

Proposal to encode Javanese and Sundanese Arabic characters

Denis Moyogo Jacquerye <moyogo@gmail.com>
2019-10-03

1. Introduction

The Arabic alphabet has been adapted and used to write languages in the Malay Archipelago since at least the 14th century. The most used adaptation is the Jawi alphabet, used in Classical Malay and several Modern Malayic languages such as Malaysian Malay (zms), Indonesian Malay (id), Brunei Malay (kxd), Acehnese (ace), Banjar (bjn), Bengkulu (pse), Minangkabau (min), Musi (mui) and others.

Other languages in the region also use adaptations of the Arabic alphabet, sharing letters with Jawi but also using variants or additional letters and diacritics:

- Pegon is the Arabic alphabet used for Javanese (jv), Sundanese (su) and Madurese (mad), with some variation between those languages or between works,
- Buri Wolio is the Arabic alphabet used in Wolio (wlo),
- Sulu or Sulat Sug is the one used for the Tausug language (tsg),
- Serang is the one used for Makassarese (mak).

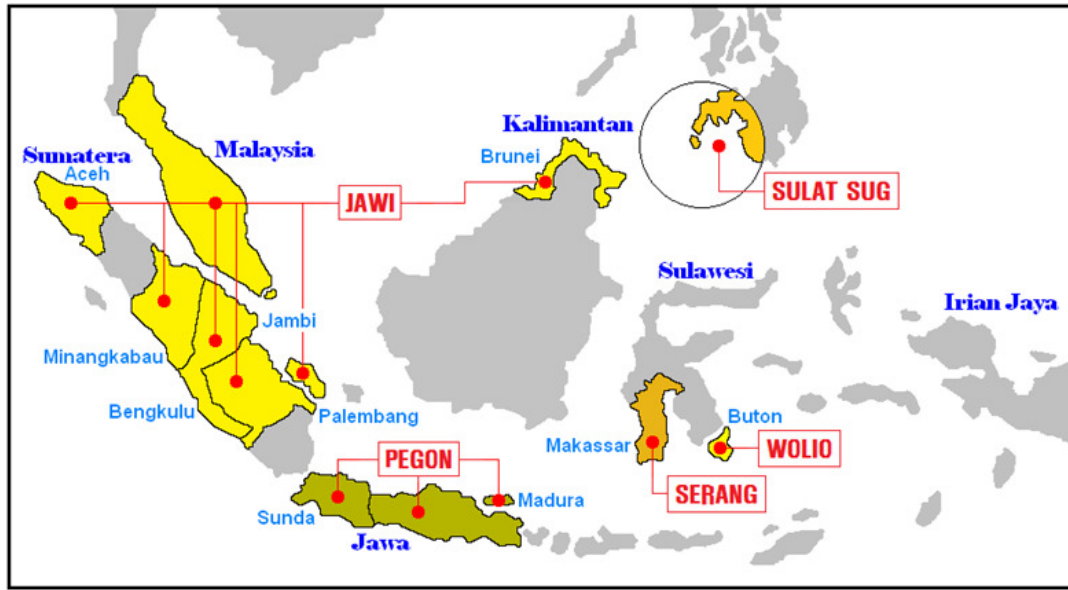


Figure 1 : Map of languages of the Malay Archipelago using the Arabic alphabet in Cho (2018), showing Jawi, Pegon, Serang, Wolio and Sulat Sug (Tausug).

Javanese Pegon uses ARABIC LETTER TAH WITH THREE DOTS BELOW or ARABIC LETTER TAH WITH DOT BELOW, which are proposed to be encoded. Malagasy (mg) Sorabe, the Arabic alphabet used in Malagasy in Madagascar, also uses ARABIC LETTER TAH WITH DOT BELOW.

Sundanese Pegon uses the ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW which is proposed to be encoded.

In Gallop et al. (2015) p. 14, it is noted that “the Malay *ga* usually has one or three dots above or below”. Werndly (1736), p. 6, describes such reduction of three dots noting it doesn’t occur with letters that would clash with letters in the same shape group that have a single dot, such as 0686 ARABIC LETTER TCHEH چ and 062C ARABIC LETTER JEEM ج, 06A4 ARABIC LETTER VEH ق and 0641 ARABIC LETTER FEH ف, or 06A0 ARABIC LETTER AIN WITH THREE DOTS ABOVE ع and 063A ARABIC LETTER GHAIN غ. This variation of number of dots and their position in Jawi manuscripts can also be observed in some Javanese or Sundanese Pegon works, for *ga* (KAF- or KEHE-shape), *dha* (DAL-shape) or *tha* (TAH-shape), with three dots below or one dot below.

Some variations :

Final *ta**jahat* جاهت *bērkāt* بركة *'adat* عادةFinal *mim**-nam* منم نم نم*Jim and ha**jajahan* جاجان *hijjah* حجه *hajji* حج*Kaf kēchil* ك ك ك*Ga* گ گ گFigure 2 : Variants of *ga* in Jawi in Lewis (1954). p. 56.

