



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

Draft Policy Document

for

INTERNATIONALIZED DOMAIN NAMES

Language: TELUGU



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	20/11/09	Whole Document	M	Language Specific Policy Document for TELUGU	1.5
1.1	22/11/2010	Page No 10, 23	A	Restriction rule added, ccTLD added	1.6
1.2	05/08/2013	Whole Document	A,M	Registry specific rules added and modified.	



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 Telugu IDN	7
2.	RESTRICTION RULES	12
3.	EXAMPLES	13
4.	LANGUAGE TABLE: TELUGU	14
5.	NOMENCLATRURAL DESCRIPTION TABLE OF TELUGU LANGUAGE TABLE.....	15
6.	VARIANT TABLE	20
7.	EXPERTS/BODIES CONSULTED	22
8.	PROPOSED ccTLD FOR TELUGU.....	23



1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF)

1.1 Declaration of Variables:

- Dash → Hyphen -
Digit → Indo-Arabid digits [0-9]
C → Consonant
V → Vowel
M → Matra
D → Anusvara/Sunna
B → Chandrabindu/Arasunna
X → Visarga
H → Halant/Virama
Y → Avagraha

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 Telugu 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/Sunna (D), Chandrabindu/Arasunna (B) or a Visarga (X). The number of D, B or X which can follow a V in Telugu are restricted to one.

The possibility of a Visarga following a Chandrabindu or Anusvara is ruled out.

The vowel sequence in Telugu is therefore

V[D|B|X]

Examples:

V	అ	ఆ
VD	అం	ఆం
VB	అఁ	ఆఁ
VX	అః	ఆః

1.4 The Consonant Sequence

A consonant sequence admits the following combinations:

1. A single consonant (C)

Example:

C	క	క
---	---	---

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

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

Example:

CM	కి	కి
----	----	----



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

CD	क०	कं
CB	क८	कँ
CX	कः	कः
CH	क	क् (Pure Consonant)

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

Example:

CMD	की०	कीं
CMB	का८	काँ
CMX	वीः	वीः

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

Example:

CHC	त्र	त्रि+र	त्र
CHCHC	त्र्य	त्रि+रि+य	त्र्य

Subsets:

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

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

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



Example:

CHCM	ఙ్ఁ	క+ఁ+క+ఁ	क्की
CHCD	కఙ్ఁ	క+ఁ+క+ఁ	क्कं
CHCB	కఙ్ఁ	క+ఁ+క+ఁ	क्क
CHCX	కఙ్ఁ	క+ఁ+క+ఁ	क्क:
CHCH	ఙ్ఁ	క+ఁ+క+ఁ	क्क्

3.b. *2(CH)CM may further be followed by a D, B or X.

Example:

CHCMD	ఙ్ఁ	క+ఁ+క+ఁ	क्कीं
CHCMB	ఙ్ఁ	క+ఁ+క+ఁ	क्की
CHCMX	ఙ్ఁ	క+ఁ+క+ఁ	क्की:

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

*2(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 Telugu IDN

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



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

Consonant Sequence → *2(CH)C[H|D|B|X|M|D|B|X]

Vowel Sequence → V[D|B|X]

Sequence → Consonant Sequence[Y] | Vowel Sequence[Y]

IDN-Label → (sequence|digit)*([dash](sequence|digit))



Additional Examples putting more light on Telugu ABNF:

1. H, M, D, B or X cannot occur in the beginning of a Telugu IDN.

Example:

్క క్

్క క్

ంక క్

్క క్

ంక క్

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 wherever supported by the OS.

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

Example

్క క్

ంక క్

ంక క్

ంక క్

్క క్

ంక క్



-ँ -ँ

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:

क०० कं

क०० कँ

कःः कःः

की०० कीं

का०० काँ

कीःः कीःः

अ०० अं

अ०० अँ

अःः अःः

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

Example:

की० कीी

5. M is not permitted after V.

Example:

ईँ ई



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

6. The combinations of Anuswara+Visarga, Visarga+Anuswara, Chandrabindu+Anusvara and vice-versa are not permissible
Example:

कॊः

कंः

कॊँ

कँः

कःँ

कःँ



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 Telugu the following rules apply:

1. A consonant syllable that is intended to end with Halant [H] can only be followed by Hyphen, Digit or Avagraha.

§-

§1

§2

2. The number of identical consonants joined by a Halant within a label shall not exceed two. Thus క్క (ka+halant+ka) is permitted but not క్కక (ka+halant+ka+halant+ka).
3. Consecutive Hyphens will not be permitted in a domain name.
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. (Example 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 case of Telugu.



