

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

L2/03-147

Doc Type: Working Group Document
Title: Proposal to encode the Saurashtra script in the UCS
Source: Jeyakumar Chinnakkonda Krishnamoorthy and Michael Everson
Status: Expert Contribution
Date: 2002-12-03

This is a PDF commented copy of L2/03-098

A. Administrative

1. Title

Proposal to encode the Saurashtra script in the UCS.

2. Requester's name

Jeyakumar Chinnakkonda Krishnamoorthy and Michael Everson.

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

Individual contribution.

4. Submission date

2002-12-03

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical -- General

1. Choose one of the following:

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

Yes.

Proposed name of script

Saurashtra.

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

No.

1b. Name of the existing block

2. Number of characters in proposal

78.

3. Proposed category (see section II, Character Categories)

Category A.

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 2.

4b. Is a rationale provided for the choice?

Yes.

4c. If YES, reference

Brahmic script with vowel signs.

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Jeyakumar C. K. (www.palkar.org) & Michael Everson (Everson Typography). TrueType.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Jeyakumar C. K. (www.palkar.org) & Michael Everson (Everson Typography). Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes (see below).

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. 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.

9. 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 below for ordering and shaping behaviour.

C. Technical -- Justification

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

No.

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.)?

Yes.

2b. If YES, with whom?

Mr Kubendiran, editor of *Bashabimāni*.

2c. If YES, available relevant documents

N/A

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

See below.

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

Rare.

4b. Reference

See below.

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

Yes.

5b. If YES, where?

In Tamil Nadu, Andhra Pradesh, Gujarat states of India. A monthly magazine called *Bashabimāni* is published in Madurai, India.

6a. 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?

Not necessarily.

6b. If YES, is a rationale provided?

6c. If YES, reference

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)?

Yes.

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

No.

11c. If YES, reference

Brahmic vowel signs.

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

Yes.

13b. If YES, describe in detail (include attachment if necessary)

See below and attachment, which is a primary teaching book for the language.

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

14c. If YES, reference

D. Proposal

Introduction

Saurashtra is an Indo-European language, related to Gujarati and spoken by about 85,000 people in southern India, in the area around Madurai, Salem, and Thanjavur cities. The SIL *Ethnologue* describes it as “an Indo-European island surrounded by Dravidian languages”. Since the end of the 19th century the Telugu, Tamil, Devanagari, and Saurashtra scripts have been used to publish books in Saurashtra. At present, the Tamil script is most commonly used to write Saurashtra, augmented with the use of superscript digits and a colon to indicate sounds not available in the Tamil script.

The Saurashtra script is of the Brahmic type; its modern version was developed in the late 1800s. Like Tamil, Saurashtra does not make use of complex consonant clusters, preferring instead to mark the killed vowel with a visible subscript virama. An exception to this is the conjunct ્kṣa, which is formed of KA + VIRAMA + SSA in this encoding, and which is sorted as a unique letter in older dictionaries. Early Saurashtra text makes use of additional conjuncts, which can be handled with the usual Brahmic shaping rules. So although, apart from ્kṣa, the virama is always visible in modern Saurashtra, it is best to encode modern Saurashtra with ZERO WIDTH NON-JOINER making the virama visible, rather than to consider it a general visible mark. This supports both older and newer Saurashtra text.

Collating order

The usual Sanskrit ordering applies to Saurashtra: A AA I II U UU RU RUU LU LUU E EE AI O OO AU AM AH KA KHA GA GHA NGA CA CHA JA JHA NYA TTA TTHA DDA DDHA NNA TA THA DA DHA NA PA PHA BA BHA MA YA RA LA VA SHA SSA SA HA LLA **KSSA**

Glyph placement

The vowel signs in Saurashtra follow the consonant to which they are applied. The vowels *-i* and *-ī*, however, are typographically joined to the top right corner of their consonant. Examples: **𑌕** *ka* + **𑌖** *-i* = **𑌕𑌖** *ki*, **𑌕** *ka* + **𑌗** *-ī* = **𑌕𑌗** *kī* and **𑌔** *ta* + **𑌖** *-i* = **𑌔𑌖** *ti*, **𑌕** *ta* + **𑌗** *-ī* = **𑌔𑌗** *tī*.

