



7. Indian Standard Font Code (INSFOC)

The draft layout for Devanagari Font Standard was published in the 5th issue of the TDIL news letter. Based on the feedback from user community and industries the proposed standard is slightly modified. The revised draft for the proposed Devanagari Font Standard and draft Font Standard layout for Gujarati, Punjabi and Malayalam are also published here for comments/ feedback. The proposed font standards are targeted towards the following class of users

- Data Processing
- Office Users / Word Processing
- Textbook Publishers
- Web Content Creators
- Desktop Applications

It is certainly not targeted towards professional desktop publishers, advertising agencies and highly Sanskritized text content creators.

The font is laid out such that the font remains unchanged between the character locations 0x80 to 0xFF in Monolingual and Bilingual Font layout. The monolingual font contains more compound characters and conjuncts.

Font Naming Convention

The fonts having the standard layout shall have a naming convention that will identify the script, monolingual / bilingual, type of numerals. This is implemented by attaching a preamble string to the actual name of the font. Following abbreviations will be used:

| | |
|----|-------------------|
| DV | Devanagari |
| GJ | Gujarati |
| PN | Punjabi |
| BN | Bengali |
| AS | Assamese |
| OR | Oriya |
| KN | Kannada |
| TM | Tamil |
| TL | Telugu |
| ML | Malayalam |
| M | Monolingual Font |
| B | Bilingual Font |
| E | Roman Numerals |
| L | Language Numerals |

The preamble string shall have the following format:

L1 L2 - M/B E/L - Font Name

As illustrated above the first two letters denote the script of the font as per the abbreviation given above. The M/B field denotes whether the font is

monolingual or bilingual. The E/L field denotes the type of numerals contained in the font. The field will contain E if the font contains Roman numerals and will contain L if the font contains the native numerals of the language script. The fields are separated by an underscore character. To fix this more clearly a bilingual Gujarati font 'Narsi' containing Gujarati numerals will have the preamble as GJ_BL_Narsi. A Punjabi monolingual font 'Govind' containing Roman numerals will have a preamble PN_ME_Govind.

Rules for Composing Devanagari Text

(1) The Devanagari characters lying in the codes 0x80 to 0xFF are designed to be kept in the same location for Devanagari bilingual font. Here a majority of the consonants are kept in their half form. The full consonant is formed by adding a 'kana' (Vertical stroke- 0xE0) to the half form. It is recommended that the kana located at 0xE0 be used for that purpose. For example
र (0xAA) + ि (0xE0) = र

(2) There are two *matras* (Vowel Signs) of vowel I (ई) with different overhanging spans. These *matras* are located at 0x4C and 0xE3. The *matra* at 0x4C is used for the wider letters like ka (क), fa (फ) as shown below:

क (0xA7) + ि (0x4C) = की

The *matra* at 0xE3 is used for other letters, which are not wider like Ma (म), Ra (र), Ya (य) etc. for example

म + ि (0xE3) = मी
र + ि (0xE3) = री

The *matras* shown at code points 0x4A and 0x4D are with the *rakar* (र Ra is coming in a syllable and being pronounced before the consonant to which it is applied) and with different overhanging spans. For example:

प्रार्थी, तुर्की

The *matras* shown at code points 0x4B and 0x4E are with the *rakar* & *Anuswar* and with different overhanging spans.

(3) Similarly, there are three types of the *matras* of vowel sign (इ) with different overhanging spans. These *matras* are located at locations 0xE1, 0xE2 and 0x4F. The *matra* at 0xE1 is used for normal size letters such as र (Ra), क (Ka), फ (Fa), ड (dha) etc. For example

ि (0xE1) + र (0xCA) = रि
ि (0xE1) + क (0xA7) = कि



The other form of *matra* of vowel I (इ) is used for wider letters such as स (Sa), म (Ma), य (Ya) etc. For example

$$\text{ि}(0xE2) + \text{म} (0xC8 + 0xE0) = \text{मि}$$

$$\text{ि}(0xE2) + \text{स} (0xCD + 0xE0) = \text{सि}$$

The third form of the *matra* of vowel I (इ) is used when there is a half form of a consonant in a word. In this case the *matra* is attached to the '*Kana*' (Vertical stroke) of the preceding consonant. For example in the words स्थित, शक्ति etc.

