Break it up, you two!: The zero width non-joiner
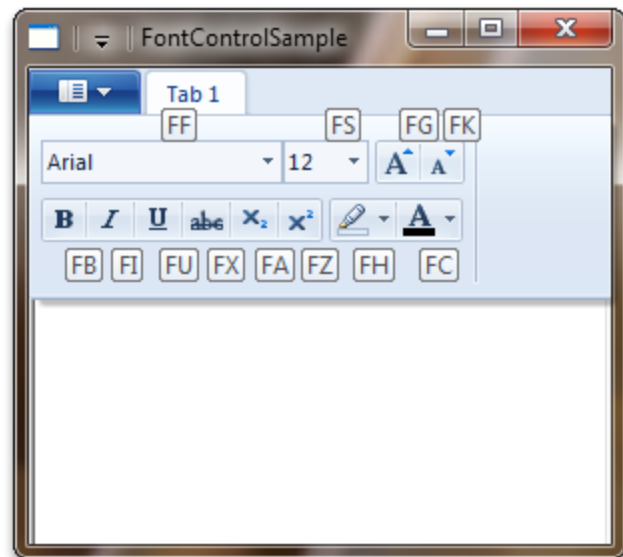
November 14, 2012

Raymond Chen

Keytips are those little pop-up keyboard accelerator thingies that appear on the Ribbon when you tap the Alt key:

A tester discovered that when a test tried to read the accessibility name for a Ribbon keytip, "an extra character appears after every keytip character." In the above example, the keytip for "Tab 1" was being read back as

```
46 00 0C 20 46 00 0C 20
----- ----- ----- -----
  F   ?????   F   ?????
```

The question marks are U+200C, formally known as ZERO WIDTH NON-JOINER. Michael Kaplan discussed the character (and its evil twin the ZERO WIDTH JOINER) some time ago.

The ZERO WIDTH NON-JOINER (or ZWNJ to his friends) is a hint to the font engine that the characters on opposite sides of the ZWNJ should not be combined into a ligature. In English, the ZWNJ would prevent two consecutive lowercase "f"s from being converted into a "ff" ligature. Ligatures are fading from use in contemporary printing, probably due to the rise of computers. Back in the old days, you saw all sorts of neat ligatures, like "st".

Breaking up the ligature is important when presenting keyboard accelerators. Imagine if the keyboard accelerator for a key sequence was "A" followed by "E". If this were displayed as "Æ", users would waste their time looking for an "Æ" key on their keyboard. Although English doesn't have many ligatures any more, many other languages still employ them

heavily. (You may have noticed that the keytip was a bit overzealous with the ZWNJ, putting one at the end of the string even though there was nothing for the second F to be unjoined from!)

So if you encounter one of these ZWNJ characters, don't be afraid. He's just there to break things up. And as Michael notes, ZWNJ and ZWJ "are supposed to be ignored in things like the Unicode Collation Algortihm."

Raymond Chen

Follow