

Report for Malaysia's Internationalized Domain Name: Jawi Language Issues



Report for Malaysia's Internationalized Domain Name: Jawi Language Issues

Abstract

This report discusses the development of table for Jawi characters currently used in Malaysia.

Seven issues have been discussed and resolved in this report. Some of the issues are highlighted by the linguists during the Internationalized Domain Name Forum held on 20 August 2008 at Sheraton Hotel Subang.

The MYNIC's language committee for IDN had gathered on 20 January 2009 at MYNIC Board Room to discuss the Jawi character table. An important issue is the introduction of a new "Jawi Letter Hamzah Three Quarter". This requires submission to the Unicode table 5.0 which will be described in this report.

Table of Contents

<u>Content</u>	<u>Page</u>
1. Background	1
2. Language Committee	2
3. Discussion Issue 1	3
- Rules of spelling	
4. Discussion Issue 2	4
- Hindi Numerals and Arabic Numerals	
5. Discussion Issue 3	5
- Jawi Letter Hamzah Three Quarter	
6. Discussion Issue 4	6 -7
- Jawi character in initial medial and final locations	
7. Discussion Issue 5	8 - 10
- Table of Arab characters , numbers and joining characters used in Jawi	
8. Discussion Issue 6	11 - 13
- Unicode table for Jawi characters and Extended Arabic Letters	
9. Discussion Issue 7	14 - 15
- Variant character , Variant Unicode Table and Extended Arabic Letters	
10. References	16

Background

MYNIC is an agency under the Ministry of Science, Technology and Innovation and is regulated by the Malaysian Communications and Multimedia Commission.

MYNIC administers the name space for the .my Top Level Domain (TLD), and has been doing so since 1987. This involves the registration of domain names and the maintenance and operation of a domain name registry (a central database for .my domain names).

Domain names are a very crucial part of using Internet technology. They use Roman or ASCII characters regardless of the worldwide spread of the internet. Other language scripts are not yet fully supported to locate resources and sites on the network. Nevertheless, using Jawi, Mandarin and Tamil scripts domain names is essential in increasing the Internet penetration in Malaysia. Supporting these three characters calls for investigating and addressing a number of language and technical issues.

In light of this, MYNIC organised the Forum on Challenges and Benefits in Introducing Internationalized Domain Name (IDN) in Malaysia. This was held on 20th August 2008 at the Sheraton Subang Hotel with the following objectives:

- Create public awareness on importance of IDN in Malaysia
- To discuss multilingual (Jawi, Chinese, Tamil) domain name implementation issues in Malaysia
- To discuss standards and technical implementation issues relating to IDN
- To create a working group in developing IDN for Malaysia

Language Committee

One of the outcomes of the IDN forum led by MYNIC Berhad is the creation of Jawi language committee comprising the following committee members:

Jawi Language Committee:

1. Tn. Hj Wan Mohd Saophy Amizul Bin Wan Mansor (DBP)
2. Tn. Hj Hamdan Bin Hj Abd Rahman (Author DBP)
3. Dr. Mohd Azidan Bin Abdul Jabar (UPM)
4. Azmah Abd Malik (MYNIC Berhad)
5. Noor Harnidda Khalid (MYNIC Berhad)
6. Madihah Md Jalal Fahmey (MYNIC Berhad)

The language committee had a follow up meeting for the Jawi language issues to discuss.

Discussion Issue 1:

