Proposals For Devanagari, Gurmukhi, And Gujarati Scripts' Root Zone Label Generation Rules

Publication Date:	20 October 2018
Prepared By:	IDN Program, ICANN Org

Public Comment Proceeding	
Open Date:	27 July 2018
Close Date:	8 October 2018
Staff Report Due Date:	19 October 2018

Important Information Links

[Announcement](#)
[Public Comment Proceeding](#)
[View Comments Submitted](#)

Staff Contact:	Sarmad Hussain	**Email:**	sarmad.hussain@icann.org

Section I: General Overview and Next Steps

The Neo-Brahmi Script Generation Panel (NBGP) was formed by nine communities that use scripts derived from the Brahmi script. NBGP is developing Root Zone Label Generation Rules (LGR) for Bengali, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Tamil and Telugu scripts. The GP has published the proposals for the LGRs of eight of these scripts in three sets, releasing proposals for the scripts which share cross-script variant code points together to the extent possible. The first set included the following: (1) *Proposal for the Devanagari Script Label Generation Rules for the Root Zone,* (2) *Proposal for the Gurmukhi Script Label Generation Rules for the Root Zone,* and (3) *Proposal for the Gujarati Script Label Generation Rules for the Root Zone.* As per the [LGR Procedure](#), these proposals had been posted for public comments to allow those who did not participate in the NBGP to make their views known. Based on the feedback, the NBGP will finalize each proposal for evaluation and integration into the Label Generation Rules for the Root Zone.

Section II: Contributors

At the time this report was prepared, a total of five (5) community submissions had been posted on the forum. The contributors, both individuals and organizations/groups, are listed below in chronological order by posting date with initials noted. To the extent that quotations are used in the foregoing narrative (Section III), such citations will reference the contributor's initials.

Organizations and Groups:

Name	Submitted by	Initials
Integration Panel	Asmus Freytag	IP
Mitsu.in	Sanjeev Goyal	SG
Nepal Power Investment Corporation	Dinesh Ghimire	NPI
Reverie Language Technologies	Vivekananda Pani	RLT

Individuals:

Name	Affiliation (if provided)	Initials
Liang Hai		LH

Section III: Summary of Comments

General Disclaimer: This section intends to summarize broadly and comprehensively the comments submitted to the public comment proceeding but does not address every specific position stated by each contributor. The preparer recommends that readers interested in specific aspects of any of the summarized comments, or the full context of others, refer directly to the specific contributions at the link referenced above (View Comments Submitted).

IP reviewed the Devanagari LGR Proposal along with the draft Bangla LGR Proposal and makes the following comments:

IP1. IP notes that the Bengali VISARGA is not listed in Appendix B of the Devanagari LGR proposal, while VISARGAs for other scripts are listed. Bangla and Devanagari share at least one consonant variant, the Bangla VISARGA could be used to form labels that only have a small difference in shape between the two VISARGAs. This should be reviewed; if this distinction is enough then it should be included in Appendix B; if this code point represents a security issue, it could be added to the list of cross-script variants.
IP2. The IP notes that the GP very properly does not consider cross-script variants for cases where only combining marks have a shared form.

SG points out the case where two domain names, with-Nukta and without-Nukta, are visually and phonetically similar, e.g. due to Hindi and Urdu language variation. SG emphasizes that this type of issue needs greater attention.

NPI acknowledges the work on LGR proposals and hopes the work is done in time.

RLT finds the proposals are commendable. RLT urges NBGP to continue this good work and extend the work beyond the scope of LGR, e.g. for font definition. RLT makes the following comments on the Devanagari proposal:

RLT1. In 3.3.2, RLT suggests text following the first three sentences is not important. It may add unnecessary confusion.

RLT2. In 3.3.3, RLT suggests that vowels U+090D/ U+0972 should not be treated as same. Marathi uses ॲ (U+0972) instead. These vowels are not taught in the native alphabet in schools and alternate forms are used to write English loan words. Also, these vowels are not used officially or unofficially and further not found frequently in Hindi corpus.

RLT3. In 3.3.4, RLT notes that the normalization definition should be provided. "While the definition in this section states the use of Anusvara with respect to the varga consonants, it doesn't define the same for the rest. Will कंस, कन्स and कम्स be variants of the same or different?"

RLT4. In 3.3.5, the sentence "Present-day Hindi users tend to replace the Chandrabindu by the Anusvara" should be removed. The common errors in spelling should not become the norm.

RLT5. RLT suggests the LGR should disallow Nukta because it can only appear under ड or ढ;. If Nukta is allowed, it should be limited to the valid combinations. Nukta is not used in conjuncts.

RLT6. RLT agrees with disallowing ZWJ. The Halant is the remover of inherent vowel but not a joiner. So, use of Halant may not form conjuncts. More work needs to be done to capture the concept of joining "Yukta" which is taught as part of the language. However, this work needs to be taken up with the Unicode consortium.

RLT7. In section 5.2, RLT points out that the U094A is mis-represented. It should show the matra form.

