



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

Policy Document For

INTERNATIONALIZED DOMAIN NAMES

Language: ASSAMESE



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

1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF)	3
1.1 Naming of Variables:	3
1.2 ABNF Operators	3
1.3 The Vowel Sequence	4
1.4 Consonant Sequence *	5
1.5 ABNF Applied to the Assamese IDN	8
2. RESTRICTION RULES	9
3. EXAMPLES	12
4. LANGUAGE TABLE: ASSAMESE	14
5. NOMENCLATRURAL DESCRIPTION TABLE OF ASSAMESE LANGUAGE TABLE	16
6. VARIANT TABLE	19
7. EXPERTS/BODIES CONSULTED	20
8. Country Code Top Level Domain (ccTLD) FOR ASSAMESE	21



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),Candrabindu (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 ঞ bengali letter ya. Represented by the sequence < U+09CD ঞ bengali sign virama, U+09AF ঞ bengali letter ya>, ya-phalaa has a special form. When combined with U+09BE ঞ 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 ঞ

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 ঞঁ
V[B] ঞঁ ঞঁ
V[X] অঃ ঞঃ

* 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[BD]	अँ	अँ
V[BX]	अंः	अंः
V[HCM]	अ (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]	अ्यां
VHCM[B]	अ्याँ
VHCM[X]	अ्याः
VHCM[BD]	अ्याँं
VHCM[BX]	अ्याँः

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 छ्य = 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
CHC[X]	क्कः	क्क्कः	क्कः	क्क्कः



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

CHC[BD] क्कँः क्कँः वक्कँः क्कँः
CHC[BX] क्कँः क्कँः वक्कँः क्कँः

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] क्कीं क्क्रीं वक्कीं क्क्रीं
CHCM[B] क्काँ क्क्रीं वक्काँ क्क्रीं
CHCM[X] क्कीः क्क्रीः वक्कीः क्क्रीः
CHCM[BD] क्काँः क्क्रीं वक्काँः क्क्रीं
CHCM[BX] क्काँः क्क्रीं वक्काँः क्क्रीं

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
১্	১্
-্	-্



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

Example

क॑ं	कं
कँ	कँ
कः	कः
कीं	कीं
कौ	कौ
कीः	कीः
अं	अं
अँ	अँ
अः	अः

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

Example:

कीी	कीी
-----	-----

5. M is not permitted after V

Example:

औ	ई
---	---

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

Example:

कः	कः
कःं	कःं

7. Khanda Ta is NOT allowed in the beginning of a IDN label.



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

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.

अग -> अ(0985) + ं(09CD) + य(09AF) + ा(09BE)

एग -> ए (098F) + ं(09CD) + य(09AF) + ा(09BE)

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



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
VD	অং	অংশ
VB	অঁ	অঁত



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

VX	आः	आः
VHCM	अ्या	अ्यासिड
Z	९	ब९स
CHZ	९	ड९सना



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

4. LANGUAGE TABLE: ASSAMESE²

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

0980

Bengali

09FF

	098	099	09A	09B	09C	09D	09E	09F
0	৭ 0980	ঐ 0990	ঔ 09A0	ৱ 09B0	ী 09C0		ঋ 09E0	ঔ 09F0
1	ঁ 0981		ড 09A1		ৗ 09C1		ঋ 09E1	ঔ 09F1
2	ং 0982		ঢ 09A2	ল 09B2	ৗ 09C2		ৗ 09E2	ৗ 09F2
3	ঃ 0983	ও 0993	ণ 09A3		ৗ 09C3		ৗ 09E3	ৗ 09F3
4		ঔ 0994	ত 09A4		ৗ 09C4			ৗ 09F4
5	অ 0985	ক 0995	খ 09A5					ৗ 09F5
6	আ 0986	খ 0996	দ 09A6	শ 09B6			ৗ 09E6	ৗ 09F6
7	ই 0987	গ 0997	ধ 09A7	ষ 09B7	ে 09C7	ী 09D7	ৗ 09E7	ৗ 09F7
8	ঈ 0988	ঘ 0998	ন 09A8	স 09B8	ৈ 09C8		ৗ 09E8	ৗ 09F8
9	উ 0989	ঙ 0999		হ 09B9			ৗ 09E9	ৗ 09F9
A	ঊ 098A	চ 099A	প 09AA				ৗ 09EA	ৗ 09FA
B	ঋ 098B	ছ 099B	ফ 09AB		ৌ 09CB		ৗ 09EB	ৗ 09FB
C	ৗ 098C	জ 099C	ব 09AC	ৗ 09BC	ৌ 09CC	ড় 09DC	ৗ 09EC	ৗ 09FC
D		ঝ 099D	ভ 09AD	ৗ 09BD	ৗ 09CD	ঢ 09DD	ৗ 09ED	ৗ 09FD
E		ঞ 099E	য 09AE	া 09BE	ং 09CE		ৗ 09EE	ৗ 09FE
F	এ 098F	ট 099F	য 09AF	ি 09BF		য় 09DF	ৗ 09EF	

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



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	ৗ	BENGALI LETTER VOCALIC L
098B	৘	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
0999	ঙ	BENGALI LETTER NGA



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

Unicode Value	Character	Description
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
09B8	স	BENGALI LETTER SA
09B9	হ	BENGALI LETTER HA



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

Unicode Value	Character	Description
09F0	৷	BENGALI LETTER RA WITH MIDDLE DIAGONAL
09F1	৸	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	ৄ	BENGALI VOWEL SIGN VOCALIC L
VIRAMA (H)		
09CD	্	BENGALI SIGN VIRAMA
KHANDATA (Z)		
09CE	৐	BENGALI LETTER KHANDATA
AVAGRAHA (Y)		
09BD	৑	BENGALI SIGN AVAGRAHA
NUKTA (N)		
09BC	়	BENGALI SIGN NUKTA



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

6. VARIANT TABLE

VARIANTS	
ॠ 09C3	ॠ 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. Country Code Top Level Domain (ccTLD) FOR ASSAMESE

India (Bhārat) localized in Assamese – **ভাৰত** (xn--45br5cyl)