Issues	Spelling in Jawi	Status	Comment
Confusion spelling of "Malaysia" in Jawi and Arabic	مليسيا	Accepted	Spelling based on pronunciation in Bahasa Melayu
Rules of spelling acronym words:			
a) If the acronym words spell such as:			a) Spelling based on pronunciation in Bahasa Melayu
i) MYNIC	ماينيك	Accepted	
ii) MOSTI	موستي	Accepted	
b) If the acronym words can't be pronounced according to syllable such as:			b)
i) The acronym word derived from Malay word eg: SKMM	س-ك-م-م	Accepted	i) Meaning of the acronym word is in Malay, the word must be spelt according to a sequence of character.
ii) The acronym word is derived from English word eg: MCMC	ع-ي-م-س-ي-ع-ي-م-س-ي	Accepted	ii) Meaning of the acronym word is in English, the word must be spelt according to the pronunciation of each character in the word.
c) If the acronym words have dots '.' as separator e.g: K.S.U	ك.س.او Replaced with ك-س-او	Accepted	c) For case of a domain name the dot '.' must replace with the hyphen '-'
Rules of the Jawi number			
a) If a domain name comprise number two and it is a reduplication word e.g: 'www.rama-rama.my'	www.راما۲.my	Accepted	a) Reduplication word
b) If a domain name comprise number two but it is not the reduplication word e.g: 'www.rama22.my'	www.راما۲۲.my	Not Accepted	b) Not a reduplication word
c) If a domain name comprise any numbers except number two e.g 'www.999.my'	www.۹۹۹.my	Not Accepted	c) Not follow the number rules in Jawi

Discussion Issue 2: Arabic Numerals and Hindi Numerals

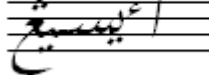
- 1) Allow all numbers in Arabic Numerals and
- 2) Allow only number 2 (٢) in Hindi numerals

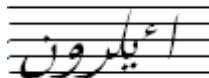
Arabic Numerals	Decision Y=Allowed N=Not Allowed	Rationale	Hindi Numerals	Decision Y=Allowed N=Not Allowed	Rationale
0	Y	Rules of Jawi the number must in Arabic Numerals.	٠	N	
1	Y	Rules of Jawi the number must in Arabic Numerals.	١	N	
2	Y	Rules of Jawi the number must in Arabic Numerals.	٢	Y	Reserve this number for any reduplication word. Must combine word but not in numbering to avoid confusion. This has been the traditio. Refer to old book <i>Daftar Kata Za'ba</i>
3	Y	Rules of Jawi the number must in Arabic Numerals	٣	N	
4	Y	Rules of Jawi the number must in Arabic Numerals.	٤	N	
5	Y	Rules of Jawi the number must in Arabic Numerals.	٥	N	
6	Y	Rules of Jawi the number must in Arabic Numerals.	٦	N	
7	Y	Rules of Jawi the number must in Arabic Numerals.	٧	N	
8	Y	Rules of Jawi the number must in Arabic Numerals.	٨	N	
9	Y	Rules of Jawi the number must in Arabic Numerals.	٩	N	

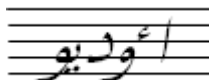
Discussion Issue 3: New “Jawi Letter Hamzah Three Quarter”

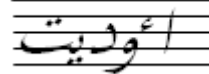
According to Dewan Bahasa and Pustaka on *Jawi Letter Hamzah Three Quarter High* is used in our daily Jawi. However this character is not in the Unicode table.

Here are some words *Jawi Letter Hamzah Three Quarter* in our daily Jawi writing:

a. *Aising* spelling in Jawi is: 

b. *Aileron* spelling in Jawi is: 

c. *Audio* spelling in Jawi is: 

d. *Audit* spelling in Jawi is: 

(Reference: *Daftar Kata Bahasa Melayu Rumi- Sebutan –Jawi Jilid 1* pg 83)

The characters displayed above are using the superscript of Arabic letter Hamzah (U+0621) character. This character is not in the Unicode Table 5.0. The linguist came up with the decision to propose the inclusion of **Jawi Letter Hamzah Three Quarter** into the Unicode table.

Discussion Issue 4: Discussion and confirmation of Jawi character in initial medial and final location

Below is the Jawi character table in initial medial and final locations that have been reviewed and agreed.