- (4) The shifted *ukar* and *ookar* (Vowel signs for u and uu) located at 0x42 and 0x43 are to be used with characters which are not having full *Kana* and in this characters the *matra* is attached to the center lower part of the characters such as in डु, टू, रू .
- (5) The *rakar* located at '0xED' is provided for characters क (Ka), फ (Pha), म (Ma), भ (Bha), व (Va), न (Na), ब (Ba), etc. This *rakar* is attached to the characters at slightly upwards shifted position (almost at the middle of *Kana*). For example as in वक्र, नम्र.
- (6) The *rakar* located at 0x50 is provided for characters ग (Ga), च (Ca), ज (Ja), थ (Tha), घ (Gha), ध (Dha), य (Ya) etc. This *rakar* is attached to the characters at slightly downwards shifted position. For example व्यग्र, वज्र, व्याघ्र .
- (7) The widths of letters क (Ka), फ (Pha), रु (RRu), रू (Roo) etc. are reduced by the width of the fixed *kern* space located at 0xFD to ensure proper anchoring of *matras*.
- (8) The widths of letters त्त (TTa), थ (Tha), द (Da), त+त (Ta+Ta), थ+थ (Tha+Tha), etc. are reduced by the width of the fixed *kern* space located at 0xFE to ensure proper anchoring of *matras*.

Rules for Composing Gujarati Text

1. The Gujarati characters lying in the codes 0x80 to 0xFF are designed to be kept in the same location for Gujarati bilingual font. Here a majority of the consonants are kept in their half form. The full consonant is formed by adding a '*kana*' to the half form. It is recommended that the *kana* located at 0xFA be used for that purpose.
2. Two I *matras* are kept with different overhanging spans. It is recommended that for simple consonant, the first one at 0xEC be used whereas for wider akshars consisting of one or two half consonants and one full consonant the second one at 0x6B be used.

3. The widths of letters 'Ka', 'Fa', 'Tta', 'Dda', 'Roo' etc. are reduced by the width of the fixed space located at 0xFD to ensure proper anchoring of *matras*. The fixed space at 0xFD should be added after outputting the bottom *matra*.
4. The widths of letters 'Ja', 'Ja+Ra', 'Ru' etc. are reduced by the width of the fixed space located at 0xFE to ensure proper anchoring of *matras*. The fixed space at 0xFE should be added after outputting the top or the bottom *matra*.
5. The shifted '*ukar*' and '*ookar*' located at 0x61 and 0x62 are to be used with vertical conjuncts such as 0x54 to 0x58 etc.
6. The shifted '*Anuswara*' located at 0xFC is be used if there is another top *matra* such as 0XF1, 0xF2, 0xF3. It should also be used if there is a side *matra* 0xED. It may be used with vowel 0xA4.
7. The shifted '*Anuswara*' located at 0xF9 is be used with vowel 0xA4.

Rules for Composing Punjabi Text

1. All Punjabi characters (except numerals) could be accommodated in the right half (0xA1 to 0xFF) codes.
2. The widths of letters 'thha' and 'na' are reduced by the width of the fixed space located at 0xFE to ensure proper anchoring of *matras*. The fixed space at 0xFE should be added after outputting the top *matra*.
3. The '*ukar*' and '*ookar*' located at 0xE4 and 0xE5 respectively are to be used if the akshar contains characters from 0xD6 to 0xDD.

Rules for Composing Malayalam Text

1. Characters line in the codes of 0x80 to 0x9F has been sifted to A1 onwards. So that we can have common Font for Windows and Linux.
2. Characters like Bba, Vva, Yya, Chha can be composed with full consonant and characters at location 0xF8.
3. In Bilingual Font most of the conjuncts can be formed with help of Chandrakala for eg:
Ka-halant+Ta, Ka+hlanat+Sha, Ga+halant+Na, Ga+halant+Ma, Chha+halant+Chha, Ja+halant+Ja, Ja+halant+Nja, Nna+halant+Ma, Ta+halant+Bha, Da+halant+Dha etc.
4. More frequently used conjuncts are kept on Monolingual side that is in the range of 0x41 to 0X74.

(Courtesy: Consortium for Innovation and Language Technology (COIL-Tech), MAIT

Tel : 011-6855487 E-mail : mait@vsnl.com

&

TDIL Programme, DIT, E-mail : tdilinfo@mit.gov.in)



