

Doc Type: Working Group Document  
Title: Proposal to Encode 22 Characters for Arabic Pedagogical Use  
Source: Attash Durrani (Pakistan NLA), Kamal Mansour, Rick McGowan  
Status: Individual Contribution  
Action: For National Body Review and Feedback  
Date: April 18, 2008

References: L2/06-039 L2/06-240 L2/07-094 L2/07-174  
<http://www.unicode.org/L2/L2006/06039-urdu-arabic-prelim.pdf>  
<http://www.unicode.org/L2/L2006/06240-durrani.pdf>  
<http://www.unicode.org/L2/L2007/07094-regarding-nuqta.pdf>  
<http://www.unicode.org/L2/L2007/07174-nuqta-disc.pdf>

There is a need to encode dot patterns used in Arabic letters for pedagogical and other purposes, as exemplified in the original proposal L2/06-039 and related documents.

That proposal requested 22 characters to be encoded within the 0600 block, but there is a better open area of right-to-left codepoints tentatively reserved for Arabic additions in the range from 0880 through 08BF. This range is probably the best place to put a set of spacing dot pattern characters.

A good deal of required information and examples are contained in the previous proposal L2/06-039, as well as discussion documents L2/06-240, L2/07-094 and L2/07-174. That information will not be re-iterated here.

The proposed 22 characters, their glyphs, names, and annotations are shown on the accompanying chart and nameslist.

These would be given characters properties as follows (in the syntax of the UnicodeData.txt file):

**08nn ; ---NAME--- ; Sk ; 0 ; AL ; ; ; ; N ; ; ; ;**

The line-breaking behavior of these characters would be similar to ordinary Arabic letters or symbols. They should also be allowed as parts of identifiers.

	x88	x89
0	◊ xx80	= xx90
1	◊ xx81	= xx91
2	◈ xx82	◦ xx92
3	◈ xx83	◦ xx93
4	◈ xx84	◌ xx94
5	◈ xx85	◌ xx95
6	◈ xx86	
7	◈ xx87	
8	◈ xx88	
9	◈ xx89	
A	 xx8A	
B	 xx8B	
C	‡ xx8C	
D	‡ xx8D	
E	— xx8E	
F	— xx8F	

0880	◊	ARABIC SINGLE NUQTA ABOVE
0881	◊	ARABIC SINGLE NUQTA BELOW
0882	◈	ARABIC DOUBLE NUQTA ABOVE
0883	◈	ARABIC DOUBLE NUQTA BELOW
0884	◈	ARABIC TRIPLE NUQTA ABOVE
0885	◈	ARABIC TRIPLE NUQTA BELOW
0886	◈	ARABIC TRIPLE INVERTED NUQTA ABOVE
0887	◈	ARABIC TRIPLE INVERTED NUQTA BELOW
0888	◈	ARABIC QUADRUPLE NUQTA ABOVE • Sindhi
0889	◈	ARABIC QUADRUPLE NUQTA BELOW • Sindhi
088A		ARABIC DOUBLE DANDA ABOVE • Sindhi
088B		ARABIC DOUBLE DANDA BELOW • Sindhi
088C	‡	ARABIC DOUBLE NUQTA VERTICAL ABOVE • Sindhi
088D	‡	ARABIC DOUBLE NUQTA VERTICAL BELOW • Sindhi
088E	—	ARABIC SINGLE KASHIDA ABOVE • Urdu
088F	—	ARABIC SINGLE KASHIDA BELOW • Urdu
0890	=	ARABIC DOUBLE KASHIDA ABOVE • Urdu
0891	=	ARABIC DOUBLE KASHIDA BELOW • Urdu
0892	◦	ARABIC SINGLE CIRCLE ABOVE • Pashto
0893	◦	ARABIC SINGLE CIRCLE BELOW • Pashto
0894	◌	ARABIC TOTA ABOVE • Urdu
0895	◌	ARABIC TOTA BELOW • Urdu

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

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

Please read Principles and Procedures Document (P & P) from <http://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

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

**A. Administrative**

1. Title: **Proposal to encode 22 characters for Arabic pedagogical use**

2. Requester's name: **Dr. Attash Durrani (NLA, Pakistan) ; Kamal Mansour ; Rick McGowan**

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

4. Submission date: **April 20, 2008**

5. Requester's reference (if applicable): **L2/06-039 L2/06-240 L2/07-094 L2/07-174**

6. Choose one of the following:

This is a complete proposal: **YES**

(or) More information will be provided later: \_\_\_\_\_

**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)**

2. Number of characters in proposal: **22**

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

A-Contemporary	<b>A</b>	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" in Annex L of P&P document? **YES**

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

5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? **Kamal Mansour (Monotype Imaging)**

If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: **TrueType font – used to create this proposal – Kamal Mansour**

6. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? **See L2/06-039**

b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? **See L2/06-039, and othe references above**

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)? **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> for such information on other scripts. Also see <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

<sup>1</sup> Form number: N3102-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)

### C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	YES
If YES explain	<i>This is a significant revision of earlier proposals submitted to UTC. See references above.</i>
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?	<i>National Language Authority, Pakistan</i>
If YES, available relevant documents:	<i>See references above.</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	NO
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	<i>Common</i>
Reference:	
5. Are the proposed characters in current use by the user community?	YES
If YES, where? Reference:	<i>Pedagogical materials</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?	NO
If YES, reference:	
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 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?	NO
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?	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 character(s)?	NO
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	