#	Character	Isolated	Initial	Medial	Final	In Unicode	Name
1	ا	ا	ا	ا	ا	0627	alif
2	ب	ب	ب	ب	ب	0628	ba
3	ت	ت	ت	ت	ت	062A	ta
4	ة	ة	-	-	ة	0629	ta marbutah
5	ث	ث	ث	ث	ث	062B	sa [tha]
6	ج	ج	ج	ج	ج	062C	jim
7	چ	چ	چ	چ	چ	0686	ca
8	ح	ح	ح	ح	ح	062D	ha [ħa]
9	خ	خ	خ	خ	خ	062E	kha [khO]
10	د	د	د	د	د	062F	dal [dhal]
11	ذ	ذ	ذ	ذ	ذ	0630	dzal
12	ر	ر	ر	ر	ر	0631	ra [rO]
13	ز	ز	ز	ز	ز	0632	zai
14	س	س	س	س	س	0633	sin
15	ش	ش	ش	ش	ش	0634	syin
16	ص	ص	ص	ص	ص	0635	sad [şOd]
17	ض	ض	ض	ض	ض	0636	dad [dOd]
18	ط	ط	ط	ط	ط	0637	to [tO]
19	ظ	ظ	ظ	ظ	ظ	0638	zo [zO]
20	ع	ع	ع	ع	ع	0639	ain
21	غ	غ	غ	غ	غ	063A	ghain
22	ڠ	ڠ	ڠ	ڠ	ڠ	06A0	nga
23	ف	ف	ف	ف	ف	0641	fa
24	ڤ	ڤ	ڤ	ڤ	ڤ	06A4	pa
25	ق	ق	ق	ق	ق	0642	qOf

(continued next page...)

(Continued from previous page)

#	Character	Isolated	Initial	Medial	Final	In Unicode	Name
26	ک	ک	ک	ک	ک	06A9	kaf
27	گ	گ	گ	گ	گ	0762	ga
28	ل	ل	ل	ل	ل	0644	lam
29	م	م	م	م	م	0645	mim
30	ن	ن	ن	ن	ن	0646	nun
31	و	و	و	و	و	0648	wau
32	ف	ف	ف	ف	ف	06CF	va
33	ه	ه	ه	ه	ه	0647	ha
34	ء	ء	ء	ء / ؤ	ء	0621	hamzah
35	ي	ي	ي	ي	ي	064A	ya
36	ى	ى	-	-	ى	0649	ye [ye]
37	ن	ن	پ	پ	ن	06BD	nya

Source: DBP (*Daftar Kata Bahasa Melayu-Rumi- Sebutan- Jawi - Jilid1* -Pg 71-72)

Discussion Issue 5: Jawi characters taken from Arabic Unicode Table

Table for Jawi characters, Arab characters, numbering and joining character used in Jawi.

After review the linguist have agreed with the table below:

Malaysia's Jawi Characters

#	Char Jawi- Arab	Pronunciation in Malay	Char Info	Hexa decimal Unicode	Variant Char	Puny code	Decimal HTML Code
1	ا	Alif	ARABIC LETTER ALEF	0627		xn--mgb	ا
2	ب	Ba	ARABIC LETTER BEH	0628		xn--ngb	ب
3	ت	Ta	ARABIC LETTER TEH	062A		xn--pgb	ت
4	ة	ta marbutah	ARABIC LETTER TEH MARBUTA	0629	Variant is 06C3	xn--ogb	ة
5	ث	sa [tha]	ARABIC LETTER THEH	062B		xn--qgb	ث
6	ج	Jim	ARABIC LETTER JEEM	062C		xn--rgb	ج
7	چ	Ca	ARABIC LETTER TCHEH	0686		xn--cjb	چ
8	ح	ha [ha]	ARABIC LETTER HAH	062D		xn--sgb	ح
9	خ	kha [khO]	ARABIC LETTER KHAH	062E		xn--tgb	خ
10	د	dal [dhal]	ARABIC LETTER DAL	062F		xn--ugb	د
11	ذ	dzal	ARABIC LETTER THAL	0630		xn--vgb	ذ
12	ر	ra [rO]	ARABIC LETTER REH	0631		xn--wgb	ر
13	ز	Zai	ARABIC LETTER ZAIN	0632		xn--xgb	ز
14	س	Sin	ARABIC LETTER SEEN	0633		xn--ygb	س
15	ش	Syin	ARABIC LETTER SHEEN	0634		xn--zgb	ش
16	ص	sad [sOd]	ARABIC LETTER SAD	0635		xn--0gb	ص
17	ض	dad [dOd]	ARABIC LETTER DAD	0636		xn--1gb	ض
18	ط	to [tO]	ARABIC LETTER TAH	0637		xn--2gb	ط
19	ظ	zo [zO]	ARABIC LETTER ZAH	0638		xn--3gb	ظ

(Continued next page...)

