

J A W I 表記マレー語エディタの設計

米田政明*¹ 長谷博行*¹ 酒井充*¹ 光山和哉*² ローマド・ビン・ファケー*³

*¹ 富山大学工学部

*² (株)CSK

*³ マレーシアRTM

概要

マレー語には2種類の表記法がある。一つはRUMIと呼ばれ、ローマ文字が用いられる。他の一つはJAWIと呼ばれ、これにはアラビア文字とよく似た文字が用いられる。RUMIはマレーシアの日常生活や官公庁において広く用いられており、通常の英文エディタが使用できる。一方、JAWIは主に宗教(イスラム教)目的に用いられているが、使用できるエディタがあまりない。本稿では、JAWI表記のためのエディタおよびFEP(フロント・エンド・プロセッサ)を設計・制作したので報告する。

A DESIGN OF AN EDITOR FOR JAWI WRITING OF MALAY LANGUAGE

Masaaki Yoneda*¹ Hiroyuki Hase*¹ Mitsuru Sakai*¹

Kazuya Mitsuyama*² Rohmad Bin Fakeh*³

*¹ Faculty of Engineering, Toyama University
3190 Gofuku, Toyama 930, Japan

*² CSK Inc.

*³ Malaysia RTM

ABSTRACT

Malay language uses two types of writing letters. One is called Rumi and the other Jawi. Rumi letters are the same as the Roman letters, and are used widely in daily communication and in most of government offices. Ordinary English Editors are available for Rumi writing. On the other hand, Jawi letters originating from Arabic letters are used for Islamic religion. And so far very few editors are available for Jawi writing. In this report we develop an editor and a FEP (Front End Processor) for Jawi writing.

1. はじめに

マレー語は公式には「Bahasa Malaysia」と呼ばれるマレーシアにおける国語である。言語体系は発音も含めて一つであるが、表記法に2種類ある。一つはRUMIと呼ばれ、ローマ文字を用いて左から右へ表記する。RUMIは日常生活や官公庁において広く用いられており、Word Starなどの通常の英文ワープロを用いて作文することができる。他の一つはJAWIと呼ばれ、アラビア文字とよく似た文字を用いて「右から左へ」表記する。JAWIは主に宗教（イスラム教）目的に用いられている。このことはアラビア語で書かれているコーランが歴史的に他国語に翻訳することを禁じられていたことによるものと考えられる。JAWI表記の大きな特徴としては、上記の右から左へ書かれること以外に、文字の形が単語中の位置によって変わることがあげられる。アラビア語のワープロについての報告はあるが⁽¹⁾⁽²⁾、JAWI表記のマレー語ワープロに関する報告は筆者らの知る限りあまりない⁽³⁾。本稿ではJAWI表記のためのFEP（フロント・エンド・プロセッサ）とエディタを設計・制作したので報告する。

以下、2.でJAWI文字とJAWI表記について述べ、3.でJAWI文字の入力方式について述べる。4.でJAWIフォントについて述べ、5.でJAWI文字入力のためのFEP「JAWI太郎」について述べる。更に、6.でJAWI表記用マレー語エディタ「eK」について述べる。

2. JAWI文字とJAWI表記

JAWI文字は図1に示す36種類の文字からなる。これらの文字は単語中の位置によって字形が変化する。即ち、先頭に位置するときは「先頭形（Beginning shape）」、中間に位置するときは「中間形（Middle shape）」、末尾に位置するときは「末尾形（Ending shape）」となる。また、単独で存在するときは「孤立形（Isolated shape）」となる。

JAWI文字は、これらの字形が適宜使い分けられ、基線（Base line）上に左右から手を伸ばした形で連結されて右から左へ表記される。JAWI表記における基線はローマ文字を筆記体で書くときの基準線と同様のものである。なお、単語と単語の間にはスペースが置かれる。また、一つの単語が2行にまたがって表記されることはない。

