PROPOSAL TO ENCODE SIXTEEN QURANIC ARABIC CHARACTERS

Rikza F. Sh.

r  .S . f a

rikzafsh@gmail.com

5 June 2024

1. Introduction
The Quran is a holy book for Moslems that was revealed by Allah to the Prophet Muhammad through the intercession of the angel Jibril. At first, the Quran was only memorized by the companions of the Prophet and was not written into a complete book. Until the caliph Uthman bin ‘Affan codified and established a standard version, which became the basis for writing most of the Qurans that exist today.

Currently, most of the characters needed to type the Quran text are well-encoded in Unicode. However, there are several special characters used in various mushafs to represent the ten reading of the Quran that are not yet encoded. The author has made a list of those characters in L2/22-281. This document proposes sixteen of those characters that appeared in the mushafs published by alwa7y.com.

2. Character to be encoded
2.1. ARABIC SMALL BASELINE FATHA
This character is used to mark fatha when continuing recitation from the opening letters to the next verse.

Figure 1  Quran 3:1 from Mushaf at-Tayseer bi riwayah Hafs by alwa7y.com, p. 50

2.2. ARABIC SMALL BASELINE DOTLESS HEAD OF KHAH
This character is used to mark sukun when continuing recitation from the opening letters to the next verse.
2.3. ARABIC NORTHEAST POINTING ARROWHEAD ABOVE

This character is used to mark changing of vowel to fatha or damma when starting recitation from the word. For example, if we start our recitation from 

\[
\text{ذَٰلِكَ الَّذِي نُزِّلَ إِلَيْكَ فَلَا يَنْبِيِّلُكَ تَحْكِيمٌ}
\]

even though it is really not recommended to start reciting from that word because the meaning would be incomplete.

2.4. ARABIC NORTHEAST POINTING ARROWHEAD BELOW

This character is used to mark changing of vowel to kasra when starting recitation from the word. For example, if we start our recitation from 

\[
\text{لَيْقَطَّعُ}
\]

even though it is really not recommended to start reciting from that word because the meaning would be incomplete.
2.5. ARABIC SOUTHWEST POINTING ARROWHEAD BELOW
This character is used to mark changing of vowel to kasra when pausing recitation from the word. For example, if we pause our recitation on the word عُلِّيْهِمْ, we should read it as عُلِّيْهِمْ, even though it is really not recommended to pause reciting on that word because the meaning would be incomplete.

2.6. ARABIC SMALL CIRCLE ABOVE
This character is used to mark imalah sughra when pausing recitation from the word. This character has smaller and more circular shape than U+06DF ARABIC SMALL HIGH ROUNDED ZERO and is usually used together with fatha or fathatan.

2.7. ARABIC LARGE CIRCLE ABOVE
This character is used to mark deletion of the letter below it when pausing recitation.
This character is also used in another mushaf to mark place of *dua* (saying a phrase written in the footnote).

2.8. **ARABIC SMALL LOW UPRIGHT RECTANGULAR ZERO**
This character is used to mark long /iː/ vowel when pausing recitation. Usually, long /iː/ vowel that marked with U+06E6 ARABIC SMALL YEH is not pronounced when pausing recitation.

2.9. **ARABIC SQUARE BELOW**
This character is placed below the letter reh to indicate that the letter should be pronounced lightly without pharyngealization.
2.10. ARABIC FILLED SQUARE BELOW
This character is placed below the letter lam to indicate that the letter should be pronounced with pharyngealization.

2.11. ARABIC SMALL HIGH NOON WITH FATHA
This character is used to mark naql, when there is a hamza after tanween, according to Warsh narration, the vowel of the hamza is moved to tanween so that it is pronounced as the letter noon.

2.12. ARABIC SMALL LOW NOON WITH FATHA
This character is used to mark naql, when there is a hamza after tanween, according to Warsh narration, the vowel of the hamza is moved to tanween so that it is pronounced as the letter noon. This character needs to be MCM because it appears closer to the base letter than a kasra.
2.13. ARABIC SMALL HIGH NOON WITH DAMMA
This character is used to mark *naql*, when there is a hamza after tanween, according to Warsh narration, the vowel of the hamza is moved to tanween so that it is pronounced as the letter noon.