2. Proposed characters to be encoded



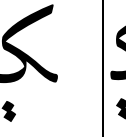

ARABIC LETTER TAH WITH DOT BELOW and ARABIC LETTER TAH WITH THREE DOTS BELOW

Many Javanese Pegon works use variants of a dotted TAH to represent a voiceless retroflex stop /t/: ARABIC LETTER TAH WITH DOT BELOW or ARABIC LETTER TAH WITH THREE DOTS BELOW, which need to be encoded. It also uses 068A ARABIC LETTER DAL WITH DOT BELOW or 08AE ARABIC LETTER DAL WITH THREE DOTS BELOW to represent a voiced retroflex stop /d/. Adelaar (1995), p. 337, notes that ARABIC LETTER TAH WITH DOT BELOW and 068A ARABIC LETTER DAL WITH DOT BELOW are also used for the same sounds in Sorabe, the Malagasy (mg) Arabic alphabet, in Madagascar.

codepoint	isolated	final	medial	initial	name
U+088B	ط	ط	ط	ط	ARABIC LETTER TAH WITH DOT BELOW
U+088C	ط	ط	ط	ط	ARABIC LETTER TAH WITH THREE DOTS BELOW

ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW

Various Sundanese Pegon works use ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW, for /g/, which needs to be encoded.

codepoint	isolated	final	medial	initial	name
U+088D					ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW

3. Unicode character properties

```
088B;ARABIC LETTER TAH WITH DOT BELOW;Lo;0;AL;;;;;N;;;;;
088C;ARABIC LETTER TAH WITH THREE DOTS BELOW;Lo;0;AL;;;;;N;;;;;
088D;ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW;Lo;0;AL;;;;;N;;;;;
```

4. Joining type and group for ArabicShaping.txt

```
088B; ARABIC LETTER TAH WITH DOT BELOW; D; TAH
088C; ARABIC LETTER TAH WITH 3 DOTS BELOW; D; TAH
088D; ARABIC LETTER KEHEH WITH VERTICAL 2 DOTS BELOW; D; GAF
```

5. Annotations

The following annotations could be updated in NameList.txt.

```
068A ARABIC LETTER DAL WITH DOT BELOW
    * Pegon, Malagasy
08AE ARABIC LETTER DAL WITH THREE DOTS BELOW
    * Belarusian
    * Pegon alternative for 068A
```

The following annotations are recommended for NameList.txt.

```
@ Additions for Pegon:
088B ARABIC LETTER TAH WITH DOT BELOW
    * Pegon, Malagasy
088C ARABIC LETTER TAH WITH THREE DOTS BELOW
    * Pegon alternative for 088B
088D ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW
    * Sundanese Pegon
```

6. Suggested collation

088B ARABIC LETTER TAH WITH DOT BELOW and 088C ARABIC LETTER TAH

WITH THREE DOTS BELOW can be sorted after 08A3 ARABIC LETTER TAH
WITH TWO DOTS ABOVE and before 0639 ARABIC LETTER AIN.

088D ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW can be
sorted after 063B ARABIC LETTER KEHEH WITH TWO DOTS ABOVE.

7. Examples

TAH WITH DOT BELOW

18.	ba	ب	ب
19.	tha	ط	ط
20.	nga	نغ	نغ

Berdasarkan padanan aksara Jawa dan *Pegon*, tampaklah bahwa aksara *Pegon* yang bentuknya huruf Arab mempunyai sistem bunyi yang sama dengan aksara Jawa. Namun, dari padanan aksara Arab dan *Pegon*, terlihat bahwa huruf *Pegon* yang bunyi dan bentuknya sama dengan huruf Arab dalam sistem tulisan Arab jumlahnya hanya tiga belas, yaitu *ba* (ب), *ta* (ت), *jim* (ج), *dal* (د), *ra* (ر), *sin* (س), *kaf* (ك), *lam* (ل), *mim* (م), *nun* (ن), *wau* (و), *alif/ha* (ا / هـ), dan *ya* (ي), sedangkan tujuh aksara *Pegon* lainnya, yaitu *ca* (چ), *pa* (پ), *dha* (ذ), *nya* (نپ), *ga* (گ), *tha* (ط), dan *nga* (نغ).

Ketujuh huruf Arab hasil modifikasi yang bentuknya mirip dengan huruf Arab, tetapi terdapat tanda titik diakritik dan bunyinya tidak dikenal dalam sistem tulisan huruf Arab, disebut huruf Arab rekaan. Ketujuh huruf Arab rekaan ini, kiranya memang diciptakan untuk mewakili bunyi-bunyi yang ada dalam bahasa Jawa, karena tidak ada padanan bunyinya dalam sistem tulisan Arab. Bagaimana ketujuh huruf Arab rekaan ini diciptakan hingga kini belum diketahui dengan pasti. Namun, bertolak dari pendapat Achadiati Ikram (1977:38) yang mengatakan bahwa *Jawi* adalah aksara Arab yang digunakan sebagai sarana rekam bahasa Melayu, maka sistem tulisan *Pegon* dipadankan dengan *Jawi*. Berdasarkan padanan itu, dapat diketahui bahwa lima dari tujuh huruf Arab rekaan *Pegon*, yakni *ca* (چ), *pa* (پ), *nya* (نپ), *ga* (گ), dan *nga* (نغ) juga terdapat di dalam tulisan *Jawi*.

Figure 3 : Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Yulianto and Pudjiastuti (2001), p. 207.

کتر اغان	حرف عرب فيكون	حرف لاتين	كاواچا	اڪسار اچاوا	نومر
حرف ريكان	د : دا	dh	dha	ڊا	۱۳
چوچوك واتاگي	ج : جيم	j	ja	جا	۱۴
" "	ي : ياء	y	ya	يا	۱۵
حرف ريكان	پ : پا	ny	nya	ڻا	۱۶
چوچوك واتاگي	م : ميم	m	ma	ما	۱۷
حرف ريكان	ك : كا	g	ga	گا	۱۸
چوچوك واتاگي	ب : باء	b	ba	با	۱۹
حرف ريكان	ط : طا	th	tha	ٲا	۲۰
" "	غ : غا	ng	nga	ڱا	۲۱

Figure 4 : Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Ghithrof Danil-Barr (20??), p. 7.

