

भारतीय मानक
सूचना अन्तरविनिमय के लिए भारतीय लिपि संहिता

Indian Standard
INDIAN SCRIPT CODE FOR INFORMATION
INTERCHANGE - ISCII

UDC 681.3

***** Notice for this soft copy ISCII91.PDF (V1.0) issued on 1st April 1999 *****

Note: (ISCII-91 or IS13194:1991 is the current National Standard used for Indian Languages)

1. This document is a soft copy of some extracts of the original document published by Bureau of Indian Standards. The tabulation and Roman transliteration is not identical to the original document due to differences in the typesetting software used earlier. The purpose is to convey the information only and no reference should be made to this soft copy as a standard document. Annex G and H are not included in this soft copy.

2. The original document is available from the following address (Group LTD-37)

BUREAU OF INDIAN STANDARDS

MANAK BHAVAN, 9 BAHADUR SHAH ZAFAR MARG

NEW DELHI 110 002

Phone: +91-11-3239402, 3232979 Fax: +91-11-3234062, 3239399, 3239382

Email: bisind@del2.vsnl.net.in

3. This soft copy must not be used to take a hard copy or for distribution as an alternative to the original standard, that must be purchased from BIS. It is meant for understanding ISCII by the International Developers who are unable to get the original document for some time. They must procure the original from the address given below within their period of use of this soft copy.

4. Being in electronic form, the correctness of this soft copy is not guaranteed and you must refer to the original document in case of doubt. Differences are also marked in Italics.

5. Anybody viewing this copy must not make any modification in this document.

6. Validity: You are advised not to use this document for more than 2 months from the date of issue given above on this page, as it may get updated/corrected for typographical errors. Check again for the new version from the same source where you got this copy.

©- BIS 1991

BUREAU OF INDIAN STANDARDS

MANAK BHAVAN, 9 BAHADUR SHAH ZAFAR MARG

NEW DELHI 110 002

Price Group 12

CONTENTS

PAGE

1.	SCOPE	
2.	TERMINOLOGY	
2.1	Alphabet/Script Terminology	
2.2	Font/Display Terminology	
2.3	Character/Coding Terminology	
2.4	Other Terminology	
3.	ISCII CODE PHILOSOPHY	
4.	NATURE OF INDIAN ALPHABET	
4.1	The Consonants	
4.2	Anuswar	
4.3	Nasalization Sign: Chandrabindu	
4.4	Visarg	
4.5	Vowel and Vowel signs	
4.6	Vowel Omission Sign: Halant	
4.7	Conjuncts	
4.8	Diacritic Mark: Nukta	
4.9	Punctuation	
4.10	Other Signs	
4.11	Numerals	
5.	LAYOUT OF ISCII CODE TABLE	
6.	STRUCTURE OF THE ISCII CODE	
6.1	Vowels and Matras	
6.2	Vowel Modifiers	
6.3	Halant	
6.4	Invisible Consonant	
6.5	The Nukta Character	
6.6	Attribute Code	
6.7	Extension Code	
6.8	Numerals	
7.	PROPERTIES OF ISCII CODE	
7.1	Phonetic Sequence	
7.2	Direct Sorting	

- 7.3 Unique Spellings
- 7.4 Display Independence
- 7.5 Transliteration
- 8. ISCII CODE SYNTAX
 - 8.1 Indian Script Word Syntax
 - 8.2 Order within a Syllable
- 9. REFERENCES

ANNEXES

- A INDIAN SCRIPT ALPHABET CORRESPONDENCE
- B PC-ISCII CODE
- C ENGLISH ALPHABET ISCII CODE: EA-ISCII
- D INSCRIPT KEYBOARD
- E ATTRIBUTE CODES
 - E-1 Display Attributes
 - E-2 Font Attributes
- F ROMAN SCRIPT TRANSLITERATION

- G EXTENDED CHARACTER SET FOR VEDIC*
 - G-1 Nature of Vedic Characters*
 - G-2 Extension Code for Vedic*
 - G-3 Structure of Vedic Characters*
 - G-4 Keyboard Overlay for Vedic*
 - G-5 Vedic Syllable Syntax*

- H ISCII TELEX/TELEPRINTERS*
 - H-1 ISSCII-83 Syntax*
 - H-2 Bilingual to Bilingual/Multilingual protocol*
 - H-3 Multilingual machines*
 - H-4 Multilingual to multilingual protocol*

Computer Media Sectional Committee, LTD 37

FOREWORD

This Indian Standard was adopted by the Bureau of Indian Standards after the draft finalized by the Computer Media Sectional Committee has been approved by the Electronics and Telecommunication Division Council.

This standard conforms to IS 10401:1982, "8-bit coded character set for information interchange" (equivalent to ISO 4873). It is intended for use in all computer and communication media which allow usage of 7 or 8 bit-character set - code extension techniques".

In an 8-bit environment, the lower 128 characters are the same as defined in IS 10315:1982 (ISO 646 IRV) "7-bit coded character set for information interchange" also known as ASCII character set. The top 128 characters cater to all the 10 Indian scripts based on the ancient Brahmi script.

In a 7-bit environment the control code SI can be used for invocation of the ISCII code set, and control code SO can be used for reselection of the ASCII code set.

There are 15 officially recognized languages in India: Hindi, Marathi, Sanskrit, Punjabi, Gujarati, Oriya, Bengali, Assamese, Telugu, Kannada, Malayalam, Tamil, Urdu, Sindhi and Kashmiri.

Out of these, Urdu, Sindhi and Kashmiri are primarily written in Perso-Aabic scripts, but get written in Devanagari too (Sindhi is also written in the Gujarati script). Apart from Perso-Arabic scripts, all the other 10 scripts used for Indian languages have evolved from the ancient Brahmi script and have a common phonetic structure, making a common character set possible. The Northern scripts are Devanagari, Punjabi, Gujarati, Oriya, Bengali and Assamese, while the Southern script are Telugu, Kannada, Malayalam and Tamil.

The official language of India, Hindi is written in the Devanagari script. Devanagari is also used for writing Marathi and Sanskrit. It is also the official script of Nepal.

As Perso-Arabic scripts have a different alphabet, a different standard is envisaged for them.

An attribute mechanism has been provided for selection of different Indian script font and display attributes. An Extension mechanism allows use of more characters along with the ISCII code. These are only meant for the environment where no other alternative selection mechanism is available.

The ISCII code table is a super-set of all the characters required in the ten Brahmi-based Indian scripts. For convenience, the alphabet of the official script Devanagari (with diacritic marks for

non-Devanagari alphabets) has been used in the standard. For notational simplicity, elsewhere, the term Indian scripts implies Brahmi-based Indian scripts.

Annex-A provides information on the shapes of the corresponding alphabet of the 10 Indian script. Annexes B and C provide information on the adaptation of the ISCII code for an IBM-PC and "English-Alphabet only" environment. Annex-D defines a suitable keyboard overlay which is common for all the Indian scripts. Annex-E defines the Attribute codes used for selection of different scripts and display attributes. Annex-F defines the Roman script transliteration scheme for all the Indian scripts. Annex-G defines the conversion mechanism between the ISCII code and the earlier ISSCII-83 code used in bilingual telex machines.

History

Since the 70s, different committees of the Department of Official Language and the Department of Electronics (DOE) have been evolving different codes and keyboards which could cater to all the Indian scripts due to their common phonetic structure. Earlier efforts could not keep the ASCII code intact.

In July 1983, DOE announced the ISCII-83 code which complied with the ISO 8-bit code recommendations. ("Report of the sub-committee on Standardization of Indian scripts and their Code for Information Processing", DOE, July 1983). While retaining the ASCII character set in the lower half, it provided the Indian script character set in the upper 06 characters. This also had the recommendation on a Phonographic based keyboard layout for all the Indian scripts.

A keyboard standard for Indian scripts was brought out by DOE in 1986 (Report of the committee for "Standard of Keyboard Layout for Indian Script Based Computers" in

Electronics-Information & Planning, Vol. 14, No. 1, Oct. 1986). The report also contained the recommendation for the corresponding 8-bit ISCII code.

There was a revision of the ISCII code by DOE in 1988 for making it more compact, in order to evolve its corresponding IBM-PC counterpart: PC-ISCII (Report of the sub-committee on "Standardization of Indian Script codes for Information Interchange". DOE, August 1988). Electronics-Information & Planning, Vol. 14, No. 1, Oct. 1986). The report also contained the recommendation for the corresponding 8-bit ISCII code.

