

Universal Multiple-Octet Coded Character Set
 International Organization for Standardization
 Organisation internationale de normalisation
 Международная организация по стандартизации

Doc Type: Working Group Document**Title: Proposal to encode four Qur'anic Arabic characters in the UCS****Source: Michael Everson and Roozbeh Pournader****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2006-11-08**

Request. This document asks for the addition of four Qur'anic Arabic characters to the UCS. The characters are used in Qur'ans published in Iran, most importantly the new official Qur'ans published by the governmental Centre for Qur'an Printing and Publishing of the Islamic Republic of Iran (*Markaz-e Tab' va Našr-e Qor'ān-e Jomhuri-ye Eslāmi-ye Irān*).

If this proposal is adopted, the following characters would exist:

◌ِ	0617	ARABIC SMALL HIGH ZAIN	* Qur'anic use
◌َ	0618	ARABIC SMALL FATHA	* Qur'anic use
◌ُ	0619	ARABIC SMALL DAMMA	* Qur'anic use
◌ِ	061A	ARABIC SMALL KASRA	* Qur'anic use

The following combining classes are suggested. The smaller version of the *harakat* share the same properties of the normal versions.

```
0617;ARABIC SMALL HIGH ZAIN;Mn;230;NSM;;;;;N;;;;;
0618;ARABIC SMALL FATHA;Mn;30;NSM;;;;;N;;;;;
0619;ARABIC SMALL DAMMA;Mn;31;NSM;;;;;N;;;;;
061A;ARABIC SMALL KASRA;Mn;32;NSM;;;;;N;;;;;
```

ARABIC SMALL HIGH ZAIN is a Qur'anic annotation sign. Semantically, it is more or less equivalent to the Qur'anic symbol used in the Arab world, U+06D6 ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA. The character is used to indicate that while a pause can be made at the point it appears in the text, it is recommended that one recite the scripture without a pause. This character is also comparable to U+0615 ARABIC SMALL HIGH TAH (which is used in Iran as roughly equivalent to U+06D7 ARABIC SMALL HIGH LIGATURE QAF WITH LAM WITH ALEF MAKSURA).

ARABIC SMALL FATAH, ARABIC SMALL DAMMA, and ARABIC SMALL KASRA are smaller versions of the simple *harakat* ARABIC FATAH, ARABIC DAMMA, and ARABIC KASRA. Readers who do not know the finer

points of the grammar of the standard Arabic language make use of these characters to pronounce some initial ALEFS correctly. For example, see Figure 2, verses 2 and 4 from *Sura* 83. Here, verse 2 begins with an ALEF bearing a SMALL FATHA, while verse 4 starts with an ALEF bearing a normal FATHA. This means that if a reader recites verse 1 and verse 2 of the *Sura* continuously, that is, without stopping at the end of verse 1, he or she should not pronounce the initial ALEF of verse 2, and should instead connect the final FATHA of verse 1 to the first letter (LAM) of verse 2. If, on the other hand, the reader pauses at the end of verse 1, he or she should pronounce the ALEF and FATHA when beginning to read from verse 2. But when reciting verse 4, the ALEF and FATHA that begins it should always be pronounced, whether the reader had paused at the end of verse 3 or not. In Qur'ans published in Arabic-speaking countries, U+0671 ARABIC LETTER ALEF WASLA is used instead to represent the combination of U+0627 ARABIC LETTER ALEF and the small vowels.

Ordering. The three vowels SMALL FATHA, SMALL DAMMA, and SMALL KASRA all sort with FATHA, DAMMA, and KASRA, with a third-level difference to break ties.

In the chart shown on page 5 below, the background color signifies the status of the character with regard to the UTC and WG2. Light yellow is for the characters already under ballot; light green is for characters proposed in WG2 N3180; and bright yellow indicates the characters proposed in this document. In the names list these are indicated with italics, with underlined italics, and with underlining respectively.

Bibliography. All examples come from the newer Qur'ans published by the Centre for Qur'an Printing and Publishing of the Islamic Republic of Iran, which are typeset using computers in a font resembling the style of Othman Taha. The specific section describing the usage of the small *harakat* and the annotation mark is from:

Qor'ān-e Karim. With translation into Persian by Abdol-Mohammad Ayati. Tehran, Madreseh, 1382 AP (= 2003 CE). ISBN 964-385-280-6.

