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 an Arabic-Letter Mark (ALM)*
2. Requester's name: *Matitiahu Allouche and Mohamed Mohie*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *2011-01-16*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Complete proposal*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters):
   Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block:
   Name of the existing block: *Addition of one character in block 20xx*
2. Number of characters in proposal: *1*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary *A* B.1-Specialized (small collection)   B.2-Specialized (large collection) *B.1*
   C-Major extinct   D-Attested extinct   E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic   G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided? *Yes*
   a. If YES, are the names in accordance with the "character naming guidelines"
      in Annex L of P&P document? *Yes*
   b. Are the character shapes attached in a legible form suitable for review? *N/A*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   N/A (this is an invisible character)
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   N/A
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *No*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
   of proposed characters attached? *No*
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*

   Proposed name and properties

   2069;ARABIC-LETTER MARK;Cf;0;AL;;;;;N;;;;;

8. 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: N3702-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11)

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,
   user groups of the script or characters, other experts, etc.)? *Yes*
       If YES, with whom? *Various bidi experts*
       If YES, available relevant documents: *Only oral communication*

3. Information on the user community for the proposed characters (for example:
   size, demographics, information technology use, or publishing use) is included?
   Reference: *All users of the Arabic script using Arabic-Indic digits*

4. The context of use for the proposed characters (type of use; common or rare) *occasional*
   Reference:

5. Are the proposed characters in current use by the user community? *No*
   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: *Only one character, similar to LRM and RLM*

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

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?
       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)
    to an existing character? *No*
        If YES, is a rationale for its inclusion provided?
        If YES, reference:

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
    control function or similar semantics? *Yes*
        If YES, describe in detail (include attachment if necessary)
    *This is an invisible character with properties similar to an Arabic letter. It is equivalent to U+200F RLM, but has
    the added effect of affecting following European digits as if they were Arabic-Indic digits.*

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:

Additional Justification

Unicode includes the LRM (U+200E) and RLM (U+200F) characters. They are invisible characters which creators of bidirectional text can use to solve display issues that the UBA (Unicode Bidirectional Algorithm) does not address adequately.
The use of these characters is mentioned and even recommended in tutorials like:
- H34: Using a Unicode right-to-left mark (RLM) or left-to-right mark (LRM) to mix text direction inline (http://www.w3.org/TR/WCAG20-TECHS/H34.html)
- Internationalization Best Practices: Handling Right-to-left Scripts in XHTML and HTML Content (http://www.w3.org/International/geo/html-tech/tech-bidi.html#ri20030726.140315918)

However, RLM may not be appropriate in an Arabic context, because while it is effective from the ordering point of view, it neutralizes the effect of preceding Arabic letters on the following Arabic-European digits.
The UBA specifies that Arabic letters form an Arabic context wherein following Arabic-European digits must be handled as Arabic-Indic digits, but the presence of an RLM, which may be needed for ordering reasons, destroys this context.
In addition, there is a need to transform Arabic-European digits into Arabic-Indic digits when these digits are positioned at the start of the text like in formulas and numbered lists.

What is needed is a new character equivalent to RLM, but with the same bidi character type as Arabic letters. Such a character could be named ALM (Arabic-Letter Mark). It will be a normally invisible character (like RLM) with a bidi character type AL (unlike the R bidi character type of RLM), and this character should be a non-joiner as related to Arabic text shaping.

Such a character must be located in the BMP, preferably in block 20xx like LRM and RLM.