

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	
1.1	22/11/20 10	Page No 8, 17	A	Restriction rule added, ccTLD added	1.6
1.2	05/08/20 13	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
	.3 The Vowel Sequence	
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
TAl	BLE	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]

 $C \rightarrow Consonant$

 $V \rightarrow Vowel$

 $M \longrightarrow Matra$

 $X \rightarrow Visarga/Aytham$

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 i	n Tamil is therefore,
V [X]	

Examples:

1.4 The Consonant Sequence

A consonant sequence admits the following combinations:

1. A single consonant (C)

Example:

C あ क

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

C[M|X|H]

Example:

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

Example:







СНС கூடி க**்ஷ**்ய கே எ ्ष

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(CH)C[H|X|M] Vowel Sequence \rightarrow V[X] Sequence \rightarrow Consonant Sequence | Vowel Sequence IDN-Label \rightarrow (Sequence|digit)*([dash] (Sequence|digit))

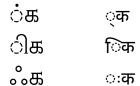






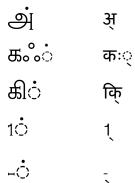
Additional Examples putting more light on Tamil ABNF:

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



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:



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

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

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







कीी कीी

6. M is not permitted after V Example:

ागा ईा







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- क-

க்1 क1

- 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
С	Ц	Шq.
СН	க்	காக்கா
СМ	கி	கிடங்கு
CX	மஃ	மஃகான்
СНС	ம்ம	அம்மா
СНСНС	ர்க்க	சேர்க்கை
V	<u> </u>	இப்ப
VX	<u></u> %،	ചു ം നിത്ത





4. LANGUAGE TABLE¹: TAMIL²

	0B8	0B9	0BA	0BB	0BC	0BD	0BE	0BF
0		සිදුම		<u>П</u>	°	@		m me
1				<u> </u>	0801			MT 08F1
2	° 0882	ල ම		လ	್ರಿಣ			æ.
3	0 00 0883	& @	600T	ണ				<u>a</u>
4		௸	৸ ৢ	9				# P
5	र्ज ≝	· В В		ഖ				®
6	क्∰			UTO 0886	୍ଦ		O OBE6	⋽
7	(S)			ஷ	ී	ാണ ₀₈₀₇	₽	6T₅u
8	□ •		₩ ₩	സ	ത		Q OBES	© Øys
9	<u>و</u>	<u></u>	60T	ബ്ബ			[<u>F</u>]_	£
Α	<u>୭୩</u>	ф #	LI 08AA		ோ ‱		日	a H
В					ோ		(F)	
С		# 			ெள		5 ≋	
D					: 0800		6T	
E	61	⊕	LD OBAE	ା∏ ଆଞ୍ଚଳ			의 #	
F	6J 088F	L	ULJ OBAF	ി			## ##	

 $^{^1}$ This language table is based on Unicode Chart for Tamil script provided by the Unicode Consortium. 2 Characters marked in yellow are not applicable to the language.







5. NOMENCLATURAL DESCRIPTION TABLE OF TAMIL LANGUAGE TABLE

VISARGA/AYTHAM (X)				
0B83	o o o	TAMIL SIGN VISARGA		
VOWEL LETT	ERS (V)			
0B85	୬	TAMIL LETTER A		
0B86	<u></u>	TAMIL LETTER AA		
0B87	<u> </u>	TAMIL LETTER I		
0B88	FF:	TAMIL LETTER II		
0B89	ഉ	TAMIL LETTER U		
0B8A	<u>ව</u> ස්	TAMIL LETTER UU		
0B8E	ଗ	TAMIL LETTER E		
0B8F	ø.	TAMIL LETTER EE		
0B90	නු	TAMIL LETTER AI		
0B92	ල	TAMIL LETTER O		
0B93	ඉ	TAMIL LETTER OO		
0B94	ଋୁଗୀ	TAMIL LETTER AU		
CONSONANTS (C)				
0B95	ಹ	TAMIL LETTER KA		





0B99	ाग्न	TAMIL LETTER NGA
0B9A	ਲ	TAMIL LETTER CA
0B9C	වියි	TAMIL LETTER JA
0B9E	6	TAMIL LETTER NYA
0B9F	L	TAMIL LETTER TTA
0BA3	ഞ്	TAMIL LETTER NNA
0BA4	5	TAMIL LETTER TA
0BA8	Щ	TAMIL LETTER NA
0BA9	ன	TAMIL LETTER NNNA
0BAA	Ц	TAMIL LETTER PA
0BAE	ΙD	TAMIL LETTER MA
0BAF	Ш	TAMIL LETTER YA
0BB0	Ţ	TAMIL LETTER RA
0BB1	Б	TAMIL LETTER RRA
0BB2	ର	TAMIL LETTER LA
0BB3	ଗୀ	TAMIL LETTER LLA
0BB4	ĥ	TAMIL LETTER LLLA
0BB5	ഖ	TAMIL LETTER VA
0BB6	ПΩ	TAMIL LETTER SHA





0BB7	ஷ	TAMIL LETTER SSA	
0BB8	സ	TAMIL LETTER SA	
0BB9	ത്ത	TAMIL LETTER HA	
VOWEL SIGNS	S (MATRAS) (M)		
0BBE	ОП	TAMIL VOWEL SIGN AA	
0BBF	ી	TAMIL VOWEL SIGN I	
0BC0	్	TAMIL VOWEL SIGN II	
0BC1	्र	TAMIL VOWEL SIGN U	
0BC2	ਾ ਉ	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	ெள	TAMIL VOWEL SIGN AU	
VIRAMA (H)			
0BCD	ċ	TAMIL SIGN VIRAMA	







6. VARIANT TABLE

VARIANT				
ලු 6 T	ලු ග්			
0B92+0BB3	_{0B94}			







7. EXPERTS/BODIES CONSULTED

Expertise provided by C-DAC Thiruvananthapuram.







8. PROPOSED ccTLD FOR TAMIL

India (Bhārat) localized in Tamil - இந்தியா

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



