L2/06-185

A. Administrative

1. Title:	**Proposal for Encoding of Vedic Characters & Symbols in Unicode (Revised - May 2006)**
2. Requester's name:	Department of Information Technology, Ministry of Communications & IT, Government of India
3. Requester type:	Member Body (Institutional Member)
4. Submission date:	**May 10th, 2006 (Revised Proposal)** February 26th , 2003 (First Submission)
5. Requester's reference	
6. (Choose one of the following:) This is a complete proposal: or, More information will be provided later:	If required, more information will be provided later. Upon detailed study of Vedas by Sanskrit scholars, more symbols & characters may be recommended for inclusion as and when completed.

B. Technical

|	
1. Choose one of the following: a. This proposal is for a new script:	**No**
Proposed Name of the Script:	**Extended Devanagari for Vedic Sanskrit**
b. The proposal is for addition of character(s) to an existing block:	**No**
c. Name of the existing block:	**NA**
2. Number of Characters in Proposal	**128 (As per proposed set at Annexure-I)** **+ Proposed 128 Characters for Grantham Script (South India) and 128 Characters for Sharda script (North India) as per Annexure-II**
3. Proposed category	
4. Proposed level of implementation. Is a rational provided for the choice? If yes, reference:	**Keyboard has to be finalized to enable Manuscripts data entry, Vedic and Laukik Sanskrit Text / HTML Pages Databases**
5. Is a repertoire including character names provided? a. if Yes, are the names in accordance with the 'character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?	**Yes**
b. Are the character shapes attached in a legible form suitable for review?	**Yes**

6.	Who will provide the appropriate computerized font (ordered preference: True Type or PostScript format) for publishing the standard? If available now, identify source(s) for the font (include address, email, ftp-site etc.) and indicate the tools used :	**Department of IT,** **Government of India** **(Attached CD Contains the True Type Font)**

7.	References :
	    a. Are references (to other character sets, dictionaries, descriptive texts etc. provided)?
	    b. Are published examples of use (Such as samples from newspapers, magazines, or other sources) of proposed characters attached?
		Published examples have been provided, if required more information may be provided.
		
		(Attached CD provides the usages examples of the proposed code points)

8.	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)?	**Information will be provided later.**
9.	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 behavior information such as line breaks, widths etc., Combining behavior, Spacing behavior, Directional behavior, Default Collation behavior relevance in Mark up contexts, Compatibility equivalence and other Unicode normalization related information	**A separate detailed write-up will be provided for the usage, display behavior and other required information.**

C. Technical - Justification

1.	Has this proposal for addition of character(s) been submitted before? If YES explain	The proposal was submitted to UTC in 2003 as document L2/03-066 and L2/03-067. It was also discussed in UTC#101.
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.)?	Yes
3.	Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	Vedic Sanskrit scholars, linguists, IT experts working in language technology, Industry representatives, Sanskrit scholars etc.
4.	The context of use for the proposed characters (type of use; common or rare)	The usage can vary from common to rare. The academic/ research community will be using Laukik Sanskrit in Devanagari more than Vedic Sanskrit. Only specialized scholars of Vedas would be using Vedic Sanskrit.
5.	Are the proposed characters in current use by the user community?	Yes
6.	After giving due considerations to the principles in *Principles and Procedures document* (a WG 2 standing document) must the proposed characters be entirely in the BMP?	Yes
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 There are phonetic variations for similar looking characters.
9.	Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	No
10.	Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	No
11.	Does the proposal include use of combining characters and/or use of composite sequences	No
12.	Does the proposal contain characters with any special properties such as control function or similar semantics?	No
13.	Does the proposal contain any Ideographic compatibility character(s)?	No

Proposed Vedic Code Set (Revised May 2006)

	088	089	08A	08B	08C	08D	08E	08F
0	0880	0890	08A0	08B0	08C0	08D0	08E0	08F0
1	0881	0891	08A1	08B1	08C1	08D1	08E1	08F1
2	0882	0892	08A2	08B2	08C2	08D2	08E2	08F2
3	0883	0893	08A3	08B3	08C3	08D3	08E3	08F3
4	0884	0894	08A4	08B4	08C4	08D4	08E4	08F4
5	0885	0895	08A5	08B5	08C5	08D5	08E5	08F5
6	0886	0896	08A6	08B6	08C6	08D6	08E6	08F6
7	0887	0897	08A7	08B7	08C7	08D7	08E7	08F7
8	0888	0898	08A8	08B8	08C8	08D8	08E8	08F8
9	0889	0899	08A9	08B9	08C9	08D9	08E9	08F9
A	088A	089A	08AA	08BA	08CA	08DA	08EA	08FA
B	088B	089B	08AB	08BB	08CB	08DB	08EB	08FB
C	088C	089C	08AC	08BC	08CC	08DC	08EC	08FC
D	088D	089D	08AD	08BD	08CD	08DD	08ED	08FD
E	088E	089E	08AE	08BE	08CE	08DE	08EE	08FE
F	088F	089F	08AF	08BF	08CF	08DF	08EF	08FF

