マレー語ワードプロセッサの設計
－入力方式の検討－

R.B.FAKEH、米田政明、長谷博行、酒井充
富山大学工学部電子情報工学科

　マレー語にはローマ文字系のRumi文字を使用するものとアラビア文字系のJawi文字を使用するものがあり、前者は広く日常使われており、後者はイスラム系の人々の間で主に宗教上の目的に使われている。Rumi文字用には一般の英文ワープロが利用できるが、Jawi文字用に利用できるワープロはあまりない。Jawi文字は右から左に書かれ、更に基本的に36文字種を持っているが使用する際にはワード中の位置や前に来る文字種によって字体が異なり、実際には100以上の字体が存在する。本研究ではこのJawi文字用のマレー語ワープロの設計、特にJawi文字の入力方式について、字体をキーに割り付けるタイプライター方式と自動選択方式の2つの方式を提案し検討を行った。

A DESIGN OF MALAY WORD PROCESSOR
— INPUTTING SYSTEM —

R.B.FAKEH, M.YONEDA, H.HASE, M.SAKAI
Faculty of Engineering, Toyama University.
3190 Gofuku, Toyama 930, Japan

　A design of Malay word processor for Jawi application has been initiated that may help many users in the Malay and Islamic community. Actually the shapes of Jawi letters depend upon the letter's position in a word. Therefore, the correct shapes of Jawi letters must be selected to match the rules of Jawi writing. In this paper we present two methods of selecting the appropriate shape of a Jawi letter, namely, the typewriter approach and the automatic selector approach.

1. INTRODUCTION

Malay language which is officially called "Bahasa Malaysia" uses two types of writing letters. One is called Rumi and the other Jawi. Rumi letters, which are the same as the Roman letters, are used very widely in most of the government offices and in daily communication. Over the past twenty years many text editors have been designed(1) , some of which represent the basis of the commercially available word processors. A design for a Malay word processor for Rumi application is not necessary due to the abundant availability of English word processors such as WordStar, which can be used very conveniently.

Jawi letters originating from Arabic letters are not widely used now. Their usage covers only the area of Malay and Islamic religion. There are very few reports or papers that show the formal design of Malay word processor for Jawi application(2). Malay language using Jawi letters has an entirely different way of writing from right to left and uses different types of letters. Therefore, a special word processing system for Jawi letters must be developed separately.

In designing a Malay word processor for Jawi application the designer faces many difficulties. The most significant difficulty is that Jawi letters have more than one shape. Most of them may have up to four distinct shapes, depending upon the letter position in a word. If a key is assigned to each shape, the data will be inputted with considerably low speed. In this paper the Jawi letters are analysed. And input design aspects will be presented to deal with the problem of inputting the Jawi letters.

Chapter 2 will explain the Jawi letters and analyze their shapes.

Chapter 3 will explain the methods used in inputting the Jawi letters. Two approaches will be presented, namely the typewriter approach and the automatic selector approach.

Chapter 4 will conclude the overall paper.

2. JAWI LETTERS

Jawi letters(3) consisting of 36 letters shown in Fig.1 are originated from Arabic letters. Each Jawi letter can be compared by analysing the shape of the letter. If the body shapes are the same, the difference can be made by comparing the dots.

The Jawi letters in Fig.1 have different shapes when compared with one another. These letters also have different shapes if used in a word. Most of the Jawi letters can have up to four distinct shapes, depending on the position and the preceding letter in a word, the beginning, the middle, or the ending, and the isolated shapes.

The utilization of the four distinct shapes of Jawi letters will be explained in the following example. For the word "Bersambung(in Rumi)" which means "Connected", the selection of Jawi letters in isolated shape are as follows:

"NGA""WAU""BA""MIM""SIN""RO""BA"

