



## Multilingual Font Creation by Mapping Unicode to ASCII

Siva Jyothi Chandra\*, Ashlesha Pandhare, Mamatha Vani  
CSE Department & JNTUH, Telangana,  
India

---

**Abstract**—This paper is to outline on creation of font by mapping ASCII characters with unicode characters. As unicode occupies two bytes of space for one character and ASCII accampains with single byte for single character. This approach of creation of user defined font with mapping comprises of 1) Extraction of glyphs 2) Mapping with ASCII set 3) Generating the font file. For Indian local languages every language has given some range of code points to it. Based on the user combination , can accommodate more than 230 characters in a single ASCII font file. In this paper unicode characters of Tamil , Telugu, and Kannada languages are used.

**Keywords**— unicode, ASCII, glyph, mapping of characters, local language, font file,

---

### I. INTRODUCTION

Microsoft added Unicode capabilities to Windows, your PC's displayable and printable characters were limited to whatever "character sets" you had installed in your operating system. For example: The Western European character set commonly installed in an American version of Windows includes the English alphabet, plus accented versions of these characters for handling characters from many Western European language. the words "font," "script," and "Unicode " is the terminology in printing industry.

### II. FONTS AND SCRIPTS

#### **Fonts**

A *font* is a set of symbols, called glyphs. For most fonts, the glyphs share common design elements so that they look visually compatible with each other. The common design elements of a font are known as the *typeface* of the font, and the font is usually named after its typeface.

#### **Scripts**

A *script* is a writing system that includes a set of symbols and rules for how to put them together into meaningful words or sets. For example, the Western script has a rule that the symbols read from left to right, the Arabic script has a rule that the symbols read from right to left and that different versions of a symbol must be used depending on whether the symbol is at the beginning of a word, the end, or in between. Some fonts, including Arial, include glyphs from more than one script. Others, like Andalus, have glyphs from only one. The font Arial Unicode MS has glyphs for all major languages and many minor languages, over 50,000 glyphs in all.

#### **Unicode**

Characters, numbers and other symbols in the character sets that we see on screen and print are encoded in the computer into "ones" and "zeros" using any of a variety of character mappings. For example, "ASCII" has long been a standard for encoding American characters. And, on Windows, various "Code Pages" have historically been used to represent numerous character sets specific to different parts of the world.

Unicode is most easily thought of as a single, giant "Code Page" designed to represent every character in almost every language in the world.

While archiving with Unicode will occupy more space. Generally PDF files are used for archiving. Converting to true type or open type font is almost accepted by most of printers. Also can convert fonts of true type to other formats like .pfa, .pfb which we can install on most of the post script printers and embed in Adobe PDF files. AFP, PCL printers has its own font format that can be converted to.

### III. CREATION OF NEW USER DEFINED FONT WITH MULTIPLE LANGUAGES CHARACTER SETS ALGORITHM:

Step 1: From Unicode character sets prepare a list of languages with character sets which you wish to do.

Step 2: From windows charmap look at the characters by giving the Unicode value

Step 3: Maximum characters are present in Arial Unicode MS font.

Step 4: Open the Arial Unicode MS font in the font editor from which characters have to copied.

Step 5: Open a new font file in the editor on which to paste.

Step 5: Select the characters from Unicode font and copy them.

Step6: Paste the selected character(glyph)s in the new font file.  
Step7: Repeat steps 5 – 6 make sure the character count is about 230

Given below is the view of the character sets of tamil, telugu, kanada glyphs which we wish to combine in a single ASCII font file.

Unicode range of

1. Tamil(0B82-0BFA) 71 from 14-85 ,
2. Telugu(0C01-0C6F) 80 from 86-166 ,
3. Kannada(0C82-0CF2) 86 from 167-253.

Start->Run->charmap. Select Arial Unicode MS font.