Acknowledgments. This project was made possible in part by a grant from the High Council of Information Dissemination of Iran (which funded the FarsiWeb Project, Sharif FarsiWeb, Tehran). The authors wish to thank Hamid Zarrabi-Zadeh for first bringing the small forms of the *harakat* to our attention, and Farzaneh Sarafraz and Araz Sarbar for helping to find examples of usage.

Figures

د (علائم وقف و سکت:
 از بین علائم وقف و سکت منداول، این علائم انتخاب شده است:
 م علامت وقف لازم (وصل به عبارت بعد جایز نیست)
 ط علامت وقف مطلق (وقف بر آن نیکوست)
 ج علامت جواز وقف
 ز علامت جواز وقف، ولی وصل آن نیکوتر است.
 ۱۱ علامت وقف معانقه (در صورت وقف بر علامت اول، وقف بر دومی
 جایز نیست و در صورت وقف بر علامت دوم، ابتدا از اولی جایز نیست)
 لا علامت عدم جواز وقف و در برخی موارد، عدم جواز ابتدا از کلمه بعد
 س علامت سکت (اگر قاری می‌خواهد وصل کند، باید، در محل علامت،
 بدون تجدید نفس، مکث کوتاهی نماید)

Figure 1. Sample text describing the use of ARABIC SMALL HIGH ZAIN.

كَلَّا بَلْ تُكْذِبُونَ بِالَّذِينَ ﴿١٠﴾ وَإِنَّ عَلَيْكُمْ لَحَافِظِينَ ﴿١١﴾ كِرَامًا
 كَاتِبِينَ ﴿١٢﴾ يَعْلَمُونَ مَا تَفْعَلُونَ ﴿١٣﴾ إِنَّ الْأَبْرَارَ لَفِي نَعِيمٍ ﴿١٤﴾ وَإِنَّ
 الْفُجَّارَ لَفِي جَحِيمٍ ﴿١٥﴾ يَصَلُونَهَا يَوْمَ الدِّينِ ﴿١٦﴾ وَمَا هُمْ عَنْهَا بِغَائِبِينَ ﴿١٧﴾
 وَمَا أَدْرَاكَ مَا يَوْمَ الدِّينِ ﴿١٨﴾ ثُمَّ مَا أَدْرَاكَ مَا يَوْمَ الدِّينِ ﴿١٩﴾
 يَوْمَ لَا تَمْلِكُ نَفْسٌ لِنَفْسٍ شَيْئًا ﴿٢٠﴾ وَالْأَمْرُ يَوْمَئِذٍ لِلَّهِ ﴿٢١﴾

آیاتها

سَبَّحُوا لِلْمُطَفِّفِينَ ﴿٢٦﴾

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
 وَيْلٌ لِلْمُطَفِّفِينَ ﴿١﴾ الَّذِينَ إِذَا أَكْتَالُوا عَلَى النَّاسِ يَسْتَوْفُونَ ﴿٢﴾
 وَإِذَا كَالُوهُمْ أَوْ وَزَنُوهُمْ يُخْسِرُونَ ﴿٣﴾ أَلَا يَظُنُّ أُولَئِكَ أَنَّهُمْ
 مَبْعُوثُونَ ﴿٤﴾ لِيَوْمٍ عَظِيمٍ ﴿٥﴾ يَوْمَ يَقُومُ النَّاسُ لِرَبِّ الْعَالَمِينَ ﴿٦﴾

Figure 2. Sample showing ARABIC SMALL HIGH ZAIN and ARABIC SMALL FATHA OVER ALEF, alongside ARABIC FATHA OVER ALEF.