Proposed Vedic Code Set (Revised May 2006)

Vedic Sanskrit Phonetic Break-up Signs

Code	Sign	Description
0880		VEDIC SANSKRIT UCCHA SWAN
0881		VEDIC SANSKRIT NIMNA SWAN
0882		VEDIC SANSKRIT BALA (STRESS) BHEDAK CHINHA 1
0883		VEDIC SANSKRIT BALA (STRESS) BHEDAK CHINHA 2
0884		VEDIC SANSKRIT KAAL (TIME) BHEDAK CHINHA ATI LAGHU
0885		VEDIC SANSKRIT KAAL (TIME) BHEDAK CHINHA LAGHU
0886		VEDIC SANSKRIT KAAL (TIME) BHEDAK CHINHA GURU
0887	[R]	Reserved
0888	[R]	Reserved

Vedic Sanskrit Anuswar

Code	Sign	Description
0889		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC (KRISHNA) ANUSVAR 1
088A		VEDIC SANSKRIT ANUSVAR –YAJURVEDIC (KRISHNA) LONG ANUSVAR 2
088B		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR MADHYA 3
088C		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR DAKSHIN 4
088D		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR 5
088E		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC (SHUKLA) ANUSVAR 6
088F		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR 7
0890		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR 8
0891		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR 9
0892	[R]	Reserved
0893		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR 11
0894		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR 12
0895		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR 13
0896		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR 14
0897		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC ANUSVAR 15
0898		VEDIC SANSKRIT ANUSVAR – ANUSHMANS ANUSVAR 16
0899		VEDIC SANSKRIT ANUSVAR – YAJURVEDIC (SHUKLA) ANUSVAR 17
089A	[R]	Reserved

Vedic Sanskrit Swaraadi Chinha – Visarga

Code	Glyph	Name
089B	◌ঃ	VEDIC SANSKRIT ARDH VISARGA
089C	[R]	Reserved
089D	÷	VEDIC SANSKRIT VISARGA 1
089E	↲	VEDIC SANSKRIT VISARGA 2
089F	˙ᵌ	VEDIC SANSKRIT VISARGA 3
08A0	ᴣ̇	VEDIC SANSKRIT VISARGA 4
08A1	˙ᴣ	VEDIC SANSKRIT VISARGA 5
08A2	ᶘ	VEDIC SANSKRIT VISARGA 6
08A3	ᶘ	VEDIC SANSKRIT VISARGA 7
08A4	ᶘ̇	VEDIC SANSKRIT VISARGA 8
08A5	[R]	Reserved
08A6	⤻	VEDIC SANSKRIT ARDHA-VISARGA – JIHVA-MULIYA 1
08A7	[R]	Reserved
08A8	[R]	Reserved
08A9	⊞	VEDIC SANSKRIT ARDHA-VISARGA – UPADHAMANIYA 1
08AA	)(	VEDIC SANSKRIT ARDHA-VISARGA – UPADHAMANIYA 2
08AB	◡	VEDIC SANSKRIT ARDHA-VISARGA – UPADHAMANIYA 3
08AC	[R]	Reserved

Vedic Sanskrit TAARATA Chinha

Code	Glyph	Name
08AD	◌	VEDIC SANSKRIT ANUDATTA
08AE	◌	VEDIC SANSKRIT UDATTA
08AF	◌	VEDIC SANSKRIT SWARIT

Vedic Sanskrit Rigvedic Svarita Chinha

Code	Glyph	Name
08B0	⸜	VEDIC SANSKRIT SVARITA - ATHARVA-VEDIC SVARITA
08B1	↑	VEDIC SANSKRIT SVARITA UDATTA
08B2	◌	VEDIC SANSKRIT SVARITA -LONG SVARITA
08B3	◌	VEDIC SANSKRIT SVARITA – MAITRAYANI
08B4	◌̄	VEDIC SANSKRIT SVARITA ANUDATTA
08B5	◌	VEDIC SANSKRIT SVARITA - SHUKLA YAJURVEDIC SVARITA

