

INTERNATIONALIZED DOMAIN NAMES

**IDN:
TECHNICAL OVERVIEW**

IDN POLICY OVER-VIEW

AGENDA

- Why This Presentation ?
- Why a Policy ?
- Overview of the Policy or How it all Gels together.

Why This Presentation

Why This Presentation ?

Corporate
Pro

Why This Presentation

The aim of this presentation is to provide a synthetic overview of the Policy which will facilitate comprehension of the different technical presentations that will follow and allow for situating them in a holistic picture.

In other words this presentation shows how things gel together and where all the nuts and bolts of the policy fit to create one coherent system.

Why a Policy ?

Why a Policy

Why not have an open-sky policy and allow the user to register his domain name without any restrictions ? In other words why evolve a policy at all?

The need for a coherent policy for IDN's in Indian languages is justifiable for the following major reasons:

1. **Security:** To ensure as far as possible spoofing, pharming and phishing attacks do not take place.
2. **Complexity of Indian Scripts:** To handle the complexities of Indian Language writing systems derived from Brahmi which follow a particular syllabic structure. In other words ensure a well-formed syllable.
3. **Language vs. Scripts:** To give priority to languages over scripts and ensure that eventually all 22 languages will be IDN enabled.
4. **Sharing of a Script:** In scripts such as Devanagari shared by more than one language, to ensure that each language is rightfully represented.

Broad Overview of the Policy

- Guiding Principles and Objectives
- Overview Of The Components Of The Policy
- Explanation Of Each of the Components

OBJECTIVES

POLICY: A BROAD OVERVIEW

Objectives-1

In a nutshell, the main objectives are as under:

- To ensure that Indian languages can have their rightful place in Internationalized Domain Names and that one can have a domain name in an Indian Language.
- Permitting such URL's in the 22 official languages including languages using Brahmi based and Perso-Arabic scripts.
- To devise a policy which will allow the Indian user to create and use seamlessly IDN's in his/her mother-tongue.

POLICY: A BROAD OVERVIEW

Objectives- 2

In the implementation of the above, to develop a policy which can ensure that rules for compliance existing .in registry are respected.

- that additional rules required for handling the complexities of Indian languages are identified.
- that correct representation/ a well-formed syllable (akshar) is implemented in the case of Brahmi languages. This is possible and handled by formalism called as Augmented Backus Naur Formalism for Indian Languages.
- that where more than one languages share the same script, pertinent and relevant characters of that script shall be implemented.
- that spoofing and pharming attacks are reduced as far as possible by introducing mechanisms such as Homograph detection and Variant tables.
- The above to be tested thoroughly on different browsers for compliance.

POLICY: A BROAD OVERVIEW

COMPONENTS OF THE POLICY

POLICY: A BROAD OVERVIEW

IDN POLICY FOR INDIAN LANGUAGES

In its broad outline, the policy comprises the following:

A. RULES

1. Generic rules in compliance with .in registry.
2. Rules specific to IDN in Indian languages.

B. DIRECTIVE PRINCIPLES

1. Augmented Backus Naur formalism and language specific restriction rules to ensure a well-formed Brahmi Syllable (Akshar).
2. Variant tables to reduce the risk of spoofing and phishing.
3. Language Tables which will ensure that only those characters pertinent within a given language are used.

Each of these will be taken up in detail in the slides that follow:

RULES

Generic & Specific

POLICY: A BROAD OVERVIEW

RULES:

Two types of rules are built into the policy:

1. **Generic Policy Rules** which are inherited from the .in

Broadly these are as under:

- Letter, Hyphen and Digit alone will be permitted
- Only digit(s) alone in a domain name will NOT be allowed.
- No two consecutive hyphens at third and fourth position will not be allowed.
- A hyphen shall not occur either at the beginning or end of a Domain name.

POLICY: A BROAD OVERVIEW

RULES:

Specific Policy Rules created specifically for IDN's in Indian languages.

These are as under:

1. CODE SET

UNICODE COMPLIANCY

The layouts shall be Unicode compliant in anticipation 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 upgradations may be visualized as and when Unicode adds new characters.

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, where one script caters to many languages e.g. Hindi, Marathi, Konkani, Nepali, Sanskrit, all use Devanagari.

POLICY: A BROAD OVERVIEW

RULES: Specific Policy Rules-2

3. Within the Brahmi syllable (akshar) the following are permissible and non-permissible entities:

PERMISSIBLE ENTITIES:

Letter-Hyphen-Digit shall be the only entities permitted.

Hyphen and Digits shall belong to the Latin set.

e.g. **1 2 3 4 5 6 7 8 9 0** and **–** will be of the Latin set.

Letters will be of the language in question.

e.g. Only Tamil characters shall be permissible within a given domain name in Tamil language.