**Select from Charmap  
TAMIL**

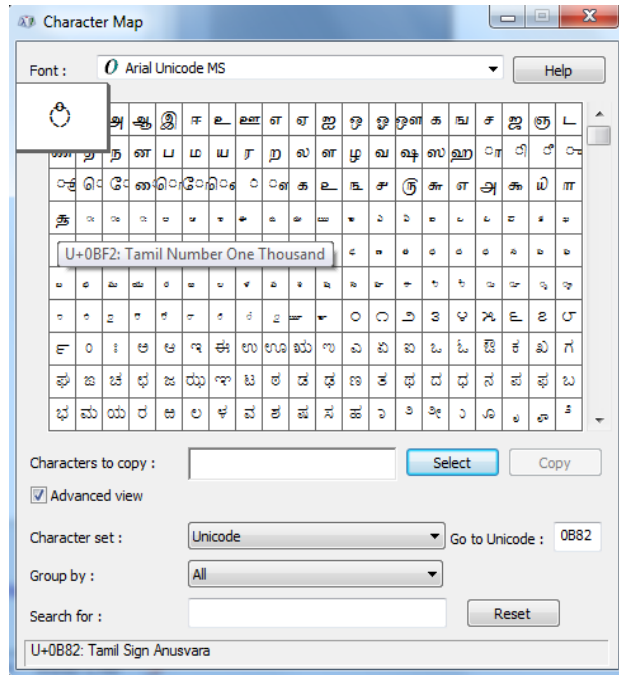


Fig 1: Tamil Characters with Unicode values

```
# TAMIL LANGUAGE UNICODE SET
0B82      ; Tamil # Mn   TAMIL SIGN ANUSVARA
0B83      ; Tamil # Lo   TAMIL SIGN VISARGA
0B85..0B8A ; Tamil # Lo   [6] TAMIL LETTER A..TAMIL LETTER UU
0B8E..0B90 ; Tamil # Lo   [3] TAMIL LETTER E..TAMIL LETTER AI
0B92..0B95 ; Tamil # Lo   [4] TAMIL LETTER O..TAMIL LETTER KA
0B99..0B9A ; Tamil # Lo   [2] TAMIL LETTER NGA..TAMIL LETTER CA
0B9C      ; Tamil # Lo   TAMIL LETTER JA
0B9E..0B9F ; Tamil # Lo   [2] TAMIL LETTER NYA..TAMIL LETTER TTA
0BA3..0BA4 ; Tamil # Lo   [2] TAMIL LETTER NNA..TAMIL LETTER TA
0BA8..0BAA ; Tamil # Lo   [3] TAMIL LETTER NA..TAMIL LETTER PA
0BAE..0BB9 ; Tamil # Lo  [12] TAMIL LETTER MA..TAMIL LETTER HA
0BBE..0BBF ; Tamil # Mc   [2] TAMIL VOWEL SIGN AA..TAMIL VOWEL SIGN I
0BC0      ; Tamil # Mn   TAMIL VOWEL SIGN II
0BC1..0BC2 ; Tamil # Mc   [2] TAMIL VOWEL SIGN U..TAMIL VOWEL SIGN UU
0BC6..0BC8 ; Tamil # Mc   [3] TAMIL VOWEL SIGN E..TAMIL VOWEL SIGN AI
0BCA..0BCC ; Tamil # Mc   [3] TAMIL VOWEL SIGN O..TAMIL VOWEL SIGN AU
0BCD      ; Tamil # Mn   TAMIL SIGN VIRAMA
0BD7      ; Tamil # Mc   TAMIL AU LENGTH MARK
0BE6..0BEF ; Tamil # Nd  [10] TAMIL DIGIT ZERO..TAMIL DIGIT NINE
0BF0..0BF2 ; Tamil # No   [3] TAMIL NUMBER TEN..TAMIL NUMBER ONE THOUSAND
0BF3..0BF8 ; Tamil # So   [6] TAMIL DAY SIGN..TAMIL AS ABOVE SIGN
0BF9      ; Tamil # Sc   TAMIL RUPEE SIGN
0BFA      ; Tamil # So   TAMIL NUMBER SIGN
```