بَعْضُهُمْ لِبَعْضٍ عَدُوٌّ إِلَّا الْمُتَّقِينَ ﴿٦٧﴾ يَلْعَابِدِ لِأَخَوْفٍ
 عَلَيْكُمْ الْيَوْمَ وَلَا أَنْتُمْ تَحْزَنُونَ ﴿٦٨﴾ الَّذِينَ ءَامَنُوا بِآيَاتِنَا
 وَكَانُوا مُسْلِمِينَ ﴿٦٩﴾ ادْخُلُوا الْجَنَّةَ أَنْتُمْ وَأَزْوَاجُكُمْ
 تُحْبَرُونَ ﴿٧٠﴾ يُطَافُ عَلَيْهِمْ بِصِحَافٍ مِنْ ذَهَبٍ وَأَكْوَابٍ
 وَفِيهَا مَا تَشْتَهِيهِ الْأَنْفُسُ وَتَلَذُّ الْأَعْيُنُ وَأَنْتُمْ فِيهَا
 خَالِدُونَ ﴿٧١﴾ وَتِلْكَ الْجَنَّةُ الَّتِي أُورِثْتُمُوهَا بِمَا كُنْتُمْ
 تَعْمَلُونَ ﴿٧٢﴾ لَكُمْ فِيهَا فَاكِهَةٌ كَثِيرَةٌ مِنْهَا تَأْكُلُونَ ﴿٧٣﴾

Figure 3. Sample showing ARABIC SMALL FATHA and ARABIC SMALL DAMMA.

وَاشْرَبُوا مِنْ رِزْقِ اللَّهِ وَلَا تَعَثُّوا فِي الْأَرْضِ مُفْسِدِينَ ﴿٦٠﴾
 وَإِذْ قُلْتُمْ يَا مُوسَى لَنْ نَصْبِرَ عَلَىٰ طَعَامٍ وَاحِدٍ فَادْعُ لَنَا رَبَّكَ
 يُخْرِجْ لَنَا مِمَّا تُنْبِتُ الْأَرْضُ مِنْ بَقْلِهَا وَقِثَّائِهَا وَفُومِهَا
 وَعَدَسِيهَا وَبَصِلِهَا قَالَ آتَسْتَبْدِلُونَ الَّذِي هُوَ أَدْنَىٰ
 بِالَّذِي هُوَ خَيْرٌ إِنْ هِيَ إِلَّا مَصْرًا فَإِنَّ لَكُمْ مَا سَأَلْتُمْ
 وَضُرِبَتْ عَلَيْهِمُ الذِّلَّةُ وَالْمَسْكَنَةُ وَبَاءَ وَبِعَضَبٍ مِنَ
 اللَّهِ ذَٰلِكَ بِأَنَّهُمْ كَانُوا يَكْفُرُونَ بِآيَاتِ اللَّهِ وَيَقْتُلُونَ
 النَّبِيَّيْنَ بِغَيْرِ الْحَقِّ ذَٰلِكَ بِمَا عَصَوْا وَكَانُوا يَعْتَدُونَ ﴿٦١﴾

Figure 4. Sample showing ARABIC SMALL KASRA.

البته در مواردی که همزه وصل در آغاز آیه یا بعد از علامت وقف واقع شده، حرکت همزه وصل ریزتر از حرکات معمولی نوشته شده است؛ مانند:

مُسْلِمِينَ ادْخُلُوا.

Figure 5. Text describing the use of the three small vowels, showing ARABIC SMALL DAMMA. The text follows an introduction which explains how other *alefs* and *harakats* are written, and reads: “Except in cases when the *wasla hamza* [*alef*] occurs in the beginning of the Ayah or after a *waqf*, the *harakat* of the *hamza* [*alef*] has been written much smaller than the normal *harakat*; thus: *moslimīn-a* (69) *odxolu*.”

