

## Response on Maximal Starting Repertoire Version 1 (MSR1)

Task Force on Arabic Script IDNs (TF-AIDN)<sup>1</sup>

Middle East Strategy Working Group

TF-AIDN welcomes the release of Maximal Starting Repertoire Version 1 for the development of LGR in the Zone. Members of TF-AIDN also form the Arabic Script Generation Panel. However, the comments are not limited to the scope of Arabic Script Generation Panel, but also include additional comments by TF-AIDN as a community based organization.

1. TF-AIDN acknowledges the work put in by the Integration Panel to produce a baseline reference for the LGR for the Root Zone very timely for the communities to consider it in their work as Generation Panels.
2. Based on its own experience, TF-AIDN re-iterates that an in-depth analysis of each code point being excluded by MSR1 is required by the relevant community. Therefore, TF-AIDN suggests that complete MSR1 should not be frozen together, but IP should consider freezing it one script at a time, after relevant script Generation Panel has been formed and given its feedback on its relevant portion.
3. TF-AIDN requests IP to share reference materials they use internally for relevant discussions and decisions for MSR and LGR, to whatever extent possible. This can be done by posting the materials at the IP or IDN project web pages. Opening up this information would help understanding the IP perspective and for community and GPs to consult and incorporate such material timely. This will also help reach consensus and serve as a repository for future revisions of the MSR and LGR.
4. TF-AIDN is also focusing on work on universal acceptance of Arabic IDNs. The task force foresees potential utility of both MSR and LGR guidelines and documents in its work on universal acceptance, and therefore appreciates this work in a more general context as well.
5. ICANN is also requested to work with IP to define a tentative timeline for releasing further versions of MSR so that language communities can organize their work accordingly, if revision in MSR1 is needed for a certain script.

---

<sup>1</sup> See <https://community.icann.org/display/MES/Task+Force+on+Arabic+Script+IDNs> for details

6. TF-AIDN intends to look at Whole Label Rules toward the end of its Arabic Generation Panel work later this year. Thus, the task force has not reviewed the file MSR-1-Repertoire+WLE-Rules-20140225.xml released along with MSR1 and reserves its comments on its contents at this time. We hope that IP will be able to incorporate feedback on this part of the work by TF-AIDN at a later stage.
7. TF-AIDN has the following specific comments on the Arabic script portion of the MSR1, in the context of its role as the Arabic Script Generation Panel.

Code Point Excluded by MSR1 (as proposed)	Code Point	Code Point, Name and Details	TF-AIDN Feedback on MSR1 Suggestion (OK = agree with MSR1; Revise = MSR1 suggestion for the code point should be changed for the reasons given)
0610	◌ْ	0610;ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM ;Mn;230;NSM;;;;;N;;;;;	OK
0611	◌ِ	0611;ARABIC SIGN ALAYHE ASSALLAM;Mn;230;NSM;;;;;N;;;;;	OK
0612	◌ِ	0612;ARABIC SIGN RAHMATULLAH ALAYHE;Mn;230;NSM;;;;;N;;;;;	OK
0613	◌ِ	0613;ARABIC SIGN RADI ALLAHOU ANHU;Mn;230;NSM;;;;;N;;;;;	OK
0614	◌ِ	0614;ARABIC SIGN TAKHALLUS;Mn;230;NSM;;;;;N;;;;;	OK
0615	◌ِ	0615;ARABIC SMALL HIGH TAH;Mn;230;NSM;;;;;N;;;;;	OK
0616	◌ِ	0616;ARABIC SMALL HIGH LIGATURE ALEF WITH LAM WITH YEH;Mn;230;NSM;;;;;N;;;;;	OK
0617	◌ِ	0617;ARABIC SMALL HIGH ZAIN;Mn;230;NSM;;;;;N;;;;;	OK
0618	◌ِ	0618;ARABIC SMALL FATHA;Mn;30;NSM;;;;;N;;;;;	OK
0619	◌ِ	0619;ARABIC SMALL DAMMA;Mn;31;NSM;;;;;N;;;;;	OK
061A	◌ِ	061A;ARABIC SMALL KASRA;Mn;32;NSM;;;;;N;;;;;	OK
063B	◌ِ	063B;ARABIC LETTER KEHEH WITH TWO DOTS ABOVE;Lo;0;AL;;;;;N;;;;;	OK



AR-RUH AL-AMIN *Kalimat Membaca & Mengetik Jari dalam masa 30 Jam*

NAMA HURUF	TUNGGAL	AWAL KATA	TENGAH KATA	AKHIR KATA
YA	ي	ي	ي	ي
C	چ	چ	چ	چ
G	گ	گ	گ	گ
NG	ع	ع	ع	ع
NY	ن	ن	ن	ن
TAMARBUTAH	ة	-	-	ة
HAMZAH	ء	ء	ء	ء
V	و	و	و	و

Figure 1: A book teaching Jawi in Malaysia published in 2006

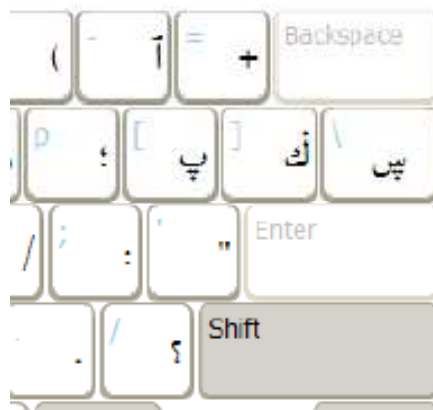


Figure2: Harari Keyboard developed in 2009.

