



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

# Draft Policy Document

for

# INTERNATIONALIZED DOMAIN NAMES

Language: TAMIL



## 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 TAMIL	1.5
1.1	22/11/2010	Page No 8, 17	A	Restriction rule added, ccTLD added	1.6
1.2	05/08/2013	Whole Document	A,M	Restriction 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 .....	6
1.6	ABNF Applied to Tamil IDN .....	6
2.	RESTRICTION RULES .....	9
3.	EXAMPLES .....	10
4.	LANGUAGE TABLE: TAMIL .....	11
5.	NOMENCLATURAL DESCRIPTION TABLE OF TAMIL LANGUAGE TABLE.....	12
6.	VARIANT TABLE .....	15
7.	EXPERTS/BODIES CONSULTED .....	16
8.	PROPOSED ccTLD FOR TAMIL.....	17



# 1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF)

## 1.1 Declaration of Variables:

- Dash → Hyphen -  
Digit → Indo-Arabic digits [0-9]  
C → Consonant  
V → Vowel  
M → Matra  
X → Visarga/Aytham  
H → Halant/Virama

## 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 Tamil 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 a Visarga (X). The number of X which can follow a V in Tamil are restricted to one.

The vowel sequence in Tamil is therefore,

V [X]

Examples:

Vowel	V	அ	ஃ
Vowel+Aytham	VX	அஃ	ஃ:

### 1.4 The Consonant Sequence

A consonant sequence admits the following combinations:

1. A single consonant (C)

Example:

C	ஃ	க
---	---	---

2. A consonant optionally followed by Dependent Vowel sign/Matra [M] or Visarga [X] or Halant/Virama [H]

C[M|X|H]

Example:

CM	கி	கி
CX	கஃ	க:
CH	க்	க் (Pure Consonant)

3. A sequence of consonants (up to 3) joined by Halant/Virama

\*2(CH)C

Example:



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

CHC	கஞ	க ஂ ஞ	க ஶ	க ஶ
CHCHC	கஞய	க ஂ ஞ ஂ ய	க ஶ ய	க ஶ ய

## 1.5 Sequence

A sequence can be made up by Consonant-sequence or Vowel-sequence.  
Thus a sequence is,

consonant-sequence | vowel-sequence

## 1.6 ABNF Applied to Tamil IDN

Consonant Sequence  $\rightarrow *2(\text{CH})\text{C}[\text{H}|\text{X}|\text{M}]$

Vowel Sequence  $\rightarrow \text{V}[\text{X}]$

Sequence  $\rightarrow$  Consonant Sequence | Vowel Sequence

IDN-Label  $\rightarrow (\text{Sequence}|\text{digit})^*([\text{dash}] (\text{Sequence}|\text{digit}))$



## Additional Examples putting more light on Tamil ABNF:

1. H or M or X cannot occur in the beginning of a Tamil IDN.

Example:

ஂக	ஃக
ஃக	ஃக
ஃக	ஃக

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, X, M, Digit or Dash.

Example:

ஂ	ஃ
கஂஂ	கஃ
கஂ	கஃ
1ஂ	1ஃ
-ஂ	-ஃ

3. Visarga/Aytham[X] is permitted after Consonant or a Vowel is restricted to one. Thus following combinations are invalidated.

Example:

கஂஂஂ	கஃஃ
அஂஂஂ	அஃஃ

4. Visarga/Aytham[X] is not permitted after a Matra.

கஂஂ	கஃ
-----	----

5. Number of M permitted after consonant is restricted to one

Example:



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

कीी कीी

6. M is not permitted after V  
Example:

फ़ाई ई





## 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 fine-tune the ABNF.

In case of Tamil the following rules apply:

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

க்-      க்-  
க்1      க்1

2. The number of identical consonants joined by a Halant within a label shall not exceed two. Thus க்ஃக் (ka+halant+ka) is permitted but not க்ஃக்ஃக் (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 Tamil.



### 3. EXAMPLES

Combination	Example	Word with combination
C	பி	பிடி
CH	க்	காக்கா
CM	கி	கிடங்கு
CX	மஃ	மஃகான்
CHC	ம்ம	அம்மா
CHCHC	ர்க்க	சேர்க்கை
V	இ	இப்ப
VX	அஃ	அஃறிணை