RLT8. In section 5.2, RLT emphasizes that the use of RRA may not be permitted. In case the panel experts decide otherwise, it must be stated clearly that use of RRA is only for a display variant.

RLT9. In section 5.5.4, RLT reiterates that the Nukta should be disallowed. If the Nukta is allowed, the list of consonants which Nukta can follow should be mentioned. RLT also notes that akshar formation information is useful for rendering implementation, and if the LGR proposal does not intend to address the rending issue this additional information may be removed.

LH reviewed all three proposals and makes the following comments:

Comments on Devanagari Proposal

LH1. LH suggests that the following points or sections should be revised.

(1) The transliteration or phonetic transcription should be reviewed in sections 2, 3.3.1, and 3.3.6, as per the specific comments by LH on these sections.
(2) In section 3.3.3, Table 5 should be reviewed to make consistently list of vocalic sounds, e.g. including vocalic l and ll if vocalic r is included, and also include Kashmiri short vowels if diphthongs are included.
(3) In section 3.3.4, LH notes that bindu/anusvara explanation to be just a sign for a certain nasal feature.
(4) Table 6 the code point repertoire should note that "Indic syllabic category" column is not the Unicode character property of the same name. In row 6, the glyph and name are incorrect.
(5) In section 3.3.6, Nukta is not only used in Perso-Arabic borrowed words but also English words, etc. The section should be updated accordingly.
(6) In section 5.5, "notation" should be used instead of "variables".
(7) In section 6.1, glyphs should be manually drawn to better illustrate the proper rendering.

LH2. LH raises the following questions and discussion points.

(1) In section 3.3.8, Unicode Standards Core Specification does not state a preference for ZWJ encoding, so unclear what is meant by the text "Earlier the ZWJ was recommended …". This should be reviewed and clarified.

(2) In section 6, the text suggesting "no characters/character sequences in Devanagari which can be created …that look exactly alike." should be reviewed as it is incorrect. Specific examples are presented by LH.
(3) In section 6.4, the disposition "blocked" may be too restrictive.
(4) The variant analysis between Devanagari and Bengali seems incomplete. Further suggestions are provided by LH for consideration.
(5) The reason 2 of Case of Eyelash Reph is unclear, as given in Section 7.
(6) The case of V preceded by H is too restrictive, as given in Section 7.

LH3. LH suggests the pattern representation of Devanagari labels as:
C[N][M[N]][B|D|X] | V[N][B|D|X] | C[N]H.

Comments on Gujarati Proposal

LH4. LH suggests using "gujarâtî": transliteration consistently throughout the document.

LH5. In section 3.4.4, LH notes that the spelling alternation is not relevant. Both functions are representations of a nasal sound.

LH6. LH questions the inclusion of U+0A8C GUJARATI LETTER VOCALIC L and U+0AC4 GUJARATI VOWEL SIGN VOCALIC RR, because vocalic rr and vocalic ll are excluded.

LH7. LH suggests the pattern representation of Gujarati labels as:
`C[N][M][B|X] | V[B|X] | C[N]H`.

LH8. In section 6, LH suggests adding the restrictions posed by WLE rules in addition to the MSR in the text of the following statement: "There are no characters/character sequences in Gujarati, which can be created … as per the [MSR] and look exactly alike"

Comments on Gurmukhi Proposal

LH9. LH suggests the following changes in the text:

(1) In section 3, remove some text on history from the proposal as it may not be relevant.
(2) In section 3.3.2, correct the statement suggesting the inclusion of implicit vowel sound with consonants.
(3) In section 3.3.4, use of the term "Suprasegmental" and explanation of nasality should be reviewed, based on the suggestions provided. Also, use of terminology and analysis in sections 3.3.4.2, 3.3.4.3, 3.3.4.4, 3.3.5, 4.1.6 and 5.3 should be reviewed as per the suggestions provided.

LH10. In section 3.3, LH suggests adding explanation on why "ya" is not in the list of consonants /h, r and v/.

LH11. LH notes that Visarga is used for marking abbreviations and it should be clarified in either section 3.3.4.5 or 4.1.3.

LH12. LH comments that the characters referred to in section 4.1.6 are excluded because these are only allowed as separated words and single character TLDs are not allowed. However, such single character words can still occur in multi-word labels.

LH13. LH notes that the GP should also consider streamlining rules and the 'CH' part for the proposals of other scripts, as done for Gurmukhi in section 5.3. Also, in this section the rule given in row 2 of the third table needs adjustment per the details provided.

LH14. LH summarizes the pattern in Gurmukhi labels as: [C[N]{HC}[M] | V] [A|B|D]

LH15. LH suggests that rule 7.6 is too restrictive and motivated by spelling conventions, and provides existing exceptions which may exist.

Section IV: Analysis of Comments

General Disclaimer: This section intends to provide an analysis and evaluation of the comments submitted along with explanations regarding the basis for any recommendations provided within the analysis.

The comments received are being submitted to the Neo-Brahmi Generation Panel for their consideration and incorporation (as required) in the final version of the LGR proposals.