

Lao Fonts

Phonpasit Phissamay, Valaxay Dalaloy
Science Technology and Environment Agency (STEA)
phonpasit@stea.gov.la, valaxay@stea.gov.la

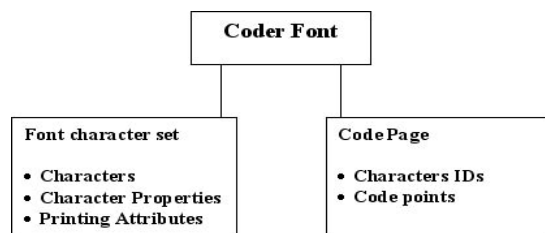
Abstract

This paper discusses font development in Lao language using Microsoft Volt technology. Different Open Type features such as positioning, substitution and kerning have been discussed.

1. Introduction

Font is collected glyphs that are used for visual depictions of character data. A font is combined with a set of parameters, including size, posture, weight, and serifness. Font has three components; character set, code page and font code, and when its set to certain values, generates a collection of imagable glyphs.¹

Font has three components: Coded font, character set and Code page.



2. Coded Font

When you type each font code it will translate your demand, for instance the text you previously entered in a computer terminal, into characters for printing. For a font code, combining with a specific code page and a specific font character, consists of two parts:

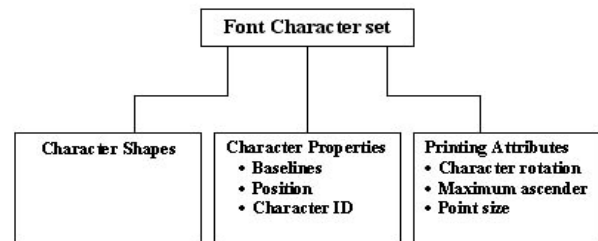
- The specific font character sets of reference
- The specific code pages of reference

Before any character is able to print it must be consist of a specified font character set and listed on the specified code page.

2.1. Font Character Set

A font character set consists of a single type family, typeface, and type size.

A font character set details *character properties* and attributes of printing.



Characters are the letters, numerals, punctuation marks, or other symbols of a font.

Properties of character are introduced in the positions of characters for instance:

- 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, for instance the ID of character A (uppercase A) will be LA20000. The aim of character ID is to decide the character from similar characters, because some characters may look the same but their IDs are completely different.²

² Reference source: (2-16-07)

<http://www.redbooks.ibm.com/redbooks/pdfs/sg244636.pdf>

¹ Reference source: <http://www.icann.org>

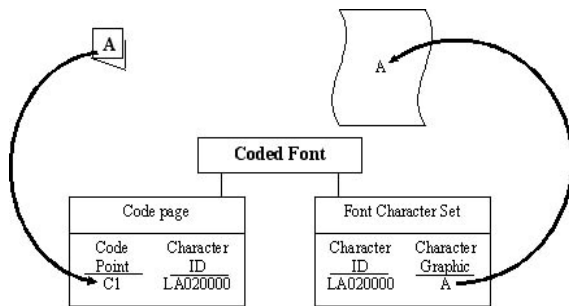
- **Minus sign (-)** Character ID SA000000;
- **Hyphen (-)** Character ID SP100000;
- **Em dash (--)** Character ID SM900000

The *printing attributes* define how the font character set will be printed. Some printing attributes include rotation of characters, maximum ascender, and point size.

2.2. Code Page

A code page will chart the text character of the font character set, and each keyboard character will interpret into a code point when you enter the text at a computer terminal. Then, each code point will be matched to its character ID on the code page when you print the text, and the character ID will also match the character image in front of the character set that you indicated.³

The image in the character set is the image that is printed.



A character ID is an 8-byte character data string. A code point is an 8-bit binary number representing one of 256 potential characters (the maximum number of characters available on a code page). Code points are usually shown as hexadecimal representations of their binary values.

Binary: 11000001; **Decimal:** 193; **Hexadecimal:** C1

3. Word in Lao

3.1. Structure of Lao syllable:

- Level 1: The character appearing in level 1 is of diacritic type. There are five diacritic namely:

◌̄; ◌̅; ◌̆; ◌̇ and ◌̈

- Level 2: Level 2 is occupied by superscript vowels only. The seven vowels of level 2 are:

◌̄̄; ◌̅̄; ◌̆̄; ◌̇̄; ◌̈̄ and ◌̉̄

- Level 3: This level is the main level of Lao word. There is always a character at level 3 at each position in a Lao word. All thirty-three consonants as well as the before and after vowels twelve and 2 special symbols are also at level 3. However some consonants and vowels are also extended into level 2 and level 4 such as:

- Consonants ມ; ນ; ພ; ຜ and before vowels ຄ; ຄ; ຄ are exceeded into level 2
- Consonants ຖ; ຊ; ງ and symbol ງ; ງ are exceeded into level 4.