There was a revision of the ISCII code by DOE in 1988 for making it more compact, in order to evolve its corresponding IBM-PC counterpart: PC-ISCII (Report of the sub-committee on "Standardization of Indian Script codes for Information Interchange". DOE, August 1988).

Indian Standard

INDIAN SCRIPT CODE FOR INFORMATION INTERCHANGE - ISCII

1. SCOPE

The ISCII code standard specifies a 7-bit code table which can be used in 7 or 8-bit ISO compatible environment. It allows English and Indian script alphabets to be used simultaneously.

It shall not be used in incompatible environments like that of IBM-PC, and with computers which do not allow 8-bit characters, or which do not follow ISO code extension techniques.

It cannot be used in the 5-bit Baudot code used for telecommunications. However transcoding to Baudot is possible as given in Annex-H.

2. TERMINOLOGY

2.1 Alphabet/Script Terminology

2.1.1 Letter: A character representing one or more of the simple or compound sounds used in speech. It can be any of the alphabetic symbols.

2.1.2 Conjunct (Ligature): A letter which is a combination of two or more basic letters. The shape of the conjunct may, or may not, give clue to the constituting letters. Example: the joint form (diagraph) of "ae".

2.1.3 Diacritic mark: A mark added to a letter which distinguishes it from the same letter without a mark, usually having a different phonetic value or stress.

2.1.4 International numerals: International numerals: The conventional 0 to 9 digits used in English for denoting numbers. these are also known as Indo-Arabic numerals (to differentiate them from the Roman numerals like IX for 9).

2.1.5 Script numerals: The 0 to 9 digits in a script, which have shapes distinct from their international counterparts.

2.1.6 Vowel: A letter representing a speech sound made with the vibration of the vocal cords, but without audible obstruction,. English examples: a, e, i, o, u.

2.1.7 Vowel sign: A graphic character associated with a letter, to indicate a vowel to be associated with that character (Matra in Hindi).

2.1.8 Diphthong: A compound vowel character, in which the articulation begins as for one vowel and moves onto another. Example: as in "coin", "loud" and "side".

2.1.9 Consonant: A letter representing a speech sound in which the breath is at least partly obstructed, and which has to be combined with a vowel to form a syllable.

2.1.10 Pure consonant: A consonant which does not have any vowel implicitly associated with it. Example: all the English consonants.

2.1.11 Nasal consonant: A consonant pronounced with the breath passing through the nose. Example: m, n, ng.

2.1.12 Nasalized vowel: A vowel pronounced with the breath passing both through the nose and the mouth. Example: French bon voyage. In Indian scripts this is denoted by a Chandrabindu, diacritic mark.

2.1.13 Aspirated consonant: A consonant which is pronounced with an extra puff of air coming out at the time of release of the oral obstruction. This has a sound of an extra "h". Example: there are two syllables in "water" and three in "inferno".

2.1.14 .Syllable: A unit of pronunciation uttered without interruption, forming whole or part of a word, and usually having one vowel or diphthong sound optionally surrounded by one or more consonants. Example: there are two syllables in "water" and three in "inferno".

2.1.15 Alphabet: A set of letters used in writing a language. Example: the English alphabet consists of upper and lower-case letters A to Z.

2.1.16 Basic alphabet: The minimal set of letters which can be used for uniquely encoding every word of a language. Example: the basic alphabet for English consists of only the upper-case letters A to Z.

2.1.17 Phonetic alphabet: An alphabet which has direct correspondence between letters and sounds. Example: the Indian scripts.

2.1.18 Latin alphabet: The alphabet used for writing the language of ancient Rome,. Also known as the Roman alphabet. Used today for writing English and some other European languages.

2.1.19 Script: A distinctive and complete set of characters used for the written form of one or more languages.

2.1.20 Roman script: The script based on the ancient Roman alphabet, with the letters A-Z and additional diacritic marks. Used for writing a language which is not usually written in the Roman alphabet.

2.1.21 Romanization: Representation of words of a script using the Roman alphabet, possibly through additions of diacritic marks. Example: Romaji is the romanized form of the Japanese script.

2.1.22 Transliteration: Representation of words with the closest corresponding letters in an alphabet of a different language.

2.2 Font/Display Terminology

2.2.1 Font: A set of symbols used for display or printing of a script in a particular style.

2.2.2 Display rendition: The process by which a string of characters is displayed (or printed). In this process several consecutive characters may combine with each other on the screen. The sequence of display of the characters may become different.

2.2.3 Display composing: The process of organizing the basic shapes available in a font in order to display (or print) a word.

2.3 Character/Coding Terminology

2.3.1 Bit: Binary digit. It can have only two values: 0 and 1.

2.3.2. Byte: A bit string that is operated upon as a unit. It usually represents a character and usually consists of eight bits.

2.3.3. Hex digit: Hexadecimal digit, where each digit has 16 values. The values above 9 are denoted by the letters A to F as shown: A(10), B(11), C(12), D(13), E(14), F(15). Four bits are needed to encode a hex digit.

2.3.4 Character: A symbol which can represent a letter, a numeral, a punctuation mark, a special symbol or even a control function.

2.3.5 Control character (control code): A character which normally has no visual form, but affects the recording, processing, transmission or interpretation of data.

2.3.6 Graphic character: A character, other than a control character, that has a visual representation. Normally handwritten, printed or displayed.

2.3.7 5-bit characters (5-bit codes): Characters, whose code has 5 bits, allowing representation of 32 characters.

2.3.8 7-bit characters (7-bit codes): Characters, whose code has 7 bits, allowing representation of 128 characters.

2.3.9 8-bit characters (8-bit codes): Characters, whose code has 8 bits, allowing representation of 256 characters.

2.3.10 Character set: A set of characters grouped together for a purpose, like that of representing a script.

2.3.11 Code table: A table showing the positions allotted to individual characters from a character set.

2.3.12 Character code: Position in the code table of the character.

2.13.13 Code extension: The techniques for encoding of characters that are not included in the character set of a given code.

2.3.14 Extended character set: Characters which are not present in the main character set, but are available through some code extension techniques.

2.3.15 ASCII code: American Standard Code for Information Interchange. 7-bit code which specifies 32 control characters and 96 graphic characters, for English language.

2.3.16 Transcoding: A set of tables and rules by which a code-table can be transformed to another code-table, such that the characters get mapped to their equivalent forms.

2.4 Other Terminology

2.4.1 Direct sorting: Sorting of words done through direct comparison of the corresponding character codes. No special heuristics or rules are used.

2.4.2 Dictionary sorting order: Order in which the letters should be organized within an alphabet, such that words can get ordered according to the language dictionaries. Special rules may have to be applied in addition to direct sorting to achieve this. Example: in English, upper and lower cases have to be transformed to a single case before direct sorting is applied.

2.4.3 Default: A value of state which is assumed when no value or state is explicitly stated.

2.4.4 Keyboard overlay: Defines the characters for each key position (unshifted, shifted etc.), which are meant to replace the standard English characters on a QWERTY keyboard.

3. ISCII CODE PHILOSOPHY

A code for all the Indian scripts is made possible by their common origin from the Brahmi script. An optimal keyboard overlay for all the Indian scripts, is made possible by the phonetic nature of the alphabet.

There are manifold advantages in having a common code and keyboard for all the Indian scripts. Any software which allows ISCII codes to be used, can be used in any Indian script, enhancing its commercial viability. Furthermore, immediate transliteration between different Indian scripts becomes

possible, just by changing the display modes. Simultaneous availability of multiple Indian languages in the computer medium will accelerate their development and facilitate national integration.

The 8-bit ISCII code retains the standard ASCII code, while the Indian script keyboard overlay is designed for the standard English QWERTY overlay. This ensures that English can co-exist with the Indian scripts. This approach also makes it feasible to use Indian scripts along with existing English computers and software, so long as 8-bit character codes are allowed.

4. NATURE OF INDIAN ALPHABET

All the Indian scripts have originated from the ancient Brahmi script which is phonetic in nature. The alphabet in each may vary somewhat, but they all share a common phonetic structure. The differences between scripts primarily are in their written forms, where different combination rules get used.

4.1 The Consonants

Indian script consonants have an implicit + (a) vowel included in them. They have been categorized according to their phonetic properties. There are 5 Vargs (Groups) and non-Varg consonants. Each Varg contains 5 consonants, the last of which is a nasal one. The first four consonants of each Varg, constitute the Primary and Secondary pair. The second consonant of each pair is the aspirated counterpart (has an additional "h" sound) of the first one.