Punctuation and digits

In the samples viewed at present, European COMMA, FULL STOP, and QUESTION MARK are used, as well as the **Devanagari DANDA** and **DOUBLE DANDA**. Digits have distinctive forms in Saurashtra.

Issues

Two fonts exist which contain an unidentified letter **𑌕** *?a*. One of them, *Sangudhari*, which includes precomposed glyphs for characters in *-i* and *-ī*, gives **𑌕𑌖** *?ī* as well. In the publication *Padārthṛṇ*, the words **𑌕𑌖𑌕𑌖** *na?āna* and **𑌕𑌖𑌕𑌖** *na?īna* are transliterated into Tamil-based orthography as **நா:ந** *nā:na* and **நீ:ந** *nī:na*. **𑌕𑌖𑌕𑌖** *ma?or* and **𑌕𑌖𑌕𑌖** *na?annō* are transliterated as **மோ:ந** *mo:r* and **ந:ந்நொ** *na:nnō*. The colon-like mark is used also in Tamil transliterations of this letter in *Bashabhimāni*, where **𑌕𑌖𑌕𑌖** *na?īttak* is written **நீ:த்தக்** *nī:ttak*. This colon-like mark is not encoded in the Tamil block, and the exact meaning of either it or the Saurashtra letter is uncertain.

In a chart from *Bashabhimāni*, four Saurashtra letters are shown in combination with this unidentified letter: **𑌕𑌖** *rh*, **𑌕𑌖** *lh*, **𑌕𑌖** *mh*, **𑌕𑌖** *nh*.

Another issue is the shape of the vowel signs for *-l* and *-ī*. In one teaching book it appears that full-size glyphs are used, with a tilde beneath marking the long vowel: **𑌕𑌖** *kl*, **𑌕𑌖** *kī* (compare **𑌕𑌖** *kr* and **𑌕𑌖** *kṛ*). In earlier sources, however, the glyph for *-l* is reduced, and length is marked with *-ā*; the *Sangudhari* font supports this practice: **𑌕𑌖** *kl*, **𑌕𑌖** *kī* (compare *Sangudhari* **𑌕𑌖** *kr* and **𑌕𑌖** *kṛ*).

Implementations

There are a number TrueType implementations of the modern Saurashtra script in circulation on the internet. To date, no computer implementation of the older version of the script has been created. Previously, a letterpress was used to print *Bhashabimani* magazine. As printing technology advanced, the need for a Saurashtra script font was felt, and was created by local publishers. The *Suresh* font, designed by Suresh Thimma Ramdas, is used to print *Bhashabimani*; a later corrected version of that font is *Kuber*, which is used here. (The name *Kuber* is derived from Kubendiran, the name of the editor of this magazine.) The fonts are **freely available** in the Internet at www.palkar.org.

References

U. Norihiko, *Language of the Saurashtrans in Tirupati*. (In Latin script.)

N. Uchida. *Saurashtra-English dictionary*. ISBN 3447030550. In Saurashtra script. (Uchida is a Japanese scholar who went to Madurai to research Saurashtra. He did his

research at Madurai Kamarajar University. The local language in Madurai is Tamil; in the course of his research, he learned Saurashtra well and is said to speak Saurashtra as well as a native Saurashtrian.)

சுப்ரமணியன், ஸ்ரீ தாடா. 2000. *ஸவ்ராஷ்ட்ர பாத்² மாலொ*.
மதுரை: ப்ரியா பதப்பகம் ப்ரியா டெக்னிகல் இன்ஸ்டிடியூட். [Śrī Tāṭā.
Cupramaṇiyaṇ. 2000. *Savrāṣṭra pātha mālo* = *Savrāṣṭra pāṭ²a mālo*. Madurai: Priyā
Patappakam Priyā Ṭekṇikal Inṣṭīṭiyuṭ]

Bashabhimāni.

