Why doesn't findstr use the standard regular expression library?

December 9, 2015

Raymond Chen

Tim wonders <u>why there isn't a standard library that was settled on and used by everyone by now</u>. (While you're at it, <u>why isn't there a standard for electrical outlets that was settled on and used by everyone by now</u>?) And the answer is the same: Things started out with everybody doing their own thing, and by the time a standard emerged, it was too late.

The `findstr` program was written in 1990 by a colleague of mine who retired in the year 2000. Let's call him Bob. It was originally written for MS-DOS and was called `qgrep`. I don't know what the `q` stands for, but it couldn't call itself `grep` because that name was already taken. And since this was Bob's little program, he got to choose which regular expression language it accepted.

The `qgrep` program sat in Bob's bag of tricks, and he shared it with his closest friends, who shared it with their friends, and so on. Meanwhile, Bob ported `qgrep` to OS/2 (because he needed a version that ran on OS/2) and eventually Windows NT (because he needed a version that ran on Windows NT).

At this point, `qgrep` caught the attention of the people in charge of the Windows Resource Kit. They were on the prowl for handy little utilities that could be tossed onto the CD. They said, "Hey, can we put `qgrep` on the Resource Kit CD?" Bob said, "Sure, here you go." And then the Resource Kit people said, "Okay, but we are afraid to call it `qgrep` because that might create licensing or trademark problems, so we're going to have to call it something else. We'll call it `findstr`! Also, we'd like to change the command line switches to match the other Resource Kit tools, and we'd like to change the help text based on this feedback from our editors."

"Whatever," said Bob. "I'm going to keep calling mine `qgrep`, thanks the same. Here you go: I created a clone of `qgrep`, renamed it `findstr`, and made the changes you requested."

That's how things stood for several years. Bob had `qgrep`. The Resource Kit had `findstr`, a mutant offshoot of `qgrep`.

One piece of common feedback from system administrators was that a lot of the Resource Kit tools were really handy, but it was pain to have to install them on every computer. And since they aren't part of the core Windows installation, the tools aren't available for use in logon scripts either.

And that's how `findstr` ended up part of Windows. It came in through on the coattails of the Resource Kit. I remember when they were added to Windows because becoming part of the core product meant another round of security reviews.

Okay, that's a nice story, but it doesn't answer the question. Why wasn't `findstr` upgraded to use a newer regular expression engine?

Recall that Bob retired in 2000. And since `qgrep` was Bob's baby, all development on `qgrep` stopped when he retired. When Bob gave the `findstr` project to the Resource Kit team, they got the source code, but there was no knowledge hand-off so that somebody on the Resource Kit team understood how the program worked, in case they needed to fix a bug or add a feature. Not that there was anybody on the Resource Kit team available to receive said knowledge. The Resource Kit was primarily a *book*, so the Resource Kit team consisted mostly of writers and editors, not programmers. (That's probably why they were so excited about changing the help text.) The CD filled with tools was considered a bonus feature, not the primary product. I guess they figured that if they needed a bug fixed or a feature added, they'd just ask Bob.

Besides, you can't change the regular expression language accepted by a program after it has been released, because that would break all the scripts that used the old language. Remember those logon scripts that use `findstr` ? If any of them used a regular expression whose meaning changed between the old syntax and the new syntax, those scripts would subtly stop working properly. A change in the regular expression syntax would require a new switch to opt into the new behavior.

I don't recall Bob ever mentioning to me that somebody asked him to upgrade the regular expression engine in `qgrep` . I suspect nobody asked, seeing as perl-style regular expressions didn't become popular until long after Bob retired. Also, Bob is not a lawyer, so he doesn't want to have to read the license for a third-party library and figure out how to remain in compliance with it.

(From reading the PCRE license, it appears that if your program uses PCRE, you must reproduce "the above copyright notice", but there are *three* copyright notices on that page. Does the program need to reproduce all of them? Or just the last one? It seems to me that nearly everybody just ignores the license requirements. For example, Safari uses PCRE, but the PCRE copyright, licensing terms, and disclaimer do not appear in the Safari EULA or any other Safari documentation I can find.)

Raymond Chen

Follow