Document L2/98-274

Title: Proposal to add 25 Arabic characters to the BMP
Source: Mark Davis & Kamal Mansour
Status: Expert Contribution
Action: For consideration by JTC1/SC2/WG2

This document contains the proposal summary and a full proposal for the encoding of the new characters in the BMP of ISO/IEC 10646.

A. Administrative

<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Proposal to add 25 Arabic characters to the BMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>Requester's names</td>
<td>Mark Davis, Kamal Mansour</td>
</tr>
<tr>
<td>3.</td>
<td>Requester type</td>
<td>Expert contribution</td>
</tr>
<tr>
<td>4.</td>
<td>Submission date</td>
<td>1998-12-01</td>
</tr>
<tr>
<td>5.</td>
<td>Requester's reference</td>
<td></td>
</tr>
<tr>
<td>6a.</td>
<td>Completion</td>
<td>This is a complete proposal.</td>
</tr>
<tr>
<td>6b.</td>
<td>More information to be provided?</td>
<td>No</td>
</tr>
</tbody>
</table>
List of Proposed Characters

Part 1

1. Dotless Beh:  ُ

   Usage
   a) Ancient Arabic
      This letter was used to represent BEH, TEH, THEH, and NUN without distinction.

2. Dotless Feh-Qaf:  ﻕ

   Usage
   a) Ancient Arabic
      This letter was used to represent, without distinction, both FEH and QAF in initial and medial positions. In most cases, the final and separate shapes of FEH and QAF were distinct.

3. Dotless Qaf:  ق

   Usage
   a) Ancient Arabic
      The final shape of DOTLESS QAF was most often distinguishable from that of DOTLESS FEH.

   b) Minority languages
      This character can be used for minority languages which combine QAF with non-standard diacritics.
Part 2

For each of the following non-spacing characters, examples are given of usage by Arabic and non-Arabic languages. In each case, at least one graphic example is shown.

1. Single dot above:  ُ
   Languages: Arabic, Jawi, Moplah, ...
   Examples in Unicode: U+0696, 069A, 06A7, 06AC, 06B6, 06CF
   Graphic sample: ُ U+06CF (Jawi)

2. Single dot below: ِ
   Languages: Arabic, Sindhi, Pashto, Kurdish
   Examples in Unicode: U+068A, 068B, 0694, 0696, 069A, 06A3
   Graphic sample: ِ U+068A (Sindhi)

3. Two dots above: ْ
   Languages: Arabic, Sindhi, Dargwa, Kurdish, Judeo-Arabic
   Examples in Unicode: U+0697, 06CA
   Graphic sample: ْ U+068C (Sindhi); ْ (Judeo-Arabic)

4. Two dots below: ْ
   Languages: Arabic, Sindhi, Turkic
   Examples in Unicode: U+0683, 069D, 06B2
   Graphic sample: ْ U+0683 (Sindhi)

5. Two vertical dots above: ٍ
   Languages: Pashto
   Examples in Unicode: U+0682
   Graphic sample: ٍ U+0682 (Pashto)
6. Two vertical dots below: Ḿ
Languages: Sindhi, Pashto, Uighur
Examples in Unicode: U+067B, 06B3, 06D0
Graphic sample: Ḽ U+067B (Sindhi)

7. Three dots above: Ṣ
Languages: Arabic, Jawi, Judeo-Arabic
Examples in Unicode: U+067D, 0685, 068F, 069C, 069E, 069F, 06A0, 06A8, 06B4, 06B7, 06BD
Graphic sample: Ṣ U+06A0 (Jawi), Ṣ (Judeo-Arabic)

8. Three dots below: ṫ
Languages: Arabic, Farsi, Pashto, Urdu, Uighur
Examples in Unicode: U+0686, 069B, 069C, 067E, 06A5, 06AE, 06D1
Graphic sample: ṫ U+067E (Farsi et al.)

9. Three dots reversed above: ṫ
Languages: Sindhi
Examples in Unicode: U+067D, 068F
Graphic sample: ṫ U+067D, ṫ U+068F (Sindhi)

10. Four dots above: ṫ
Languages: Sindhi
Examples in Unicode: U+067F, 0690, 0699, 06A6
Graphic sample: ṫ U+067F (Sindhi)
11. Four dots below:  
*Languages*: Sindhi  
*Examples in Unicode*: U+0680, 0687  
*Graphic sample*: Ṣ U+0680 (Sindhi)

12. Wavy Hamza above:  
*Languages*: Baluchi, Kashmiri  
*Examples in Unicode*: U+0672  
*Graphic sample*: ṭ U+0672

13. Wavy Hamza below:  
*Languages*: Baluchi, Kashmiri  
*Examples in Unicode*: U+0673  
*Graphic sample*: ṭ U+0673

14. Small Tah above:  
*Languages*: Urdu, Lahnda  
*Examples in Unicode*: U+0679, 0688, 068B, 0691  
*Graphic sample*: ṭ U+0688 (Urdu)

