



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

Policy Document For

INTERNATIONALIZED DOMAIN NAMES

Language:
SANTALI
DEVANAGARI SCRIPT



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

Contents

1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF).....	3
1.1 Declaration of variables	3
1.2 ABNF Operators	3
1.3 The Vowel Sequence	3
1.4 Consonant Sequence	4
1.5 ABNF Applied to the SANTALI (DEVANAGARI SCRIPT) IDN	6
2. RESTRICTION RULES.....	7
3. EXAMPLES.....	10
4. LANGUAGE TABLE: SANTALI (DEVANAGARI)	12
5. NOMENCLATRURAL DESCRIPTION TABLE OF SANTALI(DEVANAGARI) LANGUAGE TABLE.....	13
7. EXPERTISE/BODIES CONSULTED.....	17
8. Country Code Top Level Domain (ccTLD) FOR SANTALI in DEVANAGARI script.....	18

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
B	→ Chandrabindu
X	→ Visarga
H	→ Halanta
N	→ Nukta

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 SANTALI in DEVANAGARI script are given.

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)



or a Nukta (N). The number of D, B, X or N which can follow a V in SANTALI in DEVANAGARI script should be restricted to one.

The vowel sequence in SANTALI IN DEVANAGARI SCRIPT is therefore
V [D | B | X | N]

Examples :

Vowel	V	अ
Vowel + anusvara	V[D]	अं
Vowel + Chandrabindu	V[B]	अँ
Vowel + Visarga	V[X]	अः
Vowel + Nukta	V[N]	अ़

Note : *The permissible nukta combinations are only अ़ ओ़*

1.4 Consonant Sequence

A consonant sequence admits the following shapes:

1. A single consonant (C)

Example:

C	क
C[N]	ङ

2. A consonant optionally followed by dependent vowel sign[M] or anusvara[D]
Chandrabindu[B] or visarga[X] or Halanta [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| X |N]
 (CM)[D|B|X |N]

Example:

CM[D]	कीं
CM[B]	काँ
CM[X]	वीः
CM[N]	का

Note : The permissible nukta combinations are only ाँ, ीँ

3. A CMN sequence can optionally be followed by Anuswara[D] or Chandrabindu [B]

CMN[D|B]

Example:

CMN[D]	कां
CMN[B]	काँ

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

Example:

CHCHC न्द्र न+्+द+्+र

Subsets

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

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

Example:

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



4.b. (CH)CM may be followed by a [D | B | X | N]

Example:

CHCM[D]	क्की	क्कीं
CHCM[B]	क्कीँ	क्कीँँ
CHCM[X]	क्कीः	क्कीःः
CHCM[N]	क्का	क्काः

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

$*2(C[N]H)C[N][H|D|B|X|M[D|B|X|N[D|B]]]$

1.5 ABNF Applied to the SANTALI (DEVANAGARI SCRIPT) IDN

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

Vowel-sequence $\rightarrow V [D | B | X | N]$

Consonant-sequence $\rightarrow *2(C[N]H)C[N][H|D|B|X|M[D|B|X|N[D|B]]]$

Sequence \rightarrow consonant-sequence | vowel-sequence

IDN-label \rightarrow (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 SANTALI in DEVANAGARI script 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/circle” marking an invalid character. This is an intrinsic property of the Indic syllable and is quasi automatically applied.

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

Example:

अ्
कं्
कँ्
कः्
क्ि
1्
-्

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

Example

कंं



कँ
कः
कीं
काँ
कीः
अं
अँ
अः

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

Example

कीी

5. M is not permitted after V

Example

ई

6. The combination of anusvara+Visarga, chandrabindu+anusvara, chandrabindu+visarga and vice-versa is not permissible

Example

कः
कँ
कँः

7. Nukta which can be generated out as an Akhand form or Consonant + Nukta, shall be treated as part of a variant table. For SANTALI IN DEVANAGARI SCRIPT nukta shall be allowed after the following:

ड (0921)

ढ (0922)

ा (093E)

ो (094B)

आ (0906)

ओ (0913)



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

Nukta shall not be allowed for any other consonant, vowel or matra.