غ و ب م س ر ب

However, the correct Jawi writing, utilizing the beginning, middle, ending,

	Jawi names	Isolated shape	Beginning shape	Middle shape	Ending shape
1	ALIF	ا			ـا
2	BA	ب	بـ	ـبـ	ـب
3	TA	ت	تـ	ـتـ	ـت
4	THA	ث	ثـ	ـثـ	ـث
5	JIM	ج	جـ	ـجـ	ـج
6	HA	ح	حـ	ـحـ	ـح
7	KHO	خ	خـ	ـخـ	ـخ
8	CA	چ	چـ	ـچـ	ـچ
9	DAL	د			ـد
10	ZAL	ذ			ـذ
11	RO	ر			ـر
12	ZAI	ز			ـز
13	SIN	س	سـ	ـسـ	ـس
14	SYIM	ش	شـ	ـشـ	ـش
15	SAD	ص	صـ	ـصـ	ـص
16	DAD	ض	ضـ	ـضـ	ـض
17	TO	ط	طـ	ـطـ	ـط
18	ZO	ظ	ظـ	ـظـ	ـظ
19	AIN	ع	عـ	ـعـ	ـع
20	GHAIN	غ	غـ	ـغـ	ـغ
21	NGA	ڠ	ڠـ	ـڠـ	ـڠ
22	FA	ف	فـ	ـفـ	ـف
23	PA	ڤ	ڤـ	ـڤـ	ـڤ
24	QAF	ق	قـ	ـقـ	ـق
25	KAP	ک	کـ	ـکـ	ـک
26	GA	ݢ	ݢـ	ـݢـ	ـݢ
27	LAM	ل	لـ	ـلـ	ـل
28	MIM	م	مـ	ـمـ	ـم
29	NUN	ن	نـ	ـنـ	ـن
30	WAU	و			ـو
31	VAU	ۏ			ـۏ
32	HEA	ه	هـ	ـهـ	ـه
33	LAM_ALIF	لا			ـلا
34	AMZAH	ء			
35	YA	ي	يـ	ـيـ	ـي
36	NYA	ڽ	ڽـ	ـڽـ	ـڽ

Fig. 1 List of Jawi letter shapes.

and isolated shapes is as follows:
"NGA""WAU""BA""MIM""SIN""RO""BA"

غ و ب ـم ـس ر ب

The first letter BA(ب) when it exists at the beginning of a word is written using the beginning shape. The second letter RO(ر), which is in the middle position of the word, does not have the middle shape, because this letter can not be connected to the following letter but it can be connected to the preceding letter. In this case, the ending shape is chosen.

The third letter of SIN which appears in the middle position of the word, has all the four shapes. But in this case, the preceding letter RO is not able to be connected to the following letter. The SIN letter has to be chosen from the beginning shape. The next leter MIM is taken from the middle shape, since the connection to the preceding and the following letter is possible. The next letter BA is also from the middle shape. The next letter WAU has to be in the ending shape because there is no middle shape for this letter. NGA which appears in the ending position is chosen from the isolated shape because the preceding letter WAU is not able to be connected to the following letter (in this case NGA).

The final Jawi writing performed by connecting the above shapes becomes:

برسمبوغ

3. INPUT SYSTEM OF JAWI LETTERS

There are two approaches proposed in this paper that can be applied in the design of Malay word processor for Jawi application. Research is being initiated on all the two approaches, which will act as a beginning to the proper design of a Malay word processing system. The two approaches are:

    (a) The typewriter approach, which uses the concept of using a typewriter keyboard with the fixed shapes of Jawi letters.

    (b) The automatic selector approach, which uses the concept of selecting the correct shapes of Jawi letters automatically, with respect to the basic shape of 36 Jawi letters.

3.1 Typewriter approach

Initially the idea of designing a typewriter approach to input the Jawi letters came from the same concept of using a Jawi typewriter.

The number of Jawi letters in Fig. 1 that have four distinct shapes are further analysed, so that the number of Jawi characters can be reduced to suit the typewriter approach.

3.1.1 Font design

A matrix size of 8x16 has been used in the Jawi font design. There are 78 shapes that have been designed, as shown in Fig.2, that will be sufficient to be used in writing Malay language using Jawi letters.

Referring to Fig.2, the analysis of the reduction of Jawi letters is explained using an example of "BA".

Fig. 2 List of reduced Jawi letters for typewriter approach.