TABLE XXX - Row 06: ARABIC

	060	061	062	063	064	065	066	067
0	ا	آ		ذ	-	ـ	٠	١
1	ب	بٲ	ء	ر	ف	ٲ	١	أ
2	م	مٲ	ا	ز	ق	ٲ	٢	أ
3	ص	صٲ	أ	س	ك	آ	٣	إ
4		بٲ	ؤ	ش	ل	ٲ	٤	ء
5		طٲ	إ	ص	م	ٲ	٥	أ
6	٣	آ	ئ	ض	ن	ٲ	٦	ؤ
7	٤	زٲ	ا	ط	ه	ٲ	٧	و
8	٥	ٲ	ب	ظ	و	ٲ	٨	ئ
9	٦	ٲ	ة	ع	ى	آ	٩	ط
A	٧	ٲ	ت	غ	ي	ٲ	١٠	ث
B	فٲ	؛	ث	تى	ٲ	ٲ	ر	ب
C	،		ج	كٲ	ٲ	ٲ	،	ت
D	ٲ		ح	ئى	ٲ	ٲ	★	ت
E	م	؛	خ	تى	ٲ	ٲ	س	پ
F	ع	؟	د	ثى	ٲ		ف	ث

TABLE XXX - Row 06: ARABIC

hex	Name
00	ARABIC NUMBER SIGN
01	ARABIC SIGN SANAH
02	ARABIC FOOTNOTE MARKER
03	ARABIC SIGN SAFHA
04	(This position shall not be used)
05	(This position shall not be used)
06	<i>ARABIC-INDIC CUBE ROOT</i>
07	<i>ARABIC-INDIC FOURTH ROOT</i>
08	<i>ARABIC RAY</i>
09	<i>ARABIC-INDIC PER MILLE SIGN</i>
0A	<i>ARABIC-INDIC PER TEN THOUSAND SIGN</i>
0B	AFGHANI SIGN
0C	ARABIC COMMA
0D	ARABIC DATE SEPARATOR
0E	ARABIC POETIC VERSE SIGN
0F	ARABIC SIGN MISRA
10	ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM
11	ARABIC SIGN ALAYHE ASSALLAM
12	ARABIC SIGN RAHMATULLAH ALAYHE
13	ARABIC SIGN RADI ALLAHOU ANHU
14	ARABIC SIGN TAKHALLUS
15	ARABIC SMALL HIGH TAH
16	<i>ARABIC SMALL HIGH ALEF-LAM-YEH</i>
17	<i>ARABIC SMALL HIGH ZAIN</i>
18	<i>ARABIC SMALL FATHA</i>
19	<i>ARABIC SMALL DAMMA</i>
1A	<i>ARABIC SMALL KASRA</i>
1B	ARABIC SEMICOLON
1C	(This position shall not be used)
1D	(This position shall not be used)
1E	ARABIC TRIPLE DOT PUNCTUATION MARK
1F	ARABIC QUESTION MARK
20	(This position shall not be used)
21	ARABIC LETTER HAMZA
22	ARABIC LETTER ALEF WITH MADDA ABOVE
23	ARABIC LETTER ALEF WITH HAMZA ABOVE
24	ARABIC LETTER WAW WITH HAMZA ABOVE
25	ARABIC LETTER ALEF WITH HAMZA BELOW
26	ARABIC LETTER YEH WITH HAMZA ABOVE
27	ARABIC LETTER ALEF
28	ARABIC LETTER BEH
29	ARABIC LETTER TEH MARBUTA
2A	ARABIC LETTER TEH
2B	ARABIC LETTER THEH
2C	ARABIC LETTER JEEM
2D	ARABIC LETTER HAH
2E	ARABIC LETTER KHAH
2F	ARABIC LETTER DAL
30	ARABIC LETTER THAL
31	ARABIC LETTER REH
32	ARABIC LETTER ZAIN
33	ARABIC LETTER SEEN
34	ARABIC LETTER SHEEN
35	ARABIC LETTER SAD
36	ARABIC LETTER DAD
37	ARABIC LETTER TAH
38	ARABIC LETTER ZAH
39	ARABIC LETTER AIN
3A	ARABIC LETTER GHAIN
3B	<i>ARABIC LETTER KEHEH WITH TWO DOTS ABOVE</i>
3C	<i>ARABIC LETTER KEHEH WITH THREE DOTS BELOW</i>
3D	<i>ARABIC LETTER YEH WITH INVERTED V</i>
3E	<i>ARABIC LETTER YEH WITH TWO DOTS ABOVE</i>
3F	<i>ARABIC LETTER YEH WITH THREE DOTS ABOVE</i>
40	ARABIC TATWEEL
41	ARABIC LETTER FEH
42	ARABIC LETTER QAF
43	ARABIC LETTER KAF
44	ARABIC LETTER LAM
45	ARABIC LETTER MEEM
46	ARABIC LETTER NOON
47	ARABIC LETTER HEH
48	ARABIC LETTER WAW
49	ARABIC LETTER ALEF MAKSURA
4A	ARABIC LETTER YEH
4B	ARABIC FATHATAN
4C	ARABIC DAMMATAN
4D	ARABIC KASRATAN
4E	ARABIC FATHA
4F	ARABIC DAMMA
50	ARABIC KASRA
51	ARABIC SHADDA
52	ARABIC SUKUN
53	ARABIC MADDAH ABOVE
54	ARABIC HAMZA ABOVE
55	ARABIC HAMZA BELOW
56	ARABIC SUBSCRIPT ALEF
57	ARABIC INVERTED DAMMA
58	ARABIC MARK NOON GHUNNA

