To: UTC and ISO/IEC JTC1/SC2 WG2

Title: Proposal to encode Warsh-based Arabic script characters

From: Lorna Evans (SIL International)

Date: 15 August 2014

Reference: L2/14-104, L2/14-207R

Introduction

This document is requesting the encoding of three characters for modern Warsh-based orthographies in Africa. Warsh is "the most widespread tradition in North and West Africa" (The Arabic Script in Africa, p 7).

Three characters are proposed for adding to Unicode:

> 08BB ڥ ARABIC LETTER AFRICAN FEH

> 08BC ڦ ARABIC LETTER AFRICAN QAF

> 08BD ڧ ARABIC LETTER AFRICAN NOON

The character names UTC has recommended use the word "AFRICAN" rather than "WARSH". However, the word "African" would include Egypt, Sudan and East Africa where Warsh would be unknown and thus does NOT represent those regions. It is true that the modern orthographies using this tradition are in West Africa. However, the Warsh tradition is also used indigenously in Spain, as well as by many Sufi communities the world over. Our preference would be to use "WARSH" rather than "AFRICAN".

Nonetheless, this proposal uses the world "AFRICAN" as UTC has approved those names.

These characters are unusual in that isolate and final forms are dotless and initial and medial forms have a dot. The isolate, initial, medial and final forms of these characters are shown below:

Final	Medial	Initial	Isolate
ڥ	ڢ	ڡ	ڡ
ڧ	ڧ	ڧ	ڧ
ں	ٮ	ٮ	ں

These characters have been accepted by UTC and are requested for approval by ISO 10646.

Unicode character Properties

```
08BB;ARABIC LETTER AFRICAN FEH;Lo;0;AL;;;;;N;;;;;
08BC;ARABIC LETTER AFRICAN QAF;Lo;0;AL;;;;;N;;;;;
08BD;ARABIC LETTER AFRICAN NOON;Lo;0;AL;;;;;N;;;;;
```

Joining type and group for ArabicShaping.txt

The properties for ArabicShaping.txt would be as follows:

```
08BB; AFRICAN FEH; D; AFRICAN FEH
08BC; AFRICAN QAF; D; AFRICAN QAF
08BD; AFRICAN NOON; D; AFRICAN NOON
```

Annotations

The following annotations are recommended for NamesList.txt:

```
@ Additions for African orthographies
08BB ARABIC LETTER AFRICAN FEH
     * initial and medial forms of this letter have one dot below
     x (arabic letter dotless feh - 06A1)
     x (arabic letter feh with dot moved below - 06A2)
08BC ARABIC LETTER AFRICAN QAF
     * initial and medial forms of this letter have one dot above
     x (arabic letter dotless qaf - 066F)
     x (arabic letter qaf with dot above - 06A7)

08BD ARABIC LETTER AFRICAN NOON
     * initial and medial forms of this letter have one dot above
     x (arabic letter noon ghunna - 06BA)
     x (arabic letter noon - 0646)
```

Normalization and Confusability Issues

There should not be any normalization issues.

These characters could be confused with 06A1 and 06A2 (08BB), 066F and 06A7 (08BC), 06BA and 0646 (08BD).
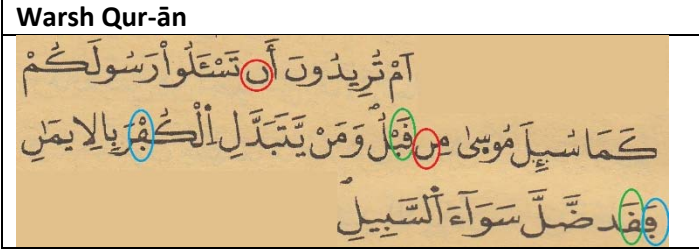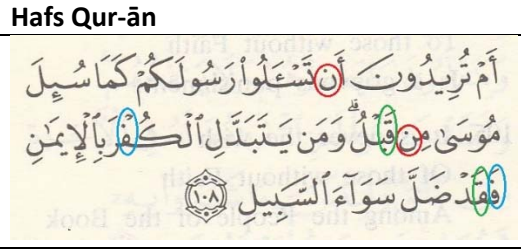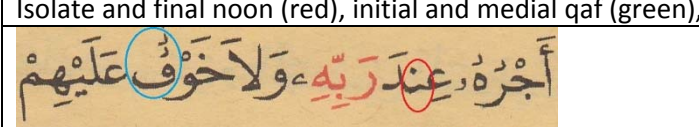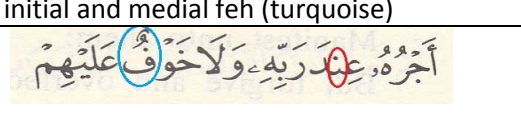
Suggested Collation
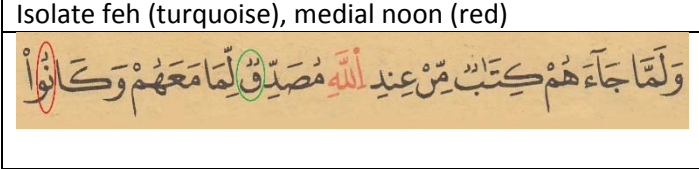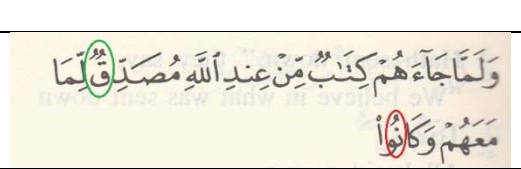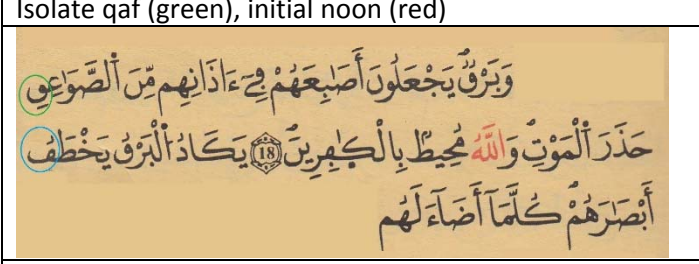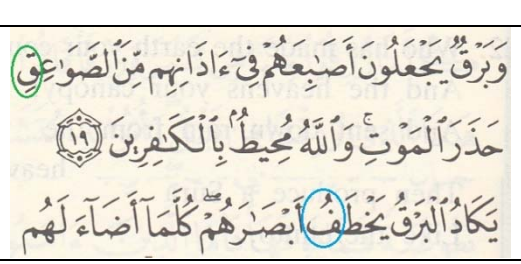
Suggested collation is as follows:

> 06A1 < 08BB < 1EE1E < 1EE7E < 06A2
>
> 066F < 08BC < 1EE1F < 1EE5F < 0642
>
> 06BA < 08BD < 1EE1D < 1EE5D < FB9F < FB9E < 06BB

Characters proposed in this document are colored in red with the black characters already in the standard.

Samples

Below are sample texts from a Warsh Qur-ān and a more standard Hafs Qur-ān:

Warsh Qur-ān	Hafs Qur-ān
Isolate and final noon (red), initial and medial qaf (green), initial and medial feh (turquoise)	
Isolate feh (turquoise), medial noon (red)	
Isolate qaf (green), initial noon (red)	
Final qaf (green), final feh (turquoise)	

References

The Holy Qur-ān. English translation of the meanings and commentary. Revised & Edited By The Presidency of Islamic Researches, IFTA, Call and Guidance. Published 1989 by King Fahd Holy Qur-ān Printing Complex in Al-Madinah. Arabic text and English translation in parallel columns.

The Holy Qur-ān. Warsh Qur'an for Tijaniyya sect. Used in Cameroon, Nigeria and Niger.

Evans, Lorna. 2014. "Supporting the Warsh orthography for Arabic script." UTC Document Register L2/14-104. The Unicode Consortium. http://www.unicode.org/L2/L2014/14104-warsh.pdf

Mumin, Meikal and Kees Versteegh (editors). 2014. The Arabic Script in Africa: studies in the use of a writing system. Studies in Semitic Languages and Linguistics. Vol. 71. Koninkligjke Brill NV, Leiden, The Netherlands.

Pournader, Roozbeh. 2012. Initial and medial forms of Arabic Letter Noon Ghunna (L2/12-381 http://www.unicode.org/L2/L2012/12381-dotless-noon.pdf)

Pournader, Roozbeh. 2012. Implications of the Unicode Arabic model for Warsh-based orthographies. (L2/14-207R http://www.unicode.org/cgi-bin/GetMatchingDocs.pl?L2/14-207)

UTC. 2012. Approved Minutes of UTC Meeting 133. (L2/12-343 http://www.unicode.org/L2/L2012/12343.htm)

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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 Warsh-based Arabic script characters
2. Requester's name: Lorna Evans
3. Requester type (Member body/Liaison/Individual contribution):   Individual contribution
4. Submission date: 14 August 2014
5. Requester's reference (if applicable):   L2/14-104, L2/14-207R
6. Choose one of the following:
This is a complete proposal: Yes
or,   More information will be provided later: L2/14-207R

B. Technical – General

1. Choose one of the following:
        a. This proposal is for a new script (set of characters):   No
        Proposed name of script:
        b. The proposal is for addition of character(s) to an existing block:
                Name of the existing block: Arabic Extended-A
2. Number of characters in proposal: 3
3. Proposed category (select one from below - see section 2.2 of P&P document):
  A-Contemporary   x   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. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document):   2
        Is a rationale provided for the choice?
        If Yes, reference: Arabic is always 2
5. Is a repertoire including character names provided?   Yes
        a. If YES, are the names in accordance with the "character naming guidelines"   Yes
                in Annex L of P&P document?
        b. Are the character shapes attached in a legible form suitable for review?   Yes
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
        publishing the standard?   Michael Everson
        If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
        used:
7. References:
        a. Are references (to other character sets, dictionaries, descriptive texts etc.)   Yes
provided?
        b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
                of proposed characters attached?   Yes
8. 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, suggested character properties and shaping information is included
9. 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.

---

[1] Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?   No
    If YES explain _____
2. Has contact been made to members of the user community (for example: National Body,   Yes
    user groups of the script or characters, other experts, etc.)?
        If YES, with whom?   SIL and UBS linguists
        If YES, available relevant documents: Personal communication
3. Information on the user community for the proposed characters (for example:
    size, demographics, information technology use, or publishing use) is included?
    Reference: See proposal
4. The context of use for the proposed characters (type of use; common or rare)   Common
    Reference: _____
5. Are the proposed characters in current use by the user community?   Yes
    If YES, where? Reference: _____
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?
            If YES, reference: This is an existing script in the BMP and these should be in same
                                plane
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
8. Can any of the proposed characters be considered a presentation form of an existing   No
    character or character sequence?
        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)   Yes
    to an existing character?
        If YES, is a rationale for its inclusion provided?
            If YES, reference: All information is in proposal
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?
        If YES, reference: _____
12. Does the proposal contain characters with any special properties such as   No
    control function or similar semantics?
        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: _____