Design and Implementation of Jawi keyboard layout: A Proposal

Mohd Zamri Murah
Pattern Recognition Group
Center for Artificial Intelligence Technology
Fakulti Teknologi Dan Sains Maklumat
Universiti Kebangsaan Malaysia
zamri@ftsm.ukm.my

Abstract

This paper presents a proposal for a Jawi keyboard layout based on Arabic(Saudi Arabia) keyboard layout. This layout allows the users to enter the six additional Jawi characters while maintaining the standard layout for the other Arabic characters. The main different between the two keyboards layout is that we change the positions for Arabic diacritical markings into the positions to enter the six Jawi characters (چ ,ڠ ,ک ,ۏ ,ڤ ,ڽ). The layout would be familiar to those which Arabic background. The layout is not based on phonetic of the Jawi characters, for instance the arabic character alef \ is not map into the character 'a', thus may make it difficut for new users.

1. Introduction

There is currently no Malaysian standard for Jawi keyboard layout. Users who would like to write Jawi face a difficult problem because the current keyboard layout for Arabic doesn't support the additional six Jawi characters. Eventhough, there are various Jawi mapping available from various hardware manufactures for Jawi keyboard stickers and keyboards, each manufactures have their own Jawi layout mapping. This has make it difficult for Jawi users to write Jawi effectively.

In this paper, we propose a standard layout for Jawi based on Arabic(Saudi Arabia) layout. This layout is choosen because;

1. it is the standard keyboard layout for the Arabic-writing world
2. it has mapping for all the Arabic characters used in Jawi
3. it is familiar to those with Arabic education background
4. it has minimal changes to the Arabic(Saudi Arabia) layout

2. Keyboard layout

2.1. The normal keyboard

The propose Jawi keyboard layout for normal position is shown in figure 1. The position is similar to the Arabic(Saudi Arabia) keyboard layout.

2.2. The shift-position

The propose Jawi keyboard layout for shift position is shown in figure 2. There are 7 changes in the Arabic(Saudi Arabia) keyboard layout positions; 6 diacritical markings are change into Jawi characters (see on the upper left hand side of the keyboard) and 1 diacritical marking for ZWNJ (Zero-Width None Joiner)(refer 3.3. All other positions remain unchanged.

Figure 1: Position of Jawi characters in keyboard layout. It is the same with Arabic(Saudi Arabia) keyboard layout.

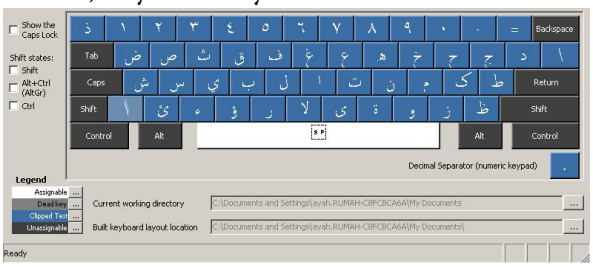
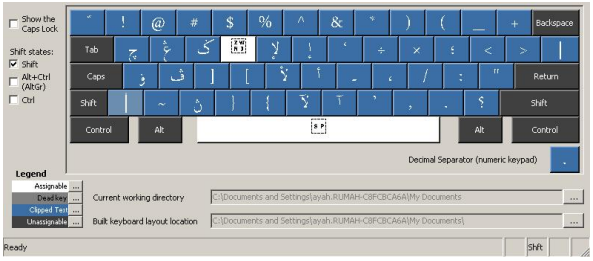


Figure 2: Position of Jawi characters in shift-position. There are six changes where 6 diacritical markings are changes into Jawi characters چ ڤ ك غ ڠ and 1 diacritical marking for ZWNJ.



3. Discussion

3.1. Advantages

The advantages of this jawi keyboard layout;

1. it is based on the standard Arabic keyboard layout. Since the standard Arabic keyboard layout is available on all computer platform, it is easier for users to install and use the Jawi keyboard layout.

2. it has mapping for all the Arabic characters used in Jawi, thus reducing time to design a new layout for Jawi.

3. it is familiar to those with Arabic education background.

4. it make only 7 changes to the Arabic(Saudi Arabia) layout.

3.2. Disadvantages

The disadvantages of this layout are;

1. not based on phonetic of the Jawi characters (e.g the character 'a' for arabic character ا) or character 'b' for arabic character ب). This probably will it difficult for new users to know the locations of each Jawi characters.

2. users may need to buy keyboard stickers or a new Arabic keyboard to use the mapping effectively.

3.3. ZWNJ

ZWNJ is an invinsible coding to separate Jawi words from joining. This is important in some words spelling. Example;

1. the word 'sains' is spell س ساءين, not ساءينس.