JAWI表記の例を単語「bersambung（RUMI表記）」で見よう。この単語は「バーサムブング」と発音され、英語の「connected」と同義である。この単語を構成するJAWI文字（孤立形）を右から左へ並べると、

「NGA」	「WAU」	「BA」	「MIM」	「SIN」	「RO」	「BA」
غ	و	ب	م	س	ر	ب

であるが、これらの文字は

「NGA」 「WAU」 「BA」 「MIM」 「SIN」 「RO」 「BA」
 غ و ب ه س ر ب

孤立形 末尾形 中間形 中間形 先頭形 末尾形 先頭形

のように変形されて、 بر سه بوغ と表記される。

ZAI	RO	ZAL	DAL	CA	KHO	HA	JIM	THA	TA	BA	ALIF
ز	ر	ذ	د	چ	خ	ح	ج	ث	ت	ب	ا
QAF	PA	FA	NGA	GHAIN	AIN	ZO	TO	DAD	SAD	SYIM	SIN
ق	ف	ق	غ	ع	ظ	ط	ض	ص	س	ش	س
NYA	YA	AMZAH	LAM-ALIF	HEA	VAU	WAU	NUN	MIM	LAM	GA	KAP
ن	ي	ء	لا	ه	و	و	ن	م	ل	ك	ك

図1. JAWI 文字のアルファベット (----- は基線を表す)

一般に JAWI 文字は

末尾形 - 中間形 - . . . - 中間形 - 中間形 - 先頭形

のように単語中の位置による変形を受けるが、「RO」や「WAU」にはもともと中間形がない（即ち、次の文字と接続しない）ため、上例のように、中間に位置しているにも拘らずこれらは末尾形になっている。そのため、次に続く「SIN」は中間に位置しているにも拘らず先頭形となり、「NGA」は末尾に位置しているにも拘らず孤立形となっている。

36 種の JAWI 文字の字形を図2に示す。「RO」や「WAU」のように次の文字と接続しない文字は 8 種あり、(「ALIF」, 「DAL」, 「ZAL」, 「RO」, 「ZAI」, 「WAU」, 「VAU」, 「LAM_ALIF」), これらの文字は孤立形と末尾形しかもたない。この 8 種類の文字をスペシャル・レターと呼ぶ。スペシャル・レターは単語の先頭に位置するときおよびスペシャル・レターの次に位置するときは孤立形となり、それ以外では末尾形となる。スペシャル

文字名	孤立形	末尾形	中間形	先頭形	文字名	孤立形	末尾形	中間形	先頭形
ALIF	ا	ا			AIN	ع	ع	ع	ع
BA	ب	ب	ب	ب	GHAIN	غ	غ	غ	غ
TA	ت	ت	ت	ت	NGA	ڠ	ڠ	ڠ	ڠ
THA	ث	ث	ث	ث	FA	ف	ف	ف	ف
JIM	ج	ج	ج	ج	PA	ڤ	ڤ	ڤ	ڤ
HA	ح	ح	ح	ح	QAF	ق	ق	ق	ق
KHO	خ	خ	خ	خ	KAP	ك	ك	ك	ك
CA	چ	چ	چ	چ	GA	گ	گ	گ	گ
DAL	د	د			LAM	ل	ل	ل	ل
ZAL	ذ	ذ			MIM	م	م	م	م
RO	ر	ر			NUN	ن	ن	ن	ن
ZAI	ز	ز			WAU	و	و		
SIN	س	س	س	س	VAU	و	و		
SYIM	ش	ش	ش	ش	HEA	ه	ه	ه	ه
SAD	ص	ص	ص	ص	LAM-ALIF	لا	لا		
DAD	ض	ض	ض	ض	AMZAH	ز			
TO	ط	ط	ط	ط	YA	ي	ي	ي	ي
ZO	ظ	ظ	ظ	ظ	NYA	ن	ن	ن	ن

図2. JAWI 文字の字形 (---- は基線を表す)

