

# Lao Keyboard

Phonpasit Phissamay, Valaxay Dalolay, Thonglor Duangsavanh, Vethsouvanh Phengchanh,  
Khamkeo Kommadam

*Science Technology and Environment Agency (STEA)*

[phonpasit@stea.gov.la](mailto:phonpasit@stea.gov.la), [valaxay@stea.gov.la](mailto:valaxay@stea.gov.la)

## Abstract

*This paper documents the research conducted while creating Lao keyboard file in XKB to have Lao keyboard support in Windows Operating system.*

## 1. Definition

A keyboard layout is the collection of data for each keystroke and shift state combination within a particular keyboard driver. It is not the physical keyboard that user types on, but rather, the software that the hardware calls to output text streams to applications.

Keyboard input starts at the hardware level. The keys on the physical keyboard each have a value assigned to them called a scan code, and these scan codes are sent wherever you type a key. As we progress from the hardware and move to the software level, what becomes crucial is the Virtual Key (VK) value.<sup>1</sup> These values fit within a byte (0x00 to 0xff) and are defined in winuser.h: the Platform SDK header file that contains procedure declaration, constant definitions and macros for USER subsystem of Windows.<sup>2</sup>

When a Windows message loop handles a VK in the WM\_KEYDOWN message, it can pass the Virtual Key to the DefWindowProc API. To handle the message, the code in the USER subsystem will process the keystroke and convert it to a character, passed as a WM\_CHAR message. This processing requires a great deal of information: the shift state, the virtual key and the current keyboard layout.<sup>3</sup>

When developing a keyboard, we should consider some of factors as following:

- When developing a keyboard, we should consider some of the following questions:

- Does each country or region need one keyboard standard?
- What language will be supported in the keyboard?
- Should it include Linguistic characters for a specific language script? This requirement can be found only when a user uses dead keys or an additional shift state. Linguistic characters should be positioned where it is easy to type, ideally is by using a non shifted state for typing.
- Does the keyboard focus on code points and not be displayed as glyphs? It will be crucial if we do not change or place the display of the shape on a keyboard and should be displayed by fonts and the display engine is needed.
- Do all Unicode characters exist on the keyboard? Because, most inputs on windows are base on Unicode and can not be handled if the code points are not encoded to Unicode.
- Keyboard behavior should be concordant with the composite characters.

## 2. Lao Keyboard layout

### 2.1. The character existed in the Lao keyboard layout

The Lao language is composed of 65 unit characters and signs which contained 27 characters of single consonants, 06 mixed consonants characters, 14 single vowels characters, 04 tone marks characters, 03 special symbols characters and 10 characters of Lao digit numbers. The number of symbol in the keyboard is larger than English symbols, but doesn't have upper case and lower case characters. Lao characters in any characters set code are assigned key positions on standard keyboard without a problem. The remaining symbols are activated by using the shift function for typing.

Because the number of buttons existed in Standard English keyboard layout are having grater number then

<sup>1</sup><http://win32.reb00t.com/system/Microsoft%20Keyboard%20Layout%20Creator/Unicode-KbdsonWindows.pdf>

<sup>2</sup><http://www.bhashaindia.com/Developers/KnowHow/KeyboardLayout/index.aspx>

<sup>3</sup><http://www.bhashaindia.com/Developers/KnowHow/KeyboardLayout/index.aspx>

number of Lao characters, therefore many developers have added the symbol that already existed in the English keyboard into the Lao keyboard as well, Also some developers has created the new glyph in both or either in the keyboard layout or in the internal code table.

Usually the number of characters in the Keyboard is equal with the number of character set in the internal code table. However some keyboard layout might have grater or lesser numbers then internal code table. The following table is representing on how developers coding the characters:

Option	Hand writing	Keyboard	Character set
1	$\overset{\sim}{X} = \overset{\sim}{x} + \overset{\sim}{x}$	$\overset{\sim}{X} + \overset{\sim}{x}$	$\overset{\sim}{X} + \overset{\sim}{x}$
2	$\overset{\sim}{X} = \overset{\sim}{x} + \overset{\sim}{x}$	$\overset{\sim}{X}$	$\overset{\sim}{X} + \overset{\sim}{x}$
3	$\overset{\sim}{X} = \overset{\sim}{x} + \overset{\sim}{x}$	$\overset{\sim}{X} + \overset{\sim}{x}$	$\overset{\sim}{X}$
4	$\overset{\sim}{X} = \overset{\sim}{x} + \overset{\sim}{x}$	$\overset{\sim}{X}$	$\overset{\sim}{X}$

Only the option 1<sup>st</sup> can applied to the Unicode structure and rule. However many font developer using the option 2<sup>nd</sup>,3<sup>rd</sup> and 4<sup>th</sup> in the ASCII code table structure for their own reason and convenient, therefore it made the number of keyboard layout and internal code table are not equal. This will facing the problem for the counting the character in the text, converting the text from one code table to another code table and for using the font in the different platform or application that not specified in the rule.