2.14. ARABIC SMALL LOW NOON WITH DAMMA
This character is used to mark *naql*, when there is a hamza after tanween, according to Warsh narration, the vowel of the hamza is moved to tanween so that it is pronounced as the letter noon. This character needs to be MCM because it appears closer to the base letter than a kasra.

2.15. ARABIC SMALL HIGH HEH INITIAL FORM
This character is used to indicate the end of the word should be pronounced with /h/ sound when pausing recitation.
This character is also used in some Quran manuscript to mark every five verses.

2.16. ARABIC SMALL HIGH WORD KABBIR
This character is used to mark place of takbeer (saying “Allahu akbar”) at the end of the chapter.

3. Character Data
Characters

<table>
<thead>
<tr>
<th>No.</th>
<th>Codepoint</th>
<th>Glyph</th>
<th>Name</th>
<th>Reference in L2/22-284</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10EC9</td>
<td></td>
<td>ARABIC SMALL BASELINE FATHA</td>
<td>§3.2:1, p. 4</td>
</tr>
<tr>
<td>2</td>
<td>10ECA</td>
<td></td>
<td>ARABIC SMALL BASELINE DOTLESS HEAD OF KHAH</td>
<td>§3.2:2, p. 4</td>
</tr>
<tr>
<td>3</td>
<td>10ECB</td>
<td></td>
<td>ARABIC NORTHEAST POINTING ARROWHEAD ABOVE</td>
<td>§3.4:5, p. 10</td>
</tr>
<tr>
<td>4</td>
<td>10ECC</td>
<td></td>
<td>ARABIC NORTHEAST POINTING ARROWHEAD BELOW</td>
<td>§3.4:6, p. 10</td>
</tr>
<tr>
<td>5</td>
<td>10ECD</td>
<td></td>
<td>ARABIC SOUTHWEST POINTING ARROWHEAD BELOW</td>
<td>§3.4:7, p. 10</td>
</tr>
<tr>
<td>Code Point</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10ECE</td>
<td>ARABIC SMALL CIRCLE ABOVE</td>
<td>§3.4:8, p. 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10ECF</td>
<td>ARABIC LARGE CIRCLE ABOVE</td>
<td>§3.4:9, p. 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10EF0</td>
<td>ARABIC SMALL UPRIGHT RECTANGULAR ZERO</td>
<td>§3.4:13, p. 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10EF1</td>
<td>ARABIC SQUARE BELOW</td>
<td>§3.4:14, p. 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10EF2</td>
<td>ARABIC FILLED SQUARE BELOW</td>
<td>§3.4:15, p. 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10EF3</td>
<td>ARABIC SMALL HIGH NOON WITH FATHA</td>
<td>§3.5:106, p. 31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10EF4</td>
<td>ARABIC SMALL LOW NOON WITH FATHA</td>
<td>§3.5:108, p. 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10EF5</td>
<td>ARABIC SMALL HIGH NOON WITH DAMMA</td>
<td>§3.5:107, p. 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10EF6</td>
<td>ARABIC SMALL LOW NOON WITH DAMMA</td>
<td>§3.5:109, p. 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10EF7</td>
<td>ARABIC SMALL HEH INITIAL FORM</td>
<td>§3.5:113, p. 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10EF8</td>
<td>ARABIC HIGH WORD KABBIR</td>
<td>§3.6:16, p. 36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Unicode character properties**