(Continued from previous page)

#	Char Jawi- Arab	Pronunciation in Malay	Char Info	Hexa decimal Unicode	Variant Char	Punycode	Decimal HTML Code
20	ع	ain	ARABIC LETTER AIN	0639		xn--4gb	ع
21	غ	Ghain	ARABIC LETTER GHAIN	063A		xn--5gb	غ
22	غ	Nga	ARABIC LETTER AIN WITH THREE DOTS ABOVE	06A0		xn--2jb	ڠ
23	ف	Fa	ARABIC LETTER FEH	0641		xn--dhh	ف
24	ف	Pa	ARABIC LETTER VEH	06A4		xn--6jb	ڤ
25	ق	qOf	ARABIC LETTER QAF	0642		xn--ehb	ق
26	ك	Kaf	ARABIC LETTER KEHEH	06A9	Variant is 0643	xn--ckb	ک
27	ك	Ga	ARABIC LETTER KEHEH WITH DOT ABOVE	0762	Variant is 06AC	xn--mpb	ݢ
28	ل	Lam	ARABIC LETTER LAM	0644		xn--ghb	ل
29	م	Mim	ARABIC LETTER MEEM	0645		xn--hhb	م
30	ن	Nun	ARABIC LETTER NOON	0646		xn--ihb	ن
31	و	Wau	ARABIC LETTER WAW	0648		xn--khh	و
32	و	va	ARABIC LETTER WAW WITH DOT ABOVE	06CF		xn--flb	ۏ
33	ه	ha	ARABIC LETTER HEH	0647	Variant are 06BE, 06C1 and 06D5	xn--jhb	ه
34	ء	hamzah	ARABIC LETTER HAMZA	0621		xn--ggb	ء
35	ي	ya	ARABIC LETTER YEH	064A		xn--mhb	ي
36	ى	ye [ye]	ARABIC LETTER ALEF MAKSURA	0649	Variant is 06CC	xn--lhb	ى
37	ن	nya	ARABIC LETTER NOON WITH THREE DOTS ABOVE	06BD		xn--wkb	ڽ

(Continued next page...)

(Continued from previous page)

Arabic Characters Used for Malay Words Written in Jawi

#	Char Jawi-Arab	Pronunciation in Malay	Char Info	Hexa decimal Unicode	Variant Char	Punycode	Decimal HTML Code
1	ا	alif with hamzah above	ARABIC LETTER ALIF WITH HAMZAH ABOVE	0623	Variant is 0672	xn--igb	أ
2	ا	alif with hamzah below	ARABIC LETTER ALIF WITH HAMZAH BELOW	0625	Variant is 0673	xn--kgb	إ
3	ي	ye with hamzah above	ARABIC LETTER YEH WITH HAMZAH ABOVE	0626		xn--lgb	ئ

Jawi Number

#	Char Jawi-Arab	Pronunciation in Malay	Char Info	Hexa decimal Unicode	Variant Char	Punycode	Decimal HTML Code
1	٢	number two	ARABIC INDIC DIGIT TWO	0662	06F2	xn--bib	٢

Joining Character

#	Char Jawi-Arab	Pronunciation in Malay	Char Info	Hexa decimal Unicode	Variant Char	Punycode	Decimal HTML Code
1	ي	lam alif	ARABIC LIGATURE LAM WITH ALEF ISOLATED FORM	0644+ 0627		xn--mq7c	ﻻ

Note:

Char : stands for "character"

Source : *Daftar Kata Bahasa Melayu Edisi 1 (DBP)*

Discussion Issue 6: Unicode table for Jawi characters and Extended Arabic Letters

Below is the Unicode table for Jawi characters and Extended Arabic Letters that have been reviewed and agreed.

Unicode table for Arabic characters.

