



Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

Draft Policy Document For

INTERNATIONALIZED DOMAIN NAMES

Language: SANSKRIT



Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

RECORD OF CHANGES

***A - ADDED M - MODIFIED D - DELETED**

VERSION NUMBER	DATE	PAGES AFFECTED	A* M D	TITLE OR BRIEF DESCRIPTION	COMPLIANCE VERSION OF MAIN POLICY DOCUMENT
1.0	20/11/09	Whole Document	M	Language Specific Policy Document for SANSKRIT	1.5
1.1	22/11/2010	Page No 8, 14, 16	A, M	Restriction rule added, Variant modified, ccTLD added	1.6



Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF).....	4
1.1 Declaration of variables.....	4
1.2 ABNF Operators	4
1.3 The Vowel Sequence.....	4
1.4 Consonant Sequence.....	5
1.5 ABNF Applied to the SANSKRIT IDN.....	6
2. RESTRICTION RULES.....	7
3. EXAMPLES.....	9
5. NOMENCLATURAL DESCRIPTION TABLE OF SANSKRIT LANGUAGE TABLE	
.....	11
6. VARIANT TABLE.....	14
7. EXPERTISE/BODIES CONSULTED.....	15
8. PROPOSED ccTLD FOR SANSKRIT.....	16

1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF)

1.1 Declaration of variables

Dash	→ Hyphen -
Digit	→ Indo-Arabic digits [0-9]
C	→ Consonant
M	→ Matra
V	→ Vowel
D	→ Anusvara / Bindu
B	→ Chandrabindu / Anunasika
X	→ Visarga
Y	→ Avagraha
H	→ Halant / Virama

1.2 ABNF Operators

S. No.	Symbols	Functions
1	“/”	Alternative
2	“[]”	Optional
3	“*”	Variable Repetition
4	“()”	Sequence Group

In what follows the Vowel Sequence and the Consonant Sequence pertinent to Sanskrit (Laukik) are given. Since only Laukik Sanskrit is supported at present, the stress markers such as Swarita and Udatta will not be introduced.

1.3 The Vowel Sequence

A vowel sequence is made up of a single vowel. It may be followed but not necessarily (optionally) by an Anusvara (D), Chandrabindu (B) or a Visarga (X)

The number of D,B or X which can follow a V in Sanskrit (Laukik) may be restricted to one.

The possibility of a Visarga following a Chandrabindu or Anusvara is ruled out, since it is used only in Vedic and in Bangla script.

The vowel sequence in Sanskrit (Laukik) is therefore V [D | B |X]

Examples :

V	अ
V[D]	अं
V[B]	अँ
V[X]	अः

1.4 Consonant Sequence

A consonant sequence admits the following shapes:

1. A single consonant (C)

Example:

C	क
---	---

2. A consonant optionally followed by dependent vowel sign[M] or Anusvara[D] Chandrabindu[B] or visarga[X] or Halant [H]

C[M|D|B|X|H]

Example:

C[M]	कि
C[D]	कं
C[B]	कँ
C[X]	कः
C[H]	क् (Pure Consonant)

2.a. A CM sequence can be optionally followed by D, B or X
(CM)[D|B|X]

Example:

CM[D]	कीं
CM[B]	काँ
CM[X]	वीः

3. A sequence of consonants (up to 5) joined by Halant *4(CH)C

Example:

CHCHCHCHC त्स्न्य र+्+त+्+स+्+न+्+य

Subsets

3.a. The combination may be followed by M , D,B or X

Example:

CHC[M]	क्की	क्	क्	ी
CHC[D]	क्कं	क्	क्	ं
CHC[B]	क्कँ	क्	क्	ँ
CHC[X]	क्कः	क्	क्	ः

3.b. *4(CH)CM may be followed by a D, B or X

Example:

CHCM[D]	क्कीं	क्	क्	ी	ं
CHCM[B]	क्कीँ	क्	क्	ी	ँ
CHCM[X]	क्कीः	क्	क्	ी	ः

The final canonical structure of the consonant sequence in IDN can be defined in ABNF as:

*4(CH)C [H|D|B|X |M[D|B|X]]

as exemplified in the word:

कात्स्न्य

1.5 ABNF Applied to the SANSKRIT IDN

The formalism can be applied to create/validate IDN labels. So a valid IDN label can be defined as follows.