© State Library of Victoria and the Australian Say Harari Association

However, the IDN table submitted by MyNIC in 2009 (see [https://www.iana.org/domains/idn-tables/tables/my\\_ms-my\\_1.0.pdf](https://www.iana.org/domains/idn-tables/tables/my_ms-my_1.0.pdf)) and the Malaysian Standard for Jawi (MS2443, 2012) does not include U+06AC but U+0762. Kew (2005) suggests that both variations are in use, though the latter should be preferred.

06D6	06D6;ARABIC SMALL HIGH LIGATURE	OK
------	---------------------------------	----



06EA	◌◌	06EA;ARABIC EMPTY CENTRE LOW STOP;Mn;220;NSM;;;;;N;;;;;	OK
06EB	◌◌◌	06EB;ARABIC EMPTY CENTRE HIGH STOP;Mn;230;NSM;;;;;N;;;;;	OK
06EC	◌◌◌◌	06EC;ARABIC ROUNDED HIGH STOP WITH FILLED CENTRE;Mn;230;NSM;;;;;N;;;;;	OK
06ED	◌◌◌◌◌	06ED;ARABIC SMALL LOW MEEM;Mn;220;NSM;;;;;N;;;;;	OK
06F2	٢	06F2;EXTENDED ARABIC-INDIC DIGIT TWO;Nd;0;EN;;2;2;N;EASTERN ARABIC-INDIC DIGIT TWO;;;;;	Revise; inclusion should be limited to only the case below, through Whole Label Rules and supporting documentation. General use of the code point should not be allowed.  Used in Jawi for re-duplication

Jawi uses the code point U+06F2 to indicate reduplication in the language. The plural is formed through this reduplication process, as are some other words. See [https://www.iana.org/domains/idn-tables/tables/my\\_ms-my\\_1.0.pdf](https://www.iana.org/domains/idn-tables/tables/my_ms-my_1.0.pdf) for some additional explanation. This is a productive process and no alternate way exists in representing this mechanism in Jawi.

According to the Malaysian Jawi Standard (MS2443, 2012; pg. 5):

**“5.2 Coded character set for Extended Arabic-indic digit two ٢**

This standard also specifies the extended Arabic-Indic digit two ٢. The Arabic-indic digit two is used in Jawi spelling system to indicate repetition of a Jawi word. It is an important part of Jawi. For instance, the word rumah-rumah is written in Jawi as رومه رومه, not رومه-رومه.”

Some other examples of the usage of Arabic-indic numeral ٢ in writing Jawi words include:

- ٢ راما Rama-rama (meaning in English: Butterfly)
- ٢ فلاجر Pelajar-pelajar (meaning in English: Students)
- ٢ رومه Rumah-rumah (meaning in English: Houses)

As discussed in this standard, inclusion should be limited to only this case, through Whole Label Rules and supporting documentation. General use of the code point should not be allowed.

06FD	۽	06FD;ARABIC SIGN SINDHI AMPERSAND;So;0;AL;;;;;N;;;;;	Revise  Actively used in Sindhi to mean “and”
06FE	۽	06FE;ARABIC SIGN SINDHI POSTPOSITION MEN;So;0;AL;;;;;N;;;;;	Revise  Actively used in Sindhi to mean “in” or “into”

Sindhi uses ۽ for the conjunction “and” and ۽ for the post-position “in” or “into”. Both words are frequently used in Sindhi and no alternate spellings exist in writing them in Sindhi.

As an example, introductory page of a recently published dictionary of Sindhi by Sindhi Language Authority is given below (Baloch 2006). Frequent use of both characters can be observed.

