

Research Report on Pashto Font Development

Rafiq Ullah Kakar, Habib-ur-Rahman Ghaznawe Afghan Computer Science Association (ACSA), Kabul, Afghanistan rafiqullah.kakar@acsa.org.af, habib.ghaznawe@acsa.org.af

Abstract

This document is a research report on Pashto Font (PaykanRawan) Development conducted by Afghan Computer Science Association. This document gives a brief picture of the Font Development process using FontLab and Ms. Volt technology.

1. Introduction

Fonts play special role for the development of rendering any computer software. Unfortunately, no fonts were developed from the scratch for Pashto so far. In most scenarios, Farsi or Arabic fonts have been modified to meet the basic needs of the end users. With the development of PaykanRawan special fonts developed from the scratch by ACSA Calligrapher, Mr. Paykan Rawan, meets all the grammar and other relevant needs in Pashto Language.

Background information on True Type and Open Type Fonts and how they are being supported in Linux operating system is provided here.

Font is collection of glyphs (basic element of the font) to show data characters (minimal unit of the written language) images visually. Each font has a structure made of:

- Character Set

- Code Page
- Font Code

2. Methods:

Font Character Set:

In computer and machine-based telecommunications terminology, a character is a unit of information that roughly corresponds to a grapheme, or symbol, in the written form of a natural language.

An example of a character is a letter, numeral, or punctuation mark. The concept also includes control characters, which do not correspond to natural language symbols but to other bits of information used to process texts of the language, such as instructions to printers or other devices that display such texts

Character properties are seen in the positions of characters i.e.:

Project No: 102042 Ref. No: PANL10n/Admn/PR/0705AF

A character baseline demonstrates an alignment on the line for writing.

- The way the character will be printed from its space dimension.
- The character position in its space.
- Each character has its character ID.

A character encoding consists of a code that pairs a set of characters (representations of graphemes or grapheme-like units, as might be found in an alphabet or syllabify for the communication of a natural language) with a set of something else, such as numbers or electrical pulses, in order to facilitate the storage and transmission of text in computers and through telecommunication networks.

Character Set is the string processing and the logic with which you manipulate text and it can also be defined as the way to store and transmit data. In brief encoding is mapping a code number to a character.

There are many types of encoding systems:

- DOS character sets, also known as IBM code pages: CP437, CP737, CP850, CP852, CP855, CP857, CP858, CP860, CP861, CP863, CP865, CP866, CP869
- Windows character sets: Windows-1250-1251-1252-1253-1254-1255-1256-1257 ...
- UTF-16/UCS-2, UTF-32/UCS-4, UTF-7, UTF-8, UTF-9 and UTF-18, UTF-EBCDIC...

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	(<u></u>) 0600	م ٥٤١٥		ذ ₀₆₃₀	-	়	• 0650	۱ ٥٥٦٥	ي ‱	<mark>ت</mark> ٥٤٩٥	ش 0640	گ ‱	5 06C0	ې	06E0	• 06F0
1	0601	ء 0611	\$ 0621	ر 0631	ف ⁰⁶⁴¹	് 0651	\ 0651	٦ 0571	ځ	ڑ 0591	ف 06A1	ڴ	► 06C1	ي ٥٤٥٦	े 06E1	\ 06F1
2	0602	で 0612	آ 0622	ز 0632	ق 0642	் 0652	Y 0652	٦ 0572	<u>خ</u> 0582	<u>ر</u> 0692	ب ۵۶۸2	میں 0682	% 06C2	<u>ک</u>	් 06E2	Y 06F2
3	<u>مر</u>] 0603	হ । ।	٦ 0623	س ***	<u>ح</u> 0643	ॅ	۳ 0653	0573	<u>ج</u> 0583	→ 0593	ف 06A3	ر مروقع	× 06C3	<u>ح</u> 05D3	ु ०६६३	٣ 06F3
4		ें 0614	ؤ 0624	ش 6634	J 0614	ి 0654	* ٤ 0654	¢ 0574	<u>ج</u> 0584	<u>)</u> .	ث 6641	گ	3 06C4	 06D4	<u>ँ</u> 06E4	¢ 06F4
5		් 0815	ا 0625	ص ٥٤٦٥	0645	ূ	0 0655	0575	<u>څ</u>)	پ «۸۵	ڵ 0585	9 0605	0 06D5	ر 06E5	۵ 06F5
6			ئ 0626	ض ۵636	ن 0646		٦ 0656	ۇ 0576	چ 0686	<u>ب</u> 0696	ق ۵۶۸6	J 0686	ۆ ٥٥٥٥	ل ا ۵۹۵۵	ے۔ 06E6	9 06F6
7			0627	ط 0637	هر 0617	් 0657	V 0657	وُ 0577	<u>7</u> 0587	ز ₀₅₉₇	ف	Ĵ 0687	ر و 06C7	يغ 0507	ے 06E7	V 06F7
8			ب	ظ	و	ð	٨	ئى	ڈ	ژ	ڨ	Ļ	ۈ	ó	ိ	٨
9			0628 Ö 0629	0638 0639	0648 کی 0649	0558 Ō 0659	0658 ٩ 0659	0578 ط 0579	0588 2 0689	0698 رت	06A8	۰ دی ۵	06C8 9 06C9	V V OSD9	06E8	06F8 4 06F9
A			ت 062A	0634	ور ي 0644	٥65A	•/• 066A	<u>ن</u> ۵67A	2 0585	بن ۵69۸	0644	U 0684	<u>و</u> ۵۶۵۸	د د ٥٤D4		رش ٥٤٢٨
в	م 0608	6 061B	ث 062B		් 664B	் 0658	ر 0668	<u>ب</u> 0678	ذ 0588	پس ۵69B	S 0648	ڻ 0688	ۋ 06CB	் 06D8	் 0688	ښ ≪FB
с	6 050C		0625 0620		064C	065C	, 9	رت _{067C}	3 0680	يش 0690	J OSAC	نې ۵680	ی عنو	് സംഗം	06EC	ie-
D	0500		062C		064D	<u>ර්ග</u> ේ 065D	066D	ٽ	2 068D	<u>يص</u>	میں 0540	٥٥٥٥ ٥٥٥٥	می می 06CD		् 	06FD
E	٩	*	خ		ó	ð	ں	_{067D}	ĉ	ڞ	Ļ	æ	ێ	Ø	ŝ	ĥ
F	060E	061E	062E		064E	065E	066E	067E	058E	069E	DEAE S DEAF	OGBE	06CE 9 06CF	OGDE OGDF	OGEE	06FE

