

Title: Proposal to encode Arabic Alef with Right Hamza used in Quran published in Iran



Action: For consideration by UTC and ISO/IEC JTC1/SC2/WG2

Author: Lateef Sagar Shaikh

Date: 19-Dec-2021

Introduction

Quran published in Iran uses a special Alef which has Hamza on the right side. Initially it looked like



U+0675 but its Hamza is high in the final shape  whereas this Alef has Hamza in the middle in its final shape 

Same text also uses regular Alef with Hamza above (Figure 1) which suggests that the Alef with Right Hamza should be encoded separately. Similar combinations with Yeh and Waw are also used but U+08A8 and U+0676 can be used without any issue.







Requests

The author requests the encoding of Arabic Alef with Right Hamza

Character name and shape

Isolated Shape	Final Shape	Code Point	Name
		U+088F	ARABIC ALEF WITH RIGHT HAMZA

Yeh and Waw with Hamza which are already encoded

Code Point	Isolated Shape	Final Shape	Medial Shape	Initial Shape
U+0676				
U+08A8 (See Reference 1 for shapes)				

Properties and Arabic Shaping

The main character properties should be as follows:

088F;ARABIC LETTER ALEF WITH RIGHT HAMZA;Lo;0;AL;;;;N;;;;

The main character shaping should be as follows:

088F; ALEF WITH RIGHT HAMZA; R; ALEF

Figures and comparison



٦

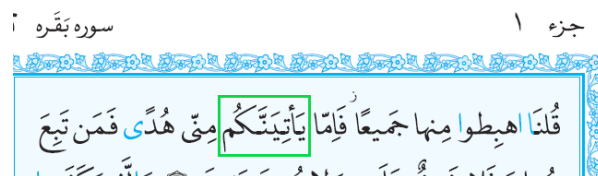


Figure 1 - Here we can see that Alef with Hamza Above and this special Alef both are used in the same text. As you can see that the final shape of this Alef has hamza lower than that of U+0675

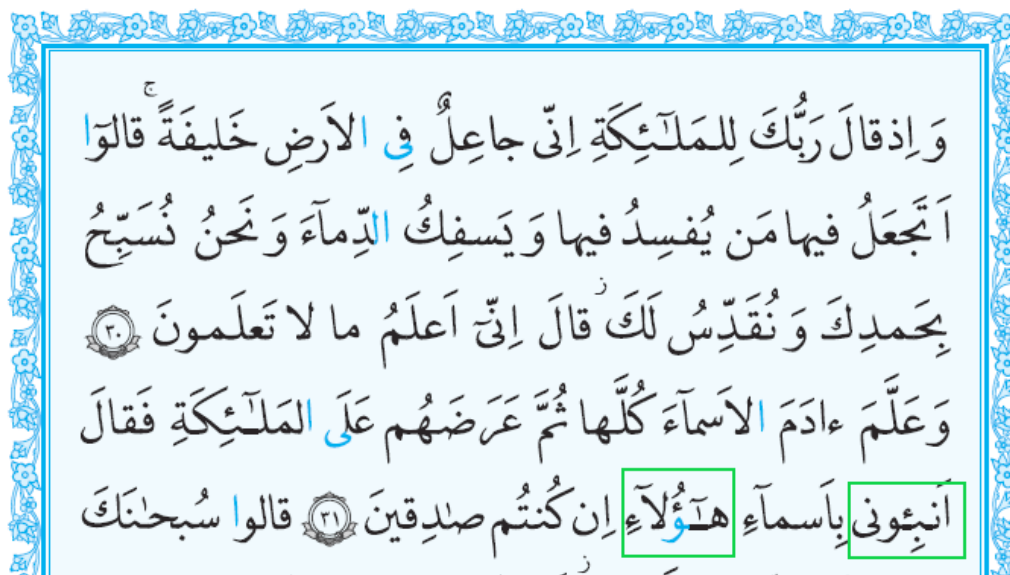


Figure 2 – In this figure I have highlighted High Hamza Waw (U+0676) and Waw with hamza above (U+0624)

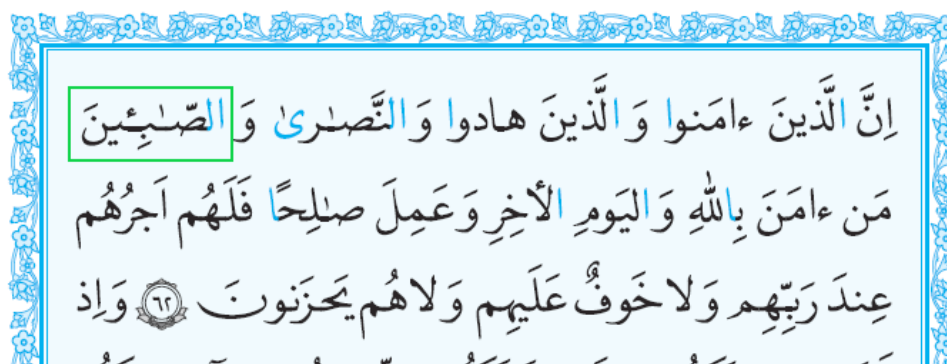


Figure 3 - In this figure you can see an example of Yeh With Two Dots Below And Hamza Above (U+08A8)

Reference

1. Proposal to add Arabic script characters for African and Asian languages, by Lorna A. Priest, Martin Hosken (SIL International) - 12 August 2010

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3882.pdf>

2. Holy Quran published in Iran – Mushaf Jamhori Islami Iran Bi Shewa Naskh Arabi (Kam Alamat)

http://moshaf.org/page/IranMoshaf_Arabi

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

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.

A. Administrative

1. Title:	Proposal to encode Arabic Alef with Right Hamza used in Quran published in Iran		
2. Requester's name:	Lateef Sagar Shaikh		
3. Requester type (Member body/Liaison/Individual contribution):	Individual Contribution		
4. Submission date:	12/19/2021		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:	<input checked="" type="checkbox"/>		
(or) More information will be provided later:	<input type="checkbox"/>		

B. Technical – General

1. Choose one of the following:			
a. This proposal is for a new script (set of characters):	<input type="checkbox"/>		
Proposed name of script:			
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/>		
Name of the existing block:	Arabic Extended-B		
2. Number of characters in proposal:	1		
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input checked="" type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>	
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>	
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>		
4. Is a repertoire including character names provided?	<input checked="" type="checkbox"/> Yes		
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/> Yes		
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/> Yes		
5. Fonts related:			
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	Lateef Sagar Shaikh		
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	Lateef Sagar Shaikh lateef_sagar@yahoo.com		
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> Yes		
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> Yes		
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 enclose information)?	<input checked="" type="checkbox"/> Yes		
	See the section titled Character name and shape		

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> 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.

¹ 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)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	No
If YES explain	
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.)?	No
If YES, with whom?	
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	Yes
Reference: <i>Muslims in Pakistan, There are around 18 Million Muslims in Pakistan</i>	
4. The context of use for the proposed characters (type of use; common or rare)	Quranic
Reference: <i>The proposed mark is very common in Quran published in Pakistan</i>	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference: <i>Most of recent published Qurans use this character</i>	
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?	
If YES, reference: <i>Character could be kept with their counterparts in Arabic block</i>	
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?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
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?	
If YES, reference:	
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?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	Combining
If YES, is a rationale for such use provided?	
If YES, reference: <i>The character is combining, as can be seen in samples</i>	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	No
If YES, reference:	
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)	
13. Does the proposal contain any Ideographic compatibility characters?	No
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	