The Pegon script has similarities and differences with Arabic script and Jawi script that developed and used in Malaysia, Brunei, Singapore and Pattani Thailand. The equation is: (1) The script is written from right to left. (2) In the script system, there are alphabets arranged with each other. (3) There is a dot sign and a harak or syakal. (4) There is a change of alphabets when in front, middle and back. (5) The alphabet connection is in one word, not between words. (6) In one word, there can be several pieces of alphabet connection. (7) Numbers are written from left to right. While the difference is: (1) System script in Arabic Pegon is based on the phonological system of Java language. (2) The vowels in the Pegon script are more than the vowels contained in the Arabic script as well as the Jawi script. (3) In the Pegon script, there is syakal, while in Jawi writing does not exist. (4) There are alphabets ج, چ, ف, د, ي, ڊ, ڳ, ڱ, which are not found in Arabic script. (5) There are alphabets ڊ and ڳ, which are not found in Jawi. (6) The sound /g/ is written with the alphabets ڳ, while in Jawi it is written with گ. (7) The sound /n/ written with the alphabet ڻ, while in Jawi script written with ن.

Figure 5 : Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW and variants of ga in Mahfudh (2017), p. z.

سُودَا سَدِينَا تَوَمَكَا ٨ تَاهُون .
 ٨ يَيْنُ سَتَاهُونُ اَنَا سَيْعُ كَوِطَاعُ
 اَتَوَا فِدَوْت ، هِيَا كُوْدُو بِالِيْنِي

Figure 6 : Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Semarang (19??), p.12.

٨
 وَفِقْ كَعْ بَادِي كَاسَبَات . لَاجَعْ دِيْفُونُ لَفِيْتِ دِي تَالِيْنِي بُوْلَاهُ
 لَاجَعْ دِي چَنْطِيْلَا كَنْ دَاتْعْ فَا كَاغِي وَيْتِ دَلِيْمَا سَا جَاوِيْنِي كَرِيَا
 مَنَاوِي بُوْتَنْ وَوَنْتَنْ وَيْتِ دَلِيْمَا كِيْتِيْعْ كَكَايُونُ سَانِيْسِي .

Figure 7 : Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Anwar (19??), p.8.

دَالَانْ جُوْرُوْسَانَ مَلْبِتْ كُوْجَا اَتُوِي كَامْفُوْعْ مُوسُوْهْ بُوْتَنْ
 سَابْتْ مَلْبِتْ كُوْجَا اَتُوِي
 كَامْفُوْعْ لَنْ نَالِيْكَادِي كِيْمَبُوْلْ

الله	مانع دافع محيط	الله
ا	و / ه / ح / ط	ا

Figure 8 : Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Anwar (19??), p.28.

اَيْفُونُ اِثْغِيَه مَنِكَ صِيَامُ تِيكَاع دِيْنَتُنْ
ثَلَاثَا، رَبُّو، خَمِيْس . سَالْتِ اَيْفُونُ
صِيَامُ سُوْفَادُوْس تَاْطَه ٢ اَكْنُ مَا هُوْس
صَلَوَاتُ مَنِكَ . لَاجِعُ سَاتْرُوْس اَيْفُونُ

Figure 9 : Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Musthofa (19??), p.3.

نَوْمَز تِيكَ صِفَة بَقَاء :
تَكْسِيْفُونُ بُوْسْتِي اِنَّهُ مَنِكَ مَسْبِي اِلَاغْبَعُ بُوْتِنُ مَاوِي
فُوْغَا سَان . دِيْنِي دَلِيْل اَيْفُونُ اِثْغِيَه مَنِكَ ، سَاء اُوْفِي

Figure 10 : Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Asnawi Kudus (190?), p.8.

تَكْسِيْفُونُ اِيْمَانُ دَاتَعُ قَدْرُ (فَسْبِيْن) اِثْغِيَه
مَنِكَ كِيْتَا فَرَجِدْ وِس بِيْلِيَه سَدَايَا كَدَادُوْسَانُ

Figure 11 : Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Asnawi Kudus (190?), p.48.

It is also necessary to draw the adaptations of the Arabic script as used for some other Indonesian languages into one's comparison. Pegon, for instance, shows remarkable agreement with Sorabe in that it uses the same graphic conventions for retroflex *d* and *t* as does Sorabe for *d* and *t*: retroflex *d* is written as a *dal* with a subscript dot (ڍ), and retroflex *t* is written as a *ta* [tɔ] with a subscript dot (ڒ). For a better appreciation of the agreement it should be added that Javanese speakers often perceive the alveo-

Figure 12 : Malagasy Sorabe ARABIC LETTER TAH WITH DOT BELOW in Adelaar (1995), p.337.

