

Re: Need help with regular expression

Re: Need help with regular expression

Source: <http://coding.derkeiler.com/Archive/Java/comp.lang.java.help/2007-02/msg00712.html>

- *From:* Hendrik Maryns <hendrik_maryns@xxxxxxxxxxxxxxx>
 - *Date:* Wed, 28 Feb 2007 16:14:10 +0100
-

-----BEGIN PGP SIGNED MESSAGE-----

Hash: SHA1

Lars Enderin schreef:

Hendrik Maryns skrev:

-----BEGIN PGP SIGNED MESSAGE-----

Hash: SHA1

Lars Enderin schreef:

Lars Enderin skrev:

heAzk skrev:

Luc Van Bogaert wrote:

Hi,

I'm trying to
construct a
regular
expression
to check the
occurrence
of
two
substrings
in a String,
but I haven't
yet found
something
that
works.
Could
someone
please help?
This is used

Re: Need help with regular expression

to parse
lines of
HTML
code.

I'm using
`String.matches(regex)`
to find out
if the string
matches
anything
like this :

```
...<!--  
%%...%%  
-->...
```

The dots
can be
replaced
with
anything or
even
nothing.

This doesn't
seem to
work :

```
String.matches(".*"  
+ "<!--"  
+ "%%" + ".*"  
+ "%%"  
+ "-->" + ".*")
```

Thanks,

The `.*` is used to match
zero or more occurrences of
any character.

The `*` modifier is greedy
however. This means that
it'll match as
many times as it can. You
need to match only zero to
three times,
which can be accomplished
by `{lower, upper}`, so in
your case:
`{0,3}`.

Re: Need help with regular expression

I don't think Luc implied that there can be at most three characters between the %--s, and I don't think that String.matches(regex) is the right method to use. It would be better to use something like

```
static Pattern commentPtrn =  
Pattern.compile("<!-- %%(.*)%% -->");
```

Remove the extra ": The pattern should be "<!-- %%(.*)%% -->".

Almost: "<!-- %%(.*)%% -->"; but I do not really see why you would need the parens, "<!-- %%.*?%% -->" should suffice.

The parentheses are not *necessary*, but useful if you want to know what the special comment contains.

/I/ know that, but I wanted to point it out to the OP.

(Your signature should start with a line containing only "-- " (two dashes and a single space). This will enable news/mail clients to recognize it as a signature and, for example, remove it automatically from quoted material in replies.)

(I know, it has to do with Enigmail and Thunderbird messing up PGP signed messages. There is a bug for it somewhere.)

H.

--

Hendrik Maryns
<http://tcl.sfs.uni-tuebingen.de/~hendrik/>

=====

<http://aouw.org>
Ask smart questions, get good answers:
<http://www.catb.org/~esr/faqs/smart-questions.html>

-----BEGIN PGP SIGNATURE-----

Version: GnuPG v1.4.2 (GNU/Linux)

iD8DBQFF5ZxCe+7xMGD3itQRAmOXAJ9I8vSXaKOqmB1Gg5+et6ejZEg8UgCeK1AL
i0EpHyoWfbNyB8Z1D7IIwkY=

Re: Need help with regular expression

=Lzsa

-----END PGP SIGNATURE-----

.