सत्यमेव जयते

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

Draft Policy Document
For

INTERNATIONALIZED DOMAIN NAMES

Language: HINDI

सी डैक
CDAC
Gist

nixi

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 HINDI	1.5
1.1	22/11/2010	Page No. 10, 17, 19	A, M	Restriction rule added, Variant Modified, ccTLD added	1.6
1.2	01/03/2013	All	M D	Modified character repertoire as per IDNA 2008	1.8

Table of Contents

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 Sequence 7
1.6 ABNF Applied to the HINDI IDN 7
2. RESTRICTION RULES 10
3. EXAMPLES 12
4. LANGUAGE TABLE: HINDI 13
5. NOMENCLATURAL DESCRIPTION TABLE OF HINDI LANGUAGE TABLE ... 14
6. VARIANT TABLE FOR HINDI 18
7. EXPERTISE/BODIES CONSULTED 19
8. PROPOSED ccTLD FOR HINDI 20

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

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 Hindi 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) The number of D, B or X which can follow a V in Hindi 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 Hindi is therefore V [D |B | X]
Examples :

Vowel	V	अ
Vowel + Anusvara	V[D]	अं
Vowel + Chandrabindu	V[B]	अँ
Vowel + Visarga	V[X]	अः

1.4 Consonant Sequence

A consonant sequence admits the following shapes:
1. A single consonant (C)
(The consonant in Hindi shall be treated as co-terminus with the Consonant along with the Nukta sign wherever such a case is pertinent.)
Example:

C	क
C[N]	क़

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

C[M|D|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 4) joined by Halant *3(CH)C
Example:

CHCHCHC न्क्रय न+्+क+्+र+्+य

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

CHC[M] क्की क ् क ी

CHC[D] क्कं क ् क ं

CHC[B] क्कँ क ् क ँ

CHC[X] क्कः क ् क ः

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

CHCM[D] क्कीं क ् क ी ं

CHCM[B] क्कीँ क ् क ी ँ

CHCM[X] क्कीः क ् क ी ः

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

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

It should be noted that in the case of Sanskrit the possible sequence could be:
*4(CH)C [H|D|B|X|M[D|B|X]]
as exemplified in the word:
कात्स्न्र्य

Although this is possible, it is not used in neo-Prakrits unless a potential client chooses it as a possible IDN.

1.5 Sequence

1. A sequence can be made up by Consonant-sequence or Vowel-sequence.

   1.a A Consonant-sequence can optionally be followed by Avagraha[Y].
   1.b A Vowel-sequence can optionally be followed by Avagraha[Y].

1.6 ABNF Applied to the HINDI 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 → *3(C[N]H)C[N] [H|D|B|X|M[D|B|X]]
Sequence → consonant-sequence[Y] | vowel-sequence[Y]
IDN-label → (sequence | digit) * ([dash] (sequence |digit))

Additional Examples putting more light on ABNF

Below are some of the examples which will help a casual reader understand some of the rules ABNF puts in place. These are just given for reference purposes and are not meant to be comprehensive.

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

Example:

◌्क

◌ं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.

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

Example:

अ्

कं◌्

कँ◌्

कः◌्

कि्

1◌्

-◌्

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

   Example

कं◌ं

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
   ईा

6. The combination of Anusvara+Visarga, chandrabindu+Anusvara, chandrabindu+visarga and vice-versa is not permissible
   Example
   कं◌ः

   कँ◌ं

   कँ◌ः

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 Hindi the following rules apply:

1. Nukta shall be allowed after the following characters:

   क (0915)

   ख (0916)

   ग (0917)

   ज (091C)

   ड (0921)

   ढ (0922)

   फ (092B)

2. A consonant sequence that is intended to end with Halant [H] can only be followed by Hyphen, digit or Avagraha [Y]. Thus following combinations are permissible.

   क्-

   क्1

   क्ऽ

3. Consecutive hyphens will not be permitted in a domain name.

4. The number of identical consonants joined by a Halant within a label shall not exceed two. Thus त (ta+halant+ta) is permitted but not त्त (ta+halant+ta+halant+ta).

5. Wherever a variant is present in a given label, the variants shall be in a relationship of transitivity but the generation of the variant table shall be limited only to the relationship existing between the two variants. Thus given a variant त and त्त, the number of variants in label such as किताब shall be कित्ताब. कित्त्ताब generated by adding an extra त् to त्त shall not be permitted. This ensures that over generativity does not take place.

