Draft Policy Document
For

INTERNATIONALIZED DOMAIN NAMES

IN INDIAN LANGUAGES

Version History

1.0	Draft document	Feb.-march 2009	Initial draft for clarifications
1.1	Clarifications added and received	March 2009	Modified suitably
1.2.	Working Document	April 2009	Modified suitably
1.3	Working Document	July 2009	Modified suitably
1.4	Working Document	July 2009	Modified Suitably
1.5.	Working Document	October 2009	Modified Suitably
1.6	Working Document	December 2010	Modified Suitably
1.7	Working Document	April 2011	VHCM Rule added for Bangla and Assamese. page number 12, ABNF Structure
1.8	Working Document	May 2012	1. Added languages Bodo(Boro), Dogri, Maithili and Santhali 2. Added category S. 3. Added Avagraha as valid Halanta succeeding character.

POLICY DOCUMENT FOR CREATION OF LANGUAGE INFORMATION FOR INTERNATIONALIZED DOMAIN NAMES IN INDIAN LANGUAGE

This document is a white paper which attempts to lay down the policy document for the creation of internationalized domain names in Indian languages. It is the result of discussions, email exchanges as well as document formalizations over the past months in order to arrive at a working draft which is proposed in what follows.

The document comprises the following sections
1. A broad statement of IDN providing the Background of IDN.
2. Objectives.
3. Policy in brief.
4. The Broad policy statement to serve as a guideline.
5. Technicalities Statement which defines what the IDN structure shall be as based on the Backus-Naur Formalism (BNF) suitably modified to suit the requirements of the IDN.
6. Some Examples to show valid ABNF Label.
7. References.

1. BACKGROUND

In this age of Information Technology (IT) when the entire Globe is being integrated into a web-linked village with the knowledge as the sole differentiator, development of convivial (i.e. natural, convenient, and at the same time, affordable) Access Technology has gained prime importance. Especially for India, with its diverse and multi-lingual heritage and culture, Internet is expected to play dominant integrating role for integrating all most all aspects of social and economic endeavor.

Normally, in Internet operation host name of the target Web Server is submitted to the browser who then sends a request to the Domain Name System (DNS) Resolver Service for translating it into the corresponding Internet Protocol (IP) Address for establishing a physical connection to that Web Server.

Today, the above addressing mechanism supports "Simple English Latin" alphabet of the 26 "Usual" un-accented Latin letters, the 10 "Arabic" Digits (i.e. 0-9), and hyphen with dot (i.e. ".'") as field separators. In addition, capitalization of letters (i.e. upper or lower case) would be insignificant, so that the strings "Bharat", "BHARAT" and "BHArat" are all same.

However, for a number of crucial Customer-centric Applications (such as e-governance, e-learning, e-commerce) sole dependence on a single language (i.e. English) may not be sufficient to provide the requisite infrastructural support to all kinds of Internet usage in present and future. Till recently, there was no standardized method for specifying Domain Names in any non- "Simple English Latin" (say, in ASCII) Character Set.

Thus, there is a dire need for the common Internet Infrastructure to support local languages using their respective written scripts for expressing the Domain Names; such Internationalization (termed as "Internationalized Domain Names (IDNs) is being termed as the associated key process of supporting such requirements towards domain names in multilingual scripts with their respective linguistic as well as cultural sensitivities in the Internet Infrastructure.

One of the most significant innovations in the Internet since its inception will be the introduction of top level Internationalized Domain Names (IDNs). These will offer many new opportunities and benefits for Internet users around the world by allowing them to establish and use domains in their native languages and scripts.

IDNs provide a convenient mechanism for users to access Web sites in local language; for example: if a person wants to give his or her system domain name in his or her local language, say Hindi, then that will look like www.भाषा.भारत

The development of this policy culminating in this white paper has been the outcome of long discussions with linguists and experts. C-DAC GIST Pune in close collaboration with DIT and with the cooperation of C-DAC Kolkata and C-DAC Thiruvananthapuram

has evolved the policy of implementation of Internationalized Domain Names in Indian Languages and also tried to institute safe-guards and safety measures through methods such as defining the Brahmi Syllable pertinent to IDN, providing restriction rules as well as variant tables to make IDN's in Indian languages as secure as possible. This white paper has been formulated by C-DAC GIST after feedback from other bodies and experts.