- Level 4: The characters appearing in level 4 is lowered script vowels and “◌̉; ◌̉; ◌̉”. d consonant. There are following symbols:

Due to the four levels structure, the high and length of characters existed in each level are not the same. If considering the character in the level 3 is main for compare then the size of character in level2 and level4 are equivalent 50% of size of character in level3. And the size of character in level1 is equivalent 50% of size of character in level2

3.2. The type of Lao characters:

The type of Lao characters development also impacted from the country development such as regime and equipment facilities. However it can be classified into 3 groups:

1. The traditional or old typewriter: Based on MAHASILA grammar book (Old Lao Grammar) this has been developed during the royal regime (before 1975). The characteristic is rounded glyphs with thin and uniform-width strokes. Example:

ໂຄງການພັດທະນາພາສາລາວໃນຄອມພິວເຕີ

2. The new typewriter or schoolbook in present: Based on PHOUMY VONGVICHITH grammar book (new Lao grammar) this has been developed after establishment of LAO PDR (after 1975). The characteristic is glyphs with straight strokes where possible, and somewhat heavier uniform-width strokes. Example:

ໂຄງການພັດທະນາພາສາລາວໃນຄອມພິວເຕີ

³ Reference source: <http://www.redbooks.ibm.com/redbooks/pdfs/sg244636.pdf>
280

- Ornamental glyph: The new development glyph in order to make the Lao character look more beauty. The most of the modern glyphs are developed since last five year after the computer has created a big impact into the printing materials. Most of this glyphs are using in the brochure, advertisement letter or magazine. The characteristic is calligraphic strokes, handwriting styles. Example:

--
ໂຄງການພັດທະນາພາສາລາວໃນຄອມພິວເຕີ

4. Lao Fonts

4.1. Factors for considerations:

Lao font has four main factors to consider:

- The word-wrapping is important for large amounts of text and it would be much more convenience, especially for line breaking. But when the text must be edited, preventing minor changes from every subsequent line needing adjusting.

- When the text consists of Lao and roman characters in single font of Unicode there would not have a problem, however, it is a problem when the texts mixed languages are in a single entry by using ASCII font.

- Some Lao fonts use the standard codes for numbers and arithmetic symbols, for other characters can lead to program errors, especially in spreadsheet and database applications. The hyphen code is often recognized as a minus sign, and must be used with care.

- Lao fonts have a few heading signs for brochures and books but they use signs from a wide range of font styles.⁴

What style the font is drawing in must be decided before drawing even the first character so that they will all be balanced in shape and style. It is important to decide on basic width for character in reference to the showing position, especially for the tone mark and superscript vowels they have many different positions placed in the syllable.

4.2. Methodologies:

⁴ Reference source: <http://www.cicc.or.jp/english/hyoujyunka/mlit4/7-5LAOPDR/Laos1.htm>

There are 3 stages for Lao shaping engine processes text:

- Characters are analyzed for valid diacritic combinations.
- Shape is substituted glyphs with OTLS (Open Type Library Services).
- Position glyphs with OTLS.

4.2.1. Analyzing characters

The contextual analysis engine is to prove valid diacritic combinations, and its shaping engine unit is a string of Unicode characters, in a sequence. For more information please see Invalid Combing Marks.

The handling of the AM in the analysis phase is special and where an above mark does not exist on the preceding base consonant its characteristics will be use to decompose the AM into the NIGGAHITA and AA glyphs. Then its glyphs will allow to be positioned correctly above the preceding base consonant. If the tone mark is on the base consonant the analysis engine will decompose the AM and reorder the NIGGAHITA to between the base consonant and the tone mark. The NIGGAHITA glyph will be positioned correctly above the base consonant, and the tone mark to be positioned correctly above the NIGGAHITA. This kind of method cannot be tested in VOLT, as this logic is not in VOLT.

Backing stroke	'comp' form
U+0EB3	U+0ECC U+0EB2
ອ	ອ + ອ

4.2.2 Shape Glyphs

Shaping character string of Uniscribe is to map all the characters to the glyphs form. The Uniscribe uses OTLS to relate the characteristics. The processing of OTL is separated to a set of predefined characteristics, which apply one by one to the glyphs in the syllable and then the OTLS will process them.

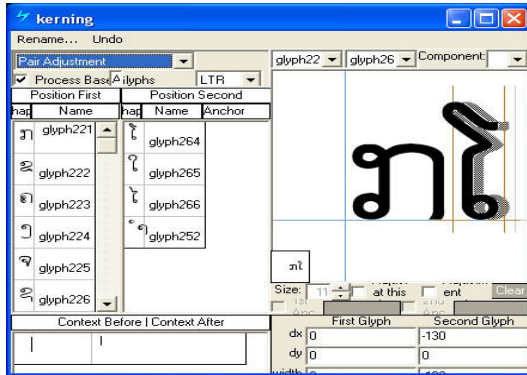
4.2.1. Position Glyphs with OTLS

The position of glyphs with OTLS to position the glyphs, Uniscribe applies to the function of OTLS for help.