・レターの次の文字は中間形や末尾形にはならず、先頭形か孤立形となる。

孤立形しかもたない「AMZAH」とスペシャル・レター以外の 27 種類の文字はそれぞれ先頭形、中間形、末尾形、孤立形の字形をもっている。

3. JAWI 文字の入力方式

JAWI 文字をコンピュータに入力する方式として、タイプライター方式と自動選択方式が考えられる。前者は機械式タイプライターに慣れている人にとって便利であり、後者は一般の利用者にとって便利であろう。

3. 1 タイプライター方式

この入力方式は全ての字形をキーに割り当てて入力するというものである。ただし、図 2 に示した字形だけでも 125 あり、更に数字 10 種類（ \backslash 、 γ 、 γ 、 ξ 、 ϕ 、 γ 、 γ 、 λ 、 η 、 \cdot ）やいくつかの記号（ $\ddot{\cdot}$ 、 $\ddot{\cdot}$ 、 $\ddot{\cdot}$ など）（数字や記号は字形変化を起こさない）を合わせると大変な数になるので、フォントを工夫することによりキーに割り当てるべき総数を圧縮した。例えば、「BA」の孤立形を末尾形としても使用できるようにし、また先頭形を中間形としても使用できるようにして本来の 4 字形のうち孤立形と先頭形の 2 字形のみをキーに割り当てた。

3. 2 自動選択方式

この入力方式はキーへの割り当てを文字種の方だけにして、状況に応じた字形が自動的に選択されるようにするものである。図 3 に字形選択アルゴリズムのフローチャートを示す。図 3 において、 $c[j]$ は位置 j に入力される文字・記号を表し、「文字」は「スペシャル」（スペシャル・レターと「AMZAH」）とその他の JAWI 文字からなり、「記号」は数字を表す記号やスペースなどからなる。便宜上、「記号」（字形変化を起こさない）の字形は孤立形とする。また、 $s[j]$ は位置 j における文字・記号の字形を表す。

4. JAWI フォント

JAWI 文字のフォントは漢字 ROM の外字領域を利用し、半角外字フォントとしてテキスト画面に出力される。後述する FEP「JAWI 太郎」、エディタ「eK」を使用する前に、外字登録をする必要がある（DOS コマンドラインから `A > jawifont [RET]` を実行する）。

5. フロント・エンド・プロセッサ 「JAWI 太郎」

JAWI 太郎は PC-9801 で JAWI 文字の入力を可能にするための TSR（常駐終了）型プロ

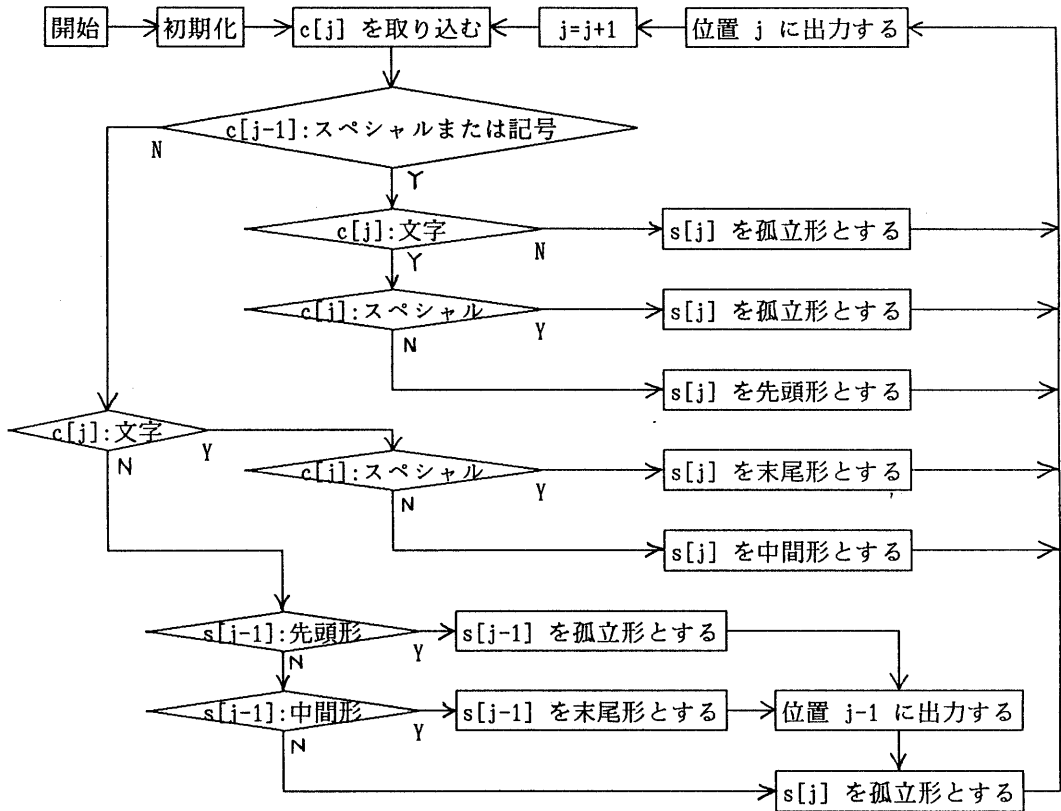


図3. 字形選択アルゴリズム

グラムである (DOS コマンドラインから A > jawitaro [RET] を実行する。以後, CTRL +GRPH で「JAWI 太郎」が ON, OFF される)。JAWI 太郎はタイプライター方式と自動選択方式の二つの入力方式をサポートし, 再変換機能をもっている。

再変換機能は自動選択方式による入力時に利用可能で, カーソル位置にある JAWI 文字の字形が不適切であればそれを適切なものに直す機能である。自動選択方式では, 図3の字形選択アルゴリズムに従って, 順次入力される JAWI 文字の適切な字形を順次選択して画面出力していく。そのため, 入力が終了した単語に後から文字を挿入したり削除したりすると, つなぎ目の部分で不適切な字形となってしまうことがある。再変換機能はこのような場合に有効である。

6. JAWI 表記用マレー語エディタ「eK」

eK (DOS コマンドラインから A > ek ファイル名 [-オプション] [RET] を実行することによって起動される) は通常のエディタのもつ機能の他に, 次のような特殊機能をもっている。

(1) Right/Left の二つの編集モードをもつ。

eK は JAWI 文字の編集を行う Right モードと英数字の編集を行う Left モードの二つの編集モードをもっている。Right モードから Left モードへ切り替えると FEP「JAWI 太郎」は自動的に OFF になり、逆の切り替えにより自動的に ON になる。

(2) 複数行にわたっての単語の記述を妨げる。

JAWI 表記では一つの単語を 2 行にまたがって記述することは許されない。よって、eK は 1 行の入力が最大桁に達すると改行を促すメッセージを出力し、それ以上の入力を妨げる。

(3) 1 行再変換機能をもつ。

Right モードで JAWI 文書を編集しているとき、入力し終えた単語に文字を挿入・削除したり、[BS] や [DEL] で二つの行が一つになったときなど、つなぎ目の部分で不適切な字形になってしまうことがある。この解決策として、前述のように FEP に再変換機能をつけたが、エディタにもカーソル行再変換機能をつけた。

(4) 挿入入力機能をもつ。

編集中に Left モードで JAWI 文字を入力、あるいは Right モードで英数字を入力した

	0~7	8	9	A	B	C	D	E	F
0	00h 7fh は 変 更 さ れ な い	ھ	ء	ب	ل	د	+	ش	ء
1		ف	۷	ب	ا	ت	ع	ش	ء
2		ق	۳	ت	غ	ي	ح	۵	d
3		ق	ا	ت	ل	ا	ع	۷	د
4		ك	ء	ق	ج	ء	ر	~	م
5		ك	*	ح	ج	ع	م	~	ء
6		ه	=	خ	د	ق	ن	ب	'
7		ه	.	ع	ز	ت	ع	°	"
8		ف	۸	غ	ر	ج	ع	ش	
9		ي	۹	ف	ز	ك	ب	س	
A		ة	.	ق	س	ع	د		
B		ة	.	ج	ش	ق	ه	"	
C		ي	:	ل	ص	ك	خ	'	
D		ك	?	م	ف	ع	ه	ء	
E		ك	۴	ن	ط	ن	ب	ش	
F		۲	\$	ي	ظ	ن	ت	ش	

(a) タイプライター方式

	0~7	8	9	A	B	C	D	E	F
0	00h 7fh は 変 更 さ れ な い		ء		\	ط	ش	+	ش
1			۷	.	۵]	ع	ش
2			۳		ح	ك	?	۵	d
3				*	ل	ا	س	۷	د
4			ء	ء	۶	ع	ف	~	م
5			*	.	ت	غ	ي	~	ء
6			=	!	د	غ	ن	ب	'
7			.	۲	خ	ل	ص	°	"
8			۸		ج	<	ز	ش	
9			۹	\$:	ر	>	س	
A			.	%	ب	ح	:	,	
B			.	&	ق	ف	ب	"	
C				۴	ت	و	م	'	
D				(ظ	=	ك	//	
E)	ض	-	"	س	
F			۲		♥	ت	ز	[ش

(b) 自動選択方式

図4. アスキーコード表 (空白は何も登録されていないコードである)

い場合がある。そのような場合、この挿入入力機能を ON にして文字を入力すると、カーソルは動かず文字だけが次々に挿入されていく。これにより、文字列を逆順に入力する不便を解消した。ただし、挿入入力後のカーソル位置はユーザが変更しなければならない。

なお、eK で扱うファイルはテキスト型ファイルであり、文字や記号はファイル内では図 4 に示したアスキーコードで表されている。タイプライター方式で入力・編集されたファイルでは一つの字形に一つのコードが対応し、自動選択方式で入力・編集されたファイルでは一つの文字種に一つのコードが対応している。

7. おわりに

JAWI 文字を入力するためのフロント・エンド・プロセッサ「JAWI 太郎」および JAWI 表記マレー語エディタ「eK」を設計し、PC-9801 (MS-DOS) 上で試作した（使用言語は Turbo C++⁽⁴⁾ および Turbo Assembler⁽⁵⁾ である）。JAWI 表記マレー語は（1）右から左へ書かれる、（2）字形が単語中の位置などにより変わるなどの大きな特徴をもっており、JAWI 太郎および eK はこれらに対処するための機能をもっている。ただし、JAWI 太郎、eK ともまだまだ完成とは言えず、（1）フォントデザインの工夫、（2）キー割り当ての最適化、（3）テキストファイル内での（入力方式によらない）文字コードの統一、（4）検索機能の充実などを図る必要がある。更に、アラビア文字系の文字を用いて表記する言語はマレー語以外にもベルシャ湾岸、北アフリカ、地中海域、東南アジアなどに多数あるので、これらに広く対応できる FEP およびエディタを IBM-PC や UNIX マシンなど世界で広く使用されている機種で作成することも意義深いことと考えられる。

ترىما كاسيه (「Terima Kasih」 in Rumi which means 「Thank you」.)

参考文献

- (1) Norman Meyrowitz and Andries Van Dam : "Interactive Editing System Part 1 & 2", Computing Surveys, Vol.14, No.3, September 1982.
- (2) K.Yajima and M.Hirashima : "Consideration for the Development of Arabic Word Processor", Proc. ICTP, pp.127-130, 1983.
- (3) S.Y.Syed Salleh : "Jawi Letters for Computer Exhibition", Journal Technology, No. 5, pp.44-69.
- (4) "Turbo C++ 全マニュアル", ボーランドジャパン.
- (5) "Turbo Assembler 全マニュアル", ボーランドジャパン.