Unicode Character Properties

The same properties as apply to Tamil and Devanagari apply to Saurashtra.

உயிரெழுத்துக்கள்

அ	ஆ	இ	ஈ	உ	ஊ	ஓ	ஔ	ஐ
ஐ	ஓ	ஔ	ஐ	ஓ	ஔ	ஐ	ஓ	ஔ

ஐ	ஓ	ஔ	ஐ	ஓ	ஔ	ஐ	ஓ	ஔ
ஐ	ஓ	ஔ	ஐ	ஓ	ஔ	ஐ	ஓ	ஔ

Saurashtra vowels, with equivalents in Tamil and Telugu scripts.

அகரமேய மெய்யெழுத்துக்கள்

க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி

உயிரெழுத்துக்கள்

க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி
க	கா	கி	கி	கி	கி	கி	கி	கி

The text in the bottom half of this figure is given in Saurashtra and Tamil script; the Saurashtra appears to be in an early form of the script as it shows conjunct consonants. It is rendered here with modern orthography and transliteration:

ஸ்ரேஷ்ட ஸௌராஷ்டிர ஸாஹித்ய ஸபா, மதுரை
 śrēṣṭha saurāṣṭra sāhitya sabhā, machuro |
 ஸ்ரேஷ்ட ஸௌராஷ்டிர ஸாஹித்ய ஸபா, மதுரை
 srēṣṭha saurāṣṭra sāhitya sapā, maturai

ஸௌராஷ்ட்ர லிபிநு

அஜுநு

அ	ஆ	இ	ஈ	உ	ஊ	஋	௠	஌
அ	ஆ	இ	ஈ	உ	ஊ	஋	-	-
ஐ	ஓ	ஔ	ஐ	ஓ	ஔ	ஐ	ஓ	ஔ
-	-	உ	உ	-	ஓ	ஓ	அ	அ:

ஒபநு

க ¹	க ²	க ³	க ⁴	ங
க	க	க	க	ங
ச ¹	ச ²	ச ³	ச ⁴	ஞ
ச	ச	ச	ச	ஞ
ட ¹	ட ²	ட ³	ட ⁴	ண ³
ட	ட	ட	ட	ண
த ¹	த ²	த ³	த ⁴	ந
த	த	த	த	ந
ப ¹	ப ²	ப ³	ப ⁴	ம
ப	ப	ப	ப	ம
ய	ர	ல	வ	ஸ
ய	ர	ல	வ	ஸ
ஷ ¹	ஸ	ஹ	ள	க்ஷ ³
ஷ	ஸ	ஹ	ள	க்ஷ

தநுநு

க	ந	க	க்லூ	நு	-
கா	ந	கா	கெ	ந	-
கி	ந	கி	கே	ந	கே
கீ	ந	கி	கை	ந	கே
கு	ந	கு	கொ	ந	-
கூ	ந	கூ	கோ	ந	கோ
க்ரு	ந	கூ	கௌ	ந	கௌ
க்ரு	ந	-	கம்	ந	கம்
க்லூ	ந	-	கஹ	ந	க:

ரோமன் எழுத்து-ஸௌராஷ்டிர மொழி

- ஜாபாலி

ரோமன் (ஆங்கில) எழுத்துகளைக் கொண்டு ஸௌராஷ்டிர மொழியை எழுதப் பயன்படுத்தும் முறை. வெளிநாடுகளில் வாழும் ஸௌராஷ்டிரர்கள் இம் முறையினைப் பயன்படுத்தி தம் உறவினர்களுக்கும் நண்பர்களுக்கும் கடிதம் மற்றும் E-mail-களை அனுப்பலாம். ஏன், இங்குள்ளவர்களும் வயது முதிர்ந்த - புது வகை வரி வடிவத்தினைக் கற்றுக் கொள்ள முடியாது என்று நம்புவர்கள் - இதனைப் பயன்படுத்தலாமே.