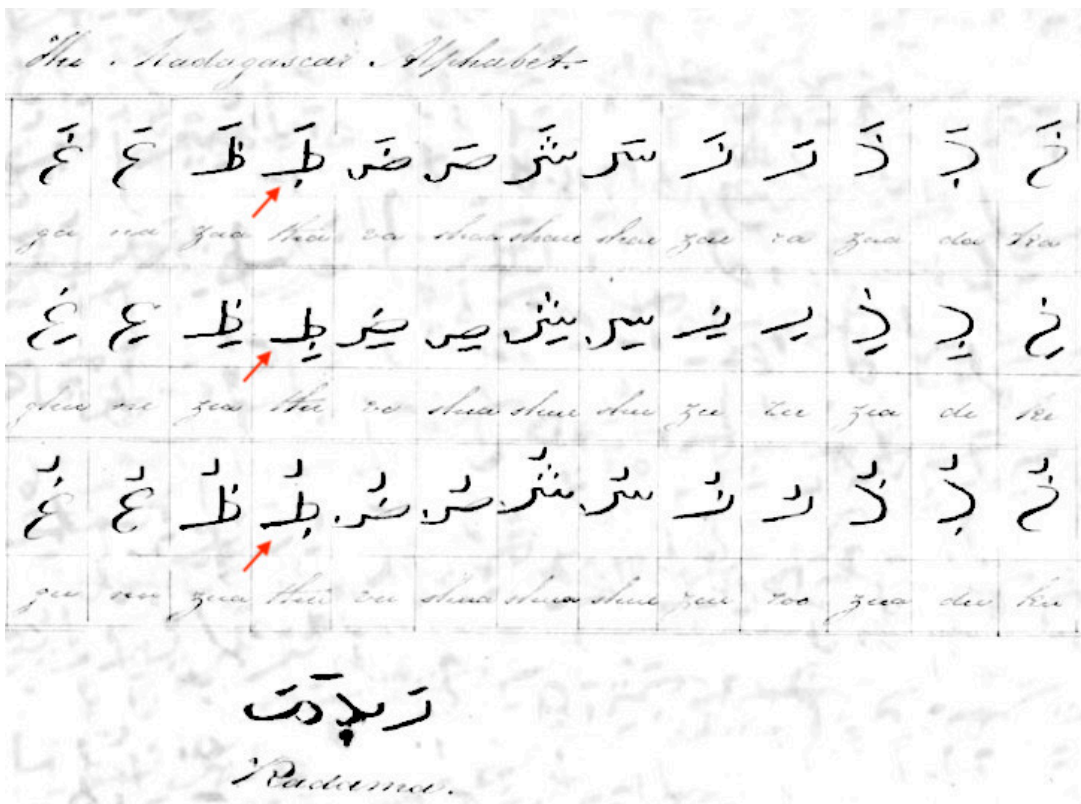


Figure 13 : Malagasy Sorabe ARABIC LETTER TAH WITH DOT BELOW as well as 068A ARABIC LETTER DAL WITH DOT BELOW (not marked) in MS 18141, f.1r.

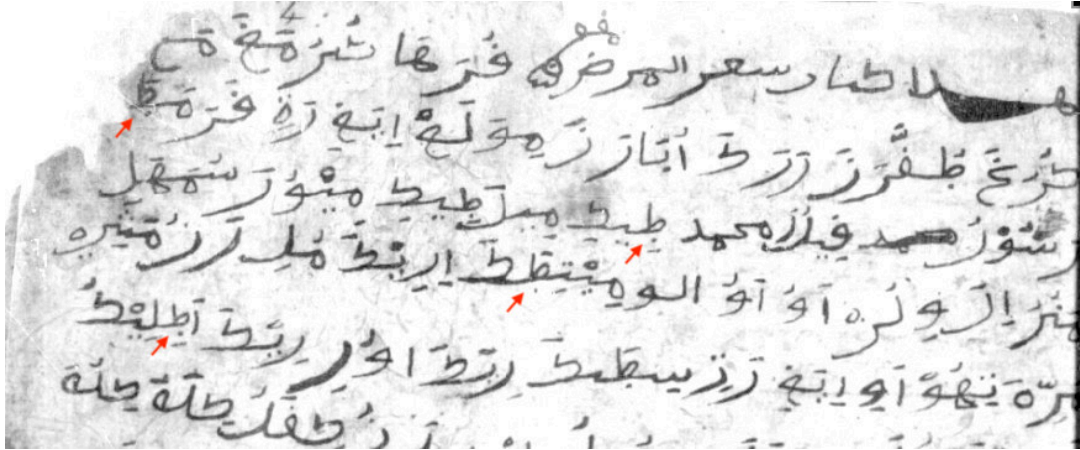


Figure 14 : Malagasy Sorabe ARABIC LETTER TAH WITH DOT BELOW as well as 068A ARABIC LETTER DAL WITH DOT BELOW (not marked) in Manuscrit Malayo-polynésien 20 f.4.

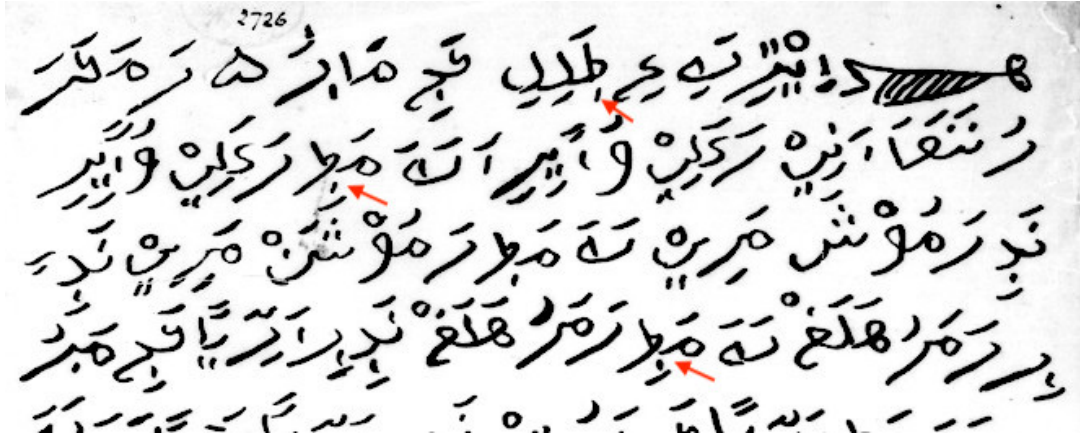


Figure 15 : Malagasy Sorabe ARABIC LETTER TAH WITH DOT BELOW as well as 068A ARABIC LETTER DAL WITH DOT BELOW (not marked) in Manuscrit Malayo-polynésien 26 f.3.