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	ا	آ		ذ	-	و	٠	١	پ	٣	ع	گ	ه	ي	٠	٠
1	س	٢	ء	ر	ف	٠	١	أ	خ	٣	ف	گ	ه	ي	٠	١
2	م	٢	آ	ز	ق	٠	٢	أ	خ	ز	ب	گ	ه	ع	٠	٢
3	ط	٣	أ	س	ك	٠	٣	إ	ج	ر	ف	ب	ه	ع	٣	٣
4		٤	و	ش	ل	٠	٤	٠	ج	ر	ث	گ	و	-	٠	٤
5		٥	إ	ص	م	٠	٥	أ	خ	ر	پ	ل	و	ه	و	٥
6	٦	٦	ئ	ض	ن	٠	٦	ؤ	چ	ن	ث	ل	و	٠	٦	٦
7	٧	٧	ا	ط	ه	٠	٧	ؤ	چ	ز	ف	ل	و	٠	٧	٧
8	٨	٨	ب	ظ	و	٠	٨	ئ	٣	ز	ق	ل	و	٠	٨	٨
9	٩	٩	ة	ع	ی	٠	٩	٣	د	ز	ک	ن	و	٠	٩	٩
A	١٠	١٠	ت	غ	ي	٠	١٠	٣	د	پ	ک	ن	و	٠	١٠	١٠
B	١١	١١	ث	ک	٠	٠	١١	ب	٣	پ	ک	ن	و	٠	٠	١١
C	١٢	١٢	ج	ک	٠	٠	١٢	ب	٣	پ	ک	ن	و	٠	٠	١٢
D	١٣	١٣	ح	م	٠	٠	١٣	ب	٣	پ	ک	ن	و	٠	٠	١٣
E	١٤	١٤	خ	م	٠	٠	١٤	ب	٣	پ	ک	ن	و	٠	٠	١٤
F	١٥	١٥	د	ث	٠	٠	١٥	ب	٣	پ	ک	ن	و	٠	٠	١٥

- Jawi character
- Used with condition
- Six additional Arabic characters used in Jawi
- No assigned character

(Continued next page...)

(Continued from previous page)

Unicode extended Arabic letters

	075	076	077
0	پ	ف	ش
1	ٹ	پ	ژ
2	پ	خ	خ
3	ت	ک	ا
4	ن	ک	ا
5	ب	م	ی
6	ل	م	ی
7	خ	ن	ی
8	چ	ن	و
9	ط	ن	و
A	د	ل	ے
B	ر	ز	ے
C	ش	ر	ح
D	غ	ش	س
E	غ	ح	ش
F	غ	چ	ک



Jawi character



Used with condition



Six additional Arabic characters used in Jawi



No assigned character

(Continued next page...)

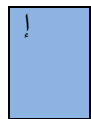
(Continued from previous page)

Notes about the table:



Character Alif with Hamzah above, is used in the following conditions:

- 1) for words that start with “di-“, for eg: دأسه = د + اسه (asah + di = diasah) from DKBM1 pg117
- 2) for words that start with “se-“, for eg: سأكام = س + اكام (agama + se = seagama) from DKBM1 pg 117
- 3) for words that start with “ke“, for eg: كإنصافن = ك + انصاف (insaf + ke = keinsafan) from DKBM1 page 118



Character Alif with Hamzah below used in condition:

- 1) Arabic words used in Malay such as the word وإما (Waimma) from DKBM2 pg 1551



Character Ye with Hamzah above used in condition:

- 1) Arabic words used in Malay such as the word ملائكة (Malaikat) from DKBM2 pg 947



Six additional Arabic Unicode characters used in Jawi:

چ (0686), غ (06A0), ف (06A4), ن (06BD), ز (06CF), ك (06CF)

Discussion Issue 7: Variant Characters and Variant Unicode Table For Jawi and Extended Arabic Letters

a) What do you know and understand about variant characters?

Different shapes of a character that have same function and sound.

Variant character is such as ك ك ك (kaf) or كى كى (ga)

b) Do variant characters need to be included in our virtual keyboard?

No. It refers to a same character and it will cause redundancy. A variant character can be declared as an alternative character or reserved character in system database.

(Continued next page...)