POLICY: A BROAD OVERVIEW

RULES: Specific Policy Rules-4

NOT PERMISSIBLE

1. CODE-PAGE MIXING

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

e.g. www.हिन्दी-Hindi.in is Not permissible since Hindi and English are mixed together.

2. DIGITS

Digits in Indian languages will NOT be allowed.

०१२३४५६७८९०

3. PUNCTUATION MARKERS

Punctuation markers present in Indian languages such as danda and double danda ।। will NOT be allowed.

POLICY: A BROAD OVERVIEW

RULES: Specific Policy Rules-5

NOT PERMISSIBLE

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 लृ ऋ as well as their matra forms ऌृ . In some languages such as Marathi the short vocalic Ll is permitted ऌ 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.

POLICY: A BROAD OVERVIEW

RULES: Specific Policy Rules-6

NOT PERMISSIBLE:

The use of 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 without ZWJ and ZWNJ

महाराष्ट्र with zero width joiner after हा

महाराष्ट्र with zero width non-joiner after म

POLICY: A BROAD OVERVIEW

ISSUE STILL REMAINS WITH IDNA2008

IDNA 2008 allows ZWJ/ZWNJ within context. still the problem remains as shown below:

Use of ZWNJ:

मिट्टी

(092E 093F 091F 094D 091F 0940)

मिट्टी

(092E 093F 091F 094D **200C** 091F 0940)

Use of ZWJ:

श्रीर

09B9 09A0 09BE **09CE**

श्रीर

09B9 09A0 09BE **09A4 09CD 200D**

पक्का

092A 0915 094D 0915 093E

पक्का

092A 0915 094D **200C** 0915 093E

POLICY: A BROAD OVERVIEW

Corporate
Pro

DIRECTIVE PRINCIPLES

POLICY: A BROAD OVERVIEW

Directive Principles

These are principles that are the foundation of the policy . These are three in number:

1. ABNF formalism and language specific restriction rules to ensure a well-formed Brahmi Syllable.
2. Language Tables which will ensure that only those characters pertinent within a given language are used.
3. Variant tables to reduce the risk of spoofing and phishing

Each will be taken up in detail

POLICY: A BROAD OVERVIEW

1. ABNF Formalism & Restriction Rules

POLICY: A BROAD OVERVIEW

1.1 ABNF Formalism

To ensure that the correct formation of the Brahmi syllable (akshar) is permissible, a study of the structure of the Brahmi syllable pertinent to the languages under survey has been undertaken and applied to the IDN's. Since the Brahmi syllable can be best described in terms of a Backus-Naur formalism, this has resulted in an ABNF (Augmented Backus-Naur formalism) in which both digits as well as Hyphens are incorporated. ABNF ensures that a malformed syllable is not acceptable:

ुददुता

1.2. Restriction Rules

Since ABNF is a generic formalism, restriction rules pertinent to each language or script have been defined. These rules ensures that the ABNF formalism becomes specific to a given language.

e.g. Hindi, all possible Nukta characters have been identified to ensure that no other character will be attached with a Nukta

Valid Nukta Characters: क़,ख़,ग़,ज़,फ़,ड़,ढ़

Invalid Combinations handled by restriction Rules: ढ़ ल़

2. VARIANT TABLES

POLICY: A BROAD OVERVIEW

2. VARIANT TABLES

1. VISUAL LOOK-ALIKES

To reduce spoofing and phishing attempts visual look-alikes within a pertinent script have been identified. These constitute variant tables.

U+0926 U+094D U+0917	रू	U+0926 U+094D U+0930	रु
U+0926 U+094D U+0918	रु	U+0926 U+094D U+0927	रु

2. FUNCTION OF THE VARIANT TABLE

Variant table will aid the registry software to take appropriate policy decision in case the requested domain name has Variant characters. .

e.g. If a user chooses समरुद्ध, the choice will automatically debar समरुद्ध, protecting against possible spoofing.

POLICY: A BROAD OVERVIEW

2. VARIANT TABLES

The following caveats have been considered while determining variant tables:

1. SPARING USE OF THE VARIANT TABLE

Since exclusion tables based on variants can debar a large number of words commonly used, the variant table have been carefully prepared. Thus where two homographs have a very high frequency, they have not been included, since such step would not have allowed registration of a large number of domain names.

2. VARIANT TABLES TO BE LIMITED TO LIGATURES ALONE

Further the variant table include largely only ligatures or conjuncts or combination of two or more consonants. Character set can be a variant of a single character. Single characters that have homographic identity have not been included in the variant table, the logic being that a native speaker can easily disambiguate single characters. It is conjunct forms that can create confusions as shown in the earlier example.