<table>
<thead>
<tr>
<th>Code Point</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>10EC9</td>
<td>ARABIC SMALL BASELINE FATHA; Lo; 0; AL; ·; ·; ·; ·; ·; ·; N; ·; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10ECA</td>
<td>ARABIC SMALL BASELINE DOTLESS HEAD OF KHAH; Lo; 0; AL; ·; ·; ·; ·; ·; ·; ·; N; ·; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10ECB</td>
<td>ARABIC NORTHEAST POINTING ARROWHEAD ABOVE; Mn; 230; NSM; ·; ·; ·; ·; ·; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10ECC</td>
<td>ARABIC NORTHEAST POINTING ARROWHEAD BELOW; Mn; 220; NSM; ·; ·; ·; ·; ·; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10ECD</td>
<td>ARABIC SOUTHWEST POINTING ARROWHEAD BELOW; Mn; 220; NSM; ·; ·; ·; ·; ·; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10ECF</td>
<td>ARABIC LARGE CIRCLE ABOVE; Mn; 230; NSM; ·; ·; ·; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10EOF0</td>
<td>ARABIC SMALL LOW UPRIGHT RECTANGULAR ZERO; Mn; 220; NSM; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10EOF1</td>
<td>ARABIC SQUARE BELOW; Mn; 220; NSM; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10EOF2</td>
<td>ARABIC FILLED SQUARE BELOW; Mn; 220; NSM; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10EOF3</td>
<td>ARABIC SMALL HIGH NOON WITH FATHA; Mn; 230; NSM; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10EOF4</td>
<td>ARABIC SMALL LOW NOON WITH FATHA; Mn; 220; NSM; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10EOF5</td>
<td>ARABIC SMALL HIGH NOON WITH DAMMA; Mn; 230; NSM; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10EOF6</td>
<td>ARABIC SMALL LOW NOON WITH DAMMA; Mn; 220; NSM; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10EOF7</td>
<td>ARABIC SMALL HIGH HEH INITIAL FORM; Mn; 230; NSM; ·; ·; ·; ·; ·; ·;</td>
</tr>
<tr>
<td>10EOF8</td>
<td>ARABIC SMALL HIGH WORD KABBIR; Mn; 230; NSM; ·; ·; ·; ·; ·; ·;</td>
</tr>
</tbody>
</table>

**PropList.txt**

```
10EF4          ; Modifier_Combining_Mark # Mn       ARABIC SMALL LOW NOON WITH FATHA
10EF6          ; Modifier_Combining_Mark # Mn       ARABIC SMALL LOW NOON WITH DAMMA
```

# Total code points: 16
Annotation
The following annotation could be updated in Namelist.txt
10ECF; ARABIC LARGE CIRCLE ABOVE
  x (arabic large circle below - 08D1)
10EF0; ARABIC SMALL LOW UPRIGHT RECTANGULAR ZERO
  x (arabic small high upright rectangular zero - 06E0)

4. Chart
Arabic Extended-C

<table>
<thead>
<tr>
<th>U+</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>10EC</td>
<td>🕑</td>
<td>Ⲥ</td>
<td>ⲥ</td>
<td>Ⲧ</td>
<td>ⲧ</td>
<td>Ⲩ</td>
<td>ⲩ</td>
<td>Ⲫ</td>
<td>ⲫ</td>
<td>Ⲭ</td>
<td>ⲭ</td>
<td>Ⲯ</td>
<td>ⲯ</td>
<td>Ⲱ</td>
<td>ⲱ</td>
<td>Ⲳ</td>
</tr>
<tr>
<td>10ED</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
</tr>
<tr>
<td>10EE</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
</tr>
<tr>
<td>10EF</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
<td>Ⲳ</td>
</tr>
</tbody>
</table>

Characters in current version (15.1)
Characters approved for publication in version 16.0
Provisionally assigned characters
Characters proposed elsewhere
Characters proposed in this proposal
Pending characters
Unassigned

5. Acknowledgement
The author would like to thank Roozbeh Pournader for the support and technical assistance.

6. Reference
7. Mushaf al-Quddus bi rasm Uthmani. Mubarokatan Thoyyibah, Kudus
7. Figures

Figure 21 The end matter of Mushaf at-Tayseer bi riwayah Warsh by alwa7y.com, p.632 explaining the usage of ARABIC SMALL BASELINE FATHA and ARABIC SMALL BASELINE DOTLESS HEAD OF KHAH

Figure 22 The end matter of Mushaf at-Tayseer bi riwayah Warsh by alwa7y.com, p.635 explaining the usage of small noons

Figure 23 The end matter of Mushaf at-Tayseer bi riwayah Ibn Kathir by alwa7y.com, p.640 explaining the usage of ARABIC SMALL HIGH WORD KABBIR
- **Figure 24** The end matter of Mushaf at-Tayseer bi riwayah Warsh by alwa7y.com, p.637 explaining the usage of ARABIC SQUARE BELOW and ARABIC FILLED SQUARE BELOW

- **Figure 25** The end matter of Mushaf at-Tayseer bi riwayah Warsh by alwa7y.com, p.639 explaining the usage of ARABIC SMALL LOW UPRIGHT RECTANGULAR ZERO

- **Figure 26** The end matter of Mushaf at-Tayseer bi riwayah Yakob by alwa7y.com, p.637 explaining the usage of ARABIC SMALL HIGH HEH INITIAL FORM

