

Why are leading digits converted to language-specific digit shapes, but not trailing digits, and how do I suppress the conversion entirely?

devblogs.microsoft.com/oldnewthing/20140226-00

February 26, 2014



Raymond Chen

If you have a string like `12345ABCDE67890`, and you render it on an Arabic system, you might get `٠١٢٤ABCDE67890`. The leading digits are rendered as Arabic-Indic digits, but the trailing digits are rendered as European digits. What's going on here? This is a feature known as *contextual digit substitution*. You can specify whether European digits are replaced with native equivalents by going to the *Region* control panel (formerly known as *Regional and Language Options*), clicking on the *Formats* tab, going to *Additional settings* (formerly known as *Customize this format*), and looking at the options under *Use native digits*. The three options there correspond to the three values for LOCAL_IDIGITSUBSTITUTION. Programmatically, you can override the user preference (if you know that you are in a special case, like an IP address) by following the instructions in MSDN.

- Uniscribe: `ScriptApplyDigitSubstitution`
- DWrite: `IDWriteTextAnalysisSink::SetNumberSubstitution`
- GDI: `ETO_NUMERICSLATIN` or `ETO_NUMERICSLLOCAL`.

As a last resort, you can stick a Unicode NODS (U+206F) at the beginning of the string to force European digits, or a Unicode NADS (U+206E) to force national digits. **Bonus chatter:** What's the point of contextual digit substitution anyway? Suppose you have the string "there are 3 items remaining." (Let's say that all text in lowercase is in Arabic.) You want this 3 to be rendered in Arabic-Indic digits because it is part of an Arabic sentence. On the other hand, if you have the string "that's a really nice BMW 350." you want the 350 to be in European digits since it is part of the brand name "BMW 350".

Contextual digit substitution chooses whether to use Arabic-Indic digits or European digits by matching them to the characters that immediately precede them. (And if no character precedes them, then it uses the ambient language.)

[Raymond Chen](#)

Follow