Vowel-sequence → V [D | B | X]

Consonant-sequence → *4(CH)C [H|D|B|X |M[D | B | X]]

Sequence → consonant-sequence[Y] | vowel-sequence[Y]

IDN-label → (sequence | digit) * ([dash] (sequence |digit))

2. RESTRICTION RULES

The ABNF is generic in nature and when applied to a specific language/script certain restriction rules apply. In other words, in a given language some of the Formalism structures do not necessarily apply. To take care of such cases restriction rules are set in place. These restrictions will help to fine-tune the ABNF. In the case of Sanskrit (Laukik) the following rules apply:

1. H | D|B|X |M cannot occur in the beginning of an IDN domain name

Example:

क्
क
ँक
ःक
िक

As can be seen they will result automatically in a “golu” marking an invalid character. This is an intrinsic property of the Indic syllable and is quasi automatically applied wherever supported by the OS.

2. H is not permitted after V, D,B, X, M, digit and dash

Example:

अ्
कं
कं
कं
कः
क्
1
-

3. Number of D,B or X permitted after consonant-sequence or vowel-sequence or M is restricted to one

Example:

कं
कः
कीं
कां
कीः
अं
अं



Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

अः

4. Number of M permitted after consonant-sequence is restricted to one

Example:

कीी

5. M is not permitted after V

Example:

ईा

6. The combination of Anusvara+Visarga as well as Visarga+Anusvara is not permissible

Example

कंः

कःं

7. A consonant syllable that is intended to end with Halant [H] can only be followed by Hyphen, digit or Avagraha [Y].



3.EXAMPLES

Following are the some examples of SANSKRIT IDN.

Combination	Example	Word With Combination
C	च	च
CH	म्	नरोत्तमम्
CM	सू	सूत
CD	णं	नारायणं
CB	जँ	राजँ
CX	लः	कालः
CMD	वीं	देवीं
CMB	भाँ	शुभाँल्लोकान्न
CMX	षेः	राजषेः
CHC	क्ष	कमलपत्राक्ष
CHCHC	न्द्र	पार्थिवेन्द्रस्य
CHCHCHC	ङ्क्ष्य	विनङ्क्ष्यन्ति
CHCHCHCHC	त्स्न्य	कात्स्न्य
V	उ	उवाच



Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

4. LANGUAGE TABLE: SANSKRIT¹

0900

Devanagari
Sanskrit

097F

	090	091	092	093	094	095	096	097
0	ॐ	ऐ	ठ	र	ी	ॐ	ॠ	ॐ
1	ॐ	ऑ	ड	र	ॐ	ॐ	ॐ	ॐ
2	ॐ	ओ	ढ	ल	ॐ	ॐ	ॐ	ॐ
3	ॐ	ओ	ण	ळ	ॐ	ॐ	ॐ	ॐ
4	ॐ	औ	त	ळ	ॐ	ॐ	ॐ	ॐ
5	अ	क	थ	व	ॐ	ॐ	ॐ	ॐ
6	आ	ख	द	श	ॐ	ॐ	ॐ	ॐ
7	इ	ग	घ	ष	ॐ	ॐ	ॐ	ॐ
8	ई	घ	न	स	ॐ	ॐ	ॐ	ॐ
9	उ	ड	न	ह	ॐ	ॐ	ॐ	ॐ
A	ऊ	च	प	ॐ	ॐ	ॐ	ॐ	ॐ
B	ॠ	छ	फ	ॐ	ॐ	ॐ	ॐ	ॐ
C	ॠ	ज	ब	ॐ	ॐ	ॐ	ॐ	ॐ
D	ॐ	झ	भ	ॐ	ॐ	ॐ	ॐ	ॐ
E	ॐ	ज	म	ॐ	ॐ	ॐ	ॐ	ॐ
F	ॐ	ट	य	ॐ	ॐ	ॐ	ॐ	ॐ

The Unicode Standard 5.1, Copyright © 1991-2008 Unicode, Inc. All rights reserved.

65

¹ Characters marked in yellow are not applicable to the language.