TAH WITH THREE DOTS BELOW

كَعْمُونَ هَع دَنِيَّاسِه هَع عَاخِرُ
 كَع عَامَعْلُو سَكَطِهَع قُوجِي نُفَيْرِن
 نَع عَالَم كَابِيَه كَع تَنَا يَكُطُو كَع عَكْرِي
 بَوْم لَنَا لَحْت مِيَع سَكَطِهَع تَيْتَه كَعَا

Figure 16 : Javanese Pegon ARABIC LETTER TAH WITH THREE DOTS BELOW in MS 12305 (late 18th, early 19th century), f.4v.

الْقُرْآنُ الْكَرِيمُ سَمْفُونُ كَاطَلَه دِيْفُونُ تَرْجَمَه دِيْنِيْعُ فَا رَا اَهْلُ تَرْجَمَه : وَوَنْتَنُ
 اِعْكَعْ مَاوِيْ بَا صَا وَالْنَدِيْ ، اِنْكِرِيْس ، جِرْمَان ، اِنْدُوْنِيْسِيَا لَنْ سَا نِيْس ٢ اِيْفُونُ
 مَا لَه اِعْكَعْ مَاوِيْ تَمْبُوْعُ دَا تَرَاه ، جَاوِيْ ، سُوْنْدَا ، لَنْ سَاءُ فَا نُوْعِيْكَ لَانْفُونُ اُوْكَ
 سَمْفُونُ كَاطَلَه ٣ كُنْطِيْ تَرْجَمَه ٢ وَا هُوْ ، اُمَّةُ اِسْلَامُ سَكِيْعُ سَدَا يَا بُوْعَصَا لَنْ سَكُوْ
 لَاجِعُ كَاطَلَه اِعْكَعْ سَا كَدُ مَا عَرْتُوْسِيْ مَعْنِيْ لَنْ تَكْسُ اِيْفُونُ .

Figure 17 : Javanese Pegon ARABIC LETTER TAH WITH THREE DOTS BELOW in Mustofa 1960, p.1.

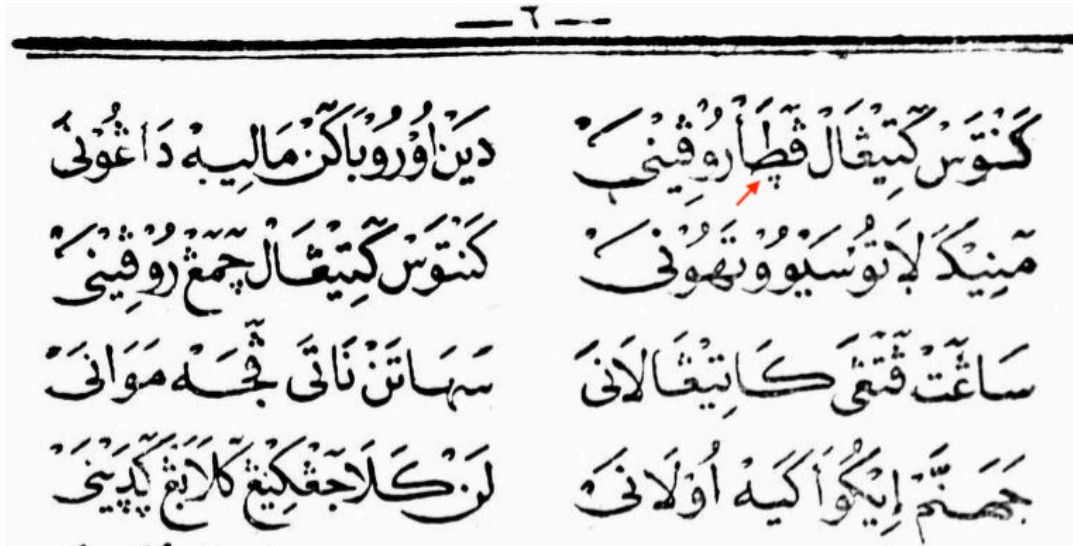


Figure 18 : Javanese Pegon ARABIC LETTER TAH WITH THREE DOTS BELOW in Al-Hajawi (19??), *Syir Sekar Cepogo*, p.6.

