

Arabic Characters: Languages of SE Asia

Rinalia Abdul Rahim
Task Force on Arabic IDNs
March 2014

Arabic Script in the Region

- Jawi is the (set of) Arabic alphabet for writing Malay, Acehnese, Banjarese, Minangkabau and several other languages (un-named) in South East Asia (Wikipedia)
- Tausūg used a version (subset) of the Malaysian Jawi script (Omniglot)
- Western Cham used the Arabic script (Omniglot); Both Eastern Cham and Western Cham used the Arabic script (Ethnologue)

Cross-Border Languages

- Malay - Austronesian language spoken in Malaysia, Indonesia, Singapore, Brunei and Thailand. Speakers: About 18 million + 170 million people who speak Indonesian, which is a form of Malay
- Minangkabau – Malayic language with about 8.5 million speakers in West and North Sumatra, Riau, Jambi, Bengkulu and Aceh in Indonesia, and in Negeri Sembilan in Malaysia.
- Acehnese – Chamic branch of Malayo-Polynesian languages. Spoken in Aceh and Sumatra (Indonesia) by about 3.5 million people. Closely related to other Chamic language such as Jarai and Cham, and also to Minangkabau, Gayo and Batak
- Banjarese – Austronesian language native to Indonesia and Malaysia spoken by 3.5 million speakers in South Kalimantan, Indonesia.
- Tausug – About 1 million speakers in parts of the Philippines, Malaysia and Indonesia, particularly on the Sulu Archipelago, the Zamboanga Peninsula, Southern Palawan (Philippines), in Sabah (Malaysia) and Kalimantan (Indonesia).
- Cham - Austronesian language spoken by about 230,000 people in Cambodia and Vietnam. Two main varieties of Cham - Western Cham, spoken in Cambodia, and Eastern Cham, spoken in Vietnam.

Arabic Alphabet for Malay/Jawi

(Omniglot)

ز	ر	ذ	د	خ	ح	چ	ج	ث	ت	ب	ا
زاي	راء	ذال	دال	خاء	حاء	چاء	جيم	ثاء	تاء	باء	ألف
zal	ra	zal	dal	kha	ha	ca	jim	tha	ta	ba	alif
z	r	z	d	kh	h	c	j	s	t	b	a/e
[z/ɖʒ]	[r]	[dz]	[d]	[x/k]	[h]	[tʃ]	[ɖʒ]	[s]	[t]	[b]	[Ø]
س	ش	ص	ض	ط	ظ	ع	غ	غ	ف	ق	ق
سين	شين	صاد	ضاد	طاء	ظاء	عين	غين	غاء	فاء	قاف	قاف
sin	syin	sad	dad	ta	za	ain	ghain	nga	fa	qaf	qaf
s	sy	s	d	t	z	ʿ	gh	ng	f	k/q	k/q
[s]	[ʃ/s]	[s]	[z/ɖʒ]	[t]	[z]	[Ø]	[y/r]	[ŋ]	[f/p]	[k]	[k]
ك	گ	ل	م	ن	و	ف	ه	لا	ء	ي	ث
كاف	گاء	لام	ميم	نون	واو	فاء	هاء	لام ألف	همزه	ياء	پاء
kaf	gaf	lam	mim	nun	wau	va	ha	lam-alif	hamzah	ya	nya
k	g	l	m	n	w	v	h	la		y	ny
[k]	[g]	[l]	[m]	[n]	[w]	[v]	[h]	[la]		[j]	[ɲ]

Arabic Numbers for Malay/Jawi

(Omniglot)















































































•	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
kosong	satu	dua	tiga	empat	lima	enam	tujuh	lapan	sembilan	sepuluh
0	1	2	3	4	5	6	7	8	9	10

Arabic Alphabet for Tausug (Omniglot)

غ	س	ر	د	ج	ت	ب	ا
gayn	sīn	rā'	dāl	jīm	tā'	bā'	alip
[ɣ]	[s]	[r]	[d]	[dʒ]	[t]	[b/β]	[a/a:]
و	ن	م	ل	گ	ک	ف	ع
wāw	nūn	mīm	lām	gāp	kāp	pā'	ngā'
[w]	[n]	[m]	[l]	[g/ɣ]	[k]	[p]	[ŋ]
لا	ئ	ؤ	أ	ء	ي	ه	
lām alip				hamja	yā'	hā'	
[la:]	[iʔ]	[uʔ]	[aʔ]	[ʔ]	[j]	[h]	

Jawi Cham - Majid Transliteration

(<http://naipaleikaohkabuak.blogspot.hk/2010/11/majid-transliteration-jawi-cham.html>)

Akhar Cham Eastern Script (Vietnam) Western Script (Cambodia)	Rumi Cham (EFEQ)	Haruf Cham / Jawi Cham (Majid Transliteration)			
 	K		 	T	
 	Kh		 	Th	
 	G		 	D	
 	Gh		 	Dh	
 	Ngâ		 	Nâ	
 	Ng		 	N	
 	C		 	Nd	
 	Ch		 	P	
 	J		 	Ph	
 	Jh		 	B	
 	Nyâ		 	Bh	
 	Ny		 	Mâ	
 	Nj		 	M	












Jawi Cham - Majid Transliteration

(<http://naipaleikaohkabuak.blogspot.hk/2010/11/majid-transliteration-jawi-cham.html>)

	Mb			_h	ه
	Y	ي		a	ا
	R	ر		e	
	L	ل		i	ي
	W	و		ei	اي
	S	س		_i	ي
	H	ح		im	يم
	A	ا		_m	م
	I	اي		u	و
	Ai	عاي		é	
	U	او		o	
	O			ao	
	É			au	او
				ia	يا

Jawi Cham - Majid Transliteration

(<http://naipaleikaohkabuak.blogspot.hk/2010/11/majid-transliteration-jawi-cham.html>)

	r	ر
	kl	
	ua	وا
	â	
	-ng	غ
	em	
	eng	

Malaysian Standards: Jawi Coded Character Set 2012

(324207_MS_2443_2012)

Developed by the Technical Committee on Multilingual Information
Technology under the authority of the Industry Standards Committee on
IT, Communications and Multimedia

Table 8: A representative glyph of each Jawi coded character set

Name in Malay	Character info	Standalone shape	Initial shape	Middle shape	Final shape
<i>alif</i>	Arabic letter <i>alef</i>	ا	ا	ا	ا
<i>ba</i>	Arabic letter <i>beh</i>	ب	ب	ب	ب
<i>ta</i>	Arabic letter <i>teh</i>	ت	ت	ت	ت
<i>ta marbutah</i>	Arabic letter <i>teh marbuta</i>	ة	ة	ة	ة
<i>sa [tha]</i>	Arabic letter <i>theh</i>	ث	ث	ث	ث
<i>jim</i>	Arabic letter <i>jeem</i>	ج	ج	ج	ج
<i>ca</i>	Arabic letter <i>tcheh</i>	چ	چ	چ	چ
<i>ha [ħa]</i>	Arabic letter <i>hah</i>	ح	ح	ح	ح
<i>kha [kho]</i>	Arabic letter <i>khah</i>	خ	خ	خ	خ
<i>dal</i>	Arabic letter <i>dal</i>	د	د	د	د
<i>zal [dhal]</i>	Arabic letter <i>thal</i>	ذ	ذ	ذ	ذ
<i>ra [ro]</i>	Arabic letter <i>reh</i>	ر	ر	ر	ر
<i>zai</i>	Arabic letter <i>zain</i>	ز	ز	ز	ز
<i>sin</i>	Arabic letter <i>seen</i>	س	س	س	س
<i>syin</i>	Arabic letter <i>sheen</i>	ش	ش	ش	ش
<i>sad [ṣod]</i>	Arabic letter <i>sad</i>	ص	ص	ص	ص
<i>dad [ḍod]</i>	Arabic letter <i>dad</i>	ض	ض	ض	ض
<i>ta [ṭo]</i>	Arabic letter <i>tah</i>	ط	ط	ط	ط
<i>za [ẓo]</i>	Arabic letter <i>zah</i>	ظ	ظ	ظ	ظ
<i>ain [ʿain]</i>	Arabic letter <i>ain</i>	ع	ع	ع	ع
<i>ghain</i>	Arabic letter <i>ghain</i>	غ	غ	غ	غ
<i>nga</i>	Arabic letter <i>ain</i> with three dots above	ڤ	ڤ	ڤ	ڤ
<i>fa</i>	Arabic letter <i>feh</i>	ف	ف	ف	ف