2. OBJECTIVES

The main objectives of this white paper are to demystify IDN policy by providing information regarding the broad policy, furnishing details of the Backus-Naur Formalism which is the backbone of the Brahmi syllable and suitably modified for IDN (termed hitherto as ABNF).

Given the complexity of Indian writing systems, pharming and spoofing can be a major threat. As palliative measures restriction rules have been introduced.

In a nutshell, the main objectives are as under:

    a. To ensure that Indian languages can have their rightful place in Internationalized Domain Names and that one can have a URL in an Indian language.

    b. To initially permit such URL's in the following major languages/scripts: Devanagari (Marathi, Hindi, Konkani, Sanskrit and Nepali, Bodo, Dogri, Maithili, Santhali, Sindhi), Gujarati, Oriya, Punjabi, Malayalam, Tamil, Telugu, Kannada, Assamese, and Bangla and subsequently to be adapted for all the 22 official languages including those using Perso-Arabic scripts: Urdu, Sindhi and Kashmiri.

    c. To limit, at present the Indian language component to the Domain Name and localize the ccTLD.

3. POLICY IN BRIEF

Following are the general policy guidelines in case of Indian domain names:

1. Only letters, digits, and hyphens will be allowed in a domain name. Names cannot begin or end with hyphens.
2. Mixing of two scripts will not be allowed.
3. Use of Zero Width Joiner/Zero Width Non Joiner will not be allowed.
4. Language numerals and punctuations will not be allowed.
5. Symbols or stress markers will not be allowed.
6. Consecutive hyphens will not be permitted in a domain name.
7. The number of identical consonants joined by a Halant within a label shall not exceed two. Thus त्त (ta+halant+ta) is permitted but not त्त्त (ta+halant+ta+halant+ta).
8. Wherever a variant is present in a given label, the variants shall be in a relationship of transitivity but the generation of the variant table shall be limited only to the relationship existing between the two variants. Thus given a variant त

and त्त, the number of variants in label such as किताब shall be क्िताब. कित्ताब generated by adding an extra त् to त्त shall not be permitted. This ensures that over generation does not take place.

9. 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)

4. BROAD POLICY

The broad policy enunciates the major guide-lines laid down for creation of IDN's in Indian languages and which will be of use to the registrars as well as all entities and organizations involved in allotment or monitoring or use of IDN. The policy guide-lines are a series of dos and don'ts which stipulate reference rules to be followed in the creation of IDN's. In addition they also handle issues such as Zero Width Joiner, Variant Tables etc. The policy guide-lines have been enunciated with the major aim of ensuring that as far as possible phishing, spoofing and pharming shall be eliminated from IDN's in Indian languages.

The broad policy guide-lines are as under:

4.1. CODE SET: UNICODE COMPLIANCY

The layouts shall be Unicode 5.1 compliant in anticipation of the same being implemented by ICANN. This will permit inclusion of Chillu Characters in Malayalam which is the latest addition to the Unicode with regard to Indian Scripts. Eventual up gradations may be visualized as and when Unicode adds new characters.

4.2. SCRIPT AND LANGUAGE: DIFFERENTIATION OF SCRIPTS AND LANGUAGES

A major decision taken is that Scripts and Languages will be differentiated at the registrar's level and that the user will be provided with keyboards which will allow him to enter an IDN in the language of his choice. Although this does not affect languages like Gujarati or Tamil where there is the relationship of One script <-> One language, it does make a difference in scripts such as Devanagari or Bangla, where one script caters to many languages e.g. Hindi, Marathi, Konkani, Nepali, Sanskrit, all use Devanagari.

4.3. BASIC STRUCTURE OR CANONICAL FORM

The basic structure of the IDN which will be discussed at length in Part II is based on the notion of the syllable, which in turn is defined by a Backus-Naur Formalism. Within this formalism, certain entities shall be permitted and others disallowed:

A. PERMISSIBLE ENTITIES
Letter-Hyphen-Digit shall be the only entities permitted. Hyphen and Digits shall belong to the Latin set. Letters will be of the language in question.
e.g. *1 2 3 4 5 6 7 8 9 0* and **–** will be of the Latin set.
All letters (characters) will be of the pertinent script.

B. NOT PERMISSIBLE
1. CODE-PAGE MIXING

No mixing of scripts at a given level will NOT be allowed

e.g. www.सोफ़ट-वे२.in or www.हिन्दी-Hindi.in is not permissible since Hindi

and Gujarati are mixed together and   Hindi and English are mixed together
respectively.