# Total code points: 71  
TELUGU

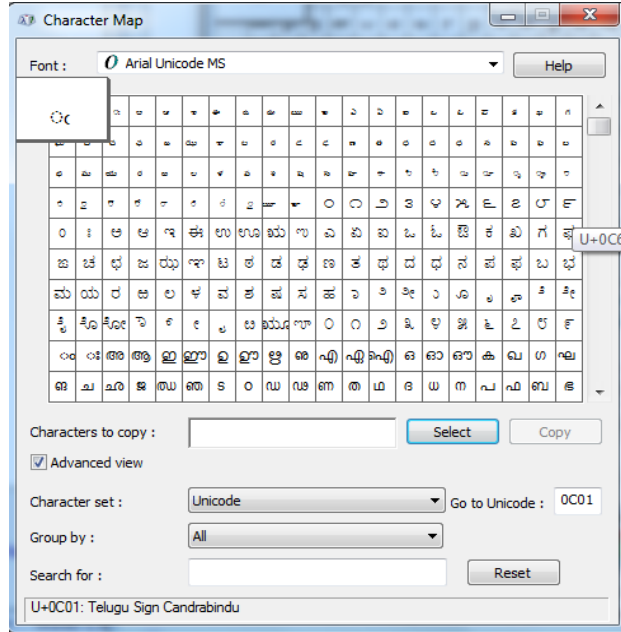


Fig 2: Telugu characters with unicode values

# TELUGU LANGUAGE UNICODE CHARACTER SET

- 0C01..0C03 ; Telugu # Mc [3] TELUGU SIGN CANDRABINDU..TELUGU SIGN VISARGA
- 0C05..0C0C ; Telugu # Lo [8] TELUGU LETTER A..TELUGU LETTER VOCALIC L
- 0C0E..0C10 ; Telugu # Lo [3] TELUGU LETTER E..TELUGU LETTER AI
- 0C12..0C28 ; Telugu # Lo [23] TELUGU LETTER O..TELUGU LETTER NA
- 0C2A..0C33 ; Telugu # Lo [10] TELUGU LETTER PA..TELUGU LETTER LLA
- 0C35..0C39 ; Telugu # Lo [5] TELUGU LETTER VA..TELUGU LETTER HA
- 0C3E..0C40 ; Telugu # Mn [3] TELUGU VOWEL SIGN AA..TELUGU VOWEL SIGN II
- 0C41..0C44 ; Telugu # Mc [4] TELUGU VOWEL SIGN U..TELUGU VOWEL SIGN VOCALIC RR
- 0C46..0C48 ; Telugu # Mn [3] TELUGU VOWEL SIGN E..TELUGU VOWEL SIGN AI
- 0C4A..0C4D ; Telugu # Mn [4] TELUGU VOWEL SIGN O..TELUGU SIGN VIRAMA
- 0C55..0C56 ; Telugu # Mn [2] TELUGU LENGTH MARK..TELUGU AI LENGTH MARK
- 0C60..0C61 ; Telugu # Lo [2] TELUGU LETTER VOCALIC RR..TELUGU LETTER VOCALIC LL
- 0C66..0C6F ; Telugu # Nd [10] TELUGU DIGIT ZERO..TELUGU DIGIT NINE

# Total code points: 80  
KANNADA

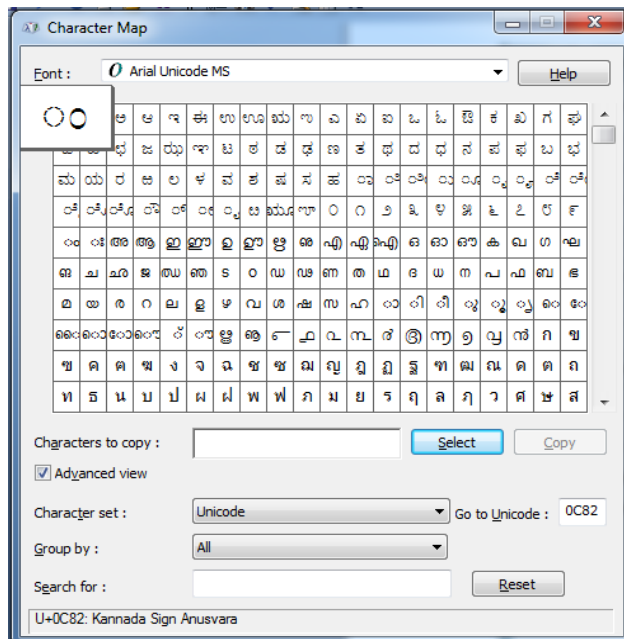


Fig 3: Kannada characters with unicode values.

