAGARI SIGN CANDRABINDU VIRAMA = <i>vaidika anusvaara candrabindu tiryak</i>	1CD1
A8F4		DEVANAGARI SIGN DOUBLE CANDRABINDU VIRAMA = <i>vaidika anusvaara dvi candrabindu tiryak</i>	1CD5
A8F5		DEVANAGARI SIGN CANDRABINDU TWO = <i>vaidika anusvaara candrabindu sa-dvi</i>	1CD2
A8F6		DEVANAGARI SIGN CANDRABINDU THREE = <i>vaidika anusvaara candrabindu sa-tri</i>	1CD3
A8F7		DEVANAGARI SIGN CANDRABINDU AVAGRAHA = <i>vaidika anusvaara candrabindu sa-avagraha</i>	1CD4
1CED		VEDIC SIGN TIRYAK = <i>vaidika tiryak</i>	1CF3
1CEE		VEDIC SIGN HEXIFORM LONG ANUSVARA = <i>vaidika anusvaara anugami</i>	1CDC
1CEF		VEDIC SIGN LONG ANUSVARA <i>vaidika anusvaara dakshinnamukha</i>	1CDD
1CF0		VEDIC SIGN RTHANG LONG ANUSVARA <i>Vaidika anusvaara ttha-sadrisha</i>	1CF6

Code point 1CEF and 1CFO are representing the same “Gumkar” sign and have been found to be glyph variation in different manuscripts. Therefore, 1CFO is recommended for encoding and 1CEF will be covered in the write-up as allographic variation and need not be encoded.

ADDITIONS IN DEVANAGARI BLOCK:

0900		DEVANAGARI SIGN INVERTED CANDRABINDU =vaidika adhomukha candrabindu	0900
------	---	--	------

Vowel Signs

094E		DEVANAGARI VOWEL SIGN PRISHTHAMATRA E • historic character	
------	---	---	--

The proposal of Prof. Peter Scharf had recommended encoding of Devanagari vowel sign of “Prishthmatra”. There was a majority consensus regarding its encoding in the last UTC meeting and Dr. Joshi’s view point could not be debated in the meeting. This issue has been thoroughly discussed within India after large number of consultations with technical and Vaidika experts. Considering the following facts, it has been strongly recommended that the four Prishthmatra characters required for representing the Prishthmatra Vedic Literature may be encoded in the Vaidika block as listed in the table below:

(i) The proposed Prishthmatra E (as recommended in the last meeting) if encoded in the Devanagari block would lead to ambiguity of data. Moreover, use of Prishthmatra is available only in the Vaidika literature and is not found in the other Devanagari based Languages.

(ii) The encoding of single matra will not be sufficient and to completely represent the Prishthmatra Vaidika texts, four Matras are required which are listed below :

		DEVANAGARI VOWEL SIGN PRISHTHAMATRA E	
		DEVANAGARI VOWEL SIGN PRISHTHAMATRA AI	
		DEVANAGARI VOWEL SIGN PRISHTHAMATRA O	
		DEVANAGARI VOWEL SIGN PRISHTHAMATRA AU	

(iii) The documentary evidence already submitted to UTC also indicates that it is not a straight line but it is slightly curved at the edge like a dagger. Since these are

recommended for encoding in the Vaidika Block, it will not pose any security problems also.

Additional vowel for Marathi

0972	अँ	DEVANAGARI LETTER CANDRA A • Marathi	
------	----	---	--

Various Signs

0973	ॐ	DEVANAGARI SIGN PUSHPIKA = <i>vaidika pushpika</i> • used as a placeholder or “filler” • often flanked by double dandas	1CF1
0974	◌ँ	DEVANAGARI CARET = <i>vaidika trutikaa</i> • zero-width character centred on the point between two orthographic syllables	1CF5

The PUSHPIKA is used in Vaidika text only so it should be shifted to the Vaidika extensions block from the proposed Devanagari block. It is used to represent end of text in Vaidika Literature.

0974 is a symbol used for editing/ proof reading of the text. It is very frequently found in the manuscripts. It was informed that there are some more signs used in manuscripts just for editing/ proof reading purpose. It was recommended that all such sign may be identified and proposal may be submitted to UTC for encoding of these signs in an appropriate block.

Additional consonants in Devanagari block

0955	ँ	DEVANAGARI VOWEL SIGN CANDRA LONG E • used in transliteration of Avestan	
0979	झ	DEVANAGARI LETTER ZHA • used in transliteration of Avestan	
097A	य़	DEVANAGARI LETTER HEAVY YA	0979

Avestan experts have been consulted and their view point is:

- (i) ***There is rare transliteration of Avestan into Devanagari. We have come across only one book which gives such a transcription scheme.***

- (ii) *The Avestan is used by Parsi's in India who have adopted Gujarati as their native script as well as the mother tongue. The Avestan literature is available in the Gujarati script. Hence, it may be desirable to encode it in the Gujarati Code page.*
- (iii) *It is strongly recommended that this should not be encoded in the Devanagari Code Page. However, it can be shifted to Devanagari Extended Block/ Vedic Code page.*

It is suggested that the name of the 097A should be DEVANAGARI LETTER HEAVY JJYA. The annotation “”used for an affricated glide – 09AF Bengali letter ya” needs to be removed because Bengla experts have confirmed that it is not used in Bengali.

Jaiminiya

FFF9	†	Vaidika Prarambha Chinha (Interlinear Annotation Anchor)	A8FE
FFFB	†	Vaidika Anta Chinha (Interlinear Annotation Terminator)	A8FF

It is felt that for representation of the Vaidika Jaiminiya Text needs thorough and detailed study of all such possible characters before finalizing or recommending any scheme. DIT will come up with the comprehensive proposal for representation of the Jaiminiya in consultation with the Jaiminiya experts.

Additional Characters Proposed by Government of India but no decision has been taken:

	卐	Vaidika swastika	1CF4
	◌ि	Vaidika maatraakaal	1CF2
	◌ँ	Vaidika Anusvaara Ubhayatomukha	1CF7
	◌ं	Vaidika Kampa	1CEF (withdrawn, see Note below)
	◌ँ	Vaidika Svarita Uraasi Rekhaa	1CF0 (withdrawn, see Note below)
	ॐ	Vaidika Letter LLha	097A (withdrawn, see Note below)

1CF4: *Swastika is being encoded in the Unicode Standard in the Tibetan block. DIT took initiative to get it encoded in the Devanagari block. BIS has submitted Negative Vote to ISO Ballet against encoding of it with its present place and name.*

1CF2: *Vedic Matraakaal: It is used to add half the duration to a digit written inline with Vedic text. The digit is post-positioned after vowels A (0905) I (0907) U (0909) Vocalic R (090B) Vocalic L (090C). As per documentary evidence, it is shorter in length from vowel sign AA (093E) and suitable properties need to be assigned so that it can't be used with consonants.*

1CF7: *Evidences have been provided and it may be included in the character set.*

1CEF: *This may be represented using combination of two code points. Hence there is no need to encode it.*

1CF0: *This may be represented using combination of two code points. Hence there is no need to encode it.*

097A: *This may be represented as the conjunct. Special glyph may be handled at font level. In the Vaidika Sanskrit write up, this glyph will be defined as a combination of letter LLA (0933) i.e. Marathi LA, Halant (Viram) and Devanagari letter HA (0939).*