2. DIGITS
Digits in Indian languages will NOT be allowed.

०१२३४५६७८९

3. PUNCTUATION MARKERS
Punctuation markers present in Indian languages such as danda and double

danda ।॥ will NOT be allowed.

4. OTHER SYMBOLS AND ABBREVIATIONS
Since IDN deals only with basic characters, abbreviations and other iconic

characters like Isshar (ॕ), Abbreviation sign (॰) etc. will NOT be allowed.

5. RARE AND OBSOLETE CHARACTERS

Characters which have been added to code-charts to accommodate rare forms

especially long vocalic RR and long vocalic LL ॡ ॠ as well as their matra

forms ॄ and ॣ. In some languages such as Marathi the short vocalic L is

permitted ॢ and used especially as a Matra. This will be permitted for Marathi.

6. STRESS MARKERS OF CLASSICAL SANSKRIT AND VEDIC

Stress markers e.g. Swarita ◌॑ and Udatta ◌॒ will NOT be allowed.

7. SINGLE DIGIT AND COMBINATION OF TWO DIGITS
Single digit (e.g. 1,2,3,4 etc.) and Combination of two digits (for ex 12, 23, 34
etc.) will NOT be allowed as per the .in registry. According to .in registry ".IN
domain names may be between 3 and 63 characters in length". All other rules
pertaining to .in will be followed.

4.4. SAFEGUARDS

To protect as far as possible against spoofing, phishing and pharming attacks the
following safeguards have been introduced. These attacks take place by
substituting an address which looks visually alike but which in fact is a fake URL.
It should be remembered that the browser window allows for a font size which is
relatively small and hence can lead to visual spoofing.

Three such cases of visual identity are possible and safeguards have been instituted against each:

A. DISALLOWING ZERO WIDTH JOINER AND NON-JOINER
The use of Zero width Joiner / Zero width non Joiner (vide RFC 3454 Zero width non joiner (200C)/ zero width joiner (200D) shall NOT be permitted. This is done to avoid spoofing. Use of ZWJ/ZWNJ can result in the following cases, all of which look visually alike.

महाराष्ट्र

महाराष्ट्र with zero width joiner after हा

महाराष्ट्र with zero width non-joiner after म

B. VARIANT TABLE
The aim of the Variant table is to identify visual look-alikes or homographs and ensure that such homographs shall not be permitted. A common case of visual look-alike is the case of द्ध ध्द

It is precisely to protect against such visual identity or homographs that a variant table has been instituted. The function of the variant table is to allow one of the homographs, debarring the other one. First use of either one of the characters shall automatically disallow the other in the case of a given word.

Thus if a user chooses समरुद्धी, the choice will automatically debar समरुध्दी,

protecting against possible spoofing.

The following rules determine variant tables:

1. Since exclusion tables based on variants can debar a large number of words commonly used, the variant table shall be used sparingly and only when absolutely necessary.
2. Further the variant table shall apply only to ligatures or conjuncts or combination of two or more consonants and single characters that have homographic identity shall not be part of the variant table, the logic being that a native speaker can easily disambiguate single characters. It is the conjunct forms that pose the maximum problems

4.5. LANGUAGES

This document only caters to the following languages:

Assamese	Maithili
Bangla	Malayalam
Bodo (Boro)	Marathi
Dogri	Nepali
Gujarati	Oriya
Punjabi	Sanskrit (Laukik)
Hindi	Sindhi-Devanagari
Kannada	Tamil
Konkani	Telugu

Languages using Perso-Arabic script: Urdu, Sindhi, Kashmiri shall be handled later, because of intrinsic complexities of the diacritics.

For each language a detailed layout will be presented as explained in 6. below:

4.6. LAYOUT

Each language will have the following layout:

1 The generic syllable structure provided by the BN Formalism shall be suitably modified to respective language.
2 Restriction rules if such rules apply.
3 Sample Examples.
4 A Code-chart for each language as per Unicode 5.1 shall be provided. Characters which are not in consonance with the LHD policy and which are to be excluded shall be clearly marked on the code-chart
5 A map of the above code-chart specifying accurately the above code-chart shall be provided.
6 Finally to reduce the risk of spoofing a variant Table will be provided where the possible variants shall be listed. **As far as possible these variants shall not be individual characters but ligatures that are close homographs.**

4.7. DIALOGUE WITH OTHER NATIONS USING THE SAME SCRIPT

Some Indian scripts are shared across geographical boundaries (i.e. Bangla is used in Bangladesh and India, Urdu in Pakistan and India, Tamil in Singapore and India, Nepali in Nepal and India) a collaborative effort can be done to avoid the confusion among the users of the corresponding language or script community.