3. EXAMPLES

Combination	Example	Word with combination
C	క	కల
CH	ఠా	భారత్
CM	కా	కాకి
CD	కం	కందకం
CX	నః	పునః దర్శనం
CMD	కాం	కాంకర్యం
CMX	దుః	దుఃఖ
CHC	ఠ్ఠా	బెర్
CHCHC	ఠ్ఠ్ఠ	అస్త్రి
V	అ	అల
VB	ఠ్ఠఠ	ఠ్ఠఠ
VD	అం	అంకె



4. LANGUAGE TABLE¹: TELUGU²

	0C0	0C1	0C2	0C3	0C4	0C5	0C6	0C7
0		బ	ల	ర	్		ఋ	
		0C10	0C20	0C30	0C40		0C50	
1	ౠ		క్ష	జ	ౡ		ౣ	
	0C01		0C21	0C31	0C41		0C51	
2	ౢ	ఙ	ఛ	ల	ౣ		౥	
	0C02	0C12	0C22	0C32	0C42		0C52	
3	౤	ఙ	ణ	శ	౥		౧	
	0C03	0C13	0C23	0C33	0C43		0C53	
4		జ్ఞ	త		౦			
		0C14	0C24		0C44			
5	౧	ర	క	వ		ౡ		
	0C05	0C15	0C25	0C35		0C55		
6	ౡ	ఖ	క	ఖ	ౠ	ౡ	ౢ	
	0C06	0C16	0C26	0C36	0C46	0C56	0C66	
7	ౣ	ర	క	వ	ౡ		ౢ	
	0C07	0C17	0C27	0C37	0C47		0C57	
8	౥	ఙ	ర	ర	ౠ	ౡ	ౢ	ౣ
	0C08	0C18	0C28	0C38	0C48	0C58	0C68	0C78
9	౦	ఙ		ౡ		ౡ	ౢ	ౣ
	0C09	0C19		0C39		0C59	0C69	0C79
A	౥	వ	వ		ౡ		ౣ	౤
	0C0A	0C1A	0C2A		0C4A		0C5A	0C7A
B	ఋ	ఋ	ఋ		ౡ		ౣ	౤
	0C0B	0C1B	0C2B		0C4B		0C5B	0C7B
C	ౠ	జ	బ		ౡ		ౣ	౤
	0C0C	0C1C	0C2C		0C4C		0C5C	0C7C
D		ౡ	క	ౡ	ౠ		ౣ	౤
		0C1D	0C2D	0C3D	0C4D		0C5D	0C7D
E	ౡ	ౡ	ౡ	ౡ			ౣ	౤
	0C0E	0C1E	0C2E	0C3E			0C5E	0C7E
F	ౡ	ౡ	ౡ	ౡ			ౣ	౤
	0C0F	0C1F	0C2F	0C3F			0C5F	0C7F

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

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



5. NOMENCLATURAL DESCRIPTION TABLE OF TELUGU LANGUAGE TABLE

CHANDRABINDU (C)		
0C01	◌ఱ	TELUGU SIGN CANDRABINDU
ANUSVARA (D)		
0C02	◌ం	TELUGU SIGN ANUSVARA
VISARGA (X)		
0C03	◌ః	TELUGU SIGN VISARGA
VOWEL LETTERS (V)		
0C05	అ	TELUGU LETTER A
0C06	ఆ	TELUGU LETTER AA
0C07	ఇ	TELUGU LETTER I
0C08	ఈ	TELUGU LETTER II
0C09	ఉ	TELUGU LETTER U
0C0A	ఊ	TELUGU LETTER UU
0C0B	ఋ	TELUGU LETTER VOCALIC R
0C0E	ఎ	TELUGU LETTER E
0C0F	ఏ	TELUGU LETTER EE



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

0C10	ఐ	TELUGU LETTER AI
0C12	ఒ	TELUGU LETTER O
0C13	ఓ	TELUGU LETTER OO
0C14	ఔ	TELUGU LETTER AU
0C60	ఋ	TELUGU LETTER VOCALIC RR
CONSONANTS (C)		
0C15	క	TELUGU LETTER KA
0C16	ఖ	TELUGU LETTER KHA
0C17	గ	TELUGU LETTER GA
0C18	ఘ	TELUGU LETTER GHA
0C19	ఙ	TELUGU LETTER NGA
0C1A	చ	TELUGU LETTER CA
0C1B	ఛ	TELUGU LETTER CHA
0C1C	జ	TELUGU LETTER JA
0C1D	ఝ	TELUGU LETTER JHA
0C1E	ఞ	TELUGU LETTER NYA