15. Ring above:  
*Languages*: Pashto, Lahnda, Kashmiri  
*Examples in Unicode*: U+06AB, 06B0, 06C4  
*Graphic sample*: ḫ U+06AB (Pashto)

16. Ring below:  
*Languages*: Pashto  
*Examples in Unicode*: U+067C, 0689, 0693, 06BC  
*Graphic sample*: Ḩ U+0689
17. Seven-like sign above: ☋
   Languages: Kazakh, Kurdish, Uighur
   Examples in Unicode: U+0692, 06B5, 06C6, 06CE
   Graphic sample: ŋ U+0692 (Kurdish)

18. Seven-like sign below: ☋
   Languages: Kurdish
   Examples in Unicode: U+0695
   Graphic sample: ŋ U+0695

19. Small separate Heh above: й
   Languages: Kashmiri
   Examples in Unicode: none
   Graphic sample: й

20. Wasla above: ڇ
    Languages: Arabic
    Examples in Unicode: U+0671
    Graphic sample: ڇ [U+0671]

21. Circumflex-like sign above: ⃞
    Languages: Kazakh, Kirghiz, Urdu (functionally equiv to U+0652)
    Examples in Unicode: U+06C9
    Graphic sample: ŋ U+06C9 (Kazakh, Kirghiz)

22. Breve-like sign above: ڋ
    Languages: Kashmiri
    Examples in Unicode: none
    Graphic sample: ڋ
[23. Madda above: overrides] pending allocation at U+0653
[24. Hamza above:  regulates] pending allocation at U+0654
[25. Hamza below:  regulates] pending allocation at U+0655
<table>
<thead>
<tr>
<th></th>
<th>ARABIC SINGLE DOT ABOVE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARABIC SINGLE DOT BELOW</td>
</tr>
<tr>
<td></td>
<td>ARABIC TWO DOTS ABOVE</td>
</tr>
<tr>
<td></td>
<td>ARABIC TWO DOTS BELOW</td>
</tr>
<tr>
<td></td>
<td>ARABIC TWO VERTICAL DOTS ABOVE</td>
</tr>
<tr>
<td></td>
<td>ARABIC TWO VERTICAL DOTS BELOW</td>
</tr>
<tr>
<td></td>
<td>ARABIC THREE DOTS ABOVE</td>
</tr>
<tr>
<td></td>
<td>ARABIC MADD ABOVE</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td>ARABIC HAMZA ABOVE</td>
</tr>
<tr>
<td></td>
<td>ARABIC HAMZA BELOW</td>
</tr>
<tr>
<td></td>
<td>ARABIC WAVY HAMZA ABOVE</td>
</tr>
<tr>
<td></td>
<td>ARABIC WAVY HAMZA BELOW</td>
</tr>
<tr>
<td></td>
<td>ARABIC SMALL TAH ABOVE</td>
</tr>
<tr>
<td></td>
<td>ARABIC RING ABOVE</td>
</tr>
<tr>
<td>Arabic</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>⦰⦱⦲⦳</td>
<td>Arabic three dots below</td>
</tr>
<tr>
<td>⦰⦱⦲⦳</td>
<td>Arabic three dots reversed above</td>
</tr>
<tr>
<td>⦰⦱⦲⦳</td>
<td>Arabic four dots above</td>
</tr>
<tr>
<td>⦰⦱⦲⦳</td>
<td>Arabic four dots below</td>
</tr>
<tr>
<td>⦰⦱⦲⦳</td>
<td>Arabic letter dotless beh</td>
</tr>
<tr>
<td>⦰⦱⦲⦳</td>
<td>Arabic letter dotless feh-qaf</td>
</tr>
<tr>
<td>⦰⦱⦲⦳</td>
<td>Arabic letter dotless qaf</td>
</tr>
<tr>
<td>ARABIC RING BELOW</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td></td>
</tr>
<tr>
<td>ARABIC SEVEN-LIKE SIGN ABOVE</td>
<td></td>
</tr>
<tr>
<td>ARABIC SEVEN-LIKE SIGN BELOW</td>
<td></td>
</tr>
<tr>
<td>ARABIC SMALL SEPARATE HEH ABOVE</td>
<td></td>
</tr>
<tr>
<td>ARABIC WASLA ABOVE</td>
<td></td>
</tr>
<tr>
<td>ARABIC REVERSE-BREVE-LIKE SIGN ABOVE</td>
<td></td>
</tr>
<tr>
<td>ARABIC BREVE-LIKE SIGN ABOVE</td>
<td></td>
</tr>
</tbody>
</table>
### B. Technical -- General