4.8. PUBLIC REVIEW

The final document so prepared for all major languages shall be put up on a site for comments and also circulated to obtain maximum feedback.

4.9. VARIANT TABLES TO BE DETERMINED AS PER THE SCRIPT

The Variant tables will be determined by the Script. However for the sake of convenience and ease of reference, it is desirable that wherever more than one languages use a single script, separate variant tables be provided for each language.

5. TECHNICALITIES

1. BACKUS-NAUR FORMALISM EXPLAINED WITH EXAMPLES

Based on the behavior of the Brahmi syllable, the formalism with suitable emendations has bee successfully deployed both in ISCII and in UNICODE and has been adopted by both standards for explaining the behavior of the syllable. Suitably modified to suit the requirements of IDN norms of Letter-Hyphen-Digit, the BN formalism becomes the backbone on which the IDN structure relies.

2. ABNF Formalism:

The ABNF (Augmented Backus-Naur Formalism) is made generic to support all the languages coming under IDN Project currently. The restriction rules take the ABNF from generic form to language specific form so that it fully satisfies the language specific norms.

When applied to IDN, the Backus-Naur formalism results in the following formalism, as shown step-wise:

a. Naming of Variables:

Dash   → Hyphen -
Digit   → Indo-Arabic digits [0-9]
C   → Consonant
V   → Vowel
M   → Matra
D   → Anusvara/Bindi/Tippi/Sunna
B   → Chandrabindu/Anunasika/Arasunna

X   → Visarga/Aytham
H   → Halant/Chandrakala/Virama
A   → Addak
N   → Nukta
Y   → Avagraha/Praslesham
L   → Chillu
Z   → Khanda ta
S   → Modifier Letter Apostrophe
k   → Number of Consonant Halanta Sequence

b. **Comparison Chart**

The following chart shows, whether the variable is present in the language or not.

	C	V	M	D	B	X	H	A	N	Y	L	Z	S
Hindi	✓	✓	✓	✓	✓	✓	✓		✓	✓			
Marathi	✓	✓	✓	✓	✓	✓	✓			✓			
Konkani	✓	✓	✓	✓	✓	✓	✓			✓			
Nepali	✓	✓	✓	✓	✓	✓	✓			✓			
Sanskrit	✓	✓	✓	✓	✓	✓	✓			✓			
Oriya	✓	✓	✓	✓	✓	✓	✓		✓	✓			
Punjabi	✓	✓	✓	✓		✓	✓	✓	✓				
Gujarati	✓	✓	✓	✓	✓	✓	✓			✓			
Bangla	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	
Assamese	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	
Tamil	✓	✓	✓			✓	✓						
Telugu	✓	✓	✓	✓	✓	✓	✓			✓			
Kannada	✓	✓	✓	✓		✓	✓			✓			
Malayalam	✓	✓	✓	✓		✓	✓			✓	✓		
Bodo(Boro)	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓
Dogri	✓	✓	✓	✓			✓		✓	✓			✓
Maithili	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓
Sindhi-Devanagari	✓	✓	✓	✓			✓		✓				

c. ABNF Operators:

Sr. No.	Symbols	Functions
1	" | "	Alternative
2	" [ ] "	Optional
3	" * "	Variable Repetition
4	" ( ) "	Sequence Group

 d. ABNF Structure:

consonant-syllable→

*k(C[N]H) C[N] $[$H|D|B|X|S|BD|BX|M[D|B|X|BD|BX]$]$

$|$ [CH]Z

$|$ L$[$HC[D|H|M[D]]$]$

$|$ AC[D|X|M[D|X]]

vowel-syllable → V$[$D|B|X|S|BD|BX|N|HCM[D|B|X|S|BD|BX]$]$

Syllable → consonant-syllable [Y] $|$ vowel-syllable[Y]

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

e. Restriction Rules:

Restriction rules are the additional filters which when applied to generic ABNF, it results to a Language specific ABNF.
Language wise restrictions are as follows:

Hindi:
1. Maximum permissible number of consonants to form a syllable up to 4 hence k = 3.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).
5. Nukta will be allowed only after following characters:

    क (0915)

ख (0916)

ग (0917)

ज (091C)

ड (0921)

ढ (0922)

फ (092B)

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

Marathi:
1. Maximum permissible number of consonants to form a syllable up to 4 i.e k = 3.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).
5. With consonant र्‍(0931), only following combinations will be allowed.