08B6	ॐ	VEDIC SANSKRIT SVARITA – MAITRAYANI SVARITA ONE
08B7	॒	VEDIC SANSKRIT SVARITA – YAJURVEDIC SVARITA
08B8	॒	VEDIC SANSKRIT SVARITA – KATTHAK/MAITRAYANI SAMHITA JATYA SVARITA
08B9	[R]	Reserved
08BA	[R]	Reserved
08BB	[R]	Reserved
08BC	[R]	Reserved
08BD	हॗ	VEDIC SANSKRIT SVARITA DEVENAGARI H WITH HORIZONTAL CROSSLINE (comes in Maitrayani Samhita jatya)
08BE	△	VEDIC SANSKRIT SVARITA – ABHINIHIT 1
08BF	ॐ	VEDIC SANSKRIT SVARITA – ABHINIHIT 2

Vedic Sanskrit Svarita Kamp

08C0	॒	VEDIC SANSKRIT SVARITA HRASVA KAMP
08C1	॒	VEDIC SANSKRIT SVARITA DEERGH KAMP
08C2	[R]	Reserved
08C3	[R]	Reserved
08C4	[R]	Reserved
08C5	[R]	Reserved
08C6	[R]	Reserved

Vedic Sanskrit Samavedic Swarochchar Chinha

08C7	॑	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA ONE
08C8	॒	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA TWO
08C9	॒	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA THREE
08CA	॒	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA FOUR
08CB	॒	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA FIVE
08CC	॒	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA SIX
08CD	॒	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA SEVEN
08CE	॒	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA EIGHT
08CF	॒	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA NINE
08D0	॑	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA KAMPA
08D1	॒	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA R
08D2	॒	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA ONE R

08D3	ऋ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA TWO R
08D4	३ऋ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA THREE R
08D5	४ऋ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA FOUR R
08D6	५ऋ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA FIVE R
08D7	ऋ॒ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA R KAMPA
08D8	१ऋ॒ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA 1R KAMPA
08D9	२ऋ॒ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA 2R KAMPA
08DA	३ऋ॒ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA 3R KAMPA
08DB	४ऋ॒ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA 4R KAMPA
08DC	५ऋ॒ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA 5R KAMPA
08DD	उ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA U
08DE	२उ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA 2U
08DF	क् ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA K
08E0	३क् ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA 3K
08E1	⋀ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA URDHVA
08E2	ऽ ◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA SMALL AVAGRAHA
08E3	◌	VEDIC SANSKRIT SAMAVEDIC SWAROCHCHAR-CHINHA Stress
08E4	R	Reserved

Vedic Sanskrit Special Symbols

08E6	卐	VEDIC SANSKRIT SYMBOL SWASTIK
08E7	य	VEDIC SANSKRIT LETTER JA
08E8	R	Reserved
08E9	R	Reserved
08EA	R	Reserved
08EB	...	VEDIC SANSKRIT LOPA CHINHA
08EC	人	VEDIC SANSKRIT KAAKAPADA
08ED	–„–	VEDIC SANSKRIT TATHAIVA
08EE	×	VEDIC SANSKRIT GUNAKA CHINHA
08EF	.	VEDIC SANSKRIT DASHANSH CHINHA

1. The present proposal for encoding of the Vedic Characters and Symbols has been prepared considering only those Vedas which were written in the Devanagari script.

2. The Vedas are also written in some other scripts such as Bengali, Kannada, Telugu, Tamil, Malayalam, Gujarati, Assamese, and Oriya which are already represented in the Unicode Standard. Studies are being conducted to find if any additional symbols are needed to be represented in the Unicode for writing Vedic Sanskrit in these scripts.

3. The Vedas are also written in some other scripts such as Grantham (south India) and Sharada (North India). These scripts are to be represented in the Unicode. Keeping this in view, it is proposed to allocate two code pages (256 Code points) in the Basic Multilingual Plane (BMP) at U+0800 to U+08FF for the Vedic Symbols and Characters. The details of these scripts and required symbols are being worked out with the consultation of the experts. A separate proposal for encoding of these additional symbols and characters will be submitted to Unicode Consortium.