<table>
<thead>
<tr>
<th></th>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>New script? Name?</td>
<td>No</td>
</tr>
<tr>
<td>1b</td>
<td>Addition of characters to existing block? Name?</td>
<td>Yes, to Arabic. Suggested locations are U+0610-061A, U+063D-063F, U+0653-065E, U+066E, U+066F.</td>
</tr>
<tr>
<td>2</td>
<td>Number of characters</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>Proposed category</td>
<td>Category A</td>
</tr>
<tr>
<td>4</td>
<td>Proposed level of implementation and rationale</td>
<td>Level 1</td>
</tr>
<tr>
<td>5a</td>
<td>Character names included in proposal?</td>
<td>Yes</td>
</tr>
<tr>
<td>5b</td>
<td>Character names in accordance with guidelines?</td>
<td>Yes</td>
</tr>
<tr>
<td>5c</td>
<td>Character shapes reviewable?</td>
<td>Yes</td>
</tr>
<tr>
<td>6a</td>
<td>Who will provide computerized font?</td>
<td>Kamal Mansour (Monotype Typography)</td>
</tr>
<tr>
<td>6b</td>
<td>Font currently available?</td>
<td>Yes</td>
</tr>
<tr>
<td>6c</td>
<td>Font format?</td>
<td>TrueType</td>
</tr>
<tr>
<td>7a</td>
<td>Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?</td>
<td>Yes</td>
</tr>
<tr>
<td>7b</td>
<td>Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?</td>
<td>N/A--See below</td>
</tr>
<tr>
<td>8</td>
<td>Does the proposal address other aspects of character data processing?</td>
<td>No</td>
</tr>
</tbody>
</table>

### C. Technical -- Justification

<table>
<thead>
<tr>
<th></th>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Has this proposal been submitted before?</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>Contact with the user community?</td>
<td>N/A--See below</td>
</tr>
<tr>
<td>3</td>
<td>Information on the user community?</td>
<td>N/A--See below</td>
</tr>
</tbody>
</table>
APPENDIX A

Proposal To Amend Arabic Repertoire (L2/98-274)

The first part of this proposal aims to provide the characters necessary for the transcription of Ancient (undotted) Arabic text. Except for the first 3 new characters proposed in document, all other characters needed for Ancient Arabic are already in Unicode 2.1.

The second part of this proposal enumerates various non-spacing marks and diacritics that are necessary for either of the following purposes:

a) Pedagogical Use

In order to be able to refer to all constituent part of Arabic script, the various marks must be addressable as independent characters. This capability is particularly important in discussions about the evolution of Arabic script and its adaptation to non-Arabic languages.

b) Minority Languages

By taking a broad sampling of the extension of Arabic script to accommodate non-Arabic languages, it is readily observable that some marks and diacritics have been repeatedly used in a generative manner to create new characters. Based on that observation, it is reasonable to imagine that other minority languages, as well as scholarly use, could resort to new combinations that are not currently supported by Unicode.

No claim is made as to the completeness of the proposed list.

Of the 25 proposed characters listed below, the last 3 have already been passed by WG2.
U+0612;ARABIC TWO DOTS ABOVE (combining)
U+0613;ARABIC TWO DOTS BELOW (combining)
U+0614;ARABIC TWO VERTICAL DOTS ABOVE (combining)
U+0615;ARABIC TWO VERTICAL DOTS BELOW (combining)
U+0616;ARABIC THREE DOTS ABOVE (combining)
U+0617;ARABIC THREE DOTS BELOW (combining)
U+0618;ARABIC THREE DOTS REVERSED ABOVE (combining)
U+0619;ARABIC FOUR DOTS ABOVE (combining)
U+061A;ARABIC FOUR DOTS BELOW (combining)
U+063D;ARABIC LETTER DOTLESS BEH (base character)
U+063E;ARABIC LETTER DOTLESS FEH-QAF (base character)
U+063F;ARABIC LETTER DOTLESS QAF (base character)
U+0653;ARABIC MADDosphate ABOVE
U+0654;ARABIC HAMZA ABOVE
U+0655;ARABIC HAMZA BELOW
U+0656;ARABIC WAVY HAMZA ABOVE (combining)
U+0657;ARABIC WAVY HAMZA BELOW (combining)
U+0658;ARABIC SMALL TAH ABOVE (combining)
U+0659;ARABIC RING ABOVE (combining)
U+065A;ARABIC RING BELOW (combining)
U+065B;ARABIC SEVEN-LIKE SIGN ABOVE (combining)
U+065C;ARABIC SEVEN-LIKE SIGN BELOW (combining)
U+065D;ARABIC SMALL SEPARATE HEH ABOVE (combining)
U+065E;ARABIC WASLA ABOVE (combining)
U+066E;ARABIC REVERSE-BREVE-LIKE SIGN ABOVE (combining)
U+066F;ARABIC BREVE-LIKE SIGN ABOVE (combining)

(End of proposal)