   र्‍य → र्‍(0931) ्‍ (094D) य (092F)

   र्‍ह → र्‍(0931) ्‍ (094D) ह (0939)

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

Konkani:
1. Maximum permissible number of consonants to form a syllable up to 4 i.e k = 3.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).

5. With consonant र्‍(0931), only following combinations will be allowed.

   र्‍य → र्‍(0931) ्‍ (094D) य (092F)

   र्‍ह → र्‍(0931) ्‍ (094D) ह (0939)

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

Nepali:
1. Maximum permissible number of consonants to form a syllable up to 4 i.e k = 3.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).

5. With consonant र्‍ (0931), only following combinations will be allowed.

   य्‍य → र्‍(0931) ्‍ (094D) य (092F)

   ह्‍ह → र्‍(0931) ्‍ (094D) ह (0939)

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

Sanskrit:
1. Maximum permissible number of consonants to form a syllable up to 5 i.e k = 4.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).
5. A consonant syllable that is intended to end with Halant [H] can only be followed by Hyphen, digit or Avagraha.

Oriya:
1. Maximum permissible number of consonants to form a syllable up to 4 i.e k = 3.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).
5. Nukta will be allowed only after following characters:

   ଡ (0B21) and ଢ (0B22)

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

Punjabi:

1. Maximum permissible number of consonants to form a syllable up to 2 i.e. $k = 1$.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).
5. Halant shall be permitted only with the following:

   Consonant + ੍(0A4D) + ਯ(0A2F)

   Consonant + ੍(0A4D) + ਰ(0A30)

   Consonant + ੍(0A4D) + ਵ(0A35)

   Consonant + ੍(0A4D) + ਹ(0A39)

6. Addak : It is used for germination

   ਅੱਕਾ   अक्का

   4.1 Addak shall not be permitted either at the beginning or at the end

   of the word: ੱ॑ਕ

   4.2 Addak will not be permissible after Halant, Visarga or Bindi/Tippi.
7. Tippi will be used in place of Bindi if the preceding character is one of the following:
   i. A Consonant

   ii. A Consonant followed by Nukta ਼ (0A3C).

   iii. GURMUKHI VOWEL SIGN I ਿ (0A3F)

   iv. GURMUKHI VOWEL SIGN U   ੁ (0A41) and

   GURMUKHI VOWEL SIGN UU   ੂ (0A42)

   v. GURMUKHI LETTER A ਅ (0A05) and

   GURMUKHI LETTER I ਇ (0A07)

8. Nukta will be allowed only after following characters:

   ਖ (0A16)

   ਗ (0A17)

ਜ (0A1C)

ਫ (0A2B)

ਲ (0A32)

ਸ (0A38)

Gujarati:
1. Maximum permissible number of consonants to form a syllable up to 4 i.e. k = 3.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).
5. A consonant syllable that is intended to end with Halant [H] can only be followed by Hyphen, digit or Avagraha.

Bengali:
1. Maximum permissible number of consonants to form a syllable up to 4 i.e. k = 3.

2. CH will be permissible with Khanda Ta only if C is র (09B0).

3. Khanda Ta will not be allowed at the beginning of an IDN label.
4. Nukta (N) will not be permissible after a vowel (V).
5. Nukta will be allowed only after following characters:

ড (09A1)

ঢ (09A2)

য (09AF)

6. Only following combination with VHCM will be allowed.
অ্যা → অ(0985) + ্(09CD) + য(09AF) + া(09BE)

এ্যা → এ (098F) + ্(09CD) + য(09AF) + া(09BE)

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

Assamese:

1. Maximum permissible number of consonants to form a syllable up to 4 i.e. k = 3.

2. CH will be permissible with Khanda Ta only if C is ৰ (09F0).

3. Khanda Ta will not be allowed at the beginning of an IDN label.
4. Nukta (N) will not be permissible after a vowel (V).
5. Nukta will be allowed only after following characters:

   ড (09A1)

   ঢ (09A2)

   য (09AF)

6. Only following combination with VHCM will be allowed.

   অ্যা → অ(0985) + ্(09CD) + য(09AF) + া(09BE)

   এ্যা → এ (098F) + ্(09CD) + য(09AF) + া(09BE)

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

Tamil:
1. Maximum permissible number of consonants to form a syllable up to 3 i.e. k = 2.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).
5. Visarga/Aytham ஃ(0B83) will not be allowed after Matra.