Table 8: A representative glyph of each Jawi coded character set (*continued*)

Name in Malay	Character info	Standalone shape	Initial shape	Middle shape	Final shape
<i>pa</i>	Arabic letter <i>veh</i>	ڤ	ڤ	ڤ	ڤ
<i>qaf</i>	Arabic letter <i>qaf</i>	ق	ق	ق	ق
<i>kaf</i>	Arabic letter <i>keheh</i>	ك	ك	ك	ك
<i>ga</i>	Arabic letter <i>keheh</i> with dot above	ڠ	ڠ	ڠ	ڠ
<i>lam</i>	Arabic letter <i>lam</i>	ل	ل	ل	ل
<i>mim</i>	Arabic letter <i>meem</i>	م	م	م	م
<i>nun</i>	Arabic letter <i>noon</i>	ن	ن	ن	ن
<i>wau</i>	Arabic letter <i>waw</i>	و	و	و	و
<i>va</i>	Arabic letter <i>waw</i> with dot above	ڤ	ڤ	ڤ	ڤ
<i>ha</i>	Arabic letter <i>heh</i>	ه	ه	ه	ه
<i>hamzah</i>	Arabic letter <i>hamza</i>	ء	ء	ء	ء
<i>ya</i>	Arabic letter <i>yeh</i>	ي	ي	ي	ي
<i>ye</i>	Arabic letter <i>alef maksura</i>	ى			
<i>nya</i>	Arabic letter <i>noon</i> with three dots above	ښ	ښ	ښ	ښ
<i>da</i>	Arabic letter <i>dul</i>	ڌ	ڌ	ڌ	ڌ
<i>gha</i>	Arabic letter <i>ng</i>	ڱ	ڱ	ڱ	ڱ
<i>gaf</i>	Arabic letter <i>kaf</i> with dot above	ڭ	ڭ	ڭ	ڭ
<i>yah</i>	Arabic letter with <i>yeh</i> with three dots below	ښ	ښ	ښ	ښ

Aggregated Malay-Jawi Tables

Jawi Characters (Dewan Bahasa dan Pustaka – Daftar Kata Bahasa Melayu-Rumi-Sebutan Jawi-Jilid 1 – Pg 71-72) cited in Mynic Report for Malaysia's IDN: Jawi Language Issues Version 1.0 (0312009)	Malaysian Standard on IT- Jawi Coded Character Set for Information Interchange (2012)	Malay Language Table Submitted by .MY Domain Registry (IANA Repository) 09 July 2009	Aggregated Characters of Interest Based on Malaysian Standards (2012) and Malay IDN Language Table (2009)
From Unicode: U+....	MS ISO/IEC 10646 Code Point	#Characters from Unicode Arabic Table (0600–06FF)	From Unicode: U+....
(1) 0627	1) U+0627	1) U+0621 Arabic Letter HAMZA	(1) 200C ZERO WIDTH NON-JOINER
(2) 0628	2) U+0628	2) U+0623 Arabic Letter ALEF with HAMZA above	(2) 0621 HAMZA
(3) 062A	3) U+062A	3) U+0625 Arabic Letter ALEF with HAMZA below	(3) 0623 ALEF W/ HAMZA Above
(4) 0629	4) U+0629	4) U+0626 Arabic Letter YEH with HAMZA above	(4) 0625 ALEF W/ HAMZA Below
(5) 062B	5) U+062B	5) U+0627 Arabic Letter ALEF	(5) 0626 YEH W/ HAMZA above
(6) 062C	6) U+062C	6) U+0628 Arabic Letter BEH	(6) 0627 ALEF
(7) 0686	7) U+0686	7) U+0629 Arabic Letter TEH MARBUTA	(7) 0628 BEH
(8) 062D	8) U+062D	8) U+062A Arabic Letter TEH	(8) 0629 TEH MARBUTA
(9) 062E	9) U+062E	9) U+062B Arabic Letter THEH	(9) 062A TEH
(10) 062F	10) U+062F	10) U+062C Arabic Letter JEEM	(10) 062B THEH
(11) 0630	11) U+0630	11) U+062D Arabic Letter HAH	(11) 062C JEEM
(12) 0631	12) U+0631	12) U+062E Arabic Letter KHAH	(12) 062D HAH
(13) 0632	13) U+0632	13) U+062F Arabic Letter DAL	(13) 062E KHAH
(14) 0633	14) U+0633	14) U+0630 Arabic Letter THAL	(14) 062F DAL
(15) 0634	15) U+0634	15) U+0631 Arabic Letter REH	(15) 0630 THAL
(16) 0635	16) U+0635	16) U+0632 Arabic Letter ZAIN	(16) 0631 REH
(17) 0636	17) U+0636	17) U+0633 Arabic Letter SEEN	(17) 0632 ZAIN
(18) 0637	18) U+0637	18) U+0634 Arabic Letter SHEEN	(18) 0633 SEEN
(19) 0638	19) U+0638	19) U+0635 Arabic Letter SAD	(19) 0634 SHEEN
(20) 0639	20) U+0639	20) U+0636 Arabic Letter DAD	(20) 0635 SAD
(21) 063A	21) U+063A	21) U+0637 Arabic Letter TAH	(21) 0636 DAD
(22) 06A0	22) U+06A0	22) U+0638 Arabic Letter ZAH	(22) 0637 TAH
(23) 0641	23) U+0641	23) U+0639 Arabic Letter AIN	(23) 0638 ZAH
(24) 06A4	24) U+06A4	24) U+063A Arabic Letter GHAIN	(24) 0639 AIN
(25) 0642	25) U+0642	25) U+0641 Arabic Letter FEH	(25) 063A GHAIN
(26) 06A9	26) U+06A9	26) U+0642 Arabic Letter QAF	(26) 0641 FEH
(27) 0762	27) U+0762	27) U+06A9 Arabic Letter KEHEH	(27) 0642 QAF
(28) 0644	28) U+0644	28) U+0644 Arabic Letter LAM	(28) 06A9 KEHEH
(29) 0645	29) U+0645		(29) 0644 LAM
(30) 0646	30) U+0646		(30) 0645 MEEM
(31) 0648	31) U+0648		(31) 0646 NOON