ڊاڪٽر نبي بخش خان بلوچ جي تيار ڪيل ۽ سنڌي ادبي بورڊ جي ڇپايل پنجن جلدن تي ٻڌل جامع سنڌي لغات سنڌي ٻوليءَ جي لفظن کي سهيڙڻ ۽ ساڃهڻ (Lexicography) جي حوالي سان نهايت وڏو علمي ڪم آهي. پر مختلف علمي ۽ ادبي حلقن مان مختلف لفظن جي داخلا ۽ انهن جي معنائن تي وقت به وقت اعتراض به سامهون پئي آيا آهن. خود ڊاڪٽر نبي بخش بلوچ صاحب لکت توڙي علمي گڏجاڻين ۾ ان طرف ڌيان ڇڪائيندا رهيا آهن ته، پهرئين جلد کان سواءِ ٻيو مواد جيڪو ادبي بورڊ وارن چئن جلدن ۾ ڇپيو آهي، ان ۾ هڪيترائي نقص آهن. ڊاڪٽر بلوچ صاحب ان مواد تي وڌيڪ ڪم ڪرڻ چاهيو پئي. پر نه ڄاڻ ڪهڙن سببن ڪري ابتدائي مسودي کي ئي ڇاپيو ويو. جڏهن ته ڪجهه جلدن تي نظر ثاني ڪندڙن جا نالا به ڏنل آهن. پر ڊاڪٽر بلوچ صاحب مطابق انهن تي نظر ثاني ٿيل ئي ڪو نه آهي.

اهڙي معلومات ظاهر ٿيڻ کانپوءِ ضروري هو ته باريڪ بينيءَ سان ڪم ڪندي سنڌي ٻوليءَ جي جامع لغت نشين سر تيار ڪري شايع ڪرائي وڃي. يقيناً سنڌ جي عالمن مان ڊاڪٽر نبي بخش خان بلوچ ئي آهي جيڪو پنهنجي اڳ تيار ڪيل لغت جو ٻيهر باريڪ بينيءَ سان جائزو وٺي سنڌا (ا ۽ وا) را ڪري تيار ڪري پئي سگهيو. ٻئي طرف اڳ ڇپيل لغات پراڻي طريقي تائين سبت تي ڪمپوز ڪري ڇپايل هئي. هيٺ ڪمپيوٽر جي استعمال ڪري ضروري هو ته جديد سهوليتن، سنڌي ٻوليءَ جي يونيڪوڊ جو حصول ۽ ا ب وار سارنگ جي ڪاميابيءَ کي استعمال ڪندي هيءَ لغت نه صرف ڪتابي صورت ۾ برسي ڏي (CD) تي پڻ مهيا ڪئي وڃي. اهڙي تقاضا مختلف علمي ادبي حلقن طرفان وقت به وقت ٿيندي رهي آهي ته اهو ڪم سنڌي لئنگويج اٿارٽي ڪري. ان صورتحال کي نظر ۾ رکندي، هن علمي ڪم کي بهتر شڪل ۾ رهندڙن تائين پهچائڻ لاءِ اسان ڊاڪٽر بلوچ صاحب جن کي گذارش ڪئي ته لغت کي نشين سر سڌاري ۽ وا ارا ڪري اداري کي ڏين ته جيئن سنڌي ٻوليءَ جي ترقي ۽ استعمال لاءِ هي عام ماڻهن تائين پهچائي سگهجي.

جيئن ته لغت نويسي مستقل ۽ ڪڏهن به مڪمل نه ٿيندڙ عمل آهي، جو ٻوليون هڪ ٻئي سان رابطو ۽ سماجي تبديلين هيٺ وڌنديون، ويجهنديون، هڪٻئي جا لفظ ۽ اصطلاح وٺنديون. ترقي ڪنديون رهنديون آهن. ان ڪري ضروري آهي ته گذرندڙ وقت ۽ سماجي حالتن جي تبديليءَ ڪري ٻوليءَ تي ٿيندڙ اثرن کي محفوظ رکڻ لاءِ لغتن تي

Figure 3: Frequent usage of letters ۽ and ڳ in normal Sindhi Text

077E	077E;ARABIC LETTER SEEN WITH INVERTED V; Lo;0;AL;;;;;N;;;;;	OK
077F	077F;ARABIC LETTER KAF WITH TWO DOTS ABOVE;Lo;0;AL;;;;;N;;;;;	OK
08F0	08F0;ARABIC OPEN FATHATAN;Mn;27;NSM;;;;;N;;;;;	OK
08F1	08F1;ARABIC OPEN DAMMATAN;Mn;28;NSM;;;;;N;;;;;	OK
08F2	08F2;ARABIC OPEN KASRATAN;Mn;29;NSM;;;;;N;;;;;	OK



08F3	08F3;ARABIC SMALL HIGH WAW;Mn;230;NSM;;;;;N;;;;	OK
------	--	----

## References

Baloch, N. A (2006). *Hik Juldi Sindhi Lughat*. National Language Authority, Hyderabad, Pakistan.

Banafunzi, Bana & Vianello, Alessandra (2014). "Chimi:ni in Arabic Script. Examples from Brava Poetry." In Mumin, Meikal & Versteegh, Kees (eds), *The Arabic Script in Africa: Studies in the Use of a Writing System*, pp. 293-309. Leiden, Boston: Brill.

Kew, Jonathan (2005). "Notes on some Unicode Arabic characters: recommendations for usage."

MS2443 (2012). "Information technology - Jawi coded character set for information interchange," Department of Standards, Ministry of Science Technology and Innovation, Malaysia.