Thus **घ ध** have not been included as variants although in the URL bar of the browser, they have a look-alike identity.

Variants are conditioned by the writing system as well as by Linguistic and Notational considerations (Unicode)

Writing system	UNICODE		LINGUISTIC ISSUES		
	Code-points (P) (H)	Legacy Inputting (P) (H)	Positional (P) (H)	Spell-variants at character level	Look-alik
Abjad	YES	NO	YES	YES	YES
Alphabet	NO	NO	NO	YES	YES
Akshar	YES	YES	NO	NO	YES
Syllabaries	NO	NO	NO	YES	NO
Phonetic- Semantic	NO	NO	NO	NO	YES

POLICY: A BROAD OVERVIEW

2. VARIANT TABLES

3. NORMALISATION

Since Unicode allows for two different ways of entering a given character normalization has been set in place. A good example of normalization is in CodePage 900X: Devanagari where

इः can be written as इ (0921) + ः (093C) or इः (095C)

Such normalization cases have been made part of the variant table.

Although browsers/IDNA are supposed to undertake Normalisation, tests have shown that not all browsers handle normalisation for all Indian languages.

4. EXPERTISE

All variant tables have been thoroughly examined and approved by experts in the respective languages. Further awareness workshops held in different geo-linguistic regions and attended by the common man as well as those interested in IDN, create public consensus.

POLICY: A BROAD OVERVIEW

2. VARIANT TABLES

क+॰ 0915+093C	क 0958
ख+॰ 0916+093C	ख 0959
ग+॰ 0917+093C	ग 095A
ज+॰ 091C+093C	ज 095B
ड+॰ 0921+093C	ड 095C
ढ+॰ 0922+093C	ढ 095D
फ+॰ 092B+093C	फ 095E

द्र 0926+094D+0917	द्र 0926+094D+0930	द्र 0926+094D+0928
द्व 0926+094D+0927	द्व 0926+094D+0918	
ए 0937+094D+091F	ए 0937+094D+0920	
श्व 0936+094D+0935	श्व 0936+094D+0930+094D+0935	
श्र 0936+094D+0928	श्र 0936+094D+0930+094D+0928	
श्व 0936+094D+091A	श्व 0936+094D+0930+094D+091A	
श्र 0936+094D+0932	श्र 0936+094D+0930+094D+0932	
त 0924+094D+0924	त 0924	
ँ 0901	ँँ 0945+0902	
द्व 0926+094D+0935	द्व 0926+094D+092C	

3. LANGUAGE SPECIFIC TABLES

POLICY: A BROAD OVERVIEW

3. LANGUAGE SPECIFIC TABLES

The ABNF policy will be further reinforced by language-specific tables in which ONLY those characters which are pertinent have been permitted. In it, Indian digits, symbols, rare characters such as the Sanskrit long vowels have not been admitted. This will result in a language specific chart in which only those characters which are pertinent will be permissible.

Each language table has been vetted by experts in the respective language.

Language wise charts are provided which specify clearly which characters are permissible in the Unicode 5.1 code-page. The code page will be modified for Unicode 5.2 and later in future.

In addition a chart with classification as per the different entities defined in the ABNF is also appended for ease of use of the developer.

POLICY: A BROAD OVERVIEW

3. LANGUAGE SPECIFIC TABLE FOR MARATHI

The language specific table will comprise the following:

DETAILED LAYOUTS FOR LANGUAGES

DEVANAGARI: MARATHI

This proposal comprises the following Parts:

1. ABNF Formalism
2. Restriction Tables if any
3. Examples
4. The Code-chart for Marathi is compliant with the Unicode 5.1 Code-chart Characters marked in yellow will NOT be considered since they do not fall within the purview of the IDN policies as laid down.
5. The nomenclatural description of the above code-chart, in conformity with the above,
6. A variant table with the most important homographs being identified.
7. Experts/Bodies consulted.

0900

Devanagari

097F

	090	091	092	093	094	095	096	097
0	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
1	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
2	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
3	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
4	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
5	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
6	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
7	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
8	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
9	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
A	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
B	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
C	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
D	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
E	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
F	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ

Marathi

POLICY: A BROAD OVERVIEW

SUMMING-UP

The IDN Policy will therefore have the following lay out:

1. Generic rules in compliance with .in registry.
2. Rules specific to IDN in Indian languages.
3. ABNF formalism and language specific restriction rules to ensure a well-formed Brahmi Syllable (Akshar)
4. Language Tables along with code-chart and clarification of the same in Unicode terminology to ensure that only those characters pertinent within a given language are used.
5. Variant tables to reduce the risk of spoofing and phishing.

The implementation of these will ensure that the Middle and Top level domain registration is a peace of mind solution both to the registrar and the registrant.

THANK YOU