Subject: Regexp is not working
Posted by crydev on Mon, 22 Oct 2012 19:48:05 GMