

Encoding of Vaidika Characters & Symbols in Unicode

L2/09-067

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

Source: Ministry of Communications and Information Technology
Department of Information Technology
Government of India

Status: Institutional Member's Contribution

Date: 28 January 2009

This document consolidates the final recommended code charts for the representation of the Vaidika characters and symbols in the Unicode Standard by Department of Information Technology, Ministry of Communications & IT, Government of India.

UTC recommended encoding of various characters and symbols required for representation of Vaidika text in three different blocks i.e. Devanagari (U+900 – U+97F), Vaidika Extensions (U+1CD0 – U+1CF1) and Devanagari Extensions (U+A8E0 – U+A8F9).




































In the Vaidika Extensions block (1CD0 – 1CFF), 35 characters have been recommended for encoding. 26 characters have been recommended for encoding in the Devanagari Extended block (A8E0 – A8FF). Four characters were recommended for encoding in the Devanagari block.

The charts shown below shows the various characters accepted for encoding for the representation of Vaidika characters and symbols.





























1. Devanagari (U+900 – U+97F)

	090	091	092	093	094	095	096	097
0	ॐ 0900	ऐ 0910	ठ 0920	र 0930	ी 0940	ॐ 0950	ऋ 0960	० 0970
1	ँ 0901	ऑ 0912	ड 0921	ऱ 0931	ु 0941	ं 0951	ृ 0961	। 0971
2	ं 0902	ओ 0912	ढ 0922	ल 0932	ू 0942	्र 0952	ृ 0962	ँ 0972
3	ः 0903	ओ 0913	ण 0923	ळ 0933	ु 0943	े 0953	ृ 0963	
4	अ 0904	औ 0914	त 0924	ळ 0934	ु 0944	ँ 0954	। 0964	
5	अ 0905	क 0915	थ 0925	व 0935	ँ 0945	ँ 0955	॥ 0965	
6	आ 0906	ख 0916	द 0926	श 0936	े 0946		० 0966	
7	इ 0907	ग 0917	ध 0927	ष 0937	े 0947		१ 0967	
8	ई 0908	घ 0918	न 0928	स 0938	ै 0948	ऋ 0958	२ 0968	
9	उ 0909	ड 0919	न 0929	ह 0939	ॉ 0949	ख 0959	३ 0969	ज़ 0979
A	ऊ 090A	च 091A	प 092A		ो 094A	ग 095A	४ 096A	य 097A
B	ऋ 090B	छ 091B	फ 092B		ो 094B	ज़ 095B	५ 096B	ग 097B
C	ृ 090C	ज 091C	ब 092C	। 093C	ौ 094C	ड 095C	६ 096C	ज़ 097C
D	ँ 090D	झ 091D	भ 092E	ऽ 093D	् 094D	ढ 095D	७ 096D	१ 097D
E	ऐ 090E	अ 091E	म 092F	ा 093E	ि 094E	फ़ 095E	८ 096E	ड 097E
F	ए 090F	ट 091F	य 0920	ि 093F		य 095F	९ 096F	ब 097F

2. Vedic Extensions (U+1CD0 – U+1CFX)

	1CD	1CE	1CF
0	 1CD0	 1CE0	 1CF0
1	 1CD1	 1CE1	 1CF1
2	 1CD2	 1CE2	 1CF2
3	 1CD3	 1CE3	
4	 1CD4	 1CE4	
5	 1CD5	 1CE5	
6	 1CD6	 1CE6	
7	 1CD7	 1CE7	
8	 1CD8	 1CE8	
9	 1CD9	 1CE9	
A	 1CDA	 1CEA	
B	 1CDB	 1CEB	
C	 1CDC	 1CEC	
D	 1CDD	 1CED	
E	 1CDE	 1CEE	
F	 1CDF	 1CEF	

3. Devanagari Extended (U+A8E0 – U+A8FX)

	A8E	A8F
0	 A8E0	 A8F0
1	 A8E1	 A8F1
2	 A8E2	 A8F2
3	 A8E3	 A8F3
4	 A8E4	 A8F4
5	 A8E5	 A8F5
6	 A8E6	 A8F6
7	 A8E7	 A8F7
8	 A8E8	 A8F8
9	 A8E9	 A8F9
A	 A8EA	 A8FA
B	 A8EB	 A8FB
C	 A8EC	
D	 A8ED	
E	 A8EE	
F	 A8EF	

Recently proposed characters

DEVANAGARI BLOCK

Additions in Devanagari Block:

0900	◌ँ	DEVANAGARI SIGN INVERTED CANDRABINDU = <i>vaidika adhomukha candrabindu</i>
0955	◌ं	DEVANAGARI VOWEL SIGN CANDRA LONG E • <i>used in transliteration of Avestan</i>

Other Consonants

0979	ज़	DEVANAGARI LETTER ZHA • <i>used in transliteration of Avestan</i>
097A	य़	DEVANAGARI LETTER HEAVY YA • <i>used for an affricate glide JJYA</i>

Prishthamatra Orthography Vowel Signs

094E	◌ँ	DEVANAGARI VOWEL SIGN PRISHTHAMATRA E • <i>Character has historic use only</i>
------	----	---

VAIDIKA EXTENSIONS (U+1CD0 – U+1CFX)



Tone marks for the Samaveda

UTC Code	Glyph	Name of character/ symbol
1CD0	◌ँ	VEDIC TONE KARSHANA = <i>vaidika saamasvara karshanna</i>
1CD1	◌ं	VEDIC TONE SHARA = <i>vaidika svarita uurdhva shara</i>
1CD2	◌ँ	VEDIC TONE PRENKHA = <i>vaidika saamasvara prenkha</i> • <i>indicates vibrato</i>
1CD3	॥	VEDIC SIGN NIHSHVASA = <i>vaidika saamagaana yogakaala</i>


Signs for Yajurvedic

1CD4	◌ँ	VEDIC SIGN MID-CHARACTER SVARITA = <i>vaidika svarita urasi rekhaa</i>
1CD5	◌ँ	VEDIC TONE YAJURVEDIC AGGRAVATED INDEPENDENT SVARITA = <i>vaidika svarita adho nyubja</i>
1CD6	◌ँ	VEDIC TONE YAJURVEDIC INDEPENDENT SVARITA = <i>vaidika svarita adhah konna</i>
1CD7	◌ँ	VEDIC TONE YAJURVEDIC KATHAKA INDEPENDENT SVARITA = <i>vaidika svarita adho vakra rekhaa</i>
1CD8	◌ँ	VEDIC TONE CANDRA BELOW = <i>vaidika svarita adho ardha vakra</i>
1CD9	◌ँ	VEDIC TONE YAJURVEDIC KATHAKA INDEPENDENT SVARITA SCHROEDER = <i>vaidika svarita adho samyukta rekhaa</i>
1CDA	◌ँ	VEDIC TONE DOUBLE SVARITA = <i>vaidika svarita uurdhva dvi rekha</i>
1CDB	◌ँ	VEDIC TONE TRIPLE SVARITA = <i>vaidika svarita uurdhva tri rekhaa</i>
1CDC	◌ँ	VEDIC TONE KATHAKA ANUDATTA = <i>vaidika svarita adho rekhaa</i>
1CDD	◌ँ	VEDIC TONE DOT BELOW = <i>vaidika svarita adho bindu</i>

Tone marks for the Satapathabrahmana

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

Tone mark for the Rigveda

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





Tone mark for the Atharvaveda





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


Diacritics for Visarga

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

Marks of nasalization

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

1CED		VEDIC SIGN TIRYAK = <i>vaidika tiryak</i>
1CEE		VEDIC SIGN HEXIFORM LONG ANUSVARA = <i>vaidika anusvaara anugamii</i>
1CEF		VEDIC SIGN LONG ANUSVARA = <i>vaidika anusvaara dakshinnamukha</i>
1CF0		VEDIC SIGN RTHANG LONG ANUSVARA = <i>vaidika anusvaara ttha-sadrisha</i>


