INSFOC - Devanagari

| Proposed Standardisation for Devanagari DV-Monolingual | | | | | | | | | | | | | | | | |
|--|-----|-----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 0 | 16 | 32 | 48 | 64 | 80 | 96 | 112 | 128 | 144 | 160 | 176 | 192 | 208 | 224 | 240 |
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
| 0 | NUL | DLE | | 0 | @ | , | ' | 5 | | | | ८ | द | ८ | । | . |
| 1 | SOH | DC1 | ! | 1 | ' | ॠ | ॡ | प | | ' | अ | | ॢ | ह | f | |
| 2 | STX | CD2 | ' | 2 | ॡ | ॢ | श | ब | | ' | इ | छ | ॣ | ट | f | ॐ |
| 3 | ETX | CD3 | # | 3 | ॣ | । | श | भ | | " | उ | ० | । | ळ | ी | ॠ |
| 4 | EOT | CD4 | \$ | 4 | । | ॥ | ॠ | म | | " | ऊ | ॥ | फ | ळ | ॡ | ॡ |
| 5 | ENQ | NAK | % | 5 | ॥ | ० | ॡ | य | | ॡ | ए | | ॢ | ॢ | ॢ | ॢ |
| 6 | ACK | SYN | & | 6 | ० | ॠ | ॡ | ल | | ॠ | ॠ | | ॣ | ॣ | ॣ | ॣ |
| 7 | BEL | ETB | " | 7 | ॠ | ॡ | ॡ | व | | ॠ | क | • | ॣ | ॣ | ॣ | ॣ |
| 8 | BS | CAN | (| 8 | ॡ | ॢ | ख | ष | | ॢ | ख | ० | ॣ | ॣ | ॣ | ॣ |
| 9 | HT | EM |) | 9 | ॢ | ॣ | घ | स | | ॣ | ख | ॠ | ॣ | ॣ | ॣ | ॣ |
| A | LF | SUB | * | : | ॣ | । | च | । | | । | ॠ | ॠ | ॣ | ॣ | ॣ | ॣ |
| B | VT | ESC | + | : | । | [| ज | { | | । | ॢ | ॢ | ॣ | ॣ | ॣ | ॣ |
| C | FF | FS | , | < | । | \ | झ | | | । | | ॣ | ॣ | ॣ | ॣ | ॣ |
| D | CR | GS | - | = | । |] | ट | } | | । | | ॣ | ॣ | ॣ | ॣ | ॣ |
| E | SO | RS | . | > | । | ^ | थ | ~ | | । | ॣ | ॣ | ॣ | ॣ | ॣ | ॣ |
| F | SI | US | / | ? | । | _ | ध | DEL | | । | | ॣ | ॣ | ॣ | ॣ | ॣ |



INSFOC - Gujarati

| Proposed Standard for Gujarati Monolingual | | | | | | | | | | | | | | | | |
|--|---|----|----|----|----|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 0 | 16 | 32 | 48 | 64 | 80 | 96 | 112 | 128 | 144 | 160 | 176 | 192 | 208 | 224 | 240 |
| 0 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
| 0 | | | ! | ૦ | @ | શ | ' | જ્ઞ | | | | ૨ | ૩ | ૪ | ૫ | ૬ |
| 1 | | | ' | ૧ | ખ | ષ | ્ | ષ્ઠ | | | અ | | ૬ | ૭ | ૮ | ૯ |
| 2 | | | ' | ૨ | ઘ | સ | ્ | ષ્ઠ | | | ઇ | ઇ | ૬ | ૭ | ૮ | ૯ |
| 3 | | | # | ૩ | ચ | ક્ષ | ્ | ૨ | | " | ઈ | જ | ૬ | ૭ | ૮ | ૯ |
| 4 | | | \$ | ૪ | ણ | કે | ્ | શ | | " | ઉ | જ | ૬ | ૭ | ૮ | ૯ |
| 5 | | | % | ૫ | ત | ઠ | ્ | શ્ચ | | | ઊ | | ૬ | ૭ | ૮ | ૯ |
| 6 | | | & | ૬ | થ | ડ | ્ | હ | | — | ઝ | | ૬ | ૭ | ૮ | ૯ |
| 7 | | | " | ૭ | ધ | ઢ | ્ | ઢ | | — | ઙ | • | ૬ | ૭ | ૮ | ૯ |
| 8 | | | (| ૮ | ન | ટ | ્ | ઢ | | | ઙ | ઙ | ૬ | ૭ | ૮ | ૯ |
| 9 | | |) | ૯ | પ | ણ | ્ | શ્ચ | | | ૞ | ૞ | ૬ | ૭ | ૮ | ૯ |
| A | | | * | : | બ | ભ્ય | ્ | શ્ચ | | | ૟ | ૟ | ૬ | ૭ | ૮ | ૯ |
| B | | | + | : | ભ | [| ્ | { | | | ૠ | ૠ | ૬ | ૭ | ૮ | ૯ |
| C | | | , | < | મ | \ | | | | | ૡ | ૡ | ૬ | ૭ | ૮ | ૯ |
| D | | | - | = | ય |] | S | } | | | ૢ | ૢ | ૬ | ૭ | ૮ | ૯ |
| E | | | . | > | લ | ^ | ઞ | ~ | | | ૣ | ૣ | ૬ | ૭ | ૮ | ૯ |
| F | | | / | ? | વ | _ | ઙ | | | | ૤ | ૤ | ૬ | ૭ | ૮ | ૯ |



