Draft Policy Document

for

INTERNATIONALIZED DOMAIN NAMES

Language: PUNJABI (PANJABI)

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 PANJABI	1.5
1.1	22/12/2010	Page No. 5, 6, 7, 20	M	Remove Explicit Halant from the consonant syllable, ccTLD added	1.6
1.2	05/08/2013	Whole Document	A,M	Modified as per IDNA 2008, Restriction rules added and modified.	
1.3	21/11/2014	Page No 9	A	Modified rule related to Addak	

Table of Contents

1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF)

1.1 Declaration of variables

Dash   → Hyphen -

Digit   → Indo-Arabic digits [0-9]

C   → Consonant

M   → Matra

V   → Vowel

D   → Bindi / Tippi

X   → Visarga

H   → Halant

A   → Addak

N   → Nukta

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 Panjabi 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 Bindi/Tippi (D) or Visarga (X). The number of D or X which can follow a V in Panjabi are restricted to one.

The vowel sequence in Panjabi is therefore,
    V[D|X]
Examples:

V	ਅ	अ
VD	ਅੰ	अं
VD	ਅਂ	आं
VX	ਅਃ	अः

1.4 The Consonant Sequence

A Consonant Sequence admits the following combinations:

1. A single consonant (C)
   Example:

C	ਵ	क

2. A consonant optionally followed by dependent Vowel Sign/Matra [M], Bindi/Tippi [D] or Visarga[X]
        C[M|D|X]
   Example:

CM	ਵਿ	कि
CD	ਵੰ	कं
CX	ਵਃ	कः

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

| CMD | ਵੀਂ | ਕੀਂ |
| CMX | ਕਃ | ਕਾਃ |

3. A sequence of consonants (up to 2) joined by Halant
   (CH)C
   Example:

   CHC　ਕ੍ਰ　ਕ+ੰ+ਰ　ਕ+ੰ+ਰ

4.a. The combination may be followed by M, D or X.
   Example:

CHCM	ਕ੍ਕਿ	ਕ ੍ ਕ ਿ	ਕ੍ਕੀ	ਕ ੍ ਕ ੀ
CHCD	ਕ੍ਕਂ	ਕ ੍ ਕ ਂ	ਕ੍ਕਂ	ਕ ੍ ਕ ਂ
CHCX	ਕ੍ਕਃ	ਕ ੍ ਕ ਃ	ਕ੍ਕਃ	ਕ ੍ ਕ ਃ

4.b. (CH)CM may further be followed by a D or X.
   Example:

| CHCMD | ਕ੍ਕਿਂ | ਕ ੍ ਕ ਿ ਂ | ਕ੍ਕੀਂ | ਕ ੍ ਕ ੀ ਂ |
| CHCMX | ਕ੍ਕਿਃ | ਕ ੍ ਕ ਿ ਃ | ਕ੍ਕੀਃ | ਕ ੍ ਕ ੀ ਃ |

4. Addak is used for germination. It doubles the following consonant.
   Example:
   AC[1]　ੱਕ

5. A sequence of Addak consonant may be followed by
   Bindi/Tippi[D],Visarga[X] or Matra[M]
   Example:
   ACD　ੱਕਂ
   ACX　ੱਕਃ
   ACM　ੱਕਿ

6. A sequence of Addak consonant matra ACM may be followed by
   Bindi/Tippi[D], Visarga[X]
   Example:

---

[1]　Addak falls on the preceding consonant or vowel syllable.

ACMD[2]   ੱਿੰ

ACMX   ੱਿਃ

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

  (C[N]H)C[N][D|X|M[D|X]]|AC[D|X|M[D|X]]

Note: D represents either Bindi or Tippi. Combinations of the type CHCMD are possible only in case of Sanskrit words transcribed in Panjabi such as क्रांती which a user could possibly want to use.

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 the Panjabi IDN

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

Vowel-sequence → V [D|X]
Consonant-sequence → (C[N]H)C[N] [D|X|M[D|X]] | AC[D|X|M[D|X]]
Sequence → consonant-sequence | vowel-sequence
IDN-label → (sequence|digit)*([dash] (sequence|digit))

---

[2]   These combinations are extremely rare in Panjabi and not encountered very frequently.

Additional Examples putting more light on Panjabi ABNF:

Below are some of the examples which will help a casual reader understand some of the rules ABNF puts in place. These are just given for reference purposes and are not meant to be comprehensive.