## 4. LANGUAGE TABLE<sup>1</sup>: TAMIL<sup>2</sup>

	0B8	0B9	0BA	0BB	0BC	0BD	0BE	0BF
0		ஐ 0B90		ர 0BB0	ீ 0BC0	ஓ 0BD0		ய் 0BF0
1				ற 0BB1	ி 0BC1			ா 0BF1
2	ஃ 0B82	ஔ 0B92		ல 0BB2	ு 0BC2			சு 0BF2
3	ஃ 0B83	ஔ 0B93	ண 0BA3	ள 0BB3				உ 0BF3
4		ஔ 0B94	த 0BA4	ழ 0BB4				ம் 0BF4
5	அ 0B85	க 0B95		வ 0BB5				ஶ 0BF5
6	ஆ 0B86			ஸ 0BB6	ெ 0BC6		ஃ 0BE6	ய் 0BF6
7	இ 0B87			ஷ 0BB7	ே 0BC7	ள 0BD7	க 0BE7	ஶ 0BF7
8	ஈ 0B88		ந 0BA8	ஸ 0BB8	ை 0BC8		உ 0BE8	ஶ 0BF8
9	உ 0B89	ங் 0B99	ன 0BA9	ஶ 0BB9			ந் 0BE9	ஶ 0BF9
A	ஊ 0B8A	சு 0B9A	ப 0BAA		ொ 0BCA		சு 0BEA	ந் 0BFA
B					ோ 0BCB		ரு 0BED	
C		ஐ 0B9C			ெள 0BCD		சு 0BEE	
D					ஃ 0BCD		எ 0BED	
E	எ 0B8E	ஞ் 0B9E	ம 0BAE	ா 0BBE			அ 0BEE	
F	ஏ 0B8F	ழ 0B9F	ய் 0BAF	ி 0BBF			சு 0BEF	

<sup>1</sup> This language table is based on Unicode Chart for Tamil script provided by the Unicode Consortium.

<sup>2</sup> Characters marked in yellow are not applicable to the language.



## 5. NOMENCLATURAL DESCRIPTION TABLE OF TAMIL LANGUAGE TABLE

VISARGA/AYTHAM (X)		
0B83	ஃ	TAMIL SIGN VISARGA
VOWEL LETTERS (V)		
0B85	அ	TAMIL LETTER A
0B86	ஆ	TAMIL LETTER AA
0B87	இ	TAMIL LETTER I
0B88	ஈ	TAMIL LETTER II
0B89	உ	TAMIL LETTER U
0B8A	ஊ	TAMIL LETTER UU
0B8E	எ	TAMIL LETTER E
0B8F	ஏ	TAMIL LETTER EE
0B90	ஐ	TAMIL LETTER AI
0B92	ஔ	TAMIL LETTER O
0B93	ஓ	TAMIL LETTER OO
0B94	ஔ	TAMIL LETTER AU
CONSONANTS (C)		
0B95	க	TAMIL LETTER KA



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

0B99	ங	TAMIL LETTER NGA
0B9A	ச	TAMIL LETTER CA
0B9C	ஐ	TAMIL LETTER JA
0B9E	ஞ	TAMIL LETTER NYA
0B9F	ட	TAMIL LETTER TTA
0BA3	ண	TAMIL LETTER NNA
0BA4	த	TAMIL LETTER TA
0BA8	ந	TAMIL LETTER NA
0BA9	ன	TAMIL LETTER NNNA
0BAA	ப	TAMIL LETTER PA
0BAE	ம	TAMIL LETTER MA
0BAF	ய	TAMIL LETTER YA
0BB0	ர	TAMIL LETTER RA
0BB1	ற	TAMIL LETTER RRA
0BB2	ல	TAMIL LETTER LA
0BB3	ள	TAMIL LETTER LLA
0BB4	ழ	TAMIL LETTER LLLA
0BB5	வ	TAMIL LETTER VA
0BB6	லு	TAMIL LETTER SHA



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

0BB7	ஷ	TAMIL LETTER SSA
0BB8	ஸ	TAMIL LETTER SA
0BB9	ஹ	TAMIL LETTER HA
<b>VOWEL SIGNS (MATRAS) (M)</b>		
0BBE	ா	TAMIL VOWEL SIGN AA
0BBF	ி	TAMIL VOWEL SIGN I
0BC0	ீ	TAMIL VOWEL SIGN II
0BC1	ு	TAMIL VOWEL SIGN U
0BC2	ு	TAMIL VOWEL SIGN UU
0BC6	ெ	TAMIL VOWEL SIGN E
0BC7	ே	TAMIL VOWEL SIGN EE
0BC8	ை	TAMIL VOWEL SIGN AI
0BCA	ொ	TAMIL VOWEL SIGN O
0BCB	ோ	TAMIL VOWEL SIGN OO
0BCC	ௌ	TAMIL VOWEL SIGN AU
<b>VIRAMA (H)</b>		
0BCD	ஃ	TAMIL SIGN VIRAMA



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

## 6. VARIANT TABLE

VARIANT	
ॐ 0B92+0BB3	ॐ 0B94



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 TAMIL

India (Bhārat) localized in Tamil - இந்தியா

Note: You can send your feedbacks to [ids-feedback@cdac.in](mailto:ids-feedback@cdac.in)