

Re: Need help with Perl regex

Source: <http://coding.derkeiler.com/Archive/Perl/comp.lang.perl.misc/2005-01/0528.html>

From: Eric Bohlman (ebohlman_at_omsdev.com)

Date: 01/07/05

Date: 7 Jan 2005 02:45:15 GMT

surfking <bkimelman@JUNK.sympatico.ca> wrote in
news:MPG.1c4798934e49976498986f@news1.on.sympatico.ca:

>
>
> *I found this line of code*
> *which was parsing the /etc/termcap file located on a UNIX system.*
>
> *if (/^(^|\\)\$term}{:|/){*
>
> *I used the code from which this line was extracted and it successfully*
> *parsed/extracted the termcap entry for my particular type of terminal.*
>
> *I realize that the "^" character is used to anchor the pattern match*
> *to the start of a buffer and that enclosing part of a pattern match*
> *within a set of parenthesis enables you to retrieve the value of the*
> *matched segment and that "|" is used as an "logical or" operator, but*
> *given the format of entries in the /etc/termcap file, I don't see how*
> *this pattern is successful. Can anyone out there give me some ideas*
> *on this ?*

Actually, in this case the parentheses are almost certainly being used simply to set precedence.

Let's spread that regex out a bit, which we can actually do in Perl code thanks to the "x" modifier:

```
/ #start regex
( #begin group that's treated as a unit
  ^ #start of the string
  | #logical or
  \\ #a literal pipe character
) #end group
#so in order to match, it has to either be at the beginning of the line
#or preceded by a pipe symbol
${term} #treat whatever is in the variable $term as part of the regex
[ #begin a character class
  : #a literal colon
```

```
\| #a literal pipe character  
] #end character class  
#the character class matches any character that's either a colon or a  
#pipe /x #end regex; the "x" lets us put in spaces and comments
```

So we know that whatever matches has to come either at the beginning of the line or after a pipe symbol, and it has to end with a colon or a pipe. The question is, what's in between? We can't know the answer until we know what's in \$term. I can guess (only guess) that it's simply the name of your terminal and doesn't contain any regex special characters. If that's the case, then the expression will match any line in which the name of your terminal appears either at the beginning or after a pipe, and is immediately followed by either a colon or a pipe. But again, that's just a guess; if \$term contains any regex special characters, they'll be treated the same as if they had been written out in the regex.

The perlretut, perlrequick, perlre, and perlref documents that come with every Perl distribution are the definitive reference for Perl regexes. Start with:

perldoc perlretut

and work your way through them.