1CF1		VEDIC SIGN ANUSVARA UBHAYATOMUKHA = <i>vaidika anusvaara ubhayatomukha</i>
------	---	---

Ardhavisarga








1CF2		VEDIC SIGN ARDHAVISARGA = <i>vaidika jihvaamuliiya upadhamaaniiya</i>
------	---	--

DEVANAGARI EXTENSIONS (U+A8E0 – U+A8FX)

Cantillation marks for the Samaveda

A8E0		COMBINING DEVANAGARI DIGIT ZERO = <i>vaidika saamasvara anka shuunya</i>
A8E1		COMBINING DEVANAGARI DIGIT ONE = <i>vaidika saamasvara anka eka udaatta</i>
A8E2		COMBINING DEVANAGARI DIGIT TWO = <i>vaidika saamasvara anka dvi svarita</i>
A8E3		COMBINING DEVANAGARI DIGIT THREE = <i>vaidika saamasvara anka tri anudaatta</i>
A8E4		COMBINING DEVANAGARI DIGIT FOUR = <i>vaidika saamasvara anka chatur</i>
A8E5		COMBINING DEVANAGARI DIGIT FIVE = <i>vaidika saamasvara anka pancha</i>
A8E6		COMBINING DEVANAGARI DIGIT SIX = <i>vaidika saamasvara anka shatt</i>
A8E7		COMBINING DEVANAGARI DIGIT SEVEN = <i>vaidika saamasvara anka sapta</i>
A8E8		COMBINING DEVANAGARI DIGIT EIGHT = <i>vaidika saamasvara anka ashta</i>
A8E9		COMBINING DEVANAGARI DIGIT NINE = <i>vaidika saamasvara anka nava</i>
A8EA		COMBINING DEVANAGARI LETTER A = <i>vaidika saamasvara abhinihita</i>
A8EB		COMBINING DEVANAGARI LETTER U = <i>vaidika saamasvara u</i>
A8EC		COMBINING DEVANAGARI LETTER KA = <i>vaidika saamasvara ka</i>
A8ED		COMBINING DEVANAGARI LETTER NA = <i>vaidika saamasvara namana</i>
A8EE		COMBINING DEVANAGARI LETTER PA = <i>vaidika saamasvara prannatam</i>
A8EF		COMBINING DEVANAGARI LETTER RA = <i>vaidika saamasvara ra</i>
A8F0		COMBINING DEVANAGARI LETTER VI = <i>vaidika saamasvara vinata</i>
A8F1		COMBINING DEVANAGARI SIGN AVAGRAHA = <i>vaidika saamasvara diirghibhaava</i>

Marks of Nasalization

A8F2		DEVANAGARI SIGN SPACING CANDRABINDU = <i>vaidika candrabindu</i>
A8F3		DEVANAGARI SIGN CANDRABINDU VIRAMA = <i>vaidika anusvaara candrabindu tiryak</i>
A8F4		DEVANAGARI SIGN DOUBLE CANDRABINDU VIRAMA = <i>vaidika anusvaara dvi candrabindu tiryak</i>
A8F5		DEVANAGARI SIGN CANDRABINDU TWO = <i>vaidika anusvaara candrabindu sa-dvi</i>
A8F6		DEVANAGARI SIGN CANDRABINDU THREE = <i>vaidika anusvaara candrabindu sa-tri</i>
A8F7		DEVANAGARI SIGN CANDRABINDU AVAGRAHA = <i>vaidika anusvaara candrabindu sa-avagraha</i>
A8F8		DEVANAGARI SIGN PUSHPIKA = <i>vaidika pushpika</i> • used as a placeholder or “filler” • often flanked by double dandas

A8F9	↪	DEVANAGARI GAP FILLER used to indicate that a gap in a manuscript or text is not a lacuna
A8FA	∨	DEVANAGARI CARET = <i>vaidika trutikaa</i> for insertion of character • zero-width character centred on the point between two orthographic syllables
A8FB	–	DEVANAGARI HEADSTROKE = <i>shirorekhaa</i> used to indicate uncertain manuscript reading