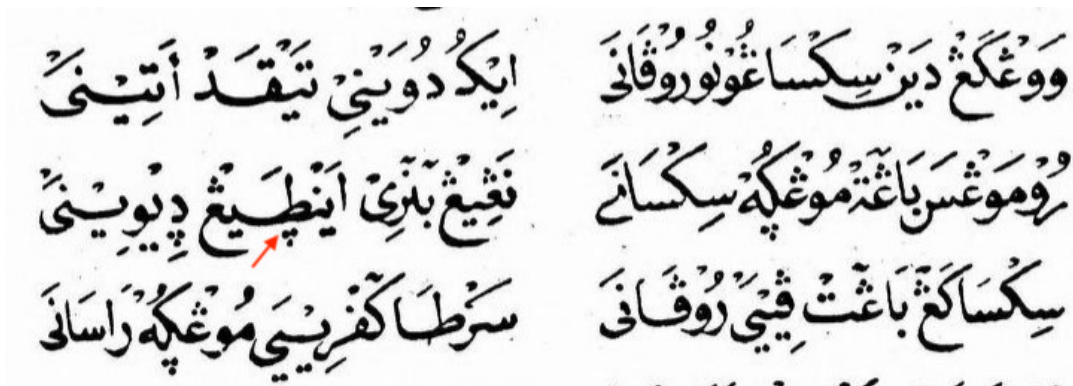


Figure 19 : Javanese Pegon ARABIC LETTER TAH WITH THREE DOTS BELOW in Al-Hajawi (19??), *Syir Sekar Cepogo*, p.9.

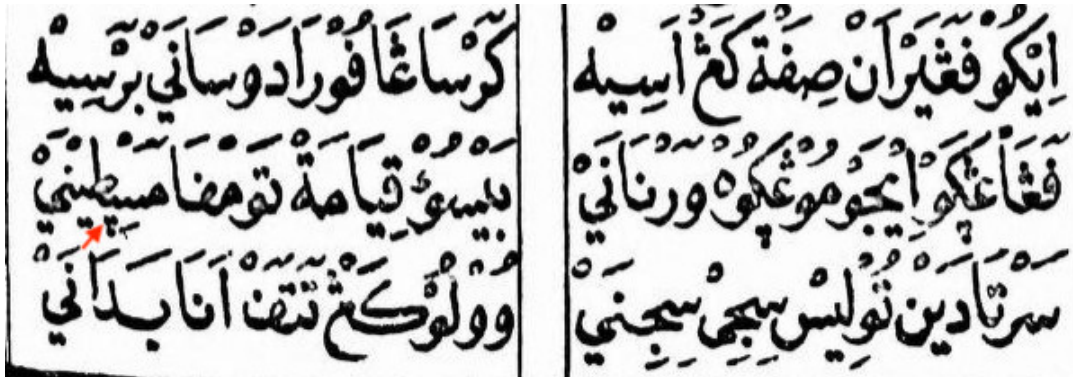


Figure 20 : Javanese Pegon ARABIC LETTER TAH WITH THREE DOTS BELOW in Al-Hajawi (19??), *Syair Sekar Kedathon*, p.12.

KEHEH WITH TWO DOTS VERTICALLY BELOW

HURUF SUNDA ARAB

Nama	Terikat	Terikat ke kiri	Terikat ke kanan	Terikat ke kiri & kanan	Sebutan oleh orang Sunda	Ditulis dengan huruf Latin	Catatan	Tanda Erab Vokal		
								Bentuk	Nama	Nilai
Alif	ا	ا	ا	ا	Alif	a atau vokal lain				
Ba	ب	ب	ب	ب	Be	b				
Ta	ت	ت	ت	ت	ta	t		ا	Pungjabar	a
Tsa	ث	ث	ث	ث	sa	ts	Tak ada dalam bahasa Sunda	ب	Pangjeer	i
Jim	ج	ج	ج	ج	jim	j		ا	Pamees	u
Ca	چ	چ	چ	چ	ca	c		ا	Paneleng	e
Ha	ح	ح	ح	ح	he leutik	h	H = ح	ا	Panolong	o
Kha	خ	خ	خ	خ	he gēde	kh	Tak ada dalam bahasa Sunda	ا	Pamēpēt	ě atau eu
Dal	د	د	د	د	dal	d				
Dzal	ذ	ذ	ذ	ذ	dal	dz	Tak ada dalam bahasa Sunda	Tanda-tanda lain yang termasuk Erab		
Ra	ر	ر	ر	ر	ěre	r	Tak ada dalam bahasa Sunda			
Za	ز	ز	ز	ز	je	z				
Sin	س	س	س	س	sin leutik	s	Tak ada dalam bahasa Sunda	ا	Ta	t akhir
Syin	ش	ش	ش	ش	sin gēde	sy		ا	Ra ha	r h
Tsad	ص	ص	ص	ص	sad	ş atau ç	Tak ada dalam bahasa Sunda	ا	Lamalif	la
Dlad	ض	ض	ض	ض	dad	dl	Tak ada dalam bahasa Sunda	ا	Waoe ta	w...t
Tha	ط	ط	ط	ط	te	th		ا	Waoe-ha	w...h
Thla	ظ	ظ	ظ	ظ	le	thl	Tak ada dalam bahasa Sunda	ا	Hamjah	Penggantian untuk menghindari penumpukan
Ain	ع	ع	ع	ع	ain	a atau vokal lain		ا		
Ghain	غ	غ	غ	غ	gin	g		ا	Pamaeh	Menutup setiap suku terbuka atau tertutup
Nga	ڠ	ڠ	ڠ	ڠ	nga	ng	E yang diubah	ا		
Fa	ف	ف	ف	ف	fe	f	Tak ada dalam bahasa Sunda	ا	Pangrajek	Tanda ulang
Pa	پ	پ	پ	پ	pe	p	و yang diubah	ا	Tasdid	Tanda penggandaan
Kaf	ک	ک	ک	ک	kaf	k atau q	ی yang diubah	ا	Pangnam-jin	Tanda pemisah, istirahat
Kap	ک	ک	ک	ک	kap	k				
Ga	گ	گ	گ	گ	ga	g	ی yang diubah			
Lam	ل	ل	ل	ل	lam	l				
Mim	م	م	م	م	ēmim	m				
Nun	ن	ن	ن	ن	ěnun	n				
Wau	و	و	و	و	wau	w atau u	Vokal dan konsonan			
Ha	ه	ه	ه	ه	he	h				
Ya	ی	ی	ی	ی	iya	y atau i	ی yang diubah			
Nya	ی	ی	ی	ی	nya	ny				

CATATAN:
Beberapa huruf yang di sini bertitik dua di bawahnya, kadang-kadang juga bertitik tiga.

