

What is Brahmi?

- An ancient script
- Most of the modern scripts in Indian subcontinent have been derived from Brahmi.
 - Geographically the scripts being used in Central Asia, South Asia and South-East Asia
- These scripts are used by multiple language families: Largely by Indo-Aryan and Dravidian

Why Brahmi?

- Despite their variations in the visual forms, the basic philosophy in their usage is common
- They all are "akshar" driven, and follow a specific syntax
 - Analogical reference can be made to Indian National standard, IS 13194:1991 – Section 8
- This syntax being the implicit foundation in representation of these scripts in the digital medium, adherence to the structure acts as a obligatory security consideration even in the case of Internationalized Domain Names.

Why Neo-Brahmi?

- Of all the scripts derived from "Brahmi", not all are in modern usage
- Approach is in consonance with the *"Conservatism Principle"* of the LGR procedure.



Neo Brahmi GP - Current Status

• Mixed bag expertise like linguistic, Unicode

- (Chair) Udaya Narayana Singh Bangla, Maithili,Hindi, English
- ⊙ Anupam Agrawal Hindi, Bangla
- ⊙ Akshat Joshi Hindi, Marathi
- ⊙ Abhijit Dutta Bengali, Hindi
- Mahesh Kulkarni Marathi, Hindi
- ⊙ Neha Gupta Hindi
- ⊙ Nishit Jain Hindi
- Prabhakar Pandey Hindi
- Raiomond Doctor English, Hindi, Marathi, Gujarati
- ⊙ N. DeivaSundaram Tamil
- Shantaram Walawalikar Konkani
- Bal Krishna Bal Nepali
- ⊙ Ganesh Murmu Santali
- Balaram Prasain Nepali

- Rajib Chakraborty Bangla
- Gurpreet Singh Lehal Panjabi
- ⊙ Saroja Bhate Sanskrit
- Shambhu Kumar Singh Maithili
- SwarnaPrabha Chainary Bodo
- ⊙ Ghanashyam Nepal Nepali
- ⊙ Kalyan Vasudeo Kale Marathi
- Shashi Pathania Dogri
- Santhosh Thottingal Malayalam, Sourashtra, Tamil
- Uma Maheshwar G Telugu
- ⊙ Girish Chandra Mishra Odia
- ⊙ K. C. Tikayat ray Odia
- Debajit Sharma Assamese
- ⊙ Basanta Kumar Panda Odia
- ⊙ Arvind Bhandari Gujarati
- Harish Chowdhary Hindi

Neo Brahmi GP – Approach

- There are cases of
 - one script one language
 - one script multiple languages
 - In this case, multiple sub-groups may exist to ensure proper representation of each language
- Each sub-group ideally would comprise of
 - Language expert(s)
 - Community representative(s)

Neo Brahmi GP – Approach





Neo Brahmi GP – Outreach Efforts

- Conducted a workshop in AprIGF-2014 for awareness and call for participation in LGR procedure.
 - Topic: "Bringing diverse linguistic communities together for a unified IDN ruleset"
 - The panel discussion touched upon the various aspects of creation of the LGR for the Neo-Brahmi scripts
 - <u>http://2014.rigf.asia/agenda/workshop-proposals/workshop-proposal-13/</u>
- Participation and presentation in 49th ICANN Public meeting at Singapore
- Participation and presentation in 50th ICANN Public meeting at London

Root LGR procedure

- Fundamental Blocks:
 - Code point repertoire



– Variant Rules

ICANN String Similarity Assessment Tool

– Whole Label Evaluation rules



Previous similar work

- For IDN version of ".in" ccTLD, (.bharat) equivalent in 22 Official Indian Languages, similar exercise had been carried out
- Following things were finalized for each language
 - Permissible set of code points
 - Visually similar variant strings
 - Complex whole label evaluation rules

Revisiting the rules in context of LGR framework

- LGR work is different in following contexts
 - Wider stakeholder group
 - Overarching principles in the LGR procedure
 - Especially *Simplicity* and *Predictability* principles
- This revision however would not change
 - the need for the well-formedness of the label in terms of Akshar formalism

Status of current work

• In the process of finalizing the code-point repertoires:



Future undertakings: Cross-Script Similarities

DEVANĀGARĪ SCRIPT	COGNATE SCRIPT	CODEPOINT IN COGNATE SCRIPT
घ	Gujarati	ધ
U+0918		U+0A98
उ	Gurmukhi	ਤ
U+0909		U+0A24
र	Gujarati	5
U+0930		U+0AAE

- Code point similarity across scripts
- Cases where Devanagari-Gujarati and Devanagari-Gurumukhi strings look similar.