	Primary	Secondary	Nasal					
Varg 1	क	ख	ग	घ	ङ			
	ka	kha	ga	gha	ṅa			
Varg 2	च	छ	ज	झ	ञ			
	ca	cha	ja	jha	ña			
Varg 3	ट	ठ	ड	ढ	ण			
	ṭa	ṭha	ḍa	ḍha	ṇa			
Varg 4	त	थ	द	ध	न			
	ṭa	ṭha	ḍa	ḍha	ṇa			
Varg 5	प	फ	ब	भ	म			
	pa	pha	ba	bha	ma			
non-Varg								
	य	र	ल	व	श	ष	स	ह
	ya	ra	la	va	śa	ṣa	sa	ha

Note that the consonants श (śa) ष (ṣa) are pronounced identically today.

Apart from these consonants, there are some other consonants used in some specific Indian scripts:

न (ṇa) Comes instead of न (ṇ) at middle and end of Tamil words except in the न्त (ṇṭ) conjunct.

य (ya) Used in Oriya, Bengali and Assamese. This is pronounced as "ja", while य get pronounced as "ya".

र (ra) Is an extra trilled "ra" used in Tamil, Telugu and Malayalam. In Marathi it is used for forming the half-ra as in "वाऱ्या" (वाऱ्या).

ळ (ḷa) Used in Tamil, Telugu, Kannada, Malayalam, Oriya, Gujarati and Marathi.

ळ (ḷa) Used in Tamil and Malayalam.

4.2 Anuswar

Anuswar indicates a nasal consonant sound. When an Anuswar comes before a consonant belonging to any of the 5 Vargs, then it represents the nasal consonant belonging to the Varg. Before a non-Varg consonant however the anuswar represents a different nasal sound. Some Hindi examples:

Varg 1 अङ्क=अंक aṅka	पङ्ख=पंख paṅkha	गङ्गा=गंगा gaṅgā	सङ्घ=संघ saṅgha
Varg 2 मञ्च=मंच mañca	पञ्छी=पंछी pañchī	पञ्जा=पंजा pañjā	साञ्ज=सांज sāñjha
Varg 3 घण्टा=घंटा ghaṅṭā	कण्ठ=कंठ kaṅṭha	झण्डा=झंडा jhaṅḍā	ढूण्ड=ढूंढ ḍhūṅḍha
Varg 4 सन्त=संत santa	पन्थ=पंथ pantha	बन्द=बंद banda	गन्ध=गंध gandha
Varg 5 चम्पा=चंपा santa	गुम्फ=गुंफ pantha	खम्बा=खंबा banda	स्तम्भ=स्तंभ gandha

4.3 Nasalization Sign: Chandrabindu ँ

The ँ denotes nasalization of the preceding vowel (can be implicit अ vowel within a consonant). Example: आँख, पाँच, हुमायूँ, हैँ, मेँ.

In Devanagari script it often gets substituted with Anuswar, as the latter is more convenient for writing. In some words, however, Anuswar and Chandrabindu can give different meanings. Hindi example: हैँस (Laugh), हंस (Swan).

4.4 Visarg: ृ

Comes after a vowel sound, and represents a sound similar to "h". This also represents the Aytham ृ character in Tamil.

4.5 Vowels and Vowel signs (Matras)

There are separate symbols for all the vowels in Indian scripts which are pronounced independently (either at the beginning of a word, or after a vowel sound). The consonants in the Indian script themselves have an implicit vowel + (a). To indicate a vowel sound other than the implicit one, a vowel-sign (Matra) is attached to the consonant. Thus there are equivalent Matras for all the vowels, excepting the + vowel.

Roman	ā	i	ī	u	ū	ṛ	e
Vowel	आ	इ	ई	उ	ऊ	ऋ	ए
Matra	ा	ि	ी	ु	ू	ृ	े
Matra on क	का	कि	की	कु	कू	कृ	के
Roman	ē	ai	ē	o	ō	au	ô
Vowel	ए	ऐ	ऎ	ओ	औ	औ	ऑ
Matra	े	ै	ँ	ो	ौ	ौ	ॉ
Matra on क	के	के	कै	को	को	कौ	कॉ

The original pronunciation of the vowel ऋ is now lost; it gets pronounced mostly as "ri" or "ru".

The vowels ऐ and औ are used in Southern scripts for denoting vowels shorter than ए and ओ respectively.

The vowels ऐ (ai) and औ (au) are actually diphthongs, although in Hindi they also get pronounced as longer vowel forms of ए and ओ respectively.

Vowels ऎ and ऑ are used in modern Devanagari for representing the English vowel sounds as in "bai" and "ball" respectively.

Sanskrit infrequently uses three other vowels, which are obsolete today in other Indian scripts. These are:

Vowels	ठ	ड	ण
Matras	ठ		

4.6 Vowel Omission Sign: Halant ्

In Indian scripts consonants are assumed to have an implicit vowel + "a" within them unless an explicit Matra (vowel-sign) is attached. Thus a special sign Halant (्) is needed for indicating that the consonant does not have the implicit + vowel in it.

In Northern languages, the Halant at the end of a word generally gets dropped, though the ending still gets pronounced without a vowel. Example: Ashok = अशोक => अशोक

This doesn't happen in Southern languages and Sanskrit, where a Halant is always used to indicate a vowel-less ending. Example: param=परम् (Sanskrit word).

4.7 Conjuncts

Indian scripts contain numerous conjuncts, which essentially are clusters of upto four consonants

without the intervening implicit vowels. The shape of these conjuncts can differ from those of the constituting consonants. These conjuncts are formed in the ISCII code by putting the Halant (◌̣) character, between the constituent consonants.

Example: क्षत्रिय = क् ष त् र ि य
 कर्म = क र् म्
 क्रम = क् र म

4.8 Diacritic Mark: Nukta-

The Nukta is used for ङ and ढ characters, in some Northern scripts. It is also used for deriving 5 other consonants in the Devanagari and Punjabi scripts, required for Urdu.

क	ख	ग	ज	ड	ढ	फ
क़	ख़	ग़	ज़	ड़	ढ़	फ़

4.9 Punctuation

All punctuation marks used in Indian scripts are borrowed from English, except for the full-stop, instead of which a Viram (।) is used in the Northern scripts. The Viram is, however, being increasingly substituted by a gull-stop. A double Viram (।।) is also used in Sanskrit texts for indicating a verse ending.

4.10 Other Signs

4.10.1 Avagrah ऽ is primarily used in Sanskrit texts. It creates an extra stress on the preceding vowel. Two Avagrahs can be used for creating further extra stress. Avagrah is not used in modern Indian scripts.

4.10.2 Om ॐ is a Hindu religious symbol.

4.11 Numerals

Many Indian scripts today use only the international numerals. Even in others, the usage of international numerals instead of the original forms is increasing. Although the Devanagari script has its own numerals, the official numeral system is the international one.

5. LAYOUT OF ISCII CODE TABLE

The 8-bit Code for Latin and Indian script alphabets is given in Table-1. It consists of 256 positions, arranged in 16 rows and 16 columns. The rows are numbered in decimal as 0 to 15, and in hex as 0 to F. The columns are numbered in decimal as 0 to 240 in increments of 16, and in hex as 0 to F. The lower 128 characters of this table contain the ASCII character set.

The 7-bit Code for Indian script alphabets is given in Table-2. It is meant for an ISO compatible 7/8 bit environment. It consists of 94 positions, arranged in 8 columns and 16 rows.

A position in the Code table is identified in decimal as well as hex notation. A character located at decimal column x and row y will have its decimal position as x+y. A character located at hex column x and row y, will have its hex position as xy.