# KANNADA LANGUAGE UNICODE SET

0C82..0C83 ; Kannada # Mc [2] KANNADA SIGN ANUSVARA..KANNADA SIGN VISARGA  
 0C85..0C8C ; Kannada # Lo [8] KANNADA LETTER A..KANNADA LETTER VOCALIC L  
 0C8E..0C90 ; Kannada # Lo [3] KANNADA LETTER E..KANNADA LETTER AI  
 0C92..0CA8 ; Kannada # Lo [23] KANNADA LETTER O..KANNADA LETTER NA  
 0CAA..0CB3 ; Kannada # Lo [10] KANNADA LETTER PA..KANNADA LETTER LLA  
 0CB5..0CB9 ; Kannada # Lo [5] KANNADA LETTER VA..KANNADA LETTER HA  
 0CBC ; Kannada # Mn KANNADA SIGN NUKTA  
 0CBD ; Kannada # Lo KANNADA SIGN AVAGRAHA  
 0CBE ; Kannada # Mc KANNADA VOWEL SIGN AA  
 0CBF ; Kannada # Mn KANNADA VOWEL SIGN I  
 0CC0..0CC4 ; Kannada # Mc [5] KANNADA VOWEL SIGN II..KANNADA VOWEL SIGN VOCALIC RR  
 0CC6 ; Kannada # Mn KANNADA VOWEL SIGN E  
 0CC7..0CC8 ; Kannada # Mc [2] KANNADA VOWEL SIGN EE..KANNADA VOWEL SIGN AI  
 0CCA..0CCB ; Kannada # Mc [2] KANNADA VOWEL SIGN O..KANNADA VOWEL SIGN OO  
 0CCC..0CCD ; Kannada # Mn [2] KANNADA VOWEL SIGN AU..KANNADA SIGN VIRAMA  
 0CD5..0CD6 ; Kannada # Mc [2] KANNADA LENGTH MARK..KANNADA AI LENGTH MARK  
 0CDE ; Kannada # Lo KANNADA LETTER FA  
 0CE0..0CE1 ; Kannada # Lo [2] KANNADA LETTER VOCALIC RR..KANNADA LETTER VOCALIC LL  
 0CE2..0CE3 ; Kannada # Mn [2] KANNADA VOWEL SIGN VOCALIC L..KANNADA VOWEL SIGN VOCALIC LL  
 0CE6..0CEF ; Kannada # Nd [10] KANNADA DIGIT ZERO..KANNADA DIGIT NINE  
 0CF1..0CF2 ; Kannada # So [2] KANNADA SIGN JIHVAMULIYA..KANNADA SIGN UPADHMANYIA

# Total code points: 86

First we need to select which characters we wish to copy. Later copy them in to the new font at selected positions of space, a-z,A-Z,0-9 etc., which leads to mapping as shown below.

Open a font editor where mapping of characters or glyphs is provided by so many tools in market.

Fig 5. Mapping of Tamil and Telugu characters

Tamil characters {space(‘ ’) to backslash(‘\’)}  
 Telugu characters {bracket right(‘]’) logical not(‘-’)}

Fig 6. Mapping of Kanada characters from Hyphen(‘-’) to Uderisis(‘ü’)

Once mapping is finished, generate the required font from the options.

With this we can create a multi-lingual font. Also if needed only some glyphs of the language character sets can be taken and made a new font. Required editing’s can be done with out changing the actual glyph which may lead to discrepancies. In creation of font we can either go with Font Lab studio or other tool that offers high quality.

After creating the font the next is the usage of the font in real time.

Install the font in System Fonts folder

#### IV. APPLICATION AREAS

Some of the application areas are:

- Access of fonts in editors like notepad, word pad etc.
- In creating PDF Documents and archiving the Font Resource.
- For creating local language documents of our choice with the editors.
- File size can be reduced with user defined fonts.
- Customization of fonts is achieved.
- Business people, bank documents contain information with local native language for communication with its clients.
- In web blog design , personal profiles , highlighting important info.
- User-defined font (UDF) is a custom graphic you can use in graphic fields.
- Generate open fonts, true type fonts, PFA (compatible printable ascii font ),PFB (compatible printable binary font).

#### V. CONCLUSIONS

In this paper, a comprehensive approach on creation of multilingual character sets is provided. Usage of Indian languages is shown, can be extended to CJK(chinese, Japanese, Korean ) fonts. Character sets of different languages are obtained from Unicode set. **Rules** can be prepared with the help of mapping of single character (Indian local languages ex: मातृ, ఎన్నకొ etc., ) with a combination of unicode characters.

#### REFERENCES

- [1] <http://old.fontlab.com/font-editor/fontlab-studio/>