Figure 21 : Sundanese Pegon ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW in Coolsma (1985), p. 10.

لَمُونَهَتْ لَتُونَاتْ مَدَهَتْ اَصْحَ صَلَاتِنَا : جَع
 تَكَلْ اَوِيَه سَلَامْ اِمَامْ تُولُوِيْ عَدْرْ مَامُومْ مَسْبُوقْ
 مَدَ مَاجْ تَكْبِيرْ كَرْنْ عَدْرْ لَمُونْ مَنَعْنَا عَدْرْ اِيْتْ
 تِنَا مَحَلْ لَقُكُهِنَا مَامُومْ مَسْبُوقْ . لَمُونْ مَنَعْنَا
 عَدْرْ اِيْتْ لَايْنِ تِنَا مَحَلْ لَقُكُهِنَا مَامُومْ مَسْبُوقْ

Figure 22 : Sundanese Pegon ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW in EAP211/1/3/2 p. 198.

كَرَنْ سَكِيْتُو مَيْسِهَنْ اَللّهُ كَا اَلَمْ دُونِبَا مُوْعَبَكِهْ دَوْعَكَفْ كَا
 مَفَرِنْ فُوْتَرَهْ نُوْ نُوْعَبَكِلْ سُوْفِيَا اُوْعَبَكِلْ ۲ جَلَمْ نُوْ فَرَجِيَا كَا
 دِيْبَا هَمُوْ جَلَكْ سَرَتْ مَنَعْ هِرُوْفْ نُوْ لَقُكِهْ

Figure 23 : Sundanese Pegon ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW in British and Foreign Bible Society (1930), *The Gospel in many tongues*, p.128.

12. بَرَعُ كَتَقْدِلْ كُوْ جَكْرِيَا سَكْبُ أَنْجَنْ نَا عَرَيْنَجَكْبُ سَرَتَ سِيْنَنْ *
 13. لَجْعُ مَلَايِكَةُ تَيْبَةُ عَلَهْرُ كَا أَنْجَنْ نَا مَوْنَتَوُعُ سِيْنَنْ ۲ جَكْرِيَا سَبَبُ
 فَئَدَ سَمْفِيْنَنْ كَيْسُ دِئَرَمَ سَرَتَ اَيْلِسَبِيَّةُ فَمَجِكَنْ سَمْفِيْنَنْ بَكْلُ

Figure 24 : Sundanese Pegon ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW in *Nederlandsch Bijbelgenootschap* (1877), p.2.

ڊڇيڪ دي ڪنٽور چيٽڪ (اِخْتِيَارَ) فزراگن كالير
 ڪڻغ ڪفتيهن نومر ۱۳ (بوڳور)

Figure 25 : Sundanese Pegon ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW in Nuh Cianjur (19??)

8. Bibliography

- Anonymous, *Manuscrit Malayo-polynésien 20*, 18th century, Bibliothèque nationale de France. Département des Manuscrits. <https://gallica.bnf.fr/ark:/12148/btv1b10087100x>
- Anonymous, *Manuscrit Malayo-polynésien 26*, 18th century, Bibliothèque nationale de France. Département des Manuscrits. <https://gallica.bnf.fr/ark:/12148/btv1b10086253c>
- Anonymous, *Papers in Malagasy and Malay*, Early 19th century, British Library, MS 18141, http://www.bl.uk/manuscripts/Viewer.aspx?ref=add_ms_18141
- Anonymous, *Sĕrat Nawawi*, Late 18th century-early 19th century, British Library, MS 12305, http://www.bl.uk/manuscripts/Viewer.aspx?ref=add_ms_12305
- Anonymous, *Sabilussa'adah Volume I (Tarjamatul Mukhtasar Sajarah Ghoyatul Ikhtishar)*, British Library, EAP211/1/3/2, <https://eap.bl.uk/archive-file/EAP211-1-3-2>
- Alexander Adelaar (1995). "Asian roots of the Malagasy; A linguistic perspective", *Bijdragen tot de taal-, land- en volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia*, volume 151: issue 3. p. 325–356. <https://doi.org/10.1163/22134379-90003036>
- Abi Muhammad Sholih Al-Hajawi (19??). *Syiir Sekar Cepogo*. <https://archive.org/details/SyiirSekarCepogoAbiSholeh22Mar2016095253>
- Abi Muhammad Sholih Al-Hajawi (19??). *Syiir Sekar Kedathon*. <https://archive.org/>