Table-1 8-bit Code Table of the Latin and Indian Script Alphabet

	Hex01	2	3	4	5	6	7	8	9	A	B	C	D	E	F		
Hex	Dec	16	32	48	64	80	96	112	128	144	160	176	192	208	224	240	
0	0	NUL	DLE	SP	0	@	P	`	p				ओ	ढ	र	े	EXT
1	1	SOH	DC1	!	1	A	Q	a	q			ँ	औ	ण	ल	े	०
2	2	STX	DC2	"	2	B	R	b	r			ं	ऑ	त	ळ	ै	१
3	3	ETX	DC3	#	3	C	S	c	s			ः	क	थ	ळ	ॅ	२
4	4	EOT	DC4	\$	4	D	T	d	t			अ	ख	द	व	ो	३
5	5	ENQ	NAK	%	5	E	U	e	u			आ	ग	ध	श	ो	४
6	6	ACK	SYN	&	6	F	V	f	v			इ	घ	न	ष	ौ	५
7	7	BEL	ETB	'	7	G	W	g	w			ई	ड	न	स	ॉ	६
8	8	BS	CAN	(8	H	X	h	x			उ	च	प	ह	्	७
9	9	HT	EM)	9	I	Y	i	y			ऊ	छ	फ	INV	र्	८
A	10	LF	SUB	*	:	J	Z	j	z			ऋ	ज	ब	ा	।	९
B	11	VT	ESC	+	;	K	[k	{			ऐ	झ	भ	ि		
C	12	FF	FS	,	<	L	\	l				ए	व	म	ी		
D	13	CR	GS	-	=	M]	m	}			ऐ	ट	य	ु		
E	14	SO	RS	.	>	N	^	n	~			ँ	ठ	य	ू		
F	15	SI	US	/	?	O	_	o	DEL			ओ	ड	र	ृ		ATR

Table-2 7-bit Code Table of the Indian Script Alphabet

	A	B	C	D	E	F	
Hex	Dec	160	176	192	208	224	240
0	0		ओ	ढ	र	े	EXT
1	1	ँ	औ	ण	ल	े	०
2	2	ं	ऑ	त	ळ	ै	१
3	3	ः	क	थ	ळ	ॅ	२
4	4	अ	ख	द	व	ो	३
5	5	आ	ग	ध	श	ो	४
6	6	इ	घ	न	ष	ौ	५
7	7	ई	ड	न	स	ॉ	६
8	8	उ	च	प	ह	्	७
9	9	ऊ	छ	फ	INV	र्	८
A	10	ऋ	ज	ब	ा	।	९

B	11	ऐ	झ	भ	ि	
C	12	ए	ञ	म	ी	
D	13	ऐ	ट	य	ु	
E	14	ऐँ	ठ	य	ू	
F	15	ओ	ड	र	्	ATR

Table-3: ISCII Character set - Coded representation

Position

Hex	Dec	Char	Name
A1	161	ँ	Vowel-modifier CHANDRABINDU
A2	162	ं	Vowel-modifier ANUSWAR
A3	163	ः	Vowel-modifier VISARG
A4	164	अ	Vowel A
A5	165	आ	Vowel AA
A6	166	इ	Vowel I
A7	167	ई	Vowel II
A8	168	उ	Vowel U
A9	169	ऊ	Vowel UU
AA	170	ऋ	Vowel RI
AB	171	ऐ	Vowel E (Southern Scripts)
AC	172	ए	Vowel EY
AD	173	ऐ	Vowel AI
AE	174	ऐँ	Vowel AYE (Devanagari Script)
AF	175	ओ	Vowel O (Southern Scripts)
B0	176	ओ	Vowel OW
B1	177	औ	Vowel AU
B2	178	औँ	Vowel AWE (Devanagari Script)
B3	179	क	Consonant KA
B4	180	ख	Consonant KHA
B5	181	ग	Consonant GA
B6	182	घ	Consonant GHA
B7	183	ङ	Consonant NGA
B8	184	च	Consonant CHA
B9	185	छ	Consonant CHHA
BA	186	ज	Consonant JA
BB	187	झ	Consonant JHA
BC	188	ञ	Consonant JNA
BD	189	ट	Consonant Hard TA
BE	190	ठ	Consonant Hard THA

BF	191	ड	Consonant Hard DA
C0	192	ढ	Consonant Hard DHA
C1	193	ण	Consonant Hard NA
C2	194	त	Consonant Soft TA
C3	195	थ	Consonant Soft THA
C4	196	द	Consonant Soft DA
C5	197	ध	Consonant Soft DHA
C6	198	न	Consonant Soft NA
C7	199	न	Consonant NA (Tamil)
C8	200	प	Consonant PA
C9	201	फ	Consonant PHA
CA	202	ब	Consonant BA
CB	203	भ	Consonant BHA
CC	204	म	Consonant MA
CD	205	य	Consonant YA
CE	206	य	Consonant JYA (Bengali, Assamese & Oriya)
CF	207	र	Consonant RA
D0	208	र	Consonant Hard RA (Southern Scripts)
D1	209	ल	Consonant LA
D2	210	ळ	Consonant Hard LA
D3	211	ळ	Consonant ZHA (Tamil & Malayalam)
D4	212	व	Consonant VA
D5	213	श	Consonant SHA
D6	214	ष	Consonant Hard SHA
D7	215	स	Consonant SA
D8	216	ह	Consonant HA
D9	217	INV	Consonant INVISIBLE
DA	218	ा	Vowel Sign AA
DA	219	ि	Vowel Sign I
DC	220	ी	Vowel Sign II
DD	221	ु	Vowel Sign U
DE	222	ू	Vowel Sign UU
DF	223	ृ	Vowel Sign RI
E0	224	े	Vowel Sign E (Southern Scripts)
E1	225	े	Vowel Sign EY
E2	226	ै	Vowel Sign AI
E3	227	ँ	Vowel Sign AYE (Devanagari Script)
E4	228	ो	Vowel Sign O (Southern Scripts)
E5	229	ो	Vowel Sign OW
E6	230	ौ	Vowel Sign AU
E7	231	ॉ	Vowel Sign AWE (Devanagari Script)
E8	232	्	Vowel Omission Sign (Halant)

E9	233	.	Diacritic Sign (Nukta)
EA	234		Full Stop (Viram, Northern Scripts)
EB	235		This position shall not be used
EC	236		This position shall not be used
ED	237		This position shall not be used
EE	238		This position shall not be used
EF	239	ATR	Attribute Code
F0	240	EXT	Extension Code
F1	241	०	Digit 0
F2	242	१	Digit 1
F3	243	२	Digit 2
F4	244	३	Digit 3
F5	245	४	Digit 4
F6	246	५	Digit 5
F7	247	६	Digit 6
F8	248	७	Digit 7
F9	249	८	Digit 8
FA	250	९	Digit 9
FB	251		This position shall not be used
FC	252		This position shall not be used
FD	253		This position shall not be used
FE	254		This position shall not be used

- Note: 1. The positions EB-EE and FB-FE are reserved for future expansion of the code.
2. Scripts corresponding to other Indian languages are given in Annex-A.

6. STRUCTURE OF THE ISCII CODE

A common alphabet for all the Indian scripts is made possible by their common origin from the same ancient Brahmi script. The ISCII code contains only the basic alphabet required by the Indian scripts. All the composite characters are formed through combinations of these basic characters.

6.1 Vowels and Matras

The ISCII code contains separate vowels and Matras (Vowel signs). While a vowel sign can be used independently, the Matra sign is valid only after a consonant. Thus:

कई = क ई, की = क ी

6.2 Vowel Modifiers ँ, ं, ः

After a consonant, vowel or Matra character, a character can be used which modifies the vowel sound and is called a "Vowel Modifier". This can be a Chandrabindu (◌ं), Anuswar (◌ं) or Visarg (◌ः).