INSFOC - Punjabi

| Proposed Standard for Punjabi Monolingual | | | | | | | | | | | | | | | | |
|---|---|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 0 | 16 | 32 | 48 | 64 | 80 | 96 | 112 | 128 | 144 | 160 | 176 | 192 | 208 | 224 | 240 |
| 0 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
| 0 | | | ! | ० | @ | | , | | | | | ਅ | ਛ | ਲ | ੜ | ਫ |
| 1 | | | ! | ੧ | | | | | | ' | ਅ | | ਣ | ਲ | f | ਯ |
| 2 | | | ' | ੨ | | | | | | ' | ੲ | ਛ | ਤ | ਵ | ੀ | - |
| 3 | | | # | ੩ | | | | | | " | ਉ | ਚ | ਥ | ਸ਼ | ਥੰ | , |
| 4 | | | \$ | ੪ | | | | | | " | ਉ | ਛ | ਦ | ਸ | ਠ | ਠਿ |
| 5 | | | % | ੫ | | | | | | — | ਉ | | ਧ | ਹ | : | । |
| 6 | | | & | ੬ | | | | | | — | ਉ | | ਨ | ਕ | । | ਠ |
| 7 | | | " | ੭ | | | | | | — | ਉ | | ਪ | ਕ | । | ਠ |
| 8 | | | (| ੮ | | | | | | — | ਉ | • | ਫ | ਕ | । | ਠ |
| 9 | | |) | ੯ | | | | | | | ਕ | ਜ | ਛ | ਕ | । | ਠ |
| A | | | * | : | | | | | | | ਖ | ਸ਼ | ਛ | ਕ | । | ਠ |
| B | | | + | : | | | | | | | ਗ | ਛ | ਛ | ਕ | । | ਠ |
| C | | | . | < | | | | { | | | ਗ | ਛ | ਛ | ਕ | । | ਠ |
| D | | | - | = | | | | | | | | ਛ | ਛ | ਕ | । | ਠ |
| E | | | . | > | | | | } | | | ਗ | ਛ | ਛ | ਕ | । | ਠ |
| F | | | / | ? | | | | ~ | | | | ਛ | ਛ | ਕ | । | ਠ |

Proposed Standardization for ML- Monolingual

| | 0 | 16 | 32 | 48 | 64 | 80 | 96 | 112 | 128 | 144 | 160 | 176 | 192 | 20 | 22 | 24 | |
|---|-----|-----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|----|----|----|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | E | C | F | |
| 0 | NUL | DLE | | 0 | @ | P | | p | | | | ° | À | Đ | à | đ | |
| 1 | SOH | DC1 | ! | 1 | A | Q | a | q | | | | ı | ± | Á | Ñ | á | ñ |
| 2 | STX | CD2 | " | 2 | B | R | b | r | | | | ¢ | ² | Â | Ò | â | ò |
| 3 | ETX | CD3 | # | 3 | C | S | c | s | | | | £ | ³ | Ã | Ó | ã | ó |
| 4 | EOT | CD4 | \$ | 4 | D | T | d | t | | | | ¤ | ' | Ä | Ô | ä | ô |
| 5 | ENQ | NAK | % | 5 | E | U | e | | | | | ¥ | µ | Å | Õ | å | õ |
| 6 | ACK | SYN | & | 6 | F | V | f | - | | | | ı | | Æ | Ö | æ | ö |
| 7 | BEL | ETB | ' | 7 | G | W | g | | | | | § | | Ç | × | ç | ÷ |
| 8 | BS | CAN | (| 8 | H | X | h | - | | | | ¨ | , | È | Ø | è | ø |
| 9 | HT | EM |) | 9 | I | Y | î | | | | | © | ¹ | É | Ù | é | ù |
| A | LF | SUB | * | : | J | Z | j | | | | | ª | º | Ê | Ú | ê | ú |
| B | VT | ESC | + | ; | K | | k | | | | | « | » | Ë | Û | ë | û |
| C | FF | FS | , | < | L | | l | | | | | | ¼ | Ì | Ü | ì | ü |
| D | CR | GS | - | = | M | | m | | | | | | ½ | Í | Ý | í | ý |
| E | SO | RS | . | > | N | | n | | | | | ® | ¾ | Î | Þ | î | þ |
| F | SI | US | / | ? | O | | o | | | | | - | ¿ | Ï | ß | ï | # |