5. NOMENCLATURAL DESCRIPTION TABLE OF SANSKRIT LANGUAGE TABLE

Chandrabindu (B)		
0901	ँ	DEVANAGARI SIGN CANDRABINDU = anunasika
Anusvara (D)		
0902	ं	DEVANAGARI SIGN ANUSVARA = bindu
Visarga (X)		
0903	ः	DEVANAGARI SIGN VISARGA
Independent vowels (V)		
0905	अ	DEVANAGARI LETTER A
0906	आ	DEVANAGARI LETTER AA
0907	इ	DEVANAGARI LETTER I
0908	ई	DEVANAGARI LETTER II
0909	उ	DEVANAGARI LETTER U
090A	ऊ	DEVANAGARI LETTER UU
090B	ऋ	DEVANAGARI LETTER VOCALIC R
090C	ॠ	DEVANAGARI LETTER VOCALIC L
090F	ए	DEVANAGARI LETTER E
0910	ऐ	DEVANAGARI LETTER AI
0913	ओ	DEVANAGARI LETTER O
0914	औ	DEVANAGARI LETTER AU
0960	ॠ	DEVANAGARI LETTER VOCALIC RR
0961	ॡ	DEVANAGARI LETTER VOCALIC LL
Consonants (C)		
0915	क	DEVANAGARI LETTER KA
0916	ख	DEVANAGARI LETTER KHA
0917	ग	DEVANAGARI LETTER GA
0918	घ	DEVANAGARI LETTER GH
0919	ङ	DEVANAGARI LETTER NGA
091A	च	DEVANAGARI LETTER CA
091B	छ	DEVANAGARI LETTER CHA
091C	ज	DEVANAGARI LETTER JA
091D	झ	DEVANAGARI LETTER JHA
091E	ञ	DEVANAGARI LETTER NYA
091F	ट	DEVANAGARI LETTER TTA
0920	ठ	DEVANAGARI LETTER TTHA



Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

0921	ड	DEVANAGARI LETTER DDA
0922	ढ	DEVANAGARI LETTER DDHA
0923	ण	DEVANAGARI LETTER NNA
0924	त	DEVANAGARI LETTER TA
0925	थ	DEVANAGARI LETTER THA
0926	द	DEVANAGARI LETTER DA
0927	ध	DEVANAGARI LETTER DHA
0928	न	DEVANAGARI LETTER NA
092A	प	DEVANAGARI LETTER PA
092B	फ	DEVANAGARI LETTER PHA
092C	ब	DEVANAGARI LETTER BA
092D	भ	DEVANAGARI LETTER BHA
092E	म	DEVANAGARI LETTER MA
092F	य	DEVANAGARI LETTER YA
0930	र	DEVANAGARI LETTER RA
0932	ल	DEVANAGARI LETTER LA
0933	ळ	DEVANAGARI LETTER LLA
0935	व	DEVANAGARI LETTER VA
0936	श	DEVANAGARI LETTER SHA
0937	ष	DEVANAGARI LETTER SSA
0938	स	DEVANAGARI LETTER SA
0939	ह	DEVANAGARI LETTER HA
Dependent vowel signs (Matras) (M)		
093E	ा	DEVANAGARI VOWEL SIGN AA
093F	ि	DEVANAGARI VOWEL SIGN I • stands to the left of the consonant
0940	ी	DEVANAGARI VOWEL SIGN II
0941	ु	DEVANAGARI VOWEL SIGN U
0942	ू	DEVANAGARI VOWEL SIGN UU
0943	ृ	DEVANAGARI VOWEL SIGN VOCALIC R
0944	्र	DEVANAGARI VOWEL SIGN VOCALIC RR
0947	े	DEVANAGARI VOWEL SIGN E
0948	ै	DEVANAGARI VOWEL SIGN AI
094B	ो	DEVANAGARI VOWEL SIGN O
094C	ौ	DEVANAGARI VOWEL SIGN AU
0962	ॠ	DEVANAGARI VOWEL SIGN VOCALIC L
0963	ॡ	DEVANAGARI VOWEL SIGN VOCALIC LL
Halant (H)		
094D	्	DEVANAGARI SIGN VIRAMA = halant (the preferred



Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

		Hindi name) • suppresses inherent vowel
Avagraha (Y)		
093D	s	DEVANAGARI SIGN AVAGRAHA



Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

6. VARIANT TABLE

VARIANTS		
ऋ 090B	ऋ 0960	
ॠ 090C	ॠ 0961	
ॡ 0943	ॡ 0944	
ॢ 0962	ॢ 0963	
ॣ 0926+094D+0917	ॣ 0926+094D+0930	ॣ 0926+094D+0928
। 0926+094D+0927	। 0926+094D+0918	
॥ 0937+094D+091F	॥ 0937+094D+0920	
० 0936+094D+0935	० 0936+094D+0930++094D0935	
१ 0936+094D+0928	१ 0936+094D+0930+094D+0928	
२ 0936+094D+091A	२ 0936+094D+0930+094D+091A	
३ 0936+094D+0932	३ 0936+094D+0930+094D+0932	
४ 0924+094D+0924	४ 0924	
५ 0926+094D+0935	५ 0926+094D+092C	



Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

7. EXPERTISE/BODIES CONSULTED

At the Initial stages, the expertise of Dr John Smith, Reader in Sanskrit Cambridge University was utilised for this analysis. The layout for Laukik Sanskrit was further checked and ratified at PRSG's held in both Delhi and Pune.



Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

8. PROPOSED ccTLD FOR SANSKRIT

India (Bhārat) localized in Sanskrit - भारतम्

Note: You can send your feedbacks to ids-feedback@cdac.in