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

0C1F	ట	TELUGU LETTER TTA
0C20	థ	TELUGU LETTER TTHA
0C21	డ	TELUGU LETTER DDA
0C22	ఢ	TELUGU LETTER DDHA
0C23	ణ	TELUGU LETTER NNA
0C24	త	TELUGU LETTER TA
0C25	థ	TELUGU LETTER THA
0C26	ద	TELUGU LETTER DA
0C27	ధ	TELUGU LETTER DHA
0C28	న	TELUGU LETTER NA
0C2A	ప	TELUGU LETTER PA
0C2B	ఫ	TELUGU LETTER PHA
0C2C	బ	TELUGU LETTER BA
0C2D	భ	TELUGU LETTER BHA
0C2E	మ	TELUGU LETTER MA
0C2F	య	TELUGU LETTER YA



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

0C30	ర	TELUGU LETTER RA
0C31	రి	TELUGU LETTER RRA
0C32	ల	TELUGU LETTER LA
0C33	లి	TELUGU LETTER LLA
0C35	వ	TELUGU LETTER VA
0C36	వి	TELUGU LETTER SHA
0C37	ష	TELUGU LETTER SSA
0C38	స	TELUGU LETTER SA
0C39	హ	TELUGU LETTER HA
0C58	ట	TELUGU LETTER TSA
0C59	డ	TELUGU LETTER DZA
VOWEL SIGNS (MATRAS) (M)		
0C3E	ా	TELUGU VOWEL SIGN AA
0C3F	ి	TELUGU VOWEL SIGN I
0C40	ీ	TELUGU VOWEL SIGN II
0C41	ు	TELUGU VOWEL SIGN U



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

0C42	ూ	TELUGU VOWEL SIGN UU
0C43	ృ	TELUGU VOWEL SIGN VOCALIC R
0C44	ౄ	TELUGU VOWEL SIGN VOCALIC RR
0C46	ె	TELUGU VOWEL SIGN E
0C47	ే	TELUGU VOWEL SIGN EE
0C48	ై	TELUGU VOWEL SIGN AI
0C4A	ొ	TELUGU VOWEL SIGN O
0C4B	ో	TELUGU VOWEL SIGN OO
0C4C	ౌ	TELUGU VOWEL SIGN AU
HALANT (H)		
0C4D	ం	TELUGU SIGN VIRAMA
AVAGRAHA (Y)		
0C3D	ౡ	TELUGU SIGN AVAGRAHA



6. VARIANT TABLE

VARIANTS			
గ 0C17+0C4D+0C26		గ 0C17+0C4D+0C27	
గ 0C17+0C4D+0C17		గ 0C17+0C4D+0C32	
ఝ 0C26+0C4D+0C26		ఝ 0C26+0C4D+0C27	
ఞ 0C30+0C4D+0C26	ఞ 0C30+0C4D+0C27	ఞ 0C30+0C4D+0C25	ఞ 0C30+0C4D+0C23
ఞ 0C38+0C4D+0C26		ఞ 0C38+0C4D+0C27	
ఞ 0C30+0C4D+0C2C		ఞ 0C30+0C4D+0C2D	
ఞ 0C30+0C4D+0C1A		ఞ 0C30+0C4D+0C1B	
ఞ 0C30+0C4D+0C2A	ఞ 0C30+0C4D+0C2B		ఞ 0C30+0C4D+0C38
ఞ 0C1A+0C4D+0C1A		ఞ 0C1A+0C4D+0C1B	



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

ప 0C2A+0C4D+0C2A	ప 0C2A+0C4D+0C2B	ప 0C2A+0C4D+0C38
బి 0C2C+0C4D+0C26	బి 0C2C+0C4D+0C27	బి 0C2C+0C4D+0C32
స 0C38+0C4D+0C2A	స 0C38+0C4D+0C2B	స 0C38+0C4D+0C38



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

7. EXPERTS/BODIES CONSULTED

Expertise provided by C-DAC Thiruvananthapuram.



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

8. PROPOSED ccTLD FOR TELUGU

India (Bhārat) localized in Telugu - భారత్

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