BA -- Two shapes of the isolated and the beginning shapes are designed. A modification of the isolated shape is made so that utilization by ending shape is possible. The beginning shape is modified to suit the middle shape. Modification is achieved as shown in Fig. 3.

In the font design, a matrix size of 8x16 has been used. Generally, there are seven rules that should to be followed in order to produce a good design of the Jawi letters.

Rule 1: Line number 7 inside the matrix 8x16 must be used as a reference line for all shapes. This line which is also called the base line, must be used as a reference when designing shapes that are connected at the left and the right sides.

Rule 2: Three lines below the reference line (13-15) from the matrix 8x16 should be left empty for line spacing. This means that every shape must have a maximum width of 8 dots and a maximum height of 13 dots.

Rule 3: For ending shapes, and for isolated shapes which can be used for ending shapes; these shapes must be drawn from column 0 to any column depending on the visibility of the shapes.

Rule 4: For Isolated shapes only; these shapes can be drawn from any column 0 to 7 depending on the visibility of the shapes.

Rule 5: For middle shapes, and for beginning shapes that can be used for middle shapes; these shapes must be drawn from column 0 and end at column 7.

Rule 6: For beginning shapes only; these shapes can be drawn from any column 0 or 1 and must end at column 7.

Rule 7: For "tail letters" of SIN, SYIM, SAD and DAD; these shapes of the letter must be drawn from column 0 to column 7, and the tail shape drawn from column 0 to 6.

3.1.2 Keyboard Layout

We designed a keyboard layout for NEC-PC9801 for typewriter approach. Most of the Jawi characters are assigned to the Kana keys and shift Kana keys.

3.2 Automatic selector approach

An automatic selector has been designed to select the correct shape of the letter that depends upon the position and the preceding letter in a word. The keyboard assignment is only for the 36 letters that represent the basic shapes of Jawi letters. It is the role of the automatic selector to analyze the typed character and display the correct shape of the character.

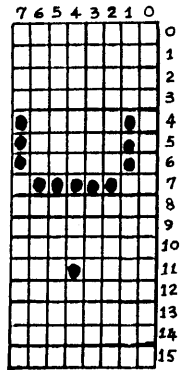## 3.2.1 Flow chart

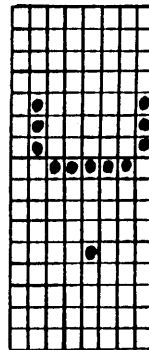The Jawi character shapes are arranged in one dimensional array using the following relation:

$$code = type + value * N$$

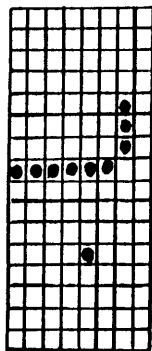where, code is character shape number which ranges from 1 to 180,

type is letter number or symbol· number which ranges from 1 to 36,
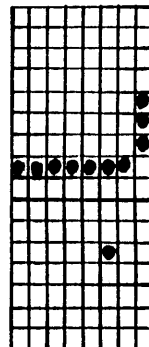
Isolated shape
of BA.

Modified shape
of BA.

Beginning shape
of BA.

Modified shape
of BA.

Fig. 3 Modification of Jawi letters for
typewriter approach.

N (=36) is the number of letters or symbols*,
and value is

                0   if the letter shape is isolated,
                1   if the letter shape is beginning,
                2   if the letter shape is middle,
                3   if the letter shape is ending,
        and   4   if the character is symbol.

   * Besides 36 letters, 36 symbols consisting of operators, delimiters,
      digits and space are used in Jawi writing.

   A flow chart to determine the value and the type of output character is
shown in Fig.4. In the chart,

                c(j)   : inputted character at position j,
                letter : any Jawi letter,
                symbol : any symbol,
                special: any special letter of ALIF, DAL, ZAL, RO, ZAI, WAU, VAU,
                         LAM_ALIF and AMZAH that cannot be connected to any
                         following letter,
                tail   : any tail letter of SIN, SYIM, SAD and DAD that needs a
                         tail shape when used in the isolated or ending shape.

   This automatic selecter is implimented by using C-programing language(4).

3.2.2 Sample text
   For instance, to display a Malay word "Terima Kasih(in Rumi)" which means
"Thank You", the selection of Jawi letters in isolated shapes are as follows:
     space"HEA""YA""SIN""ALIF""KAP"space"ALIF""MIM""YA""RO""TA"space
           x   ي   س   ١   ک   ١   م   ي   ر   ت   space

   The number of selection steps required by the automatic selector are
explained below:
step   c(j)   c(j-1)   (type,value)   code   comment
 1)   TA   space   (3,1)   39   display beginning shape
                                                of TA at (j)

 2)   RO   TA   (11,2)   83   display ending shape of
                                                RO at (j)

 3)   YA   RO   (35,1)   71   special letter of RO at
                                                (j-1), display beginning
                                                shape of YA at (j)

 4)   MIM   YA   (28,2)   100   display middle shape of
                                                MIM at (j)

 5)   ALIF   MIM   (1,2)   73   display ending shape of ALIF at (j)

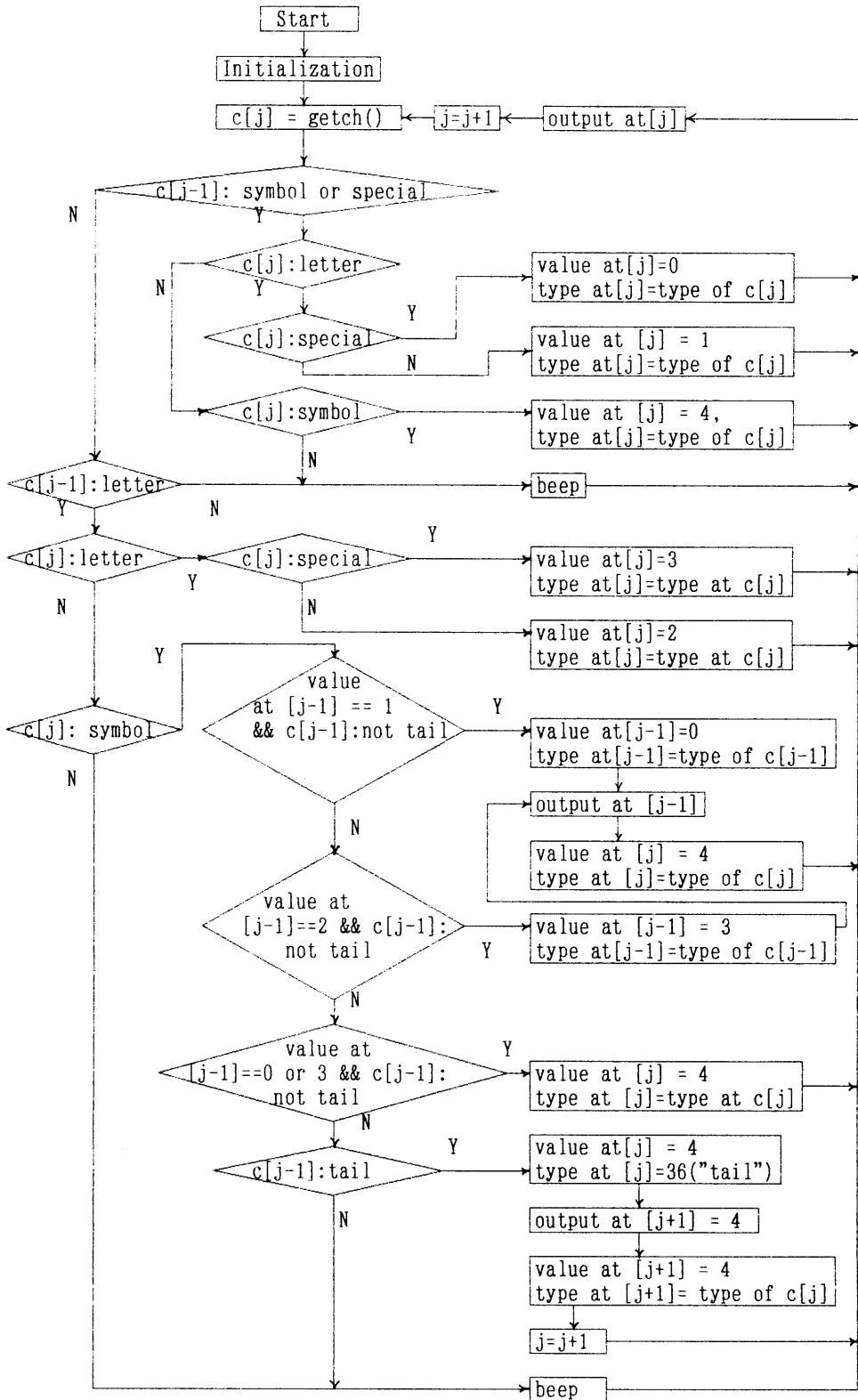 6)   space   ALIF   (35,4)   179   special letter of ALIF
                                                at (j-1), display space
                                                at (j)

