



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

Draft Policy Document For INTERNATIONALIZED DOMAIN NAMES

Language: ASSAMESE



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	19/11/09	Whole Document	M	Language Specific Policy Document for ASSAMESE	1.5
1.1	22/11/10	Page No 11, 21	A	Restriction rule and ccTLD added	1.6
1.2	05/04/11	Whole document	A	Vowel-sequence modified.	1.7



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

1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF).....	4
1.1 Naming of Variables:.....	4
1.2 ABNF Operators	4
1.3 The Vowel Sequence.....	5
1.4 Consonant Sequence *	6
1.5 ABNF Applied to the Assamese IDN.....	9
2. RESTRICTION RULES.....	10
3. EXAMPLES.....	13
4. LANGUAGE TABLE: ASSAMESE.....	15
5. NOMENCLATURAL DESCRIPTION TABLE OF ASSAMESE LANGUAGE TABLE.....	16
6. VARIANT TABLE.....	20
7. EXPERTS/BODIES CONSULTED.....	21
8. PROPOSED ccTLD FOR ASSAMESE	22

1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF)

1.1 Naming of Variables:

Dash	→ Hyphen -
Digit	→ Indo-Arabic digits [0-9]
C	→ Consonant
V	→ Vowel
M	→ Matra
D	→ Anusvara
B	→ Chandrabindu
X	→ Visarga
H	→ Virama/Hasant/Halant
N	→ Nukta
Y	→ Avagraha
Z	→ Khanda Ta

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 Assamese are given. To facilitate use , equivalents in Devanagari are provided.

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 Bangla may not be restricted to one. The possibility of a Visarga or Anusvara following a Chandrabindu exist in Bangla. Vowel can optionally be followed by a combination of Halant/ Hasant/ Virama [H], Consonant [C] to form a Ya-phalaa. “Ya-phalaa is a presentation form of U+09AF য bengali letter ya. Represented by the sequence < U+09CD ঞ bengali sign virama, U+09AF য bengali letter ya>, ya-phalaa has a special form. When combined with U+09BE া bengali vowel sign aa, it is used for transcribing [æ] as in the “a” in the English word “bat.””¹

The vowel sequence in Bangla is therefore
 V [D | B | X | BD | BX | HCM [D|B|X|BD|BX]]

A Vowel-sequence admits the following shapes:

1. A Single Vowel

Examples :

V	অ	অ
---	---	---

2. A Vowel can optionally be followed by Anusvara[D] or Chandrabindu[B] or visarga[X] or Chandrabindu+Anusvara[BD] or Chandrabindu+Visarga[BX] or combination of Halant / Hasant / Virama [H], Consonant[C], Matra [M]

Examples :

V[D]	অং	অঁ
V[B]	অঁ	অঁ
V[X]	অঃ	অঃ
V[BD]	অঁং	অঁঁ

* Combinations where Anusvara / Visarga followed by Chandrabindu, though permitted in the language, currently are not rendered because of the underlying shaping engine and the dialog needs to be opened with OS vendors so that problem can be solved in next implementation.

1 Refer chapter 9: <http://www.unicode.org/versions/Unicode5.0.0/ch09.pdf>



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

V[BX] अँः अँः
V[HCM] अग (अ + ः + य + ः)

3. A VHCM sequence can optionally be followed by Anusvara[D] or Chandrabindu[B] or visarga[X] or Chandrabindu+Anusvara[BD] or Chandrabindu+Visarga[BX].

Examples:

VHCM[D] अगं
VHCM[B] अगँ
VHCM[X] अगः
VHCM[BD] अगँं
VHCM[BX] अगँः

1.4 Consonant Sequence *

A consonant sequence admits the following shapes:

1. A single consonant (C)

Example:

C क क

2. A consonant optionally followed by dependent vowel sign[M] or Anusvara[D] or Chandrabindu[B] or Visarga[X] or Virama/Hasant/Halant [H] or Chandrabindu+Anusvara [BD] or Chandrabindu+Visarga [BX]

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

Example:

C[M] कि कि
C[D] कं कं



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

C[B]	कँ	कँ
C[X]	कः	कः
C[H]	क्	क् (Pure Consonant)
C[BD]	कँं	कँं
C[BX]	कँः	कँः

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

Example:

CM[D]	कीं	कीं
CM[D]	काँ	काँ
CM[X]	वीः	वीः
CM[BD]	काँं	काँं
CM[BX]	काँः	काँः

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

Example:

CHCHCHC क्ष = न् + त् + र् + य न्+त्+र्+य

Subsets

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

Example:

CHC[M]	क्की	क्क्की	क्की	क्क्की
CHC[D]	क्कं	क्क्कं	क्कं	क्क्कं
CHC[B]	क्कँ	क्क्कँ	क्कँ	क्क्कँ



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

CHC[X]	क्कः	क्कः	क्कः	क्कः
CHC[BD]	क्कँ	क्कँ	क्कँ	क्कँ
CHC[BX]	क्कं	क्कं	क्कं	क्कं

3.b. *3(CH)CM may be followed by a *1 D, B, X, BD or BX
(CHCM) *1[D|B|X|BD|BX]

Example:

CHCM[D]	क्की	क्की	क्की	क्की
CHCM[B]	क्काँ	क्काँ	क्काँ	क्काँ
CHCM[X]	क्कीः	क्कीः	क्कीः	क्कीः
CHCM[BD]	क्कँ	क्कँ	क्कँ	क्कँ
CHCM[BX]	क्कं	क्कं	क्कं	क्कं

4. A single Khanda Ta(Z)

Example:

Z ९

5. A Khanda Ta can be preceded by a consonant and Virama/Hasant/Halant

[CH]Z

Example:

CHZ ब् ९ ९

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|BD|BX|M[D|B|X|BD|BX]] | [CH]Z



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

1.5 ABNF Applied to the Assamese 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 | BD | BX | HCM[D|B|X|BD|BX]]

Consonant-sequence \rightarrow

*3(C[N]H)C[N][H|D|B|X|BD|BX|M[D|B|X|BD|BX]] | [CH]Z

Sequence \rightarrow consonant-sequence[Y] | vowel-sequence[Y]

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 Assamese the following rules apply:

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

Example:

্‌ক	্ক
িক	িক
ংক	ংক
ঁক	ঁক
ঃক	ঃক

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

অ্	অ্
কং্	কং্
কঁ্	কঁ্
কঃ্	কঃ্
ক্	ক্
1্	1্



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

-ॠ

-ॡ

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

Example

कॠं	कंं
कँँ	कँँ
कःॠः	कःॡः
कीॠं	कींं
काँँ	काँँ
कीःॠः	कीःॡः
अॠं	अंं
अँँ	अँँ
अःॠः	अःॡः

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

Example:

कीी कीी

5. M is not permitted after V

Example:

त्रैा ईा

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

Example:

कॠः कंः



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

कःं

कःं

7. Khanda Ta is NOT allowed in the beginning of a IDN label.
8. CH can come with Khanda Ta in only the case where C is च (09F0).
9. Nukta can be only allowed after following characters:
 - ड (09A1)
 - ढ (09A2)
 - य (09AF)
10. Only following combination with VHCM will be allowed.
 - अा → अ(0985) + ्(09CD) + य(09AF) + ाँ(09BE)
 - एा → ए(098F) + ्(09CD) + य(09AF) + ाँ(09BE)
11. A consonant syllable that is intended to end with Halant [H] can only be followed by Hyphen, digit or Avagraha [Y].



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

3. EXAMPLES

Following are the some examples to show ASSAMESE IDN Label.

C	ক	কল
CN	ড়	গুড়
CH	প্	গিলাপ্
CM	চা	চাল
CD	ভং	অভং
CB	গঁ	গঁতা
CX	ধঃ	অধঃ
CBD	কঁং	কঁং
CMD	চাং	চাং
CMB	চাঁ	চাঁত
CMX	দুঃ	দুঃখ
CHC	দ্ব	পদ্ব
CHCHCHC	স্ট্ৰ্য	স্ট্ৰ্যাপ
V	আ	আকাশ



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

VD	অং	অংশ
VB	অঁ	অঁত
VX	আঃ	আঃ
VHCM	অ্যা	অ্যাসিড
Z	ৎ	বৎস
CHZ	ৎ	ভৎসনা



































































































































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

4. LANGUAGE TABLE: ASSAMESE²

0980

Bengali

09FF

	098	099	09A	09B	09C	09D	09E	09F
0								
1								
2								
3								
4								
5								
6								
7								
8								
9								
A								
B								
C								
D								
E								
F								

² Characters marked in yellow are not applicable to the language.



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

5. NOMENCLATURAL DESCRIPTION TABLE OF ASSAMESE LANGUAGE TABLE

Unicode Value	Character	Description
CHANDRABINDU (B)		
0981	ঁ	BENGALI SIGN CANDRABINDU
ANUSVARA (D)		
0982	ং	BENGALI SIGN ANUSVARA
VISARGA (X)		
0983	ঃ	BENGALI SIGN VISARGA
INDEPENDENT VOWELS (V)		
0985	অ	BENGALI LETTER A
0986	আ	BENGALI LETTER AA
0987	ই	BENGALI LETTER I
0988	ঈ	BENGALI LETTER II
0989	উ	BENGALI LETTER U
098A	ঊ	BENGALI LETTER UU
098F	এ	BENGALI LETTER E
0990	ঐ	BENGALI LETTER AI
0993	ও	BENGALI LETTER O
0994	ঔ	BENGALI LETTER AU
098C	ল	BENGALI LETTER VOCALIC L
098B	র	BENGALI LETTER VOCALIC R
Consonants (C)		
0995	ক	BENGALI LETTER KA
0996	খ	BENGALI LETTER KHA
0997	গ	BENGALI LETTER GA
0998	ঘ	BENGALI LETTER GHA



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

0999	ঙ	BENGALI LETTER NGA
099A	চ	BENGALI LETTER CA
099B	ছ	BENGALI LETTER CHA
099C	জ	BENGALI LETTER JA
099D	ঝ	BENGALI LETTER JHA
099E	ঞ	BENGALI LETTER NYA
099F	ট	BENGALI LETTER TTA
09A0	ঠ	BENGALI LETTER TTHA
09A1	ড	BENGALI LETTER DDA
09A2	ঢ	BENGALI LETTER DDHA
09A3	ণ	BENGALI LETTER NNA
09A4	ত	BENGALI LETTER TA
09A5	থ	BENGALI LETTER THA
09A6	দ	BENGALI LETTER DA
09A7	ধ	BENGALI LETTER DHA
09A8	ন	BENGALI LETTER NA
09AA	প	BENGALI LETTER PA
09AB	ফ	BENGALI LETTER PHA
09AC	ব	BENGALI LETTER BA
09AD	ভ	BENGALI LETTER BHA
09AE	ম	BENGALI LETTER MA
09AF	য	BENGALI LETTER YA
09B2	ল	BENGALI LETTER LA
09B6	শ	BENGALI LETTER SHA
09B7	ষ	BENGALI LETTER SSA



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

09B8	স	BENGALI LETTER SA
09B9	হ	BENGALI LETTER HA
09DC	ড়	BENGALI LETTER RRA
09DD	ঢ	BENGALI LETTER RHA
09DF	য়	BENGALI LETTER YYA
09F0	ষ	BENGALI LETTER RA WITH MIDDLE DIAGONAL
09F1	ষ	BENGALI LETTER RA WITH LOWER DIAGONAL
VOWEL SIGNS(MATRAS) (M)		
09BE	া	BENGALI VOWEL SIGN AA
09BF	ি	BENGALI VOWEL SIGN I
09C0	ী	BENGALI VOWEL SIGN II
09C1	ু	BENGALI VOWEL SIGN U
09C2	ূ	BENGALI VOWEL SIGN UU
09C3	্	BENGALI VOWEL SIGN VOCALIC R
09C7	ে	BENGALI VOWEL SIGN E
09C8	ৈ	BENGALI VOWEL SIGN AI
09CB	ো	BENGALI VOWEL SIGN O
09CC	ৌ	BENGALI VOWEL SIGN AU
09E2	্ল	BENGALI VOWEL SIGN VOCALIC L
VIRAMA (H)		
09CD	্	BENGALI SIGN VIRAMA
KHANDATA (Z)		
09CE	ৎ	BENGALI LETTER KHANDATA
AVAGRAHA (Y)		
09BD	৳	BENGALI SIGN AVAGRAHA



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

NUKTA (N)		
09BC	ॐ	BENGALI SIGN NUKTA



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

6. VARIANT TABLE

VARIANTS	
ड + ़ 09A1 + 09BC	ड़ 09DC
ढ + ़ 09A2 + 09BC	ढ़ 09DD
य + ़ 09AF + 09BC	य़ 09DF
् 09C3	्र 09C2
ी+ँ 09C0 + 0981	ी 09C0
ौ+ँ 09CC + 0981	ौ 09CC



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

7. EXPERTS/BODIES CONSULTED

Expertise provided by C-DAC KOLKATA.



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

8. PROPOSED ccTLD FOR ASSAMESE

India (Bhārat) localized in Assamese - **ভাৰত**

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