Proposal to Encode Sixteen Quranic Arabic Characters
Proposal to Encode Sixteen Quranic Arabic Characters

7. علامة الوقف بعاء السكت (٢٠)، هي هاء صغيرة توضع فوق الحرف الأخير من الكلمة، للدلالة على الوقف عليها بالباء، ويختلف موضع الباء من الكلمة وفقاً حسب هذا الحرف الأخير، فإن الإكان:

- تاء تانيث، نحو: {بِيَتٍ}، تنطق الباء بدلًا منها.
- نون نسوة مسبوقة بباء ضمير، نحو: {بَيْنَهُ}، تنطق الباء بعدها.
- ضمير مفرد غائب، نحو: {بُوهُ}، {بِهِ}، تنطق الباء بعده.
- ياء متكلم، نحو: {بِيَدٍ}، تنطق الباء بعدها.
- ميم كلمة استفهامية، نحو: {فِمَ}، تنطق الباء بعدها.
- ميم كلمتي {نَمَ}، {فَكَمَ}، تنطق الباء بعدها عند رؤى.
- ألف، نحو: {بَيْنَيْنَ}، تنطق الباء بعدها مع إشاع مدهما، عند رؤى.

Figure 27 The end matter of Mushaf at-Tayseer bi riwayah Yakob by alwa7y.com, p.636 explaining the usage of ARABIC SMALL HIGH HEH INITIAL FORM

8. علامة تغيير حركة الحرف حال الابتداء (١٠)، هي علامة رأس السهم للدلالة على واجب تغيير حركة الحرف حال الابتداء بالكلمة، وذلك حسب موقع العلامة من الحرف، فإذا كانت:

- فوق الحرف ومتطرفة إلى البسار قليلًا، فمعناها تغيير حركة الحرف إلى الضم، وذلك في حاء الخلاصة في {الْحَمِيدَ}، إبراهيم، عند رؤى، فتضم الباء، حال الابتداء فقط.
- فوق الحرف مباشرة، فمعناها تغيير حركة الحرف إلى الفتح، وذلك في حرف الشاء الأول من {ثْمَارًا}، النجم، فتح الحاء الأولى بالفتح وتخفيف الثانية حال الابتداء فقط، ومتها {تَفْتَكَرُونَ}، عبد رؤى.
- تحت الحرف، فمعناها تغيير حركة الحرف إلى الكسر، وذلك {لَيْفَظُوا}، عند رؤى.

Figure 28 The end matter of Mushaf at-Tayseer bi riwayah Yakob by alwa7y.com, p.641 explaining the usage of the combining arrowheads
### A. Administrative

| 1. Title: | Proposal to Encode Sixteen Quranic Arabic Characters |
| 2. Requester's name: | Rikza F. Sh. |
| 3. Requester type (Member body/Liaison/Individual contribution): | Individual contribution |
| 4. Submission date: | 5 June 2024 |

**5. Fonts related:**
- a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
  - Rikza F. Sh. (rikzafsh@gmail.com)

**6. References:**
- a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? YES
- b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? 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)? YES

---

### C. Technical - Justification

1. **Has this proposal for addition of character(s) been submitted before?**
   - NO

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? N/A
     - If YES, available relevant documents: N/A

3. **Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?**
   - YES

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

5. **Are the proposed characters in current use by the user community?**
   - YES
     - If YES, where? Reference: Some mushafs that support the Ten Reading of the Quran

6. **After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?**
   - NO
     - If YES, is a rationale provided?
       - N/A

7. **Should the proposed characters be kept together in a contiguous range (rather than being scattered)?**
   - NO

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?
       - N/A

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?
       - N/A

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?**
    - YES
      - If YES, is a rationale for its inclusion provided?
        - NO

11. **Does the proposal include use of combining characters and/or use of composite sequences?**
   - YES
     - If YES, is a rationale for such use provided?
       - YES
     - Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
       - N/A

12. **Does the proposal contain characters with any special properties such as control function or similar semantics?**
    - YES
      - If YES, describe in detail (include attachment if necessary)
        - U+10EF4 and U+10EF6 needs to be MCM because they appear closer to the base letter than a kasra

13. **Does the proposal contain any Ideographic compatibility characters?**
    - NO
      - If YES, are the equivalent corresponding unified ideographic characters identified?
        - N/A