2. the word 'golf' is spell ف ڬول, not ڬولف

3. the word 'teks' is spell س تيک, not تيکس

3.4. tatweel

Tatweel (0640) is an arabic character to stretch characters.

For example;

1. without tatweel: ساي سوکا جاوي

2. with tatweel: ســـــاي ســـــوکـــــا جـــــاوي

3.5. The mapping of ک

There are two possible mapping of gaf (or kaf with dot above) , 06AC (kaf with dot above) and 0762 (keheh with dot above) [3, 2]. Each however produce different ligature glyph based on initial, middle and final position of the characters. [1]

The final (correct) output is application dependent, for instance `notepad.exe` produce the correct final shape for both mapping. Some version on `notepad.exe` on Window XP produces disjoint characters.

1. ك (06AC) (kaf with dot above)
   (a) initial position: كسا
   (b) final position: سك (the final ligature is wrong)
   (c) middle position: سكي
2. ک (0762) (keheh with dot above)
   (a) initial position: کسا
   (b) final position: سک
   (c) middle position: سکي

The main problem is that not many fonts support ك on the 0762 mapping.

3.6. The mapping of ک

The letter ک can be map into 06A9 (keheh) and 0643(kaf). Each produce a different final shape in a letter.

1. ك 0643 (kaf)
   (a) initial: كلس
   (b) middle: سكي
   (c) final: سيك (the final ligature is wrong)
2. ک 06A9 (keheh)
   (a) initial: کیس
   (b) middle: سکي
   (c) final: سیک

3.7. The proposal for ڭ and ک

We propose the use of keheh(06A9) and kekeh with dot(0762) since both of them produce the correct final glyph in a word such as سیک and سڭ.

3.8. The issue of hamza ء

There are two hamza encoded;

1. Arabic letter hamza (U+0621): ء
2. Arabic letter high hamza (U+0674): ٔ

The position and the size of the two hamza are different. Compare the use of the two hamza: کبڅساءن and کبڅسأن.

In certain words, the correct position of ء is 3/4 from the baseline. Compare the position of hamza using a standard font in the word کبڅساءن and the correct position of hamza in the same word کبڅساٴن . (Notice the slightly shift upward position of hamza in the second word)

There are currently no fonts with hamza in 3/4 position . However, a hack using subscript could be used, but the resulting line would be shift-up and would produce a rather imbalace document structure.

4. update 26/10/2009

4.1. TC Multilingual keyboard layout

The latest keyboard layout as agreed by TC Multilingual Group is different from this paper proposal. The keyboard layout is given in figure 3 and figure 4.

4.2. Amendments to be implemented

The TC multilingual keyboard layout shift positions have these features;

Figure 3: Position of Jawi characters in keyboard layout. It is the same with Arabic(Saudi Arabia) keyboard layout.

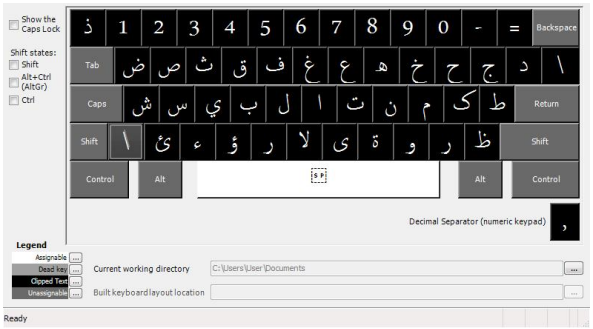
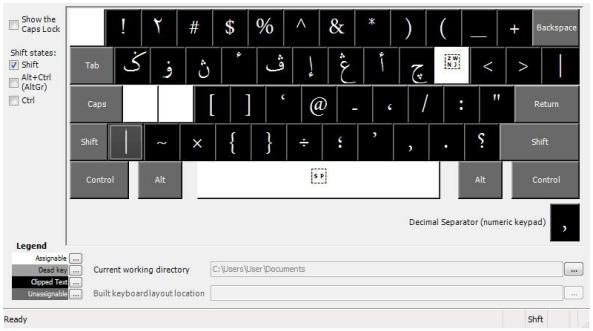


Figure 4: Position of Jawi characters in keyboard layout in shift position.



1. a position for 3/4 hamza in shift position. Currently, we do not have UNICODE encoding for 3/4 hamza. A key is reserved for the future UNICODE for 3/4 hamza.

2. the of high hamza ٔ as a temporary solution for the lack of 3/4 in all fonts.

References

[1] *Daftar ejaan Rumi-Jawi*. Dewan Bahasa dan Pustaka, Kementerian Pendidikan Malaysia, 2008.

[2] M. Davis. Moving to Unicode 5.1, 2008.

[3] B. Patterns. The Unicode Standard, Version 5.1. *Unicode, Inc*, 2008, 1991.