

Computer writing of Nepali – Achievements so far



Bal Krishna Bal

bkbal@ltk.org.np

Chief Technical Officer

Language **Technology** Kendra (LTK)

<http://ltk.org.np>

Lalitpur, PatanDhoka

Nepal

Outline

- Text Input, Internal Representation and Display
- Script, Character set, Font
- Pre-Unicode Era, Pretending to Write in Nepali
- Advent of Nepali Unicode
- Nepali Unicode and the Prevailing Issues, Limitations
- Conclusion

Text Input, Internal Representation and Display

- Text input → Input Device (Keyboard)
- Minimal unit of a text → Character
- Each character is represented internally by a character code in the storage devices(main memory, disks etc.)
- Character code – different numerical values characteristic of an encoding scheme(e.g., ASCII, UNICODE etc.).
- For example, the character code of the latin small letter A - “a” as per ASCII(decimal) is 97.
- Display → Output Device(Computer screen, disks, printer)

Script, Character set, Font

- Every language with a written tradition has a script or a writing system that it follows for writing
- For example, Devanagari script for Nepali, Hindi, Marathi and other languages
- Each script consists of a character set (a collection of characters and symbols used for the language)
- We need a font for the computer system to interpret the character codes and display on the screen.

Pre-Unicode Era

Pretending to write in Nepali

- Purely managed by some hack fonts
 - Text encoding (codes for each character of the text based on ASCII encoding).
 - For example, the Nepali character “ब” as per the Preeti font is mapped to the letter “a” in the keyboard and correspondingly assigned the latter’s code for internal representation.
 - Worse, no uniform character set mappings with character codes

Pre-Unicode Era

Pretending to write in Nepali...

Disadvantages:

- Use of the computer confined to simply typing
- Lot of compromises in terms of typing
- Source and Target computers both need to have the same font installed in their machines
 - No consistent and uniform keyboard layouts (Keyboard layouts differed according to fonts and what developers felt right in terms of character key mappings)
- Typing difficult and requiring rigorous trainings
- Typed text lacks processing benefits (Sort, Find and Replace, Arithmetic calculations etc. not possible)

Few conjuncts in Devanagari

क्क	क्ख	क्क	क्ण	क्त	क्त्य	क्त्र	क्त्र्य	क्त्व	क्न	क्न्य	क्म
kka	kkha	kca	kṇa	kta	ktya	ktra	ktrya	ktva	kna	knya	kma
क्य	क्र	क्र्य	क्ल	क्व	क्व्य	क्ष	क्ष्म	क्ष्य	क्ष्व	क्व्य	ख
kya	kra	krya	kla	kva	kvya	kṣa	kṣma	kṣya	kṣva	khya	khra
ग्य	ग्र	ग्र्य	घ्न	घ्न्य	घ्म	घ्य	घ्र	ङ्क	ङ्क	ङ्क्य	ङ्क्य
gya	gra	grya	ghna	ghnya	ghma	ghya	ghra	ṅka	ṅkta	ṅktya	ṅkya
ङ्क्ष	ङ्क्ष्व	ङ्क्ष्व	ङ्क्ष्य	ङ्ग	ङ्ग्य	ङ्घ	ङ्घ्य	ङ्घ्र	ङ्घ्र	ङ्घ्र्य	ङ्घ्र्य
ṅkṣa	ṅkṣva	ṅkṣa	ṅkṣya	ṅga	ṅgya	ṅgha	ṅghya	ṅghra	ṅghra	ṅghya	ṅghya
ञ्य	च्च	च्च	च्च	ञ्च	च्च	च्च	च्च	च्च	ञ्ज	ञ्ज	ञ्ज
ṅya	cca	ccha	cchra	cña	cma	cya	chya	chra	jja	jjha	jña
ज्ञ्य	ज्म	ज्य	ज्र	ज्व	ञ्च	ञ्च	ञ्च	ञ्च	ञ्ज	ञ्ज	ञ्च
jñya	jma	jya	jra	jva	ñca	ñcma	ñcya	ñcha	ñja	ñjya	ñca

Much more exists...

Source: <http://www.omniglot.com/writing/devanagari.htm>

Pre-Unicode era

Pretending to write in Nepali...