(Continued from previous page)

Variant Unicode Table For Jawi

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	ا	آ		ذ	-	و	و	و	پ	ڈ	غ	گ	ه	ی	و	و
1	س	و	ء	ر	ف	و	۱	آ	خ	ڑ	ف	گ	ه	ی	و	۱
2	م	و	آ	ز	ق	و	۲	أ	خ	ز	ب	گ	ه	ے	و	۲
3	ط	و	أ	س	ك	و	۳	إ	ج	ر	ف	گ	ة	ے	و	۳
4		و	و	ش	ل	و	۴		چ	ر	ف	گ	و	-	و	۴
5		و	إ	ص	م	و	۵	أ	خ	ر	پ	ل	و	ه	و	۵
6	ڤ	و	ی	ض	ن	و	۶	و	چ	ر	ف	ل	و	و	ے	۶
7	ڤ	و	ا	ط	ه	و	۷	و	چ	ز	ف	ل	و	و	و	۷
8	و	و	ب	ظ	و	و	۸	نی	ڈ	ڑ	ق	ل	و	و	و	۸
9	٪	و	ة	ع	ی	و	۹	ٹ	د	ڑ	ك	ن	و	و	و	۹
A	٪	و	ت	غ	ی	و	٪	ٹ	د	بن	ك	ن	و	و	و	بن
B	ف	؛	ث	ک	و	و	,	ب	ط	پیں	ک	ٹ	و	و	و	ض
C	،		ج	ک	و	و	،	ت	ن	پیں	ك	ن	ی	و	و	غ
D	،		ح	تی	و	و	*	ت	د	ص	ك	ٹ	ی	و	و	ح
E	م	ی	خ	تی	و	و	ل	پ	ٹ	ض	ك	ه	ی	و	و	م
F	ع	؟	د	تی	و		و	ٹ	ڈ	ظ	گ	چ	و	و	ز	ه

Extended Arabic Letters

	075	076	077
0	ا	آ	ش
1	ا	آ	ڑ
2	ا	خ	خ
3	ا	خ	ا
4	ا	ک	ا
5	ا	ن	ی
6	ا	م	ی
7	خ	ن	ی
8	چ	ٹ	و
9	ط	ن	و
A	ا	ل	ے
B	ر	ز	ے
C	ٹ	ز	ج
D	غ	ش	س
E	م	ج	ش
F	ع	چ	ك

Variant for character alif with hamzah above

Variant for character alif with hamzah below

Variant for character ta marbutah

Variant for character kaf

Variant for character hah

Variant for character ye

Variant for character ga

Variant for number two

References:

- 1) *Daftar Kata Bahasa Melayu Rumi-Sebutan-Jawi Jilid1 (A-K)- Dewan Bahasa dan Pustaka 2001*
- 2) *Daftar Kata Bahasa Melayu Rumi-Sebutan-Jawi Jilid2 (L-Z)- Dewan Bahasa dan Pustaka 2001*
- 3) The Unicode 5.0 Standard – Addison Wesley
- 4) The Unicode 5.1 Standard - <http://www.unicode.org/charts/PDF/U0600.pdf>
- 5) http://www.verisign.com/information-services/naming-services/internationalized-domain-names/idn-standards/idn-character-variants/page_001482.html
- 6) http://www.unicode.org/faq/basic_q.html
- 7) <http://www.unicodemap.org/>
- 8) <http://idn.icann.org/>
- 9) http://ms.wikipedia.org/wiki/Tulisan_Jawi#Rujukan

Confirmed By:

1) Tn Hj Hamdan Bin Hj Abdul Rahman

(Author Panduan Menulis dan Mengeja Jawi)

Dewan Bahasa Dan Pustaka

2) Tn Hj Wan Mohd Saophy Amizul Bin Wan Mansor

(Executive Language Plan)

Dewan Bahasa Dan Pustaka

3) Dr. Mohd Azidan Bin Abdul Jabar

(Head of Foreign Language Department and Lecturer)

Universiti Putra Malaysia

Prepared By:

Noor Harnidda Khalid

System Engineer

MYNIC Berhad

Verified By:

Azmah Abd Malik

Manager

MYNIC Berhad