From Kamal Mansour
Subject Left-to-right directionality in Arabic numeric expressions
In response to L2/22-261 (Part 2)
Date 2023.3.10

Summary
*In an Arabic-script context, multi-part numeric expressions do not maintain the desired left-to-right directionality that readers expect to see. For example, in an Arabic context, the numeric expression "*56 124*" can appear as "*124 56 رقم*" even when set in European-style digits. How can this problem be dealt with?*

It is common to see number-based expression that have a symbolic meaning instead of numeric value. Among the most frequently encountered strings are serial and model numbers of products, as well as phone numbers. Let's examine the case of a typical phone number in international format that can appear in a business listing:

Telephone +33 (0)4 27 63 84 15

If we type the same expression while replacing "Telephone" with its equivalent in Arabic, we obtain the following result:

هاتف +33 (0)4 27 63 84 15

If we enter the same telephone number in Arabic-style numerals, we still obtain a similar result in which the expression appears in right-to-left direction:

هاتف +٣٣ (٠)٤ ٢٧ ٦٣ ٨٤ ١٥

Even when inserting a left-to-right mark [LRM U+200E] before the expressions "+33" or "+٣٣", the displayed result may still disappoint. (We should point out that some text editors and word processors that cleverly analyze a complex string such as the telephone number, may produce the correct result with the simple insertion of LRM.)

We have demonstrated that whether we write in European- or Arabic-style numerals, the flow of the telephone number remains right to left as in a series of Arabic words. Note that each sub-expression of the number (e.g. *27*, *63*, or *15*) continues to flow from left to right. Although Arabic words proceed from right to left, numbers are expected to flow from left to right. Therefore, the expected—and desired—displayed result for the numeric express is:

+33 (0)4 27 63 84 15 هاتف
+٣٣ (٠)٤ ٢٧ ٦٣ ٨٤ ١٥ هاتف

One might justifiably wonder why the above example fails to give the expected results. Is it a failure of the Bidi Algorithm? By examining the international telephone number more analytically, one will see that it does not represent a single number, but a sequence of numeric expressions such as "27 63" separated by white space or common punctuation marks such as *(*, –,

+, or /. So while each number is appearing in left-to-right order, the whole telephone-number string flows from right to left as one would expect of any Arabic text. To obtain the desired result, it is necessary to reverse the direction of the entire telephone-number string. This can be accomplished by inserting a control character that is broader in effect than LRM; we are referring here to the left-to-right override or LRO [U+202D] character. The effect of LRO on text direction persists until it is explicitly terminated by the control character PDF [pop directional formatting, U+202C] or line ending.

By entering the control character LRO immediately before the telephone number (see highlighted position below), the direction of the telephone-number string is set to left to right, and the correct result is obtained:

هاتف 15 84 63 27 4(0) 33+
هاتف ١٥ ٨٤ ٦٣ ٢٧ ٤(٠) ٣٣+

Theoretically speaking, one should be able to utilize a less forceful control character than LRO such as LRI [U+2066], but experiments have shown that, in practice, it isn't as widely supported by applications.

When writing HTML, one should resort to using markup notation instead of control characters. For instance, for the referenced phone number, it suffices to precede the expression with <*span dir="ltr"*>. This would both isolate the expression, while also changing the direction of the text string until terminated with </*span*>.