VOWELS - உயிரெழுத்துகள்

ஐ ஐ ஐ ஐ ஐ ஐ ஐ ஐ ஐ ஐ
அ ஆ இ ஈ உ ஊ எ ஏ ஒ ஓ
A Ā I Ī U Ū E Ē O Ō

CONSONANTS - மெய்யெழுத்துகள்

க	க	க	க	ங்	ச	ச	ஜ	ஜ	ஞ்
K	Kh	G	Gh	Ng	C	Ch	J	Jh	Ņg
த	த	த	த	ண்	த்	த்	த்	த்	ந்
T	Th	D	Dh	N	T	Th	D	Dh	N
ப	ப	ப	ப	ம்	ய்	ர	ல்	வ்	ஸ்
P	Ph	B	Bh	M	Y	R	L	V	S
ஷ	ஸ்	ஹ்	ள்	க்ஷ	ர்ஹ்	ல்ஹ்	ம்ஹ்	ந்ஹ்	
Sh	S	H	L	Ksh	Rh	Lh	Mh	Nh	

CONSONANT ADDED VOWELS - உயிர் மெய்யெழுத்துகள்

க கா கி கீ கு கூ கே கோ கோ
KA KĀ KI KĪ KU KŪ KE KĒ KO KŌ

EXAMPLES - உதாரணங்கள்

பாஷாபிமானி இயெ அநுசரி அம்தெநு லேகூந் லிக்குவொ
BhĀshāBhīmānī IYE ANUCARI AMTENU LĒkhUN LIKKUVO

ஸவ்ராஷ்டிர மணி ஒண்டெதிம் சிக் குல் டுவாயி
SAVRĀshĪTRA MAṆĪ ONṬEDIM SIKKULDUVĀYĪ

Row 1x0: SAURASHTRA

	1x00	1x01	1x02	1x03	1x04	1x05
0	◌◌	இ	ஓ	ய	◌௭	4
1	◌ஃ	ஐ	௮	ப	◌௨	
2	ஐ	ஈ	ஐ	த	◌௪	
3	ஐ	ஐ	ஐ	த	◌ஈ	
4	ஐ	ஐ	ஐ		◌	
5	ஐ	ஐ	ஐ	◌7		
6	த	ஐ	ஐ	◌	0	
7	த	ஐ	ஐ	◌	9	
8	ஐ	ஐ	ஐ	◌8	௨	
9	ஐ	ஐ	ஐ	◌8	3	
A	ஐ	ஐ	ஐ	◌௭	8	
B	ஐ	ஐ	ஐ	◌௭	௭	
C	ஐ	ஐ	ஐ	◌௮	௭	
D	ஐ	ஐ	ஐ	◌௮	௭	
E	ஐ	ஐ	ஐ	◌௭	௭	
F	ஐ	ஐ	ஐ	◌௭	௭	

