

Why does my string consist of this Korean character repeated over and over?

devblogs.microsoft.com/oldnewthing/20191007-00

October 7, 2019



Raymond Chen

A customer reported that their program would sometimes print Chinese text instead of the actual desired value. Your initial reaction is probably, “Oh, I bet I know what’s going on. They’re displaying an ANSI string as if it were Unicode, amirite?”

And then you look at the screen shot.

책 책 책 책 책 책 책

Okay, first of all, that’s not Chinese text. That’s Korean.

But I’ll forgive that error, because to the uninitiated, Chinese, Japanese, and Korean characters look alike: They are all monospace complex symbols. Of course, once you’ve become initiated, you can instantly tell them apart. The hard part is the initiation.

If you look more closely, you may even recognize the character as Unicode code point U+CD.

And that’s the key to the puzzle.

The byte `0xCD` is a common fill byte. Visual Studio uses it in debug mode to represent uninitialized heap memory.

Therefore, the reason for the Korean character repeated over and over is that your so-called string is actually just uninitialized heap memory. Follow the money backward to the function which was supposed to fill it with data, and debug why that function failed. (While you’re at it, you might also want to add error checking, so that when that function fails, you don’t run ahead with uninitialized data.)

Raymond Chen

Follow