Characteristics the positioning:

The font stores a set of adjustments for pairs of glyphs, including one or more tables matching left and right classes or individual pairs. If both forms are used, the classes should be listed last; replacing any non-ideal value will result from the class tables. It will provide adjustment for larger sets of glyphs to overwrite the results of pair kerns in combinations. These should be in front of the pairs.

Example:



Using Microsoft VOLT to kern the pairs of glyphs

Before Kerning

ນໍາໃຊ້ຂົນໄກ່ໃຫ້ເປັນປະໂຫຍດ

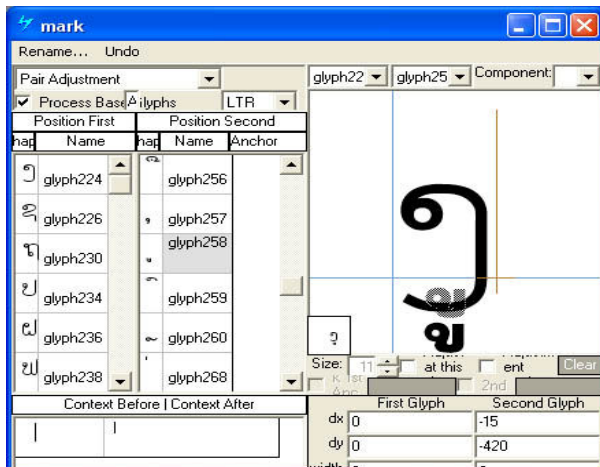
After Kerning

ນໍາໃຊ້ຂົນໄກ່ໃຫ້ເປັນປະໂຫຍດ

4.3.3. Mark to base positioning

The 'mark' characteristic positions mark glyphs that related to a ligature glyph. Its feature implements as a MarkToLigature.

Example:



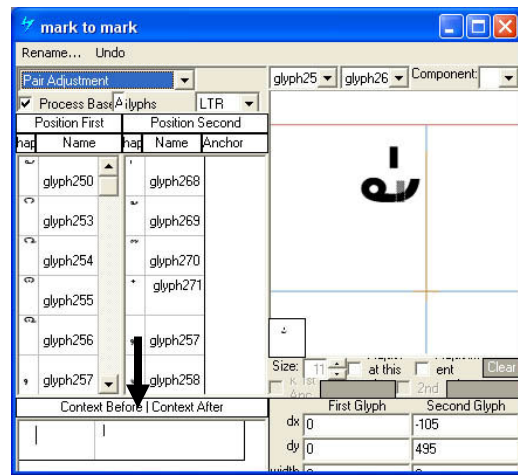
Using Microsoft VOLT to position the mark to mark

Before: ງ ຊ ຝັງ ປ້ອນ ພ້ອຍ
 After: ງ ຊ ຝັງ ປ້ອນ ພ້ອຍ

4.3.4 Positioning of mark to mark

The mark to mark is positioning marks glyphs that are related to another mark glyph. Its characteristic will work as a MarkToMark.

Example:



Positioning mark to mark using Microsoft VOLT

Before: ນີ້ ນີ້າ ສະຫຼຸບ ປີ່ງ
 After: ນີ້ ນີ້າ ສະຫຼຸບ ປີ່ງ

5. Reference:

- [1] "Microsoft Fontlap Open Type"
<http://www.microsoft.com/fontlap/OpenType%20Dev/lao/intro.msp>
- From: (5-1-03) <http://www.asia.microsoft.com/typography/otfntdev/laoot/features.htm>
- [2] "LaotianFonts"
<http://cg.scs.carleton.ca/~luc/laos.html>

⁷ Reference source: (5-1-03) <http://www.asia.microsoft.com/typography/otfntdev/laoot/features.htm>

Figure 1: The glyphs characteristic of each Lao character:

