Neo-Brahmi Generation Panel

Guiding Principles

Draft

Version 1.0

18th May 2016

Table of Contents

Neo-Brahmi Generation Panel - Guiding Principles (Draft)

PRELIMINARIES

This is a working document meant for discussion among the Neo-Brahmi Generation Panel members regarding the governing principles for code point selection in the root zone.

Although Script is favoured over Language, to begin with the Generation Panel will initially go language-wise and create the language-wise code point repertoire. For clarity's sake, the chart below sums up the script-language association among the scripts that Neo-Brahmi Generation Panel plans to work upon.

Script	ISO 15924 Code	Language	ISO 639-3 Code
Bengali	Beng	Assamese – অসমীয়া	Asm
Bengali	Beng	Bengali – বাংলা	Ben
Bengali	Beng	Manipuri – মণিপুরি	Mni
Devanāgarī	Deva	Bodo – बड़ो	Brx
Devanāgarī	Deva	Dogri – डोगरी	Dgo
Devanāgarī	Deva	Hindi – हिन्दी	Hin
Devanāgarī	Deva	Kashmiri – कॉशुर, Kạ̄šur, Koshur	Kas
Devanāgarī	Deva	Konkani – कोंकणी,	Knn
Devanāgarī	Deva	Maithili – मैथिली, মৈথিলী	Mai
Devanāgarī	Deva	Marathi – मराठी	Mar
Devanāgarī	Deva	Nepali – नेपाली	Npi
Devanāgarī	Deva	Sanskrit – संस्कृतम्, संस्कृतावाक्	San
Devanāgarī	Deva	Santali/Santhali – संथाली	Sat
Devanāgarī	Deva	Sindhi – सिंधी	Snd
Gujarati	Gujr	Gujarati – ગુજરાતી	Guj
Gurumukhi	Guru	Punjabi – ਪੰਜਾਬੀ	Pan
Kannada	Knda	Kannada – ಕನ್ನಡ	Kan
Malayalam	Mlym	Malayalam – മലയാളം	Mal
Oriya(Odia)	Orya	Odia – ଓଡ଼ିଆ	Ory
Tamil	Taml	Tamil – தமிழ், Tamizh	Tam
Telugu	Telu	Telugu – తెలుగు	Tel

Of the 21 languages as can be seen above, 3 share the Bengali script: Assamese, Bengali and Manipuri; 11 use Devanagari and Gujarati, Tamil, Punjabi, Kannada, Malayalam, Odia and Telugu have individual scripts, albeit evolved from Brahmi.

GUIDING PRINCIPLES

Without being too constraining and yet ensuring the safety and security of the eventual users, the Generation Panel plans to base its character selection on the following broad principles. The main principle is that of Environmental Limitations. These comprise protocols or standards. All further principles are in fact subsumed under these limitations but have been spelt out separately for the sake of clarity.

Neo-Brahmi Generation Panel - Guiding Principles (Draft)

1. **Environment Limitations:**

The code point repertoire for root zone being a very special case, up the ladder in the protocol hierarchies, the canvas of available characters for selection as a part of the Root Zone code point repertoire is already constrained by various protocols layers beneath it. Three main protocols/standards have been identified which act as successive filters:

i. The Unicode Chart:
Out of all the characters that are needed by the given script, if the character in question is not encoded in Unicode, it cannot be incorporated in the root zone code point repertoire. Such cases are quite rare, given the elaborate and exhaustive character inclusion efforts made by Unicode consortium.

ii. IDNA Protocol:
Unicode being the character encoding standard for providing the maximum possible representation of a given script/language, it has encoded as far as possible all the possible characters needed by the script. However the Domain name being a specialized case, it is governed by an additional protocol known as IDNA (Internationalized Domain Names in Applications) protocol. The IDNA protocol introduces exclusion of some character out of Unicode repertoire from being part of the domain names.

Example: Devanagari Letter Qa (क़) is not allowed to be a part of domain name. Its decomposed form, i.e. Devanagari Letter Ka followed by Devanagari Sign Nukta (क+ ़) can be used instead.

iii. Maximal Starting Repertoire:
The Root-zone LGR being a repertoire of the characters which are going to be used for creation of the root zone TLDs, which in turn are an even more specialized case of domain names, the ROOT LGR procedure introduces additional exclusions on IDNA allowed set of characters.

Example: Devanagari Sign Avagraha (ऽ) even if allowed by IDNA protocol, is not permitted in the Root Zone Repertoire as per the MSR.

To sum up, the restrictions start off with admitting only such characters as are part of the code-block of the given script/language. This is further narrowed down by the IDNA Protocol and finally an additional filter in the form of Maximal Starting Repertoire restricts the character set associated with the given language even more.

Since all the three above mentioned restrictions issue from environmental limitations, the Neo-Brahmi Generation Panel categorizes all such restrictions and the exclusions of characters that take place, under "Environment Limitation" condition.

2. Punctuation Marks:
The TLDs being identifiers, punctuation markers present in Indian languages such as danda (।) and double danda (॥) will not be included.

Neo-Brahmi Generation Panel - Guiding Principles (Draft)

3. Symbols and Abbreviations:

Since TLD's deal only with the basic characters of the given script/language, abbreviations, weights and measures and other such iconic characters like Isshar (ⱽ), Abbreviation sign (॰) etc. will not be included.

4. Rare and Obsolete Characters:

There are characters which have been added to Unicode to accommodate rare forms especially like long vocalic RR (ॠ) and long vocalic LL (ॡ) as well as their matra forms (ॄ) and (ॣ). All such characters will not be included. This is in consonance with the Letter principle as laid down in the Root Zone LGR procedure.

5. Stress Markers Of Classical Sanskrit And Vedic:

Stress markers for classical Sanskrit e.g. Swarita (ॉ) and Udatta (ॕ) will not be included. This is also in consonance with the Letter principle as laid down in the Root Zone LGR procedure.

Neo-Brahmi Generation Panel - Guiding Principles (Draft)