

# Unicode collation is hard

---

 [devblogs.microsoft.com/oldnewthing/20040413-00](https://devblogs.microsoft.com/oldnewthing/20040413-00)

April 13, 2004



Raymond Chen

The principle of “garbage in, garbage out” applies to Unicode collation. If you hand it a meaningless string and ask to compare it to another meaningless string, you get meaningless results.

I am not a Unicode expert; I just play one on the web. A real Unicode expert is Michael Kaplan, whose [explanation of how comparing invalid Unicode strings result in nonsensical results](#) I strongly recommend to those who attempt to generate random test strings in Unicode.

[Raymond Chen](#)

**Follow**