Fig.1. Keyboard layout for Preeti font

In Preeti	In Times New Roman	Character codes – ASCII(Decimal)
बकमन	asdg	97115100103
१२	!@	3364

Advent of Nepali Unicode

- Font Standardization Project
- Allocation of a separate block for the Devanagari script in the Unicode chart
- Development of Unicode compliant Fonts and Keyboard Layouts
- Changes
 - No more technical hassles in document transfer and exchange
 - Text processing possible for Nepali
 - Choice of fonts available

Advent of Nepali Unicode...

	090	091	092	093	094	095	096	097
0	ॐ 0900	ऐ 0901	ठ 0902	र 0903	ी 0904	ॐ 0905	ऋ 0906	० 0907
1	ँ 0908	ऑ 0909	ड 0910	र 0911	ु 0912	ं 0913	ृ 0914	ं 0915
2	ं 0916	ओ 0917	ढ 0918	ल 0919	॰ 0920	॰ 0921	॰ 0922	ँ 0923
3	ः 0924	ओ 0925	ण 0926	ळ 0927	॰ 0928	े 0929	॰ 0930	अ 0931
4	अ 0932	ओ 0933	त 0934	ळ 0935	॰ 0936	॰ 0937	। 0938	आ 0939
5	अ 0940	क 0941	थ 0942	व 0943	ँ 0944	ँ 0945	॥ 0946	औ 0947
6	आ 0948	ख 0949	द 0950	श 0951	े 0952	॰ 0953	॰ 0954	अ 0955
7	इ 0956	ग 0957	ध 0958	ष 0959	े 0960	॰ 0961	१ 0962	अ 0963
8	ई 0964	घ 0965	न 0966	स 0967	ै 0968	क्र 0969	२ 0970	
9	उ 0971	ङ 0972	न 0973	ह 0974	ॉ 0975	ख 0976	३ 0977	ज़ 0978
A	ऊ 0979	च 0980	प 0981	ं 0982	ो 0983	ग 0984	४ 0985	य 0986
B	ऋ 0987	छ 0988	फ 0989	ी 0990	ो 0991	ज 0992	५ 0993	ग 0994
C	ॠ 0995	ज 0996	ब 0997	॰ 0998	ौ 0999	ड़ 1000	६ 1001	ज़ 1002
D	ँ 1003	झ 1004	भ 1005	ऽ 1006	॰ 1007	ढ़ 1008	७ 1009	१ 1010
E	ऐ 1011	ज 1012	म 1013	ा 1014	ि 1015	फ 1016	८ 1017	इ 1018
F	ए 1019	ट 1020	य 1021	ि 1022	ौ 1023	य 1024	९ 1025	ब 1026

Fig.2. Devanagari block (U+0900 –U+097F) in the Unicode chart

Advent of Nepali Unicode...

- Has opened up myriads of opportunities
 - Blogs and social media (the medium being Nepali)
 - Localization and local language computing (Revolution in Nepali Language Computing - NepaLinux, Spell checkers, Nepali Lexicon, Grammar Analyzer, Machine Translation System, Dictionaries, Thesauri, Text-to-Speech, Font developments, Mobile applications, Internationalized Domain Names, E-governance applications etc.)
- Unicode certainly has provided a good platform to begin with but is it ideal for expressing the languages of Nepal and the neighboring regions??

Nepali Unicode and the Prevailing issues, Limitations

- The issue of the three conjuncts (क्ष,त्र,ज्ञ)
- Should/Should not they be assigned separate code spaces?
- Where should they be placed in dictionaries, telephone directories?
- Writing the languages of Nepal other than Nepali in Nepali Unicode
- What are the limitations?

Nepali Unicode and the Prevailing issues, Limitations...

- The World Wide Web is fast adopting Nepali Unicode
- But there is very low adoption in print houses
 - Formatting problems
 - Legacy issues
 - Unicode not supported in popularly used applications like Adobe PageMaker
 - Possible solutions (Use of converters)

Conclusion

- Have come a long way from where it started
- Still much to achieve in terms of adoption
- Need to go for migration tools like Font converters to ultimately switch hundred percent to Unicode
- Although Unicode is officially adopted by the Government in its policy, trainings and awareness campaigns still a must nation wide
- Continuous discourse on the current limitations of Nepali Unicode necessary

Thank You

Queries??