Future undertakings: Whole Label Evaluation Rules

• Most crucial aspect of Neo-brahmi Label Generation Ruleset

• Details in the following slides.

Before starting with the

Whole Label Evaluation Rules for LGR (The global approach)

let us take a look at

Whole Label Evaluation Rules for .bharat policy (The Indian approach)

.bharat policy

- Why understanding the .bharat policy is important?
 - It is founding work connecting IDNs and Indian languages
 - It has been demonstrated and appreciated at various National and International forums
 - It has all the basic components that are required by the "Root LGR" work, albeit in different forms.

Character classification

Components of the Syllable

-Consonants(C) :

क ख ग घ ङ च छ ज झ ज ट ठ ड ढ ण त थ द ध न न य य र र ल ळ ऴ व श ष स ह

-Vowels (V) :

अ आ इ ई उ ऊ ऋ ऎ ऐ ऍ ऒ ओ औ ऑ

-Matras (M) :

<u>ា ि ៝ ្្ ై ॆ ॓ ំ ័ ॊ ो ॏ </u>

-Vowel modifiers (D) :

ँ ं ः ः

–Halant (H) :

् –Nukta (N) :

Formalism at a glance ...



Formalism Illustrated...

• Variables :

$Dash \rightarrow$	Hyphen -	
Digit \rightarrow	Indo-Arabic digits [0-9]	
С	\rightarrow	Consonant
V	\rightarrow	Vowel
Μ	\rightarrow	Matra
D	\rightarrow	Anusvara/Bindi/Tippi/Sunna
В	\rightarrow	Chandrabindu/Anunasika/Arasunna
Х	\rightarrow	Visarga/Aytham
Η	\rightarrow	Halant/Chandrakala/Virama
А	\rightarrow	Addak
Ν	\rightarrow	Nukta
Y	\rightarrow	Avagraha/Praslesham
L	\rightarrow	Chillu
Ζ	\rightarrow	Khanda Ta
k	\rightarrow	Number of possible Consonant Halanta Sequences

Formalism Illustrated...

Formalism Operators :



Formalism Illustrated...

The Formalism:

 $Consonant-Syllable \rightarrow$

Vowel-Syllable \rightarrow V[D|B|X|BD|BX]

- Syllable \rightarrow Consonant-Syllable [Y] | Vowel-Syllable[Y]
- IDN-Label \rightarrow (Syllable | digit)*([dash](Syllable | digit))

Fundamental differences .bharat and Root LGR

.bharat zone

- It is a focused zone only the domain names under .bharat TLD
- Restricted only to Indian languages
- Policies can be strict
- Can define our own categories

Root Zone

- It is the most generic zone on the Internet. The root zone.
- Cannot be restricted. Encompasses all the scripts/languages of the world
- Policies have to be simple, yet sufficiently tight
- Have to rely on the Unicode Character properties

Character classes - Differences

.Bharat character classes

- $C \rightarrow Consonant$
- $V \rightarrow Vowel$
- $M \rightarrow Matra$

D

B

Χ

Η

Α

N

Y

L

Ζ

- → Anusvara/BindiSunna
- → Chandrabindu/Arasunna
- \rightarrow Visarga/Aytham
 - \rightarrow Halant/Chandrakala
 - \rightarrow Addak
- → Nukta
 - \rightarrow Avagraha/Praslesham
 - \rightarrow Chillu
 - \rightarrow Khanda Ta

Unicode character classes

- Mn Mark, Non-Spacing
 - 0901;DEVANAGARI SIGN CANDRABINDU
 - 093A;DEVANAGARI VOWEL SIGN OE
 - 093C;DEVANAGARI SIGN NUKTA
 - 094D;DEVANAGARI SIGN VIRAMA
- Mc Mark, Spacing Combining
 - 0903;DEVANAGARI SIGN VISARGA
 - 093E;DEVANAGARI VOWEL SIGN AA
- Lo Letter, Other
 - 0905;DEVANAGARI LETTER A
 - 0915;DEVANAGARI LETTER KA
 - 093D;DEVANAGARI SIGN AVAGRAHA

ధన్యవాదాలు

thank you

ধন্যবাদ

ଧନ୍ୟର୍ଚ୍ଚାଦ

நன்றி



धन्यवाद

ધન્યવાદ