ISO
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONAL DE NORMALISATION

ISO/IEC JTC 1/SC 2/WG 2

Universal Multiple-Octet Coded Character Set
(U C S)

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

1.	Title	Proposal to add 25 Arabic characters to the BMP
2.	Requester's names	Mark Davis, Kamal Mansour
3.	Requester type	Expert contribution
4.	Submission date	1998-12-01
5.	Requester's reference	
6a.	Completion	This is a complete proposal.
6b.	More information to be provided?	No:

List of Proposed Characters

Part 1

1. Dotless Beh: ب

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

2. Dotless Feh-Qaf: ڡ

   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: ٯ

   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: ڏ U+06CF (Jawi)

2. Single dot below: ◌̣

   Languages: Arabic, Sindhi, Pashto, Kurdish

   Examples in Unicode: U+068A, 068B, 0694, 0696, 069A, 06A3

   Graphic sample: ڊ U+068A (Sindhi)

3. Two dots above: ◌̈

   Languages: Arabic, Sindhi, Dargwa, Kurdish, Judeo-Arabic

   Examples in Unicode: U+0697, 06CA

   Graphic sample: ڌ U+068C (Sindhi); ה̈ (Judeo-Arabic)

4. Two dots below: ◌̤

   Languages: Arabic, Sindhi, Turkic

   Examples in Unicode: U+0683, 069D, 06B2

   Graphic sample: ڃ U+0683 (Sindhi)

5. Two vertical dots above: ◌̈

   Languages: Pashto

   Examples in Unicode: U+0682

   Graphic sample: ڂ U+0682 (Pashto)

4

6. Two vertical dots below: ؟

    Languages: Sindhi, Pashto, Uighur

    Examples in Unicode: U+067B, 06B3, 06D0

    Graphic sample: ٻ U+067B (Sindhi)

7. Three dots above: ؞

    Languages: Arabic, Jawi, Judeo-Arabic

    Examples in Unicode: U+067D, 0685, 068F, 069C, 069E, 069F, 06A0,
                          06A8, 06B4, 06B7, 06BD

    Graphic sample: ڠ U+06A0 (Jawi), ﬡ (Judeo-Arabic)

8. Three dots below: ؛

    Languages: Arabic, Farsi, Pashto, Urdu, Uighur

    Examples in Unicode: U+0686, 069B, 069C, 067E, 06A5, 06AE, 06D1

    Graphic sample: پ U+067E (Farsi *et al.*)

9. Three dots reversed above: ؞

    Languages: Sindhi

    Examples in Unicode: U+067D, 068F

    Graphic sample: ٽ U+067D, ڏ U+068F (Sindhi)

10. Four dots above: ؟

    Languages: Sindhi

    Examples in Unicode: U+067F, 0690, 0699, 06A6

    Graphic sample: ٿ U+067F (Sindhi)

11. Four dots below: ‎ؠ

   Languages: Sindhi

   Examples in Unicode: U+0680, 0687

   Graphic sample: ٻ   U+0680 (Sindhi)

12. Wavy Hamza above: ٴ

   Languages: Baluchi, Kashmiri

   Examples in Unicode: U+0672

   Graphic sample: ٲ U+0672

13. Wavy Hamza below: ‎

   Languages: Baluchi, Kashmiri

   Examples in Unicode: U+0673

   Graphic sample: ٳ U+0673

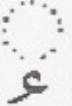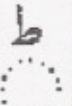14. Small Tah above: ‎

   Languages: Urdu, Lahnda

   Examples in Unicode: U+0679, 0688, 068B, 0691

   Graphic sample: ڈ U+0688 (Urdu)

15. Ring above: ‎

   Languages: Pashto, Lahnda, Kashmiri

   Examples in Unicode: U+06AB, 06B0, 06C4

   Graphic sample: ګ U+06AB (Pashto)

16. Ring below: ‎

   Languages: Pashto

   Examples in Unicode: U+067C, 0689, 0693, 06BC

   Graphic sample: ډ U+0689
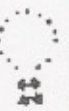
17. Seven-like sign above: ŏ

Languages: Kazakh, Kurdish, Uighur

Examples in Unicode: U+0692, 06B5, 06C6, 06CE

Graphic sample: ڒ U+0692 (Kurdish)

18. Seven-like sign below: ọ

Languages: Kurdish

Examples in Unicode: U+0695

Graphic sample: ڕ U+0695

19. Small separate Heh above: ǒ

Languages: Kashmiri

Examples in Unicode: none

Graphic sample: څ

20. Wasla above: ŏ

Languages: Arabic

Examples in Unicode: U+0671

Graphic sample: ٱ [U+0671]

21. Circumflex-like sign above: ŏ

Languages: Kazakh, Kirghiz, Urdu (functionally equiv to U+0652)

Examples in Unicode: U+06C9

Graphic sample: ۉ U+06C9 (Kazakh, Kirghiz)

22. Breve-like sign above: ŏ

Languages: Kashmiri

Examples in Unicode: none

Graphic sample: ځ
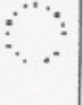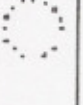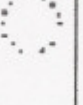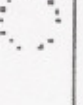
[23. Madda above: õ   ]

    pending allocation at U+0653

[24. Hamza above: ó   ]

    pending allocation at U+0654

[25. Hamza below: ọ   ]

    pending allocation at U+0655

	ARABIC SINGLE DOT ABOVE
	ARABIC SINGLE DOT BELOW
	ARABIC TWO DOTS ABOVE
	ARABIC TWO DOTS BELOW
	ARABIC TWO VERTICAL DOTS ABOVE
	ARABIC TWO VERTICAL DOTS BELOW
	ARABIC THREE DOTS ABOVE

◌ٓ	ARABIC MADDA ABOVE
◌ٔ	ARABIC HAMZA ABOVE
◌ٕ	ARABIC HAMZA BELOW
◌ۓ	ARABIC WAVY HAMZA ABOVE
◌	ARABIC WAVY HAMZA BELOW
◌ۤ	ARABIC SMALL TAH ABOVE
◌ۢ	ARABIC RING ABOVE

ۛ		ARABIC THREE DOTS BELOW
ۜ		ARABIC THREE DOTS REVERSED ABOVE
ۛ		ARABIC FOUR DOTS ABOVE
ۛ		ARABIC FOUR DOTS BELOW
ٮ		ARABIC LETTER DOTLESS BEH
ٯ		ARABIC LETTER DOTLESS FEH-QAF
ڡ		ARABIC LETTER DOTLESS QAF

ARABIC RING BELOW	
ARABIC SEVEN-LIKE SIGN ABOVE	
ARABIC SEVEN-LIKE SIGN BELOW	
ARABIC SMALL SEPARATE HEH ABOVE	
ARABIC WASLA ABOVE	
ARABIC REVERSE-BREVE-LIKE SIGN ABOVE	
ARABIC BREVE-LIKE SIGN ABOVE	

B. Technical -- General

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

C. Technical -- Justification

1.	Has this proposal been submitted before?	No
2.	Contact with the user community?	N/A--See below
3.	Information on the user community?	N/A--See below

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 MADDA 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)