VIP Webinar – Arabic Script

Fahd A. Batayneh

Arabic Script Issues

▶ Variants

▶ Diacritics

▶ Tatweel (Kashidah)

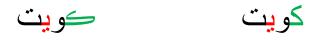▶ Numbering

Variants

- Arabic script variants are characters that almost look alike and/or are almost pronounced alike

<div dir="rtl" align="center">ا   آ   إ   أ</div>

- Variants in some Arabic script languages are not considered so in others (e.g. آ)

- Special cases could arise such as the Alf ¾ hamza of the Jawi language (not in the Unicode table yet)

- Various scenarios implemented amongst different ccTLDs to deal with variants (bundling vs. dropping)

... continue (Variants)

▸ A key recommendation for handling Arabic script variants is to minimize them as much as possible

▸ Variants within the same Arabic script language are simpler to deal with than with variants amongst different language sets under the same script

<div dir="rtl">کویت   کویت</div>

Diacritics

▸ While not used in writing extensively amongst Arabic script users, they are very essential to the enrichment of the script

▸ Sometimes can change the meaning of a word if not used

كُرة ➜ Football (Soccer)

كَرّة ➜ Attempt

▸ Prohibited at the registration level of IDNs

▸ IETF is working extensively on this matter since many scripts – including the Latin – use it

Tatweel (Kashidah)

▸ Does not change the meaning of the word, but is used as means to enrich the writing the of a word

سنغافورة   ســــنغافـــــورة

▸ Agreed upon within the Arabic script community to drop it overall at the registry level

Numbering Systems

▸ There exists three main numbering systems for usage with Arabic script IDNs:

1. Western Arabic ➔ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
2. Eastern Arabic ➔ ٠, ١, ٢, ٣, ٤, ٥, ٦, ٧, ٨, ٩
3. Persian ➔ ٠, ١, ٢, ٣, ۴, ۵, ۶, ٧, ٨, ٩

▸ Agreed upon within the Arabic script community to use the Western Arabic numbering system. The other two are strictly prohibited at the registry level

▸ Jawi community special case is to use the Eastern Arabic ٢ in words that repeat themselves twice

راما راما ➔ راما٢ ➔ Butterflies