Jawi Characters (Dewan Bahasa dan Pustaka – Daftar Kata Bahasa Melayu-Rumi-Sebutan Jawi-Jilid 1 – Pg 71-72) cited in Mynic Report for Malaysia's IDN: Jawi Language Issues Version 1.0 (0312009)	Malaysian Standard on IT- Jawi Coded Character Set for Information Interchange (2012)	Malay Language Table Submitted by .MY Domain Registry (IANA Repository) 09 July 2009	Aggregated Characters of Interest Based on Malaysian Standards (2012) and Malay IDN Language Table (2009)
(32) 06CF (33) 0647 (34) 0621 (35) 064A (36) 0649 (37) 06BD	32) U+06CF 33) U+0647 34) U+0621 35) U+064A 36) U+0649 37) U+06BD 38) U+068E Arabic Letter DUL 39) U+06AD Arabic letter NG 40) U+06AC Arabic letter KAF with dot above 41) U+06D1 Arabic letter YEH with three dots below 42) U+200C Zero Width Non-Joiner 43) U+06F2 Extended Arabic Indic digit 2 44) currently uncoded in MS ISO/IEC 10646 or MS ISO/IEC 88596) ARABIC HAMZAH THREE-QUARTER	29) U+0645 Arabic Letter MEEM 30) U+0646 Arabic Letter NOON 31) U+0647 Arabic Letter HEH 32) U+0648 Arabic Letter WAW 33) U+0649 Arabic Letter ALEF MAKSURA 34) U+064A Arabic Letter YEH 35) U+0686 Arabic Letter TCHEH 36) U+06A0 Arabic Letter AIN WITH THREE DOTS ABOVE 37) U+06A4 Arabic Letter VEH 38) U+0762 Arabic Letter KEHEH WITH DOT ABOVE 39) U+06BD Arabic Letter NOON WITH THREE DOTS ABOVE 40) U+06CF Arabic Letter WAW WITH DOT ABOVE 41) U+06F2 Arabic – Extended Arabic -Digit Two # Characters from Unicode Basic Latin Table (0000–007F): 42) U+002D Hyphen-Minus 43) U+0030 Digit Zero 44) U+0031 Digit One 45) U+0032 Digit Two 46) U+0033 Digit Three 47) U+0034 Digit Four 48) U+0035 Digit Five 49) U+0036 Digit Six 50) U+0037 Digit Seven 51) U+0038 Digit Eight 52) U+0039 Digit Nine	(32) 0647 HEH (33) 0648 WAW (34) 0649 ALEF MAKSURA (35) 064A YEH (36) 0686 TCHEH (37) 068E DUL (38) 06A0 AIN WITH THREE DOTS ABOVE (39) 06A4 VEH (40) 06AC KAF W/ DOT ABOVE (41) 06AD NG (42) 06BD NOON WITH THREE DOTS ABOVE (43) 06CF WAW WITH DOT ABOVE (44) 06D1 YEH W/ THREE DOTS BELOW (45) 06F2 ARABIC-EXTENDED ARABIC - DIGIT TWO (46) 0762 (47) (currently uncoded) ARABIC HAMZAH THREE-QUARTER (48) * 0644+0627 LIGATURE LAM W/ ALEF ISOLATED FORM * Included in Mynic Report for Malaysia's IDN: Jawi Language Issues Version 1.0 (0312009) as part of Jawi Characters agreed by consulted linguists.

Unicode table for Arabic characters.

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	ـ	و		ذ	-	و	و	و	پ	ز	ع	ف	ه	ي	و	و
1	س	و	ء	ر	ف	و	ا	ا	خ	ز	م	ك	ه	ي	و	ا
2	ـ	و	ا	ز	ق	و	ب	ا	خ	ز	ب	ك	ه	ي	و	ب
3	ط	و	ا	س	ك	و	ا	ا	ج	د	ب	ك	ه	ي	و	ب
4		و	و	ش	ل	و	ا	ا	ج	د	ف	ك	و	و	و	ب
5		و	ا	ص	م	و	ا	ا	خ	د	پ	ل	و	ه	ر	ب
6	✓	و	ئ	ض	ن	و	ا	و	ج	د	ف	ل	و	و	ا	ب
7	✓	و	ا	ط	ه	و	ا	و	ج	ز	ف	ل	و	و	و	ب
8	ي	و	ب	ظ	و	و	ا	ا	ك	ز	ف	ل	و	و	و	ب
9	٪	و	ة	ع	ي	و	ا	ا	د	ز	ك	ن	و	و	و	ب
A	٪	و	ت	غ	ي	و	ا	ا	د	ن	ك	ن	و	و	و	ب
B	ف	ا	ت	ك	و	و	ا	ب	د	ن	ك	ن	و	و	و	ب
C	ا		ج	ك	و	و	ا	ب	د	ن	ك	ن	و	و	و	ب
D	ا		ح	م	و	و	ا	ب	د	ن	ك	ن	و	و	و	ب
E	ـ	ا	خ	م	و	و	ا	ب	د	ن	ك	ن	و	و	و	ب
F	ع	ا	د	م	و		و	ا	د	ن	ك	ن	و	و	و	ب

Unicode extended Arabic letters

	075	076	077
0	پ	ف	ش
1	ب	ف	ژ
2	پ	گ	خ
3	پ	نگ	ا
4	پ	ک	ا
5	ب	م	م
6	ن	م	م
7	خ	ن	ي
8	چ	ن	و
9	ڀ	ن	و
A	د	ل	ے
B	و	ز	ے
C	ش	ژ	ج
D	غ	ش	س
E	غ	ج	س
F	غ	چ	ل



32 Base Characters



3 Characters Used With Condition



6 Additional Characters



4 Characters Added by MS (2012)



+ Zero Width Non-Joiner

+ Three Quarter Hamzah (Uncoded)

TOTAL of 47 characters of interest in
Jawi based on Malaysian Standards
2012 and Malay IDN Table 2009

References

- MYNIC Report for Malaysia's Internationalized Domain Name: Jawi Language Issues, Version 1.0 (0312009)
- Malaysian Standard MS 2443:2012, Information Technology – Jawi Coded Character Set For Information Exchange (2012)
- Malaysian Network Information Center (MYNIC)/.MY Domain Registry, Jawi Language Table Submitted to IANA, 09 July 2009
- Omniglot, Wikipedia, Ethnologue pages on relevant South East Asian languages