

Proposal to encode three Arabic characters for Arwi

Roozbeh Pournader, Google (roozbeh@google.com)

August 19, 2013

Requested action










I would like to ask the WG2 to encode the following three Arabic characters, needed for Arwi, also known as Arabo-Tamil. The characters have already been accepted for encoding by the Unicode Technical Committee:

U+08B3 ARABIC LETTER AIN WITH THREE DOTS BELOW

U+08B4 ARABIC LETTER KAF WITH DOT BELOW

U+08E3 ARABIC TURNED DAMMA BELOW

Glyph shapes

code	isolated	final	medial	initial
08B3				
08B4				
08E3				

Properties

The properties for UnicodeData.txt and ArabicShaping.txt will be as follows. Other properties will be the same as existing similar letters and combining marks already encoded in the Arabic blocks.

UnicodeData.txt

```
08B3;ARABIC LETTER AIN WITH THREE DOTS BELOW;Lo;0;AL;;;;;N;;;;;
08B4;ARABIC LETTER KAF WITH DOT BELOW;Lo;0;AL;;;;;N;;;;;
08E3;ARABIC TURNED DAMMA BELOW;Mn;220;NSM;;;;;N;;;;;
```

ArabicShaping.txt

```
08B3; AIN WITH 3 DOTS BELOW; D; AIN
08B4; KAF WITH DOT BELOW; D; KAF
```

Background

There was an original proposal by Seyed Mohamed Buhari, who had presented his work at the Unicode Conference in 2008. The UTC asked for an updated proposal asking for the glyphs in a reviewable format, but no new proposal appeared. The original proposal, including samples, is available at [L2/09-143](#). The turned damma was called “inverted” damma in the original proposal, but “turned” is a better description of the transformation applied to the shape, and is consistent with more Unicode character names.

There are other potential character candidates for Arwi, a merged *shadda-fatha* (which may be just a ligature) and a *sad* with dot below (which I haven’t seen outside of alphabet charts. Tschacher talks about it as a letter that wasn’t used beyond the texts that introduced it). I am not proposing them until more information is found or provided.

Bibliography

1. Buhari, Seyed Mohamed. 2009. “Additional Arabic letters needed for Arwi Script.” *UTC Document Registry*, L2/09-143.
<http://www.unicode.org/L2/L2009/09143-arwi-proposal.pdf>.
2. Wikipedia contributors. “Arwi.” *Wikipedia, The Free Encyclopedia*, <https://en.wikipedia.org/w/index.php?title=Arwi&oldid=557665630> (accessed June 24, 2013).
3. Tschacher, Torsten. “Arwi (Arabic-Tamil) — An Introduction.”
<http://www.suedasien.uni-halle.de/SAWA/Tschacher.pdf>.

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

A. Administrative

1. Title: **Three Arabic characters for Arwi**
2. Requester's name: **Roozbeh Pournader**
3. Requester Type (Member body/Liaison/Individual contribution): **Individual Contribution**
4. Submission date: **June 24, 2013**
5. Requester's reference, if applicable: **N/A**
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): **No**
 Proposed name of script: **N/A**
 - b. The proposal is for addition of character(s) to an existing block: **Yes**
 Name of existing block: **Arabic Extended-A**
2. Number of characters in proposal: **Three (3)**
3. Proposed category: **A-Contemporary**
4. Is a repertoire including character names provided? **Yes**
 - a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document? **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? **Roozbeh Pournader**
 - b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): **N/A**
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. In L2/09-143.**
7. Special encoding issues:
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please endorse information)? **No**

8. Additional information:
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> UTH for such information on other scripts. Also see Unicode Character Database (<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? **Yes**
If YES explain: **It was abandoned by the original author, probably due to too much bureaucracy involved.**
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.)? **Yes**
If YES, with whom? **See original proposal**
If YES, available relevant documents: **L2/09-143**
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? **No**
Reference: **N/A**
4. The context of use for the proposed characters (type of use; common or rare): **Common in Arwi**
Reference: **N/A**
5. Are the proposed characters in current use by the user community? **Yes**
If YES, where? Reference: **Among some speakers of Arwi/Tamil. See original proposal.**
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? **Yes**
If YES, is a rationale provided? **Yes. Need to be next to similar characters**
If YES, reference: **N/A**
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? **No**
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? **No**
If YES, is a rationale for its inclusion provided? **N/A**
If YES, reference: **N/A**
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? **No**
If YES, is a rationale for its inclusion provided? **N/A**
If YES, reference: **N/A**
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? **Yes**
If YES, is a rationale for its inclusion provided? **Yes. The proposed characters have different identities.**
If YES, reference: **See above**
11. Does the proposal include use of combining characters and/or use of composite sequences?

Yes

If YES, is a rationale for such use provided? **N/A**

If YES, reference: **N/A**

Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? **N/A**

If YES, reference: **N/A**

12. Does the proposal contain characters with any special properties such as control function or similar semantics? **No**

If YES, describe in detail (include attachment if necessary): **N/A**

13. Does the proposal contain any Ideographic compatibility characters? **No**

If YES, are the equivalent corresponding unified ideographic characters identified? **N/A**

If YES, reference: **N/A**