Example:

हँस = हँ स, अंत ् अ ं त, अतः ् अ त ः

6.3 Halant ्

The implicit vowel in a consonant can be removed by addition of a Halant sign (◌्). In the ISCII code conjuncts are formed by typing a Halant character between consonants. A conjunct may consist of upto 4 consonants joined by Halants. Example:

क्त् = क्त	श्र् = श्र
श्व् = श्व	ष्ट्र् = ष्ट्र
क्ष् = क्ष	त्र् = त्र
ज्ज् = ज्ञ	र्द्य् = र्द्य

In practice, a Halant sign is shown only if the consonants do not change their shape by joining up. Tamil script has no conjuncts, and thus an explicit Halant sign always gets used. Here are some Devanagari examples where Halant does not disappear:

ट्ठ् = ट्ठ ड्ग् = ड्ग

6.3.1 Explicit Halant

A Halant is used between consonants to form conjuncts. But many times in Sanskrit and Vedic texts, one may wish to show an Explicit Halant which would be shown on the previous consonant, and which would prevent the consonant from joining with the next one. Two consecutive Halants form an Explicit Halant. Example:

क्त्	= क्त	क्त्त्	= क्त
क्त्ति	= क्ति	क्त्त्ति	= क्ति
ड्क्कि	= ड्क्कि	ड्क्क्कि	= ड्क्कि
ट्त्रि	= ट्त्रि	ट्त्र्रि	= ट्त्रि

6.3.2 Soft Halant

A Soft Halant is formed by typing a Nukta character after a Halant. In Devanagari the Soft Halant allows retention of the "half form" for the preceding consonant, and prevents it from combining with the following consonant. Example:

श्व्	= श्व	श्व्	= श्व
क्त्ति	= क्ति	क्त्त्ति	= क्ति

Soft Halant is used in Malayalam along with some consonants to derive separate pure consonant shapes which do not show an attached Halant symbol:

ണ ് = ण ന ് = ന ര ് = ര ല ് = ल छ ् = छ
ण ् = ण न ् = न र ् = र ल ् = ल ङ ् = ङ

6.4 Invisible Consonant INV

The INV (Invisible) code is used for formation of composite characters which require consonantal base, but where the consonant itself ought to be invisible. These may be required only for some special display purposes. Example:

क ् INV= क	INV ्र = ्र
र ् INV= र	INVृ = ृ
INVि = ि	INVी = ी

6.5 The Nukta Character ◌

The Nukta consonants (क़ ख़ ग़ ज़ ड़ ढ़ फ़) get formed by adding a Nukta (◌) character immediately after the appropriate consonant.

In the ISCII code the same Nukta character is thought of as an operator to derive some of the lesser used Sanskrit characters which are not directly available on the Inscript keyboard.

A Nukta can be typed after a Halant to form a Soft Halant.

Table 4: ISCII characters derived by appending a Nukta

Char	Nukta Char	Name
क	क़	Consonant QA (Urdu)
ख	ख़	Consonant KHHA (Urdu)
ग	ग़	Consonant GHHA (Urdu)
ज	ज़	Consonant ZA (Urdu)
ड	ड़	Consonant Flapped DA
ढ	ढ़	Consonant Flapped DHA
फ	फ़	Consonant FA (Urdu)
ऋ	ठ	Vowel RII (Sanskrit)
ॠ	ॡ	Vowel Sign RII (Sanskrit)
इ	ऌ	Vowel LI (Sanskrit)
ि	ऍ	Vowel Sign LI (Sanskrit)
ई	ऎ	Vowel LII (Sanskrit)
ी	ॐ	Vowel Sign LII (Sanskrit)
ँ	ॐ	Sign OM
।	॥	Vowel Stress Sign AVAGRAH (Sanskrit)

6.6 Attribute Code (ATR)

The Attribute code, followed by a displayable ASCII character, defines a font attribute applicable for the following characters. The mechanism is meant for use in that medium where alternative font selection mechanism is not available. The details are given in the Annexd-E.

6.7 Extension Code (EXT)

The Extension code, followed by an ISCII character, defines a new character which can combine with the previous ISCII character. This provision has been primarily made for supplementing Vedic signs along with the Devanagari text. The Vedic character details are given in the Annex-G.

6.8 Numerals

In all the Indian scripts the international numerals are being used increasingly. From the software viewpoint, usage of the same numerals as given in the ASCII set allows proper handling of numerals by existing software. For display rendition purposes however, it may be sometimes desirable to have separate Indian script numerals which are given in the ISCII table.

The ATR mechanism also allows display rendition of the ASCII numerals in an Indian script form. The ISCII numerals should be used only when it is not possible to use the ATR mechanism for selecting numerals in an Indian script.

7. PROPERTIES OF ISCII CODE

7.1 Phonetic Sequence

The ISCII characters, within a word, are kept in the same order as they would get pronounced. Example:

राष्ट्रीय उ र ा ष ट् र ी य
हिन्दी उ ह ि न् द ी

As shown in the latter example, the display order may be different from the phonetic order. Having a spelling according to the phonetic order allows a name to be typed in the same way, regardless of the script it has to be displayed in.

7.2 Direct Sorting

Since there are variations in ordering of a few consonants between different Indian scripts, it is not possible to achieve perfect sorting in all Indian scripts. Special routines would be required when some characters like "Nukta" need to be ignored for the purpose of sorting. For most purposes, however, the direct sorting achieved through the ISCII code should be sufficient.

Vowel combinations and consonant combinations would get ordered as shown below:

अं अ आं आ	...	ओं औ ऑं ऑ
कं क कां का	...	कौं कौ काँ काँ क्
खं ख खां खा	...	खौं .खौ खॉ खॉ
...
हं ह हां हा	...	हौं हो हॉ हॉ

As shown in the chart above, in Indian scripts a character followed by a vowel-modifier comes before the character without it. This is ensured by keeping the vowel-modifiers in the beginning of the ISCII code table. The only exception is when a vowel-modifier comes at the end of a word; since the

comparison is now with the ASCII "space" character (32 decimal) having a lower value, the vowel-modifier character cannot come before the space. Though it is possible to devise another space character having a higher value, it will not be practically possible to type it in. However, the fact that a vowel-modifier would come after a space becomes intuitive as a longer word is expected to come after its shorter counterpart. Example:

अ < अंक < अक

The Nukta consonants are essentially separate consonants, and thus should get sorted separately. This indeed happens since Nukta is kept after all the Indian script characters (excepting Viram, which is a punctuation).

The new Vowels ऐ and औ have been kept after the long vowels ऐ and औ respectively, as the new vowels have more stress. They get substituted by the long vowels in the traditional text.

7.3 Unique Spellings

By using only the basic characters in ISCII, there is only one unique way of typing a word. This would not have been possible if conjuncts like क्ष, ञ, ज्ञ etc. had been given separate codes. The spelling of a word is now the phonetic order of the constituent basic characters. This provides a unique spelling for each word, which is not affected by the display rendition.

For obtaining unique spellings, Soft Halant, Explicit Halant, and INV characters should not be used. These have been provided only for deriving different display renditions, and are not needed normally.

The spelling of a word contains all the information necessary for display composition, which can be automatically done through display algorithms. It becomes possible to type in a text, without even looking at the display. When the tedium of composing goes away, on-line authoring becomes possible, where an author can think out new text while he is typing it.

Unique spellings are essential for making spelling checkers and dictionaries. They are also essential to facilitate finding of words in a word-processor, or for information retrieval from a data-base.

7.4 Display Independence

A word in an Indian script can be displayed in a variety of styles depending on the conjunct repertoire used. ISCII codes however allow a complete delinking of the codes from the displayed fonts.

An ISCII syllable can be displayed using combination of basic shapes. Different implementations can choose variant techniques in combination of these basic shapes. The same text can thus be seen in different font styles by using a different font composition routine.

The Inscript keyboard overlay has one-to-one correspondence with the ISCII code. This way, typing of word does not depend upon its displayed form.

7.5 Transliteration

The ISCII codes are rendered on the display device according to the display composition methodol-

ogy of the selected script. Transliteration to another script can thus be obtained by merely redisplaying the same text in a different script.

Since the display rendering process can be very flexible, it is possible to transliterate the Indian scripts to the Roman script, using diacritic marks. Similarly it is possible to transliterate them to their scripts such as Perso-Arabic.

Transliteration involves mere change of the script, in a manner that pronunciation is not affected. This is not the same as "translation" here the language itself changes.

8. ISCII CODE SYNTAX

In ISCII code some logically related sub-sets can be identified through simple range comparisons. Using these it is possible to predict a syllable boundary for an Indian script word. This may be necessary for composing fonts for display purposes, or for hyphenation at a syllable boundary.

Consonants (C)

क ख ग घ ङ च छ ज झ ञ ट ठ ड ढ ण त थ द ध न ण
प फ ब भ म य र ल ळ व श ष स ह

Vowels (V)

अ आ इ ई उ ऊ ऋ ऐ ए ऐ ऌ ओ औ औ

Matras (M)

ा ि िी ु ू ृ े ै ॅ ो ौ ॉ

Vowel modifiers (D) ॅ ं ः

Halant (H) ्

Nukta (N) ृ

8.1 Indian Script Word Syntax

An Indian script word contains one or more syllables, the syntax for which is given in the following Backus-Naur Formalism (BNF).