## 2.2. The types of Lao Keyboard layout

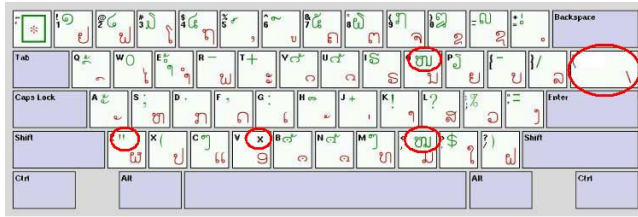
The national standards of Lao keyboard layout haven't defined yet. From the observation we can classified the existing Lao keyboard layout into 3 groups as following:

- Keyboard layout based on roman phonetic (Ex: Hongkard keyboard): On this keyboard layout was assigned the Lao and English character (or foreign language) that have similar sound pronunciation into the same keyboard map position such as: a-າ; b-ບ; c-ສ; d-ດ and etc. This solution is convenient for foreign to learn on writing Lao language, but it is not popular for Lao original people, since most of them do not understand foreign language.

- Keyboard layout based on group of characters such as consonant, vowels and tone marks: On this keyboard, the layout is divided into the character group zone such as: All the consonant character will be located in the same zone and next to each order, which usually occupied in the line 2<sup>nd</sup> and 3<sup>rd</sup>, all the vowels character are located in the line 3<sup>rd</sup> and 4<sup>th</sup> and followed by Lao symbols; All the Lao numeric number and tone marks are located in the line 1<sup>st</sup>. The advantaged of this layout is easier to find out the position of each character, but it is not convenient in the practise, because the frequently use character were located in the corner of keyboard, which the two last finger of our hand are working harder then other finger, which it is not a common strategies for any language keyboard layout.
- Keyboard layout based on typewriters: On this layout the location of each character position are depending on the frequencies used and not distinguish on whether there are consonant, vowels or tone marks. The highly used characters of consonant, vowels and tone mark group are located in the centre of keyboard layout between line 2<sup>nd</sup> to 4<sup>th</sup>, which mostly used by index and middle fingers. The neutral frequently used character will located next to the centre in the same line 2<sup>nd</sup> to 4<sup>th</sup>, the low frequently used will located in the line 1<sup>st</sup> and the less frequently used will located in the second layout, which activated by the Shift function. .



**Figure 1: The old keyboard layout based on typewriter**

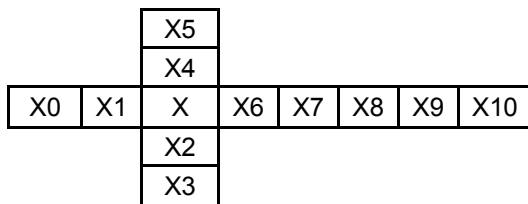


**Figure 2: The keyboard layout based on Lao Script for Window Software**

### 3. Input method and rule

In the hand writing system, the orthography of Lao characters is to write in the sequence of Consonant-Vowel-Alternate Consonant-Tone mark. But not everyone writes according to the orthography and it was found that some people thought that Consonant-Vowel-Tone mark-Alternate Consonant was the orthography.

In the computer system, the orthography of Lao character is to write from Left to Right with the following sequent order:



1. X0 represents a vowel which occurs before the nuclear consonant.
2. X1 is a combination consonant which comes before the nuclear consonant
3. X is represents the nuclear consonants.
4. X2 is a combination consonant which comes after the nuclear consonant, which placing under or next to the nuclear consonant.
5. X3 is represents a subscription vowel which occurs under the nuclear consonant.
6. X4 is represents a superscription vowel which occurs upper the nuclear consonant.
7. X5 is represents a tone marks which occurs upper the nuclear consonant or upper vowels.

8. X6 is represents consonant vowel, which occurs after nuclear consonant.
9. X7 is represents an after vowels.
10. X8 is represents alternate consonants
11. X9 is represents alternate consonant to pronounce foreign language
12. X10 represents a sign mark.

To avoid misspelling syllable, we can also define some rule for the input such as:

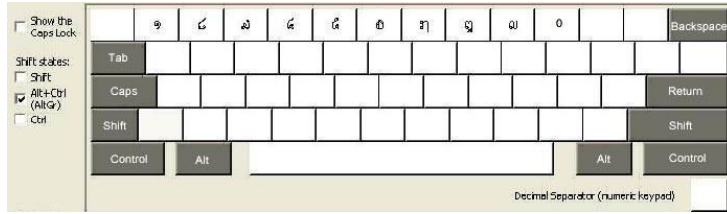
- The same Vowels and Tone Mark character cannot placing next to each order
- The same group of vowels and tone mark (except after vowels) cannot be placing next to each order.
- Etc.

#### I. The existing types of Lao Keyboard driver

The system to allowed input the Lao character are also depend of the structure of Internal code table of Lao character set. As we classified variety of Lao character sets into 3 groups as Upper ASCII, Lower ASCII and Unicode. The system to allow input of Lao character is also different based on each code table, which can elaborate in the following:

1. For upper ASCII: It is very easy to input, which just changed the English font into Lao font created based on this code table. However since the Lao language is syllable structures with have four levels, which are very much different of English structure with have only one level. Therefore to input the character that placing above or under based line are not look so nice, because that system will counted into another line of each structure, which mean that each line in Lao language will take four lines in English language. It is really not a good solution.
2. For lower ASCII: In order to input the Lao character based on this code table, it is required the add-on software (create a separate program for the operation system to activate the Lao character). Each input system is different on the quality of





**Figure 5: The keyboard layout in the third level: activate by the right alternate function**