- details/SyirSekarKedathon08Mar20161912321
- Zahwab Anwar (19??). *Doa Jaljalut*. <https://archive.org/details/DoaJaljalutKHZahwabAnwar20Mar2016063315.compressed>
- Roden Asnawi Kudus (190?). *Kitab Tauhid Jawa*. <https://archive.org/details/KitabTauhidJawanKHRAasnawi04Mar2016073558>
- British and Foreign Bible Society (1930). *The Gospel in many tongues*. https://archive.org/details/details/gospelinmanyong00brit_0
- Taeyoung Cho (2018). "Tulisan Arab: Pembina Tamadun Islam Di Nusantara [Arabic Script: The Founder Of Islamic Civilization In The Archipelago]", *Siddhayātra*, vol. 23, no 2, November 2018, p. 114–127 (DOI 10.24832/siddhayatra.v23i2.136) <http://siddhayatra.kemdikbud.go.id/index.php/siddhayatra/article/view/136>
- S. Coolsma (1985). *Tata Bahasa Sunda*, Jakarta, Djambatan, coll. "Indonesian Linguistics Development Project". <http://repositori.kemdikbud.go.id/2645/>
- Sahal Mahfudh (2017). كتابة عرب فيكون خصائصها وإسهاماتها في تطوير تعليم اللغة العربية باندونيسيا [Aksara Arab Pegon Karakteristik dan Kontribusinya dalam Pengembangan Pembelajaran Bahasa Arab di Indonesia, Jurusan Pendidikan Bahasa Arab = Arab Pegon Script Characteristics and Its Contribution in the Development of Arabic Language Learning in Indonesia], Department of Arabic Language Teaching, Postgraduate of Islamic State University Maulana Malik Ibrahim Malang.
- Nederlandsch Bijbelgenootschap (1877). *Het Evangelie van Lucas in het Soendaneesch*. Amsterdam : Nederlandsch Bijbelgenootschap. <https://babel.hathitrust.org/cgi/pt?id=hvd.hwplmq>
- Raden Muhammad Nuh Cianjur (19??). Risalah qiblat diterangkan dengan empat bahasa: Arab, Melayu, Sunda, Jawa [= Risalah Qiblat described in four languages: Arabic, Malay, Sunda, Javanese]. <https://archive.org/details/RisalahQiblat4Bahasa>
- Annabel Teh Gallop et al. (2015). "A Jawi sourcebook for the study of Malay palaeography and orthography", *Indonesia and the Malay World*, vol. 43 "A Jawi sourcebook" (DOI 10.1080/13639811.2015.1008253)
- Abu M. Ghithrof Danil-Barr (20??). الإِثْقَانُ قُدُومَانِ مَاچَالْنِ نُولِيسِ عَرَبْ قِي كَيُونُ [Al-Itqan Pêdoman Baca Tulis Arab Pegon].
- Martha Blanche Lewis (1954). *A handbook of Malay script*, London, Macmillan.
- Bisri Mustofa (1960). *Al-Ibriz li-ma'rifat tafsir al-Qur'an al-'aziz : bi-al-lughah al-jawiyah*, number 1.
- Bisri Mustofa (19??). *Al-Haqiba* <https://archive.org/details/AlHaqibah>
- Bisri Mustofa (19??). *Mitro Sejati*. <https://archive.org/details/MitroSejatiKHBisriMustofa03Mar2016100442.compressed1>
- Abdullah Umar Semarang (19??). *Ijazah Doa Jalbur Rizqi Tahunan*. <https://archive.org/details/details/IjazahDoaJalburRizqiKHAbdullahUmarSemarang20Mar20160648091>
- George Hendrik Werndly (1736). *Maleische Spraakkunst*. <https://books.google.com/books?id=AU0ahfPjKewC>
- Ninie Susanti Yulianto and Titik Pudjiastuti (2001). "Aksara" in Edi Sedyawati, I. Kuntara Wiryamartana, Sapardi Djoko Damono, Sri Sukesi Adiwimarta, *Sastra Jawa: suatu tinjauan umum*, Jakarta, Balai Pustaka, p. 199–207.
-

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from

<http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from

<http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

A. Administrative

1. Title: Proposal to encode Javanese and Sundanese Arabic characters

2. Requester's name: Denis Moyogo Jacquerye <moyogo@gmail.com>

3. Requester type (Member body/Liaison/Individual contribution): Individual contribution

4. Submission date: 2019-10-03

5. Requester's reference (if applicable): _____

6. Choose one of the following:

This is a complete proposal: YES

(or) More information will be provided later: NO

B. Technical - General

1. Choose one of the following:

a. This proposal is for a new script (set of characters): _____

Proposed name of script: _____

b. The proposal is for addition of character(s) to an existing block: YES

Name of the existing block: Arabic Extended-B

2. Number of characters in proposal: 3

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary B.1-Specialized (small collection) B.2-Specialized (large collection)

C-Major extinct D-Attested extinct E-Minor extinct

F-Archaic Hieroglyphic or Ideographic G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided? YES

a. If YES, are the names in accordance with the "character naming guidelines" YES

b. Are the character shapes attached in a legible form suitable for review? YES

5. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? Denis Moyogo Jacquerye

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): <moyogo@gmail.com>

6. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? YES

b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? YES

7. Special encoding issue

Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? YES (presentation, sorting)

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see UAX#44: <http://www.unicode.org/reports/tr44/> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	NO
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, available relevant documents:	YES
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	YES
4. The context of use for the proposed characters type of use; common or rare) Reference:	YES
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	YES
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If Yes, reference:	YES Yes See proposal
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	YES
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If Yes, reference:	NO
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If Yes, reference:	NO
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If Yes, reference:	NO
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If Yes, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If Yes, reference:	NO
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	NO
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If Yes, reference:	NO