

Why a really large dictionary is not a good thing

 devblogs.microsoft.com/oldnewthing/20040402-00

April 2, 2004



Raymond Chen

Sometimes you'll see somebody brag about how many words are in their spell-checking dictionary. It turns out that having too many words in a spell checker's dictionary is worse than having too few. Suppose you had a spell checker whose dictionary contained every word in the [Oxford English Dictionary](#). Then you hand it this sentence:

| Therf werre eyght bokes.

That sentence would pass with flying colors, because all of the words in the above sentence are valid English words, though most people would be hard-pressed to provide definitions. The English language has so many words that if you included them all, then common typographical errors would often match (by coincidence) a valid English word and therefore not be detected by the spell checker. Which would go against the whole point of a spell checker: To catch spelling errors. So be glad that your spell checker doesn't have the largest dictionary possible. If it did, it would end up doing a worse job. After I wrote this article, I found [a nice discussion of the subject of spell check dictionary size on the Wintertree Software web site](#).

[Raymond is currently on vacation; this message was pre-recorded.]

[Raymond Chen](#)

Follow