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



3. EXAMPLES

Combination	Example	Word With Combination
C	क	कम
CN	इ	पेड़
CH	त्	आत्
CM	खा	खाप
CD	चं	चंगा
CB	हँ	हँक
CX	तः	प्रातः
CMD	हिं	हिंसी , हिंरखा
CMB	चाँ	चाँदो
CMX	दुः	दुःख
CMN	फां	फांद
CMND	कां	कांत
CMNB	टाँ	टाँडी
CHC	क्य , व्य	वाक्य , द्रव्य
CHCHC	न्त्र	यन्त्रकार
V	आ	आम , आक्
VD	अं	अंगरोप , अं
VB	आँ	आँक्
VN	आ , ओ	आलू , ओत्



सत्यमेव जयते

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



4. LANGUAGE TABLE: SANTALI (DEVANAGARI) ¹

0900

Devanagari

097F

	090	091	092	093	094	095	096	097
0	ँ	ऐ	ठ	र	ी	ॐ	ऋ	०
1	ँ	ऑ	ड	र	ु	ं	ऌ	॰
2	ं	ओ	ढ	ल	२	३	४	अँ
3	ः	ओ	ण	ळ	५	६	७	अं
4	ऐ	औ	त	ळ	८	९		आ
5	अ	क	थ	व	॰	ॱ	ॱ	औ
6	आ	ख	द	श	ॲ	ॳ	ॴ	अु
7	इ	ग	ध	ष	ॵ	ॶ	ॷ	अु
8	ई	घ	न	स	ॸ	ॹ	ॺ	रु
9	उ	ड	न	ह	ँ	ख	ॳ	ज़
A	ऊ	च	प	ं	ौ	ग	ॴ	य
B	ऋ	छ	फ	ी	ो	ज	ॵ	ग
C	ऌ	ज	ब	ॱ	ौ	ड	ॶ	ज़
D	ँ	झ	भ	ॳ	ॴ	ॵ	ॶ	ॷ
E	ऐ	ज	म	ा	ि	फ	ॴ	ड
F	ए	ट	य	ि	ौ	य	ॳ	ब

The Unicode Standard 11.0, Copyright © 1991-2018 Unicode, Inc. All rights reserved.

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



5. NOMENCLATURAL DESCRIPTION TABLE OF SANTALI(DEVANAGARI) LANGUAGE TABLE

Anusvara (D)		
0901	ँ	DEVANAGARI SIGN CANDRABINDU = anunasika
Chandrabinidu(B)		
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
090F	ए	DEVANAGARI LETTER E
0910	ऐ	DEVANAGARI LETTER AI
0911	ऑ	DEVANAGARI LETTER CANDRA O
0913	ओ	DEVANAGARI LETTER O
0914	औ	DEVANAGARI LETTER AU
Consonants (C)		
0915	क	DEVANAGARI LETTER KA
0916	ख	DEVANAGARI LETTER KHA
0917	ग	DEVANAGARI LETTER GA
0918	घ	DEVANAGARI LETTER GHA
0919	ङ	DEVANAGARI LETTER NGA
091A	च	DEVANAGARI LETTER CA
091B	छ	DEVANAGARI LETTER CHA
091C	ज	DEVANAGARI LETTER JA



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

091D	झ	DEVANAGARI LETTER JHA
091E	ञ	DEVANAGARI LETTER NYA
091F	ट	DEVANAGARI LETTER TTA
0920	ठ	DEVANAGARI LETTER TTHA
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
0935	व	DEVANAGARI LETTER VA
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



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

0947	े	DEVANAGARI VOWEL SIGN E
0948	ै	DEVANAGARI VOWEL SIGN AI
094B	ो	DEVANAGARI VOWEL SIGN O
094C	ौ	DEVANAGARI VOWEL SIGN AU
Halant (H)		
094D	्	DEVANAGARI SIGN VIRAMA = halant (the preferred SANTALI IN DEVANAGARI SCRIPT name) • suppresses inherent vowel
Nukta (N)		
093C	़	DEVANAGARI SIGN NUKTA



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

6. VARIANT TABLE FOR SANTALI IN DEVANAGARI SCRIPT

झ 0926+094D+0917	झ 0926+094D+0930	झ 0926+094D+0928
झ 0926+094D+0927	झ 0926+094D+0918	
झ 0926+094D+0935	झ 0926+094D+092C	



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

7. EXPERTISE/BODIES CONSULTED

Experts from Dept. of tribal & regional languages, Ranchi University, Ranchi.



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

8. Country Code Top Level Domain (ccTLD) FOR SANTALI in DEVANAGARI script

India (Bhārat) localized in Santali in Devanagari Script - भारोत (xn--h2brj9c8c)