Word ::= {Syllable} [Cons-Syllable]
Syllable ::= Cons-Vowel-Syllable | Vowel-Syllable
Vowel-Syllable ::= V [D]
Cons-Vowel-Syllable ::= [Cons-Syllable] Full-Cons [M] [D]
Cons-Syllables ::= [Pure-Cons] [Pure-Cons] Pure-Cons
Pure-Cons ::= Full-Cons H

Full-Cons ::= C [N]

Following conventions are used in the syntax given above:

- ::= defines a relation.
- { } enclose items which may be repeated one or more times.
- [] enclose items which may not be present.
- | separates items, out of which only one can be present.

8.2 Order within a Syllable

A worst case consonant syllable can contain:

C N H C N H C N H C N M D

A worst case vowel syllable can contain:

V D

Note:

- * Nukta (N) can come after only the consonants with which it can combine.
- * The above syntax ignores the vowels derived through Nukta (ऋ, लृ and लृ) and the Avagrah sign ऽ.
- * .The INV character not mentioned here is treated as consonant.
- * The Halant + Halant (Explicit Halant) and Halant + Nukta (Soft Halant) combinations have been ignored in the above discussion.

9. REFERENCES

- IS 10315, 7-bit coded character set for information interchange, which is equivalent to ISO 646.
- IS 12326 (1987), 7-bit and 8-bit coded character sets - Code extension techniques, which is equivalent to ISO 2022.
- IS 10401 (1982), 8-bit code for information interchange - Structure and rules for implementation, which is equivalent to ISO 4873.
- ISO 2375, Procedure for registration of escape sequences.

ANNEX - A

INDIAN SCRIPT ALPHABET CORRESPONDENCE

Following mnemonics are used for Indian scripts:

DEV: Devanagari PNJ: Punjabi GJR: Gujarati ORI: Oriya BNG: Bengali

ASM: Assamese TLG: Telugu KND: Kannada MLM: Malayalam

TML: Tamil RMN: Roman

Roman script transliteration scheme is explained in Annex F.

	RMN	DEV	PNJ	GJR	ORI	BNG	ASM	TLG	KND	MLM	TML
ँ	m̄	ँ		ँ	ँ	ँ	ँ				
ं	m̄	ं	ੰ	ं	ं	ং	ং	ం	ం	ಂ	
ः	h̄	ः		ः	ଃ	ঃ	ঃ	ః	ః	ః	ஃ
अ	a	अ	ਅ	अ	ଅ	অ	অ	అ	అ	അ	அ
आ	ā	आ	ਆ	आ	ଆ	আ	আ	ఆ	ఆ	ആ	ஆ
इ	i	इ	ਇ	ઇ	ଇ	ই	ই	ఇ	ಇ	ഇ	இ
ई	ī	ई	ਈ	ઈ	ଈ	ঈ	ঈ	ఀ	ఀ	ഈ	ஈ
उ	u	उ	ਉ	ઉ	ଉ	উ	উ	ఉ	ఉ	ഉ	உ
ऊ	ū	ऊ	ਊ	ઊ	ଊ	ঊ	ঊ	ఁ	ఁ	ఁ	ఁ
ऋ	r̄	ऋ	ਰਿ	ऋ	ଠ	ঠ	ঠ	ఠ	ఠ	ఠ	
ऐ	e							ఎ	ఎ	എ	ஏ
ए	ē	ए	ਏ	એ	ଏ	এ	এ	ఎ	ఎ	ഈ	ஏ
ऐ	ai	ऐ	ਐ	ਐ	ଐ	ঐ	ঐ	ఐ	ఐ	ഐ	ஐ
ँ	ê	ँ									
ओ	o							ఒ	ఒ	ഒ	ஒ
ओ	ō	ओ	ਓ	ਓ	ଓ	ও	ও	ఓ	ఓ	ഓ	ஓ
औ	au	औ	ਔ	ਔ	ଔ	ঔ	ঔ	ఔ	ఔ	ഔ	ஔ
आँ	ô	आँ									
क	ka	क	ਕ	ક	କ	ক	ক	క	క	ക	க
क़	ka	क़	ਕ								
ख	kha	ख	ਖ	ખ	ଖ	খ	খ	ఖ	ఖ	ഖ	க

ख	<u>k</u> ha	ख	ਖ								
ग	ga	ग	ਗ	ੴ	ਗ	ਗ	ਗ	ਗ	ਗ	ਗ	
ग़	<u>gh</u> a	ग़	ਗ਼								
घ	gha	घ	ਘ	ੲ	ਘ	ਘ	ਘ	ਘ	ਘ	ਘ	ਘ
ङ	<u>ṅ</u> a	ङ	ਙ	ੳ	ਙ	ਙ	ਙ	ਙ	ਙ	ਙ	ਙ
च	ca	च	ਚ	ੴ	ਚ	ਚ	ਚ	ਚ	ਚ	ਚ	ਚ
छ	cha	छ	ਛ	ੲ	ਛ	ਛ	ਛ	ਛ	ਛ	ਛ	ਛ
ज	ja	ज	ਜ	ੴ	ਜ	ਜ	ਜ	ਜ	ਜ	ਜ	ਜ
ज़	za	ज़	ਜ਼								
झ	jha	झ	ਝ	ੲ	ਝ	ਝ	ਝ	ਝ	ਝ	ਝ	ਝ
ञ	<u>ṅ</u> a	ञ	ਞ	ੴ	ਞ	ਞ	ਞ	ਞ	ਞ	ਞ	ਞ
ट	<u>ṭ</u> a	ट	ਟ	ੲ	ਟ	ਟ	ਟ	ਟ	ਟ	ਟ	ਟ
ठ	<u>ṭ</u> ha	ठ	ਠ	ੳ	ਠ	ਠ	ਠ	ਠ	ਠ	ਠ	ਠ
ड	<u>ḍ</u> a	ड	ਡ	ੳ	ਡ	ਡ	ਡ	ਡ	ਡ	ਡ	ਡ
ड़	<u>ḍ</u> ha	ड़	ੳ								
ढ	<u>ḍ</u> ha	ढ	ਢ	ੳ	ਢ	ਢ	ਢ	ਢ	ਢ	ਢ	ਢ
ढ़	<u>ḍ</u> ha	ढ़	ਢ								
ण	<u>ṇ</u> a	ण	ਣ	ੲ	ਣ	ਣ	ਣ	ਣ	ਣ	ਣ	ਣ
त	ta	त	ਤ	ੴ	ਤ	ਤ	ਤ	ਤ	ਤ	ਤ	ਤ
थ	tha	थ	ਥ	ੲ	ਥ	ਥ	ਥ	ਥ	ਥ	ਥ	ਥ
द	da	द	ਦ	ੳ	ਦ	ਦ	ਦ	ਦ	ਦ	ਦ	ਦ
ध	dha	ध	ਧ	ੲ	ਧ	ਧ	ਧ	ਧ	ਧ	ਧ	ਧ
न	na	न	ਨ	ੴ	ਨ	ਨ	ਨ	ਨ	ਨ	ਨ	ਨ
न̄	<u>n̄</u> a										
प	pa	प	ਪ	ੲ	ਪ	ਪ	ਪ	ਪ	ਪ	ਪ	ਪ
फ	pha	फ	ਫ	ੳ	ਫ	ਫ	ਫ	ਫ	ਫ	ਫ	ਫ
फ़	fa	फ़	ਫ਼								
ब	ba	ब	ਬ	ੳ	ਬ	ਬ	ਬ	ਬ	ਬ	ਬ	ਬ
भ	bha	भ	ਭ	ੳ	ਭ	ਭ	ਭ	ਭ	ਭ	ਭ	ਭ
म	ma	म	ਮ	ੲ	ਮ	ਮ	ਮ	ਮ	ਮ	ਮ	ਮ
य	ya	य	ਯ	ੲ	ਯ	ਯ	ਯ	ਯ	ਯ	ਯ	ਯ
य़	<u>ȳ</u> a										
र	ra	र	ਰ	ੲ	ਰ	ਰ	ਰ	ਰ	ਰ	ਰ	ਰ
र̄	<u>r̄</u> a	र̄									
ल	la	ल	ਲ	ੲ	ਲ	ਲ	ਲ	ਲ	ਲ	ਲ	ਲ
ळ	<u>ḷ</u> a	ळ									
ळ	<u>ḷ</u> a										
व	va	व	ਵ	ੲ	ਵ	ਵ	ਵ	ਵ	ਵ	ਵ	ਵ
श	śa	श	ਸ਼	ੳ	ਸ਼	ਸ਼	ਸ਼	ਸ਼	ਸ਼	ਸ਼	ਸ਼
ष	ṣa	ष									
स	sa	स	ਸ	ੲ	ਸ	ਸ	ਸ	ਸ	ਸ	ਸ	ਸ
ह	ha	ह	ਹ	ੳ	ਹ	ਹ	ਹ	ਹ	ਹ	ਹ	ਹ