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

Telugu:
1. Maximum permissible number of consonants to form a syllable up to 3 i.e. k = 2.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).
5. A consonant syllable that is intended to end with Halant [H] can only be followed by Hyphen, digit or Avagraha.

Kannada:

1. Maximum permissible number of consonants to form a syllable up to 4 i.e. k = 3.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).
5. A consonant syllable that is intended to end with Halant [H] can only be followed by Hyphen, digit or Avagraha.

Malayalam:

1. Maximum permissible number of consonants to form a syllable up to 4 i.e. k = 3.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta (N) will not be permissible after a vowel (V).
5. "H" will be permitted after only one "L" i.e.ൻ (0D7B) and the following consonant must be റ (0D31).
6. A consonant syllable that is intended to end with Halant [H] can only be followed by Hyphen, digit or Avagraha.

Bodo (Boro):

1. Maximum permissible number of consonants to form a syllable up to 3 hence k = 2.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta will be allowed only after following characters:

    अ (0905)
5. A consonant syllable that is intended to end with Halant [H] can only be followed by Hyphen, digit or Avagraha.

Dogri:

1. Maximum permissible number of consonants to form a syllable up to 3 hence k = 2.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta will be allowed only after following characters:

    ज (091C)

    ड (0921)

ढ (0922)

फ (092B)

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

Maithili:

1. Maximum permissible number of consonants to form a syllable up to 3 hence k = 2.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta will be allowed only after following characters:

ड (0921)

ढ (0922)

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

Sindhi-Devanagari:

1. Maximum permissible number of consonants to form a syllable up to 3 hence k = 2.
2. BD and BX combinations will be Non-existent.
3. HCM combination in Vowel-syllable will be Non-existent.
4. Nukta will be allowed only after following characters:

क (0915)

ख (0916)

ग (0917)

ज (091C)

ड (0921)

ढ (0922)

फ (092B)

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

6. Some Examples showing valid ABNF Labels:

Below are some sample IDN Labels which are permissible.

1. Generic combination permitted in all languages.

a. Consonant+Matra (CM)

Language	IDN Label
Hindi	ताल
Gujarati	શિક્ષા
Punjabi	ਗੁਰੂ
Oriya	ଦିନ
Bengali	চাল
Tamil	கிடங்கு
Telugu	మాలా
Malayalam	കിണര്
Kannada	ಇಗ

2. Some special cases:

a. Malayalam:
   Chillu+Halant+Consonant+Matra (LHCM)

ആന്റി

b. Bengali:
   Letter Ra+Halant+Khanda Ta (CHZ)

ভর্সনা

   Consonant+Matra+Chandrabindu+Anusvar (CMBD)

হ্যাঁংচা

c. Punjabi :
 Consonant+Addak+Consonant+Matra (CHCM)

ਪੱਕਾ

7. References

Bureau of Indian Standards. *Indian Standard. Indian Script Code for Information Interchange.* IS 13194:1991 New Delhi, 1993.

Dillinger. D., *The Alphabet. A Key to the History of Mankind.* 3rd Edition in 2 Volumes. Hutchison. London. 1968.

ICANN: www.icann.org

IS 10401: *8-bit code for information interchange.* 1982

IS 10315: *7-bit coded character set for information interchange.* 1985

IS 12326: *7-bit and 8-bit coded character sets-Code extension techniques.* 1987

ISO 15919, *Information and documentation - Transliteration of Devanagari and related Indic scripts into Latin characters.* 2001

ISO 2375: *Procedure for registration of escape sequences.* 2003

ISO 8859: *8-bit single-byte coded graphic character sets -* Parts 1-13. 1998-2001

Library of Congress. *Romanization Standards..* USA. 2002

RFC 3454: www.ietf.org/rfc/rfc3454.txt

RFC 3492: RFC for PUNYCODE www.ietf.org/rfc/rfc3492.txt

Satpute. P. Open Type Font processing: some issues. *Proceedings of the NWCT.* 2007

Stone. Anthony., http://homepage.ntlworld.com/stone-catend/trind.htm*:*

Unicode Consortium. Unicode ver.3.0.

--- Unicode ver.3.2.

--- Online version of Unicode ver.4.1. (Archived).

---- Online version of Unicode ver.5.0 & 5.1. (www.unicode.org).