6. A label containing not more than three "akshara", which have got variants shall be permitted. As an example let us consider a, b, c and d as four aksharas in a given label having a', b', c' and d' as variants in which case such a label will be disallowed. (E.g. of disallowed label - abcd, acdb, cdaba and so on)

3. EXAMPLES

Combination	Example	Word With Combination
C	क	कर
CN(Required For Hindi)	ड़	पेड़
CH	म्	भारतम्
CM	ता	ताल
CD	गं	गंगा
CB	हँ	हँस
CX	तः	अतः
CMD	हिं	हिंदी
CMB	चाँ	चाँद
CMX	दुः	दुःख
CHC	व्य	काव्य
CHCHC	न्त्र	यन्त्रकार
V	आ	आज
VD	अं	अंग
VB	आँ	आँख

4. LANGUAGE TABLE: HINDI[1]

0900 **Devanagari** **097F**

	090	091	092	093	094	095	096	097
0		ऐ 0910	ठ 0920	र 0930	ो 0940	ॐ 0950	ॠ 0960	० 0970
1	ँ 0901	ऑ 0911	ड 0921	ऱ 0931	ु 0941	॑ 0951	ॡ 0961	ॱ 0971
2	ं 0902	ओ 0912	ढ 0922	ल 0932	ू 0942	॒ 0952	ॢ 0962	ॲ 0972
3	ः 0903	ओ 0913	ण 0923	ळ 0933	ृ 0943	॓ 0953	ॣ 0963	
4	अ 0904	औ 0914	त 0924	ऴ 0934	ॄ 0944	॔ 0954	। 0964	
5	अ 0905	क 0915	थ 0925	व 0935	ॅ 0945		॥ 0965	
6	आ 0906	ख 0916	द 0926	श 0936	ॆ 0946		० 0966	
7	इ 0907	ग 0917	ध 0927	ष 0937	े 0947		१ 0967	
8	ई 0908	घ 0918	न 0928	स 0938	ै 0948	क़ 0958	२ 0968	
9	उ 0909	ङ 0919	ऩ 0929	ह 0939	ॉ 0949	ख़ 0959	३ 0969	
A	ऊ 090A	च 091A	प 092A		ॊ 094A	ग़ 095A	४ 096A	
B	ऋ 090B	छ 091B	फ 092B		ो 094B	ज़ 095B	५ 096B	ग़ 097B
C	ऌ 090C	ज 091C	ब 092C	़ 093C	ौ 094C	ड़ 095C	६ 096C	ज़ 097C
D	ऍ 090D	झ 091D	भ 092D	ऽ 093D	् 094D	ढ़ 095D	७ 096D	ड़ 097D
E	ऎ 090E	ञ 091E	म 092E	ा 093E		फ़ 095E	८ 096E	ढ़ 097E
F	ए 090F	ट 091F	य 092F	ि 093F		य़ 095F	९ 096F	ब़ 097F

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 HINDI 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
090D	ऍ	DEVANAGARI LETTER CANDRA E
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
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
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
0945	ॅ	DEVANAGARI VOWEL SIGN CANDRA E = candra
0947	े	DEVANAGARI VOWEL SIGN E
0948	ै	DEVANAGARI VOWEL SIGN AI
0949	ॉ	DEVANAGARI VOWEL SIGN CANDRA O
094B	ो	DEVANAGARI VOWEL SIGN O
094C	ौ	DEVANAGARI VOWEL SIGN AU

Halant (H)		
094D	्	DEVANAGARI SIGN VIRAMA = halant (the preferred Hindi name) • suppresses inherent vowel
Nukta (N)		
093C	़	DEVANAGARI SIGN NUKTA
Avagraha (Y)		
093D	ऽ	DEVANAGARI SIGN AVAGRAHA

6. VARIANT TABLE FOR HINDI

VARIANTS		
द्र 0926+094D+O917	द्र 0926+094D+0930	द्र 0926+094D+0928
द्ध 0926+094D+0927	द्ध 0926+094D+0918	
ष्ट 0937+094D+091F	ष्ठ 0937+094D+0920	
श्य 0936+094D+0935	श्व 0936+094D+0930+094D+0935	
श्न 0936+094D+0928	श्न 0936+094D+0930+094D+0928	
श्च 0936+094D+091A	श्च 0936+094D+0930+094D+091A	
श्ल 0936+094D+0932	श्ल 0936+094D+0930+094D+0932	
त्त 0924+094D+0924	त 0924	
द्व 0926+094D+0935	द्व 0926+094D+092C	
ँ 0901	ँ ं 0945+0902	

Note : श्ल and श्ल normally not used in HINDI but are introduced since some browsers may display these shapes.

7. EXPERTISE/BODIES CONSULTED

Expertise provided by experts of Central Hindi Directorate & the late Shri R.K.Joshi at PRSG for IDN held at Delhi in 2006.

8. PROPOSED ccTLD FOR HINDI

India (Bhārat) localized in Hindi - भारत

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