1. D, X, H, A, or M cannot occur in the beginning of a Panjabi IDN.
   Example:

ਿਕ	ਿਕ
ਂਕ	ਂਕ
ਃਕ	ਃਕ
੍ਗ	੍ਗ
ਁਕ	

As can be seen such combination will result automatically in a "golu" marking it as invalid formation. This is an intrinsic property of the Indian language syllable and is quasi automatically applied wherever supported by the OS.

2. Halant (H) is not permitted after V, D, X, M, Digit or Dash
   Example:

ਅ੍	ਅ੍
ਵਂ੍	ਕਂ੍
ਵਃ੍	ਕਃ੍
ਵਿ੍	ਕਿ੍
1੍	1੍
-੍	-੍

3. Addak

3.1 Addak cannot be followed by a Halant, Visarga, Bindi/Tippi, Digit and Hyphen.

Example:

ਂ॒

ਂਃਂ

ਂਂ

ਂਂ

1ਂ

–ਂ

3.2 Addak is used for germination.

ਅੱਕਾ

4. Number of D permitted after Consonant or Vowel or a Matra is restricted to one. Thus following combinations are invalidated.

   Example:

   ਕਂ੦   ਕੰ੦

   ਕੀਂ੦   ਕੀੰ੦

   ਅਂ੦   ਅੰ੦

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

   Example:

   ਕੀੰੀ   ਕੀੰੀ

6. M is not permitted after V

   Example:

   ੲੀ   ੲ਼ੀਾ

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 Panjabi the following rules apply:

1. Tippi will be used in place of Bindi if the preceding character is one of the followings:

    A Consonant

    A Consonant followed by Nukta �਼ (0A3C).

    Gurmukhi Vowel Sign I ਿ (0A3F)

    Gurmukhi Vowel Sign U ੁ (0A41)

    Gurmukhi Vowel Sign Uu ੂ (0A42)

    Gurmukhi Letter A ਅ (0A05)

    Gurmukhi Letter I ਇ (0A07)

2. The Halant shall be permitted only with the following:

    Consonant + ੍ (0A4D) + ਯ (0A2F)

    Consonant + ੍ (0A4D) + ਰ (0A30)

    Consonant + ੍ (0A4D) + ਵ (0A35)

    Consonant + ੍ (0A4D) + ਹ (0A39)

3. Nukta shall be allowed only after following characters:

    ਖ (0A16)

    ਗ (0A17)

    ਜ (0A1C)

ਢ (0A2B)

ਲ (0A32)

ਸ (0A38)

4. Addak can neither be permitted at the beginning nor at the end of the word.
    Example:
      ੱਕ

5. Consecutive hyphens will not be permitted in a domain name.
6. 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).

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 Panjabi.

3. EXAMPLES

Combination	Example	Word with combination
C	ਕ	ਕਰ
CM	ਰਾ	ਰਾਮ
CD	ਹੰ	ਹੰਸ
CMD	ਹਿੰ , ਨੀਂ	ਹਿੰਦੀ , ਨੀਂਦ
CHC	ਗ੍ਰ	ਗ੍ਰਥ
V	ਅ	ਅਮਨ
VD	ਅੰ	ਅੰਦਰ
AC	ੱਬ	ਰੱਬ
ACM	ੱਕਾ	ਪੱਕਾ

4. LANGUAGE TABLE[1]: PANJABI[2]

	0A0	0A1	0A2	0A3	0A4	0A5	0A6	0A7
0		ਐ	ਠ	ਰ	ਗੀ			ੰ
1	ੱ		ਡ		ੁ	ੱ		ੱ
2	ਂ		ਦ	ਲ	ੂ			ਲ਼
3	ਃ	ੳ	ਣ	ਲ਼				ੲ
4		ਔ	ਤ					ੴ
5	ਅ	ਕ	ਥ	ਵ				ੵ
6	ਆ	ਖ	ਦ	ਸ਼			੦	
7	ਇ	ਗ	ਧ		ੇ		੧	
8	ਈ	ਘ	ਨ	ਸ	ੈ		੨	
9	ਉ	ਙ		ਹ		ਖ਼	੩	
A	ਊ	ਚ	ਪ			ਗ਼	੪	
B		ਛ	ਫ		ੋ	ਜ਼	੫	
C		ਜ	ਬ	ੌ	੍	ੜ	੬	
D		ਝ	ਭ		ੑ		੭	
E		ਞ	ਮ	ੇ		ੜ	੮	
F	ਐ	ਟ	ਯ	ੵ			੯	