hex	Name
59	ARABIC ZWARAKAY
5A	ARABIC VOWEL SIGN SMALL V ABOVE
5B	ARABIC VOWEL SIGN INVERTED SMALL V ABOVE
5C	ARABIC VOWEL SIGN DOT BELOW
5D	ARABIC REVERSED DAMMA
5E	ARABIC FATHA WITH TWO DOTS
5F	(This position shall not be used)
60	ARABIC-INDIC DIGIT ZERO
61	ARABIC-INDIC DIGIT ONE
62	ARABIC-INDIC DIGIT TWO
63	ARABIC-INDIC DIGIT THREE
64	ARABIC-INDIC DIGIT FOUR
65	ARABIC-INDIC DIGIT FIVE
66	ARABIC-INDIC DIGIT SIX
67	ARABIC-INDIC DIGIT SEVEN
68	ARABIC-INDIC DIGIT EIGHT
69	ARABIC-INDIC DIGIT NINE
6A	ARABIC PERCENT SIGN
6B	ARABIC DECIMAL SEPARATOR
6C	ARABIC THOUSANDS SEPARATOR
6D	ARABIC FIVE POINTED STAR
6E	ARABIC LETTER DOTLESS BEH
6F	ARABIC LETTER DOTLESS QAF
70	ARABIC LETTER SUPERSCRIPIT ALEF
71	ARABIC LETTER ALEF WASLA
72	ARABIC LETTER ALEF WITH WAVY HAMZA ABOVE
73	ARABIC LETTER ALEF WITH WAVY HAMZA BELOW
74	ARABIC LETTER HIGH HAMZA
75	ARABIC LETTER HIGH HAMZA ALEF
76	ARABIC LETTER HIGH HAMZA WAW
77	ARABIC LETTER U WITH HAMZA ABOVE
78	ARABIC LETTER HIGH HAMZA YEH
79	ARABIC LETTER TTEH
7A	ARABIC LETTER TTEHEH
7B	ARABIC LETTER BEEH
7C	ARABIC LETTER TEH WITH RING
7D	ARABIC LETTER TEH WITH THREE DOTS ABOVE DOWNWARDS
7E	ARABIC LETTER PEH
7F	ARABIC LETTER TEHEH

A. Administrative

1. Title

Proposal to encode four Qur'anic Arabic characters in the UCS

2. Requester's name

Michael Everson and Roozbeh Pournader.

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

Individual contribution.

4. Submission date

2006-11-08

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

1b. Proposed name of script

1c. The proposal is for addition of character(s) to an existing block

Yes.

1d. Name of the existing block

Arabic.

2. Number of characters in proposal

4.

3. Proposed category (A-Contemporary; 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)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

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

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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)?

Yes.

8. 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/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. 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.

2b. If YES, with whom?

Hamid Zarrabi-Zadeh.

2c. 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?

No.

4a. The context of use for the proposed characters (type of use; common or rare)

Rare.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

In editions of the Qur'an published in Iran.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Keep with other Arabic characters.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

Yes, but they are smaller and function differently.

10b. If YES, is a rationale for its inclusion provided?

Yes.

10c. If YES, reference

See the body of the proposal.

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

Yes.

11b. If YES, is a rationale for such use provided?

No.

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?