ा	ā	ा	ा	ा	ा	ा	ा	ा	ा	ा	ा	ा
ि	i	ि	ि	ि	ि	ि	ि	ि	ि	ि	ि	ि
ी	ī	ी	ी	ी	ी	ी	ी	ी	ी	ी	ी	ी
ु	u	ु	ु	ु	ु	ु	ु	ु	ु	ु	ु	ु
ू	ū	ू	ू	ू	ू	ू	ू	ू	ू	ू	ू	ू
ृ	r̄	ृ		ृ	ृ	ृ	ृ	ृ	ृ	ृ	ृ	ृ
े	e											
े	ē	े	े	े	े	े	े	े	े	े	े	े
ै	ai	ै	ै	ै	ै	ै	ै	ै	ै	ै	ै	ै
ँ	ē	ँ										
ौ	o											
ो	ō	ो	ो	ो	ो	ो	ो	ो	ो	ो	ो	ो
ौ	au	ौ	ौ	ौ	ौ	ौ	ौ	ौ	ौ	ौ	ौ	ौ
ॉ	ô	ॉ										
्		्	्	्	्	्	्	्	्	्	्	्
र्		र्	र्		र्	र्	र्					
।	.	।	।	।	।	।	।

Note: र् is used in Devanagari for representing the half र from = as in वाव्या.

Nukta consonants shown in Devanagari are used only in Hindi.

ANNEX-B PC-ISCII CODE

	Hex	8	9	A	B	C	D	E	F
Hex	Dec.	128	144	160	176	192	208	224	240
0	0	ँ	औ	ण				ATR	EXT
1	1	ं	ऑ	त				ल	े
2	2	ः	क	थ				ळ	े
3	3	अ	ख	द				ळ	ै
4	4	आ	ग	ध				व	ँ
5	5	इ	घ	न				श	ो

6	6	ई	ड	न	ष	ो
7	7	उ	च	प	स	ौ
8	8	ऊ	छ	फ	ह	ॉ
9	9	ऋ	ज	ब	INV	्
A	10	ऐ	झ	भ	ा	ं
B	11	ए	ञ	म	ि	।
C	12	ऐ	ट	य	ी	।
D	13	ँ	ठ	य़	ु	
E	14	ओ	ड	र	ू	
F	15	ओ	ढ	र	ृ	

The PC-ISCII code is the version of ISCII code defined by Centre for Development of Advanced Computing (CDAC), Pune, for compatibility with IBM-PC 8-bit character set. IBM-PC does not follow the ISO 8-bit code recommendation. It uses line-drawing character set located between B0 hex and DF hex. Since these line-drawing characters have to co-exist along with ASCII and Indian scripts, the PC-ISCII code is designed to avoid clash with them. This has been possible through a bifurcation of the ISCII character set into two halves.

The Indian script numerals defined at the end of ISCII code table are not included in the PC-ISCII code set. With PC-ISCII only the ASCII numerals should be used. These numerals themselves can be rendered in shapes of numerals in a particular script through an appropriate Attribute (ATR) character.

Although the characters are at different locations in the PC-ISCII code, their sequence remains identical to that in the ISCII code. This allows the PC-ISCII code to be functionally identical to the ISCII code, enabling the same sorting sequence.

The positions occupied by the ATR and EXT codes were left undefined in the beginning, as some IBM-PC compatibles did not allow the corresponding characters to be typed in through the keyboard. When this problem was overcome the PC-ISCII code was already in wide use, and could not be changed. These positions could not be allotted to some new characters, as the sorting order would have got affected. ATR and EXT codes (on which sorting is not defined) were therefore suitable to fill in these two positions.

The five empty character positions towards the end of the code, are reserved in ISCII, but are needed in other script codes (like Perso-Arabic code).

ANNEX-C

ENGLISH-ALPHABET ISCII CODE: EA-ISCII

EA-ISCII is meant for those computers and packages which do not allow use of 8-bit codes, or ISO compatible 7-bit codes. Here the Indian script characters have to be defined within the ASCII character set. By defining the Indian script alphabet in place of only the 52 upper and lower case English alphabet, one can ensure that the Indian scripts would be usable, wherever English alphabet can be used.

Since all the ISCII characters cannot be accommodated directly, the Nukta character is used to derive some of the lesser used ones. Only the vowel अ followed by the corresponding Matra.

Since the same codes are now used to represent both the English and Indian script alphabet, some way of distinguishing between both is needed. A small "x" (equivalent to Nukta ँ) character at the beginning of a word indicates that the word is an Indian script one. Very few words start with "x" in English. If necessary they can be written by doubling the "x" at the start. At stand-alone "x" continues to be displayed normally. Similarly "-x" after an English word gets displayed normally.

This scheme will work with all English software allowing both upper and lower cases, but will not work with software which allows usage of only one of the cases. Apart from having the same equivalent character set as ISCII, EA-ISCII also gives the same sorting order. The EA-ISCII codes can be generated by the Inscript keyboard. It is possible to even generate the Nukta automatically before the beginning of an Indian script word.

EA-ISCII has provision for Attribute (ATR) character, but does not have the Extension (EXT) character. Absence of EXT character prevents use of Vedic characters along with EA-ISCII.

EA-ISCII does not cater to separate script numerals, as included in the ISCII code. However the

ATR character can allow rendition of the ASCII numerals in the selected script form.

EA-ISCII CODE CHART: The English upper and lowercase alphabet are interpreted as the corresponding Indian script character shown in the middle of a column, when an "x" is present at the beginning of the word. The characters shown towards the right of a column are obtained by appending the Nukta code, to the corresponding Indian script character shown in the middle of a column. Vowels other than अ, obtained appending the corresponding Matra to अ.

	Hex	2	3	4	5	6	7
Hex	Dec	32	48	64	80	96	112

0	0	SP	0	@	P ण		p ळ
1	1	!	1	A ँ	Q त	a य य	ौ े
2	2	"	2	B ं ः	R थ	b र	r े
3	3	#	3	C अ	S द	c र	s ै ॅ
4	4	\$	4	D क ऋ	T ध	d ल	t ो
5	5	%	5	E ख ख़	U न न्न	e ळ	u ो
6	6	&	6	F ग ग़	V प	f व	v ौ ॉ
7	7	'	7	G घ ङ	W फ	g श	w ळ
8	8	(8	H च	X ब	h ष	x ळ
9	9)	9	I छ	Y भ	i स	y ळ ऽ
A	10	*	:	J ज ञ	Z म	j ह INV	z ATR
B	11	+	;	K झ ञ	[k ा	{
C	12	,	<	L ट	/	l ि ि	
D	13	-	=	M ठ]	m ी ी	}
E	14	.	>	N ड ङ	^	n ँ	~
F	15	/	?	O ढ ढ		o ू	DEL

ANNEX-D

INSCRIPT KEYBOARD

The Inscript (Indian Script) keyboard overlay was standardized by DOE in 1986. ("Report of the Committee for Standardization of Keyboard Layout for Indian Script Based Computers", Electronics-Information & Planning Journal, Vol. 14, No. 1 October 1986).

A revision was done in 1988 by a DOE committee, when it was decided to compact the ISCII code by deriving some characters using a separate Nukta character. This required substitution of the Nukta character in place of the earlier "Transform" key. From frequency considerations it became necessary to mutually adjust the positions of ऐ, औ, ऐँ, औँ vowels, along with their Matras.

The Inscript overlay can be used on any QWERTY keyboard. The Indian script legends should be shown in the right-hand side of a key, as the left hand side has the English legends. The Inscript overlay gets selected when Caps-Lock is active, otherwise normal lower case English overlay gets selected. It is possible to use ALT+SPACE key to toggle the Caps-Lock functionality between this new one, and the normal one (where capital English letters get selected).

Temporary selection of the other overlay can be achieved by pressing the key along with the RIGHT ALT key (In IBM Enhanced keyboard), or the SYS-REQ key (In PC-AT 88-key keyboard). This can be very convenient for embedding a single character from the other overlay.