---

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

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

5. NOMENCLATURAL DESCRIPTION TABLE OF PANJABI LANGUAGE TABLE

BINDI / TIPPI (D)		
0A02	ਂ	GURMUKHI SIGN BINDI
0A70	ਁ	GURMUKHI TIPPI
VISARGA (X)		
0A03	ਃ	GURMUKHI SIGN VISARGA
VOWEL LETTERS (V)		
0A05	ਅ	GURMUKHI LETTER A
0A06	ਆ	GURMUKHI LETTER AA
0A07	ਇ	GURMUKHI LETTER I
0A08	ਈ	GURMUKHI LETTER II
0A09	ਉ	GURMUKHI LETTER U
0A0A	ਊ	GURMUKHI LETTER UU
0A0F	ਏ	GURMUKHI LETTER EE
0A10	ਐ	GURMUKHI LETTER AI
0A13	ਓ	GURMUKHI LETTER OO
0A14	ਔ	GURMUKHI LETTER AU
CONSONANTS (C)		
0A15	ਕ	GURMUKHI LETTER KA
0A16	ਖ	GURMUKHI LETTER KHA
0A17	ਗ	GURMUKHI LETTER GA
0A18	ਘ	GURMUKHI LETTER GHA

0A19	ਙ	GURMUKHI LETTER NGA
0A1A	ਚ	GURMUKHI LETTER CA
0A1B	ਛ	GURMUKHI LETTER CHA
0A1C	ਜ	GURMUKHI LETTER JA
0A1D	ਝ	GURMUKHI LETTER JHA
0A1E	ਞ	GURMUKHI LETTER NYA
0A1F	ਟ	GURMUKHI LETTER TTA
0A20	ਠ	GURMUKHI LETTER TTHA
0A21	ਡ	GURMUKHI LETTER DDA
0A22	ਢ	GURMUKHI LETTER DDHA
0A23	ਣ	GURMUKHI LETTER NNA
0A24	ਤ	GURMUKHI LETTER TA
0A25	ਥ	GURMUKHI LETTER THA
0A26	ਦ	GURMUKHI LETTER DA
0A27	ਧ	GURMUKHI LETTER DHA
0A28	ਨ	GURMUKHI LETTER NA
0A2A	ਪ	GURMUKHI LETTER PA
0A2B	ਫ	GURMUKHI LETTER PHA
0A2C	ਬ	GURMUKHI LETTER BA
0A2D	ਭ	GURMUKHI LETTER BHA
0A2E	ਮ	GURMUKHI LETTER MA
0A2F	ਯ	GURMUKHI LETTER YA

0A30	ਰ	GURMUKHI LETTER RA
0A32	ਲ	GURMUKHI LETTER LA
0A35	ਵ	GURMUKHI LETTER VA
0A38	ਸ	GURMUKHI LETTER SA
0A39	ਹ	GURMUKHI LETTER HA
0A5C	ੜ	GURMUKHI LETTER RRA

NUKTA (N)

0A3C	਼	GURMUKHI SIGN NUKTA

VOWEL SIGNS (MATRAS) (M)

0A3E	ਾ	GURMUKHI VOWEL SIGN AA
0A3F	ਿ	GURMUKHI VOWEL SIGN I
0A40	ੀ	GURMUKHI VOWEL SIGN II
0A41	ੁ	GURMUKHI VOWEL SIGN U
0A42	ੂ	GURMUKHI VOWEL SIGN UU
0A47	ੇ	GURMUKHI VOWEL SIGN EE
0A48	ੈ	GURMUKHI VOWEL SIGN AI
0A4B	ੋ	GURMUKHI VOWEL SIGN OO
0A4C	ੌ	GURMUKHI VOWEL SIGN AU

VIRAMA (H)

0A4D	੍	GURMUKHI SIGN VIRAMA

ADDAK (A)

0A71	ੱ	GURMUKHI ADDAK

6. VARIANT TABLE

There are some variants associated with each language but there are no variants associated with Panjabi language as such. However, there were some akhand nukta characters which could be generated out by typing the pertinent consonant followed by a Nukta. These characters have been marked DISALLOWED in IDNA 2008 protocol so those characters were removed from the list of Panjabi variant set.

7. EXPERTS/BODIES CONSULTED

The policy for Gurmukhi was submitted to Dr. G.S. Lehal for comments and subsequently refined during the PRSG meeting which Dr. Lehal attended as Language Expert.

8. PROPOSED ccTLD FOR PANJABI

India (Bhārat) localized in Panjabi - ਭਾਰਤ

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