

The history of Alt+number sequences, and why Alt+9731 sometimes gives you a heart and sometimes a snowman

 devblogs.microsoft.com/oldnewthing/20240702-00

July 2, 2024



Raymond Chen

Once upon a time, the IBM PC was released.

In the IBM PC BIOS, you could enter characters that weren't present on the keyboard by holding the **Alt** key and typing the decimal value on the numeric keypad. For example, you could enter ñ by holding **Alt** and typing **Numpad1 Numpad6 Numpad4**, then releasing the **Alt** key.

For expository simplicity, I will henceforth use the notation **Alt + 164** to indicate that you press the **Alt** key, then type the specified digits in sequence on the numeric keypad, then release the **Alt** key.

Okay, so in the IBM PC BIOS, when you typed **Alt + ...**, the code numbers were treated as decimal byte values, and the result on the screen came from your video card's character generator. In the United States, the character generator's ROM showed what we today call Code Page 437.

When it was introduced, Windows in the United States used Code Page 1252 as its 8-bit character set, which it called the "ANSI character set"; the old BIOS character set was retroactively named the OEM character set. To preserve compatibility with MS-DOS, if you used the **Alt** key in conjunction with the numeric keypad, the number you typed was still looked up in OEM character set, so that your muscle-memory code numbers still worked. You could still type **Alt + 164** to get your ñ, even though the code number for ñ in Code Page 1252 is 241, not 164.

If you wanted to type a character that had no OEM equivalent, you could prefix a numeric keypad **0** to indicate that you wanted the value looked up in the ANSI code page. Therefore, you could type **Alt + 0169** to get a ©, which did not exist in the OEM code page. You could also type **Alt + 0241** to get your precious ñ, using the ANSI code point number rather than the OEM code point number.

If you entered a number larger than 255, both Windows and the IBM PC BIOS took your value mod 256, so typing **Alt + 259** was the same as typing **Alt + 3**. Both gave you OEM code point 3, which for Code Page 437 is a heart ♥.

If you ask the Internet how to type some of these non-ASCII characters on Windows, you may see people (and large language models) that tell you to type, say, **Alt + 9731** to get a Unicode snowman . Unfortunately, from what we've learned above, this doesn't work. You instead get the OEM character whose value is $9731 \bmod 256 = 3$, or the aforementioned heart ♥.

A customer reported that a recent Windows update broke their ability to type a snowman by using **Alt + 9731**. We explained that the update was not at fault; rather, **Alt + 9731** was never supposed to produce a snowman at all! But the customer insisted that it used to work.

A closer investigation of the issue revealed the reason.

You see, while it's true that the **Alt + ...** decimal value is taken mod 256, that is just the default behavior of the Windows input system. But some controls (most notably the RichEdit control) override the default handling of the **Alt + ...** sequence and parse out the decimal value *mod 65536* rather than *mod 256*.

This means that whether the **Alt + ...** value is taken mod 256 depends on what kind of control you are typing into.

By default, the value is taken mod 256, and **Alt + 9731** gives you a heart.

But if you happen to be using a RichEdit control, then the **Alt + ...** value is taken mod 65536, and **Alt + 9731** gives you a snowman.

(I don't know of anybody who takes the value mod 2097151, to support direct entry of code points outside the Basic Multilingual Plane.)