Rendering the Allah ligature

Introduction

In certain types of literature, the name *Allah* and words related to this name are given unique rendering. Unicode has a *presentation form* character (U+FDF2 ARABIC LIGATURE ALLAH ISOLATED FORM) that implements this rendering and, while this can work (in some fonts) for the word in isolation, it doesn't help users obtain special rendering in other contexts where it is desired. Due to these limitations, SIL does not recommend use of U+FDF2.

The Unicode Standard core specification provides the following recommendation:

> When the formation of the *allah* ligature is desired, the recommended way to represent the word would be <*alef, lam, lam, shadda, superscript alef, heh*> <0627, 0644, 0644, 0651, 0670, 0647>. In non-Arabic languages, other forms of *heh*, such as *heh goal* (U+06C1), may also form the ligature. Extra care should be taken not to form the ligature in the absence of the *shadda* and the *superscript alef*, as the sequences <*alef, lam, lam, heh*> and <*alef, lam, lam, shadda, heh*> exist in Persian and other languages with different meanings or pronunciations, where the formation of the ligature would be incorrect and inappropriate.[1]

Implementers desiring to support the many languages that utilize Arabic script will find it hard to know exactly when the special rendering (or ligature) should be generated and when it should not. The following set of rules provides correct results in most situations but permits suppressing the special rendering in cases where it is not desired.

The rules as implemented in SIL fonts

A sequence of *lam lam heh* should be rendered with the ligature form if all the following conditions are met:

1. One or both of the following are true:
   - the sequence is immediately preceded by an isolate *alef*
   - a *shadda* follows the second *lam*
2. Characters immediately preceding the first *lam*, if any, may have marks.
3. The first *lam* may include a *kasra* (U+0650), but no other marks.
4. The *shadda* may be followed or preceded by either a *superscript-alef* or a *fatha* but no other marks.
5. The *heh* must be final form and may have marks.

For the purposes of the above rules:

- *alef* can be any alef-like character.
  - In SIL fonts we support *alef-madda* (U+0622), *alef-hamza* (U+0623), *alef-hamza-below* (U+0625), *alef* (U+0627), *alef-wasla* (U+0671) but not *high-hamza-alef* (U+0675)
- *heh* can be either *heh* (U+0647) or *heh-goal* (U+06C1)

Additional:

- If there is an isolate *alef* but no marks on the second *lam* then *shadda superscript-alef* should be displayed on that *lam*.
- To disable the special rendering, insert a *zero-width-joiner* (U+200D) somewhere in the sequence. (See Ligatures and ZWJ for why this is the recommended character to use.)
- U+FDF2 should always display as if it were initial *alef lam lam shadda superscript-alef* final *heh*.

Sample results

Characters	→	Glyph	Comment
‍ا + ل + ل + ه	→	الله	Ligature is formed (U+0647)
‍ا + ل + ل + ہ	→	اللہ	Ligature is formed (U+06C1)
‍أَ + ل + ل +ّ + ه	→	اَللّٰه	Ligature is formed
‍ا + ل + ل +ّ +َ + ه	→	اللَّه	Ligature is formed
‍ا + ل + ل +ّ + ا + ه	→	اللّٰه	Ligature is formed
ل + ل +ّ + ه	→	لله	Ligature is formed
ل + ZWJ + ل +ّ + ه	→	للَّه	Ligature is not formed
ل + ل + ه	→	لله	Ligature is not formed
ل +ِ + ل +ّ + ه +ْ	→	لِلّٰهْ	Ligature is formed
ا + ل +ْ + ا + ل + ل +ّ + ه +ْ	→	الْاللّٰهْ	Ligature is formed
ب +ِ + ل + ل +ّ + ه	→	بِللّٰه	Ligature is formed
ف + ل + ل +ّ +َ + ه	→	فللَّه	Ligature is formed
ف + ل + ل +ّ +َ + ه	→	فللَّه	Ligature is formed
ف + ل + ل +َ + ه	→	فللَه	Ligature is not formed

1 The Unicode Consortium. The Unicode Standard, Version 14.0.0, (Mountain View, CA: The Unicode Consortium, 2021. ISBN 978-1-936213-29-0), pg 397.