Code page:

Codepages are tables that map character codes (one byte long) to the Unicode indexes. Depending on the size of the page, these tables may have 256 or 65,536 records, one for each possible character code. Long codepages are called double-byte codepages and are primarily used to represent codes used in Chinese, Japanese, Korean languages.

Codepages are necessary because we need to somehow encode text written in different languages in the onebyte code space. So when we have a text file encoded according to some codepage, we use the codepage table to find which characters were used in this text.

Codepages are used not only to identify characters, but also to simplify text sorting, conversion of lowercase to uppercase characters, spell-checking and in many other applications where it is necessary to know which characters are used in the text.

PAN Localization Project

For example Arabic script has code page 1256 True Type Fonts

Apple Computer, Inc originally designed the True Type digital font format. It was a means of avoiding per-font royalty payments to the owners of other technologies, and a solution to some of the technical limitations of Adobe's Type 1 format. Originally code named "Bass" (as they were scalable fonts) and later "Royal", the TrueType format was designed to be efficient in storage and processing, and extensible. It was also built to allow the use of hinting approaches already in use in the font industry as well as the development of new hinting techniques, enabling the easy conversion of already existing fonts to the TrueType format. This degree of flexibility in TrueType's implementation of hinting makes it extremely powerful when designing characters for display on the screen. Microsoft had also been looking for an outline format to solve similar problems, and Apple agreed to license True Type to Microsoft.

Open Type Fonts

The OpenType font format, jointly developed by Microsoft and Adobe, allows us to combine the best features of the TrueType and Type 1 font formats. For the end user there is little difference between OpenType and TrueType fonts: both are stored in a single font file, both are Unicode-encoded and Windows and Mac OS directly support both.

The real difference is the support of font features that is provided by the OpenType font format. Font features, which can be script and language dependent, can define behaviors of the font, which allow better typographic layout, and precise support of complex scripts.

For detailed OpenType specification and other documents please refer to this page:

http://www.microsoft.com/typography/otspec/default.ht m

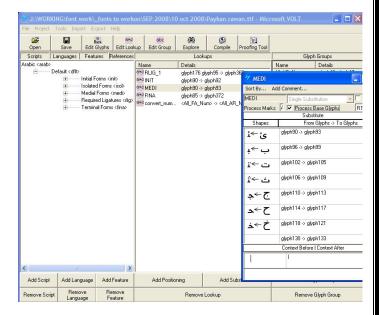
A good document covering OpenType glyph processing in Windows is:

http://www.microsoft.com/typography/developers/open type/default.htm

Font Lab

Ш.	internal 1	Jnicode		-	Window					~	HAME L				
		060B	0600	060D	0610	061B	061F	0621	0622	0623	0624	0625	0626	0627	0628
Ş		فو		,		. :	Ŷ	2	Т	Í	ۇ	ļ	ئ	I	ب
0629	062A	062B	062C	062D	062E	062F	0630	0631	0632	0633	0634	0635	0636	0637	0638
ö	ت	ث	5	۲	Ż	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ
0639	063A	0640	0641	0642	0643	0644	0645	0646	0647	0648	0649	064A	064B	064C	064D
۶	ġ.	-	ف	ق	ک	J	ግ	ن	٥	و	ى	ى		20	п
064E	064F	0650	0651	0652	0653	0654	0655	0653	0660	0661	0662	0663	0664	0665	0666
-	٩	-	-	2	-	2	2	-	•	I	۲	۳	۲.	۵	Ч
0667	0668	0663	066A	066B	066C	066D	0670	067C	067E	0681	0685	0686	0689	0693	0696
V	ື∪	٩	%	,	,	*	1	ټ	پ	حُ	څ	\$	¢	ູ	ŗ
0638	063A	06A3	06AB	06AF	06BC	06C0	0600	06CD	06D0	06E4	06EB	06EC	06F0	06F1	06F2
ژ	ښ	ک	ګ	گ	ن	5	ى	ۍ	ى				•	I	۲
06F3	06F4	06F5	06F6	06F7	06F8	06F3	200C	200D	200E	200F	FB57	FB58	FB59	FB7B	FB7C
۳	۴.	۵	4	V	^	9		1	ſ	1	Ļ	Ļ	. ÷	5₀	÷.
FB7D	FB8B	FB8F	FB30	FB91	FB93	FB94	FB35	FBFD	FDF2	FE80	FE82	FE84	FE86	FE88	FE8A
Ş	ژ ا	ک	5	2	گ	گ	ء	J	لله	2	ີເ	ĺ	ؤ	ļ	ئ
FE8B	FE8C	FESE	FE30	FE31	FE92	FE94	FE96	FE97	FE98	FEBA	FE3B	FESC	FEBE	FESF	FEA0
Ĵ	î	ι	ب	Ļ	. ÷	ä	ت	נ	ï	ث	ڎ	â	ייי	ج	ج
FEA2	FEA3	FEA4	FEA6	FEA7	FEA8	FEAA	FEAC	FEAE	FEB0	FEB2	FEB3	FEB4	FEB6	FEB7	FEB8
7	د ا	2	،	خ	خ	د	ذ	٢	ز	س	ູມ	m	ش	ش	ش
FEBA	FEBB	FEBC	FEBE	FEBF	FEC0	FEC2	FEC3	FEC4	FEC6	FEC7	FEC8	FECA	FECB	FECC	FECE
ص	ص	ص	ض	ض	ض	ط	ط	ط	ظ	ظ	ظ	5	£	*	Ż
FECF	FEDO	FED2	FED3	FED4	FED6	FED7	FED8	FEDA	FEDB	FEDC	FEDE	FEDF	FEEO	FEE2	FEE3
ż	ż	ف	ف	ف	ق	ق	ھ	ک	2	2	Դ	J	្ឋ	ግ	م
	FEE6	FEE7	FEE8	FEEA	FEEB	FEEC	FEEE	FEFO	FEF2	FEF3	FEF4	FEF5	FEF6	FEF7	FEF8
FEE4			i	4	ھ	*	و	5	5	ي_	¥.	ע	Т	Ĺ	Ľ
	ن	Ċ		~	-	-		-	-						
	じ FEFA	FEFB	FEFC	FF00	FF01	FF02	FF03	FF04	FF05	FF06	FF07	FF08	FF09	FFOA	FFOB

MS Volt



MS.Volt image for Glyph substitution

3. Development Tools

The Development tools used were FontLab and Ms. Volt Technology.

4. Results

The PaykanRawan font was successfully made and installed on MS. Windows XP. This font has the feature of obliging the typist to use the correct Yeh's used in Pashto writing language. Since there are 2 Yeh's that should not be written in the middle of a word.

The font was acceptable on the computer screen when the clear font option was enabled. In addition to this, better hinting is needed, for enhanced visibility on the screen.

5. Conclusion

The font needs to be better hinted, in order to make it look better on the computer screen.

6. References

- 1. www.panl10n.net
- 2. <u>http://partners.adobe.com/public/developer/op</u> entype/index_glyph2.html
- 3. <u>http://groups.msn.com/MicrosoftVOLTusersc</u> <u>ommunity/homepage.msnw</u>
- 4. <u>http://www.microsoft.com/typography/develo</u> <u>pers/volt/default.htm</u>
- 5. <u>http://scripts.sil.org/cms/scripts/page.php?site</u> ______id=nrsi&item__id=VOLT__Tutorial
- 6. <u>http://www.truetype-typography.com/</u>
- 7. <u>http://www.crulp.org/software/localization.ht</u> <u>m</u>