G = 00
P = 01

Row 1x0: SAURASHTRA

hex	Name	hex	Name
1x00	SAURASHTRA SIGN ANUSVARA	1x59	(This position shall not be used)
1x01	SAURASHTRA SIGN VISARGA	1x5A	(This position shall not be used)
1x02	SAURASHTRA LETTER A	1x5B	(This position shall not be used)
1x03	SAURASHTRA LETTER AA	1x5C	(This position shall not be used)
1x04	SAURASHTRA LETTER I	1x5D	(This position shall not be used)
1x05	SAURASHTRA LETTER II	1x5E	(This position shall not be used)
1x06	SAURASHTRA LETTER U	1x5F	(This position shall not be used)
1x07	SAURASHTRA LETTER UU		
1x08	SAURASHTRA LETTER VOCALIC R		
1x09	SAURASHTRA LETTER VOCALIC RR		
1x0A	SAURASHTRA LETTER VOCALIC L		
1x0B	SAURASHTRA LETTER VOCALIC LL		
1x0C	SAURASHTRA LETTER E		
1x0D	SAURASHTRA LETTER EE		
1x0E	SAURASHTRA LETTER AI		
1x0F	SAURASHTRA LETTER O		
1x10	SAURASHTRA LETTER OO		
1x11	SAURASHTRA LETTER AU		
1x12	SAURASHTRA LETTER KA		
1x13	SAURASHTRA LETTER KHA		
1x14	SAURASHTRA LETTER GA		
1x15	SAURASHTRA LETTER GHA		
1x16	SAURASHTRA LETTER NGA		
1x17	SAURASHTRA LETTER CA		
1x18	SAURASHTRA LETTER CHA		
1x19	SAURASHTRA LETTER JA		
1x1A	SAURASHTRA LETTER JHA		
1x1B	SAURASHTRA LETTER NYA		
1x1C	SAURASHTRA LETTER TTA		
1x1D	SAURASHTRA LETTER TTHA		
1x1E	SAURASHTRA LETTER DDA		
1x1F	SAURASHTRA LETTER DDHA		
1x20	SAURASHTRA LETTER NNA		
1x21	SAURASHTRA LETTER TA		
1x22	SAURASHTRA LETTER THA		
1x23	SAURASHTRA LETTER DA		
1x24	SAURASHTRA LETTER DHA		
1x25	SAURASHTRA LETTER NA		
1x26	SAURASHTRA LETTER PA		
1x27	SAURASHTRA LETTER PHA		
1x28	SAURASHTRA LETTER BA		
1x29	SAURASHTRA LETTER BHA		
1x2A	SAURASHTRA LETTER MA		
1x2B	SAURASHTRA LETTER YA		
1x2C	SAURASHTRA LETTER RA		
1x2D	SAURASHTRA LETTER LA		
1x2E	SAURASHTRA LETTER VA		
1x2F	SAURASHTRA LETTER SHA		
1x30	SAURASHTRA LETTER SSA		
1x31	SAURASHTRA LETTER SA		
1x32	SAURASHTRA LETTER HA		
1x33	SAURASHTRA LETTER LLA		
1x34	(This position shall not be used)		
1x35	SAURASHTRA VOWEL SIGN AA		
1x36	SAURASHTRA VOWEL SIGN I		
1x37	SAURASHTRA VOWEL SIGN II		
1x38	SAURASHTRA VOWEL SIGN U		
1x39	SAURASHTRA VOWEL SIGN UU		
1x3A	SAURASHTRA VOWEL SIGN VOCALIC R		
1x3B	SAURASHTRA VOWEL SIGN VOCALIC RR		
1x3C	SAURASHTRA VOWEL SIGN VOCALIC L		
1x3D	SAURASHTRA VOWEL SIGN VOCALIC LL		
1x3E	SAURASHTRA VOWEL SIGN E		
1x3F	SAURASHTRA VOWEL SIGN EE		
1x40	SAURASHTRA VOWEL SIGN AI		
1x41	SAURASHTRA VOWEL SIGN O		
1x42	SAURASHTRA VOWEL SIGN OO		
1x43	SAURASHTRA VOWEL SIGN AU		
1x44	SAURASHTRA SIGN VIRAMA		
1x45	(This position shall not be used)		
1x46	SAURASHTRA DIGIT ZERO		
1x47	SAURASHTRA DIGIT ONE		
1x48	SAURASHTRA DIGIT TWO		
1x49	SAURASHTRA DIGIT THREE		
1x4A	SAURASHTRA DIGIT FOUR		
1x4B	SAURASHTRA DIGIT FIVE		
1x4C	SAURASHTRA DIGIT SIX		
1x4D	SAURASHTRA DIGIT SEVEN		
1x4E	SAURASHTRA DIGIT EIGHT		
1x4F	SAURASHTRA DIGIT NINE		
1x50	unidentified saurashtra letter		
1x51	(This position shall not be used)		
1x52	(This position shall not be used)		
1x53	(This position shall not be used)		
1x54	(This position shall not be used)		
1x55	(This position shall not be used)		
1x56	(This position shall not be used)		
1x57	(This position shall not be used)		
1x58	(This position shall not be used)		