The Inscript overlay contains characters required for all the Indian scripts, as defined by the ISCII character set. The Indian script alphabet has a logical structure, derived from the phonetic properties. The Inscript overlay mirrors this logical structure. The overlay has also been optimized from phonetic/frequency considerations. It is divided into two parts: the vowel pad on the left hand side, and the consonant pad on the right hand side.

Within the vowel pad the vowels are given in the shift positions of the corresponding Matras. All the five main short vowels are given in the home row while their longer counterparts are located on the corresponding keys just above them. Since the vowel अ does not have a corresponding Matra, the अ vowel-omission sign, Halant, is given in the unshifted position. Halant is used for forming conjuncts, when it is typed in between consonants.

Alternate hand action gets used in typing of a conjunct; as Halant is typed from the left pad, while most of the consonants are typed from the right pad. Similarly alternate hand action occurs while typing a Matra after most of the consonants. This considerably speeds up typing of a syllable.

In the consonant pad all the primary characters of the 5 Vargs are included in the home row. The aspirated consonants are kept in the shift positions of their unaspirated counterparts. The non-nasal consonants of each Varg are contained in a part of vertically adjacent keys.

The main nasal consonants of the Vargs are contained in the bottom row of the left pad, along with the related Anuswar and Chandrabindu. The other non-Varg consonants are kept in the remaining positions of the right pad, according to their logical relations, and usage frequencies.

All the characters needed for touch typing are contained in the bottom 3 rows. The top row contains some conjuncts meant for ease in sight typing. The conjunct character keys actually send out the corresponding basic characters.

Due to the phonetic/alphabetic nature of the keyboard, a person who knows typing in one Indian script can type in any other Indian script. The logical structure allows ease in learning, while the frequency considerations allow speed in touch typing. The keyboard remains optimal both from touch-typing and sight-typing points of view, in all Indian scripts.

Note: The keyboard charts for each script as provided in the original document are not yet incorporated here. Please refer to Inscript keyboard tutorials of the LEAP software for the same.

ANNEX-E

ATTRIBUTE CODES

An ASCII character which follows the ATR character indicates a new font Attribute which is appli-

cable for the subsequent characters till the end of the row, or till another attribute code is encountered. The ASCII character, following the ATR character, can indicate 94 different attributes. Out of these the first 31 attributes are reserved for display attributes, while the rest of 63 attributes indicate selection of a font for a new script.

Hex	2	3	4	5	6	7
0		BLD	DEF			
1		ITA	RMN			ARB
2		UL	DEV			PES
3		EXP	BNG			URD
4		HLT	TML			SND
5		OTL	TLG			KSM
6		SHD	ASM			PST
7		TOP	ORI			
8		LOW	KND			
9		DBL	MLM			
A			GJR			
B			PNJ			
C						
D						
E						
F						

<-ATR Codes><-----FONT Codes----->
<-----Normal----><-Reverse->

E-1 Display Attributes (21h to 3Fh)

An Attribute code indicates a new display attribute, which is effective till the end of a line, or till the same attribute. At the beginning of a line all the display attributes are supposed to be off; subsequent occurrence of a display attribute causes toggling of the attribute on the display. Different attributes can combine together to give composite attributes.

Basic Attributes are:

Highlight Bold Outline Shadow Italics Underline Expanded

(Refer to original document for representation of attributes correctly)

These can combine together to give different effects:

Highlight + Bold = ExtraBold
 HLT Outline BLD Outline
 HLT + BLD Outline
 HLT Shadow BLD Shadow
 HLT + BLD Shadow
 Outline + Shadow = Deep Shadow

Expanded characters are of Double width.

Double Height characters can be achieved by duplicating the word on two consecutive lines and then using the TOP attribute on the top line and LOW attribute on the bottom line. Since TOP and LOW attributes also work on toggle basis, it is possible to have a mixture of double height and single height characters within the same row.

It is possible to create variety of effects using all these attributes:

The DBL, Double size row attribute makes the whole row double-height and double-width. This can be used along with TOP and LOW attributes to get quadruple size characters.

E-2 Font Attributes (40h to 7Eh)

At the beginning of a row the default display script is assumed to be active. The font attributes cause selection of a new script till the end of a row, or till another font attribute is encountered.

E-2.1 DEF (Default Font)

The Default font attribute causes re-selection of the default display script.

E-2.2 RMN (Roman Font)

A Roman script corresponding to a particular non-English script, is rendered using any English alphabet along with suitable diacritic marks. The Roman script transliteration is useful for making legible a script not known by a person. A family of scripts (like the Indian scripts, or Perso-Arabic scripts) will have a common set of diacritic marks.

The RMN font attribute selects Roman script corresponding to the currently active script. The numerals after the RMN attribute will be shown as international numerals.

E-2.3 Indian Script Fonts (42h to 4Bh)

This selects a Brahmi based Indian script. The subsequent numerals will be shown in the forms corresponding to the script, if they exist, otherwise they will be shown in their international form.

E-2.4 Perso-Arabic Fonts (71h to 76h)

These scripts are written from right to left. In general codes from 71h to 7Eh are reserved for scripts written in the reverse direction. The Perso-Arabic family contains Arabic (ARB), Persian (PRS), Urdu (URD), Sindhi (SND), Kashmiri (KSM) and Pushto (PST). Amongst these, Urdu, Sindhi and Kashmiri belonging to the Indian sub-continent have considerable similarity.

The ASCII numerals will be shown in the Perso-Arabic form, after a Perso-Arabic font attribute.

ANNEX-F

ROMAN SCRIPT TRANSLITERATION

The National Library at Calcutta standardized the diacritic marks to be used for romanization of

Indian scripts, in 1988 ("The National Library Newsletter", June 1988)

As Northern scripts do not have short ऐ and औ, the long ए and ओ can also be rendered without diacritic mark as 'e' and 'o' respectively.

Unlike Sanskrit and Southern scripts, in the Northern scripts the implicit vowel "a" at end of a word is not pronounced, and thus should be left out in the transliteration. Example: अशोक = aśōk, रमण्=ramaṇ. This also applies for nasal conjuncts where a Varg consonant is preceded by a Nasalconsonant belonging to the same Varg, Example: बन्ध=bandh, कम्प=kamp. Words ending in other conjuncts, still retain the implicit "a" vowel. Example: पुत्र=putra, राष्ट्र=rāṣṭra, मिश्र=miśra.

VOWELS

अ	आ	इ	ई	उ	ऊ	ऋ	ॠ	ऌ	ॡ
a	ā	i	ī	u	ū	ṛ	ṝ	ḷ	ḹ
ऐ	ए	ऐ	ऐ	ओ	ओ	औ	औ		
e	ē	ai	ê	o	ō	au	ô		

CONSONANTS

The Five Vargs

क	ख	ग	घ	ङ	
ka	kha	ga	gha	ṅa	
च	छ	ज	झ	ञ	
ca	cha	ja	jha	ña	
ट	ठ	ड	ढ	ण	
ṭa	ṭha	ḍa	ḍha	ṇa	
त	थ	द	ध	न	न
ta	tha	da	dha	na	ṇa
प	फ	ब	भ	म	
pa	pha	ba	bha	ma	

Non-Vargs

य	य	र	र	ल	ळ
ya	yā	ra	rā	la	ḷa
ळ	व	श	ष	स	ह
ḷa	va	śa	ṣa	sa	ha

Nukta Consonants

क	ख	ग	ज	ड़	ढ़	फ
qa	<u>kh</u> a	<u>gh</u> a	za	ḍa	ḍha	fa

VOWEL MODIFIERS

Chandrabindu	ँ	ं
Visarg	ः	़
Anuswar		ं
before क varg		ṅa
before च varg		ṅā
before ट varg		ṅa
before त varg		na
before प varg		ma
otherwise		ma

Notes:

ऋ, ॠ, ॡ are used only in Sanskrit.

ऐ = short ए in Southern scripts

औ = short ओ in Southern scripts

एँ = new vowel in Devanagari, as in "bat"

औँ = new vowel in Devanagari, as in "ball"

न n = ன in Tamil

य ya = য in Bengali and Oriya, while य y = য়

र ra = Tamil (ர) Telugu (ఱ), & Malayalam (റ)

र् ra = र in Marathi

ळ la = used in Marathi

ळ la = Tamil (ள), Malayalam (ള), Telugu (ళ) & Kannada (ಳ)

ळ za = Tamil (ழ), Malayalam (ഴ)