Glyph Numbers	Characters			Unicode	Type	High (Unit)	Length (Unit)
	Traditional	Schoolbook	Ornamental				
221	ງ	ᨡ	ᨢ	U+0E81	Simple	1168	1090
222	ຣ	ᨣ	ᨤ	U+0E82	Simple	1165	1090
223	ຄ	ᨥ	ᨦ	U+0E84	Simple	1165	1090
224	ງ	ᨧ	ᨨ	U+0E87	Simple	1586	901
225	ຈ	ᨩ	ᨪ	U+0E88	Simple	1165	1090
226	ຂ	ᨫ	ᨬ	U+0E8A	Simple	1586	1099
227	ຢ	ᨭ	ᨮ	U+0E8D	Simple	1165	1090
228	ດ	ᨯ	ᨰ	U+0E94	Simple	1165	1090
229	ຕ	ᨱ	ᨲ	U+0E95	Simple	1465	1090
230	ງ	ᨳ	ᨴ	U+0E96	Simple	1586	1090
231	ທ	ᨵ	ᨶ	U+0E97	Simple	1465	1179
232	ນ	ᨷ	ᨸ	U+0E98	Simple	1165	1090
233	ບ	ᨹ	ᨺ	U+0E9A	Simple	1165	1090
234	ປ	ᨻ	ᨼ	U+0E9B	Simple	1465	1179
235	ຜ	ᨽ	ᨿ	U+0E9C	Simple	1465	1179
236	ຝ	ᩁ	ᩂ	U+0E9D	Simple	1465	1179
237	ພ	ᩃ	ᩄ	U+0E9E	Simple	1165	1179
238	ຟ	ᩅ	ᩆ	U+0E9F	Simple	1465	1179
239	ມ	ᩇ	ᩈ	U+0EA1	Simple	1165	1090
240	ຢ	ᩉ	ᩊ	U+0EA2	Simple	1465	1090

241	ຮ	ᩋ	ᩌ	U+0EA3	Simple	1165	1090
242	ລ	ᩍ	ᩎ	U+0EA5	Simple	1165	1090
243	ວ	ᩏ	ᩐ	U+0EA7	Simple	1165	1090
244	ສ	ᩑ	ᩒ	U+0EAA	Simple	1165	1090
245	ຫ	ᩓ	ᩔ	U+0EAB	Simple	1465	1179
246	ອ	ᩕ	ᩖ	U+0EAD	Simple	1165	1090
247	ຮ	ᩗ	ᩘ	U+0EAE	Simple	1165	1090
248	ຯ	ᩙ	ᩚ	U+0EAF	Simple	1586	1090
249	ະ	ᩛ	ᩜ	U+0EB0	Simple	1165	848
250	ັ	ᩝ	ᩞ	U+0EB1	Mark	344	862
251	ຳ	᩟	᩠	U+0EB2	Simple	1165	883
252	ັ້	ᩡ	ᩢ	U+0EB3	Simple	1643	1667
253	ິ	ᩣ	ᩤ	U+0EB4	Mark	348	882
254	ີ	ᩦ	ᩧ	U+0EB5	Mark	348	1069
255	ື	ᩩ	ᩪ	U+0EB6	Mark	348	882
256	ູ	ᩬ	ᩭ	U+0EB7	Mark	348	1069
257	ຸ	ᩮ	ᩯ	U+0EB8	Mark	422	344
258	ູ	ᩱ	ᩲ	U+0EB9	Mark	372	420
259	ື	ᩴ	᩵	U+0EBB	Mark	344	862
260	ູ	᩷	᩸	U+0EBC	Mark	384	916
261	ຽ	᩺	᩻	U+0EBD	Simple	1586	1090
262	ໄ	᩼	᩽	U+0EC0	Simple	1165	438
263	ໄ	᩿	᩻	U+0EC1	Simple	1165	959
264	ໄ	᩻	᩻	U+0EC2	Simple	690	1804
265	ໄ	᩻	᩻	U+0EC3	Simple	829	1804

266	ໂ	ໂ	ໂ	U+0EC4	Simple	778	1804
267	ງ	ງ	ງ	U+0EC6	Simple	1586	1017
268	໐	໐	໐	U+0EC8	Mark to Mark	112	318
269	໐	໐	໐	U+0EC9	Mark to Mark	281	656
270	໐	໐	໐	U+0ECA	Mark to Mark	812	681
271	໐	໐	໐	U+0ECB	Mark to Mark	350	350
272	໐	໐	໐	U+0ECC	Mark to Mark	400	474
273	໐	໐	໐	U+0ECD	Mark to Mark	434	434
274	໐	໐	໐	U+0ED0	Simple	1165	1090
275	໑	໑	໑	U+0ED1	Simple	1165	1090
276	໒	໒	໒	U+0ED2	Simple	1465	1090
277	໓	໓	໓	U+0ED3	Simple	1465	1179
278	໔	໔	໔	U+0ED4	Simple	1465	1090
279	໕	໕	໕	U+0ED5	Simple	1465	1090
280	໖	໖	໖	U+0ED6	Simple	1465	1179
281	໗	໗	໗	U+0ED7	Simple	1586	1090
282	໘	໘	໘	U+0ED8	Simple	1586	1179
283	໙	໙	໙	U+0ED9	Simple	1165	1179
284	໙	໙	໙	U+0EDC	Simple	1165	1707
285	໙	໙	໙	U+0EDD	Simple	1165	1713