Start

Initialization

c[j] = getch() ← j=j+1 ← output at[j]

c[j-1]: symbol or special

N / Y

c[j]:letter
N / Y
value at[j]=0
type at[j]=type of c[j]

c[j]:special
Y / N
value at [j] = 1
type at[j]=type of c[j]

c[j]:symbol
Y / N
value at [j] = 4,
type at[j]=type of c[j]

c[j-1]:letter
Y / N
beep

c[j]:letter
N / Y
c[j]:special
Y / N
value at[j]=3
type at[j]=type at c[j]

value at[j]=2
type at[j]=type at c[j]

c[j]: symbol
Y / N

value at [j-1] == 1 && c[j-1]:not tail
Y / N
value at[j-1]=0
type at[j-1]=type of c[j-1]

output at [j-1]

value at [j] = 4
type at [j]=type of c[j]

value at [j-1]==2 && c[j-1]: not tail
Y / N
value at [j-1] = 3
type at[j-1]=type of c[j-1]

value at [j-1]==0 or 3 && c[j-1]: not tail
Y / N
value at [j] = 4
type at [j]=type at c[j]

c[j-1]:tail
Y / N
value at[j] = 4
type at [j]=36("tail")

output at [j+1] = 4

value at [j+1] = 4
type at [j+1]= type of c[j]

j=j+1

beep

Fig. 4 Flow chart for the automatic selector approach.

〔9〕

7)	KAP	space	(25, 1)	61	display beginning shape of KAP at (j)
8)	ALIF	KAP	(1, 2)	73	display ending shape of ALIF at (j)
9)	SIN	ALIF	(13, 1)	49	display beginning shape of SIN at (j)
10)	YA	SIN	(35, 2)	107	display middle shape of YA at (j)
11)	HEA	YA	(32, 2)	104	display middle shape of HEA at (j)
12)	space	HEA	(32, 3)	140	display ending shape of HEA to replace the middle shape of HEA at (j-1)
		space	(35, 4)	179	display space at (j).

The displayed Jawi writing according to the steps above is as follows:

تريما كاسيه

4. CONCLUSION

The two approaches presented in this paper, namely the typewriter approach and the automatic selector approach, are successful in generating Jawi letters according to the rules of writing Jawi text. In the typewriter approach we could achieve to reduce the number of Jawi characters from 161 (Fig. 1 and symbols) to 78 by sophisticated font design. On the other hand, the automatic selector approach is better compared to the typewriter approach since the speed of inputting the Jawi letters is much faster, because of the fewer numbers of Jawi characters used on the keyboard. The method used in the selector approach is able to select the appropriate shape of the Jawi letters according to their position in Malay words.

The above methods of inputting the Jawi letters should be included in the formal design of a special editor so that the Jawi fonts that have been designed can be utilized effectively.

References
1. Norman Meyrowtz and Andris van Dam: "Interactive Editing System Part 1 & 2", Computer Serveys, Vol. 14, No. 3, September 1982.
2. S. Y Syed Salleh: " Jawi Letters For Computer Exhibition" , Journal Technology No. 5, pages 44-69, June 1984.
3. Dewan Bahasa & Pustaka: " List of Rumi to Jawi spelling", First edition, 1988.
4. Brian W. Kernighn - Dennis M. Ritchie: " The C Programming Language ", Prentice - Hall, Inc. 1978.