

Draft Policy Document

for

**INTERNATIONALIZED
DOMAIN
NAMES**

Language: GUJARATI

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 GUJARATI	1.5
1.1	22/11/2010	Page No 9, 16, 18	A, D	Restriction rule added, Variant deleted, ccTLD added	1.6
1.2	05/08/2013	Whole Document	A,M	Restriction rules added and modified.	
1.3	07/07/2014	Page No 11	A,M	Restriction rules added.	

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.....	5
1.4	The Consonant Sequence	5
1.5	Sequence	7
1.6	ABNF Applied to the Gujarati IDN	7
2.	RESTRICTION RULES	11
3.	EXAMPLES	12
4.	LANGUAGE TABLE: GUJARATI	13
5.	NOMENCLATRUAL DESCRIPTION TABLE OF GUJARATI LANGUAGE TABLE.....	14
6.	VARIANT TABLE	17
7.	EXPERTS/BODIES CONSULTED	18
8.	PROPOSED ccTLD FOR GUJARATI.....	19

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 (Used very rarely in Gujarati)
X	→	Visarga
Y	→	Avagraha
H	→	Halant

1.2 ABNF Operators

Sr. No.	Operator	Function
1	“ ”	Alternative
2	“[]”	Optional
3	“*”	Variable Repetition
4	“()”	Sequence Group

In what follows, the Vowel Sequence and the Consonant Sequence pertinent to Gujarati are given. To facilitate understanding, 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 Gujarati are restricted to one.

The vowel sequence in Gujarati is therefore,

V[D|B|X]

Examples:

Vowel	V	અ	अ
Vowel+Anusvara	VD	અં	अं
Vowel+Chandrabindu	VB	અँ	अँ
Vowel+Visarga	VX	અઃ	अः

Standard Gujarati does not use Chandrabindu, although the same is used for Sanskrit words.

1.4 The Consonant Sequence

A consonant sequence admits the following combinations:

1. A single consonant (C)

Example:

C ક क

2. A consonant optionally followed by dependent Vowel Sign / Matra [M] or Anusvara [D] or Chandrabindu [B] or Visarga [X] or Halanta [H].

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

Example:

CM	की	की
CD	कं	कं
CB	कँ	कँ
CX	कः	कः
CH	क्	क् (Pure Consonant)

2.a. A CM sequence can be optionally followed by D, B or X.

(CM)[D|B|X]

Example:

CMD	कीं	कीं
CMB	काँ	काँ
CMX	वीः	वीः

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

Example:

CHC	चक	→ च् + क	न+च्+क
CHCHC	चक्र	→ च् + क् + र	न+च्+क+र्+र
CHCHCHC	चक्रय	→ च् + क् + र् + य	न+च्+क+र्+र+र्+य

Subsets:

While considering its subsets, as a representative example, we will consider the combination CHC only; however the same is equally applicable to CHCHC and CHCHCHC.

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

Example:

CHCM	ક્કી	ક્કી	ક્કી	ક્કી
CHCD	ક્કં	ક્કં	ક્કં	ક્કં
CHCB	ક્કૃ	ક્કૃ	ક્કૃ	ક્કૃ
CHCX	ક્કઃ	ક્કઃ	ક્કઃ	ક્કઃ
CHCH	ક્ક્	ક્ક્	ક્ક્	ક્ક્

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

Example:

CHCMD	ક્કીં	ક્કીં	ક્કીં	ક્કીં
CHCMB	ક્કૃં	ક્કૃં	ક્કૃં	ક્કૃં
CHCMX	ક્કીઃ	ક્કીઃ	ક્કીઃ	ક્કીઃ

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

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

1.5 Sequence

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

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

1.6 ABNF Applied to the Gujarati IDN

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

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

Consonant-sequence $\rightarrow$ *3(CH)C[H|D|B|X|M[D|B|X]]

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

IDN-label $\rightarrow$ (sequence | digit) * ([dash] (sequence | digit))

Additional Examples putting more light on Gujarati 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 or M cannot occur in the beginning of a Gujarati IDN

Example

્ક	્ક
િક	િક
ંક	ંક
ૅક	ૅક
ઃક	ઃક

As can be seen, such combinations will result automatically in a “golu” marking it as an invalid formation. This is an intrinsic property of the Indian language syllable and is quasi automatically applied.

2. H is not permitted after V, D, B, X, M, Digit or Dash.

Example

અ	અ
કં	કં
કૅ	કૅ
કઃ	કઃ
કી	કિ
૧	૧
-	-

3. Number of D, B or X permitted after Consonant or Vowel or a Matra is restricted to one. Thus following combinations are invalidated.

Example

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

4. Number of M permitted after Consonant is restricted to one

Example

कील	कीी
-----	-----

5. M is not permitted after V.

Example

ली	ई
----	---

6. The combinations of Anusvara + Visarga [DX], Chandrabindu + Anusvara [BD], Chandrabindu + Visarga [BX] and vice-versa are not permissible

Example

कंः	कंः
कँः	कँः
कः	कः

2. RESTRICTION RULES

The Augmented Backus Naur Formalism (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 case of Gujarati the following rules apply:

1. A Consonant-sequence that is intended to end with Halant [H] can only be followed by Hyphen, Digit or Avagraha. Thus following combinations are permissible.

ક-	ઋ-
ક1	ઋ1
કઠ	ઋઠ

2. Consecutive Hyphens will not be permitted in a domain name.
3. The number of identical consonants joined by a Halant within a label shall not exceed two. Thus ત્ (ta+halant+ta) is permitted but not ત્ત (ta+halant+ta+halant+ta).
4. 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).

Additional Note:

Wherever a variant is present in a given label, the variants shall be strictly symmetric and non-transitive. This ensures that over generativity does not take place. However the case of over generativity of variants does not exist in Gujarati.

3. EXAMPLES

Combination	Example	Word with combination
C	ર	રમ
CH	ફ	વાફ (Rarely used)
CM	તુ	તુ
CD	પં	પંપ
CX	શઃ	અક્ષરશઃ (Rarely used)
CMD	મિ	મિ
CMB	કાં	કાંચ
CMX	દુઃ	દુઃખ
CHC	ક	કમ
CHCHC	સ્ત્ર	સ્ત્રે
CHCHCHC	કસ્ત્ર	એકસ્ત્રા
V	આ	આજ
VD	અં	અંક
VB	આં	આંખ
VX	આઃ	આઃ

4. LANGUAGE TABLE¹: GUJARATI²

	0A8	0A9	0AA	0AB	0AC	0AD	0AE	0AF
0		ૐ	ઠ	૨	ી	ૐ	જી	
1	ૐ	ૐ	ડ		ૐ		જી	ૐ
2	ૐ		ઢ	લ	ૐ		જી	
3	ૐ	ૐ	ભ	ળ	ૐ		જી	
4		ૐ	ત		ૐ			
5	અ	ક	થ	વ	ૐ			
6	આ	ખ	દ	શ			ૐ	
7	ઈ	ગ	ધ	પ	ૐ		ૐ	
8	ઈ	ધ	ન	સ	ૐ		ૐ	
9	ઉ	ડ		હ	ૐ		ૐ	
A	ઊ	ચ	પ				ૐ	
B	જી	છ	ફ		ૐ		ૐ	
C	જી	જ	ઝ	ૐ	ૐ		ૐ	
D	ૐ	ઝ	ભ	ડ	ૐ		ૐ	
E		ઝ	મ	લ			ૐ	
F	ૐ	ટ	ય	િ			ૐ	

¹ This language table is based on Unicode Chart for Gujarati script provided by the Unicode Consortium.

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

5. NOMENCLATURAL DESCRIPTION TABLE OF GUJARATI LANGUAGE TABLE

CHANDRABINDU (B)		
0A81	ં	GUJARATI SIGN CANDRABINDU
ANUSVARA (D)		
0A82	ં	GUJARATI SIGN ANUSVARA
VISARGA (X)		
0A83	ઃ	GUJARATI SIGN VISARGA
VOWELS (V)		
0A85	અ	GUJARATI LETTER A
0A86	આ	GUJARATI LETTER AA
0A87	ઇ	GUJARATI LETTER I
0A88	ઈ	GUJARATI LETTER II
0A89	ઉ	GUJARATI LETTER U
0A8A	ઊ	GUJARATI LETTER UU
0A8B	ઋ	GUJARATI LETTER VOCALIC R
0A8D	એ	GUJARATI VOWEL CANDRA E
0A8F	એ	GUJARATI LETTER E
0A90	ઐ	GUJARATI LETTER AI
0A91	ઑ	GUJARATI LETTER CANDRA O
0A93	ઓ	GUJARATI LETTER O
0A94	ઔ	GUJARATI LETTER AU
CONSONANTS (C)		
0A95	ક	GUJARATI LETTER KA
0A96	ખ	GUJARATI LETTER KHA
0A97	ગ	GUJARATI LETTER GA

0A98	ઘ	GUJARATI LETTER GH A
0A99	ંગ	GUJARATI LETTER NG A
0A9A	ચ	GUJARATI LETTER CA
0A9B	છ	GUJARATI LETTER CH A
0A9C	જ	GUJARATI LETTER JA
0A9D	ઝ	GUJARATI LETTER JH A
0A9E	ઞ	GUJARATI LETTER NY A
0A9F	ટ	GUJARATI LETTER TT A
0AA0	ઠ	GUJARATI LETTER TT HA
0AA1	ડ	GUJARATI LETTER DD A
0AA2	ઢ	GUJARATI LETTER DD HA
0AA3	ણ	GUJARATI LETTER NN A
0AA4	ત	GUJARATI LETTER TA
0AA5	થ	GUJARATI LETTER TH A
0AA6	દ	GUJARATI LETTER DA
0AA7	ધ	GUJARATI LETTER DH A
0AA8	ન	GUJARATI LETTER NA
0AAA	પ	GUJARATI LETTER PA
0AAB	ફ	GUJARATI LETTER PH A
0AAC	બ	GUJARATI LETTER BA
0AAD	ભ	GUJARATI LETTER BH A
0AAE	મ	GUJARATI LETTER MA
0AAF	ય	GUJARATI LETTER YA
0AB0	ર	GUJARATI LETTER RA
0AB2	લ	GUJARATI LETTER LA

0AB3	ળ	GUJARATI LETTER LLA
0AB5	વ	GUJARATI LETTER VA
0AB6	શ	GUJARATI LETTER SHA
0AB7	ષ	GUJARATI LETTER SSA
0AB8	સ	GUJARATI LETTER SA
0AB9	હ	GUJARATI LETTER HA
DEPENDENT VOWEL SIGNS (MATRAS) (M)		
0ABE	ા	GUJARATI VOWEL SIGN AA
0ABF	િ	GUJARATI VOWEL SIGN I
0AC0	ી	GUJARATI VOWEL SIGN II
0AC1	ુ	GUJARATI VOWEL SIGN U
0AC2	ૂ	GUJARATI VOWEL SIGN UU
0AC3	ૃ	GUJARATI VOWEL SIGN VOCALIC R
0AC5	ે	GUJARATI VOWEL SIGN CANDRA E
0AC7	ૈ	GUJARATI VOWEL SIGN E
0AC8	ૌ	GUJARATI VOWEL SIGN AI
0AC9	ો	GUJARATI VOWEL SIGN CANDRA O
0ACB	ૌ	GUJARATI VOWEL SIGN O
0ACC	ૌ	GUJARATI VOWEL SIGN AU
AVAGRAHA (Y)		
0ABD	ઽ	GUJARATI SIGN AVAGRAHA
HALANT (H)		
0ACD	્	GUJARATI SIGN VIRAMA

6. VARIANT TABLE

VARIANTS		
ՏՂ 0AAB+ 0AAF	ՏՂ 0AAB+ 0ACD+ 0AAF	
Ճ 0AA6+ 0ACD+0AA7	Ճ 0AA6+ 0ACD+0A98	
Ճ 0AA6+ 0ACD+0AAC	Ճ 0AA6+ 0ACD+0AB5	
Ճ 0AA6+ 0ACD+0AB0	Ճ 0AA6+ 0ACD+0AA8	Ճ 0AA6+ 0ACD+0A97

7. EXPERTS/BODIES CONSULTED

Mr. Ashok Karania (C.E.O Magnet Technologies) in consultation with Gujarati Sahitya Parishad.

8. PROPOSED ccTLD FOR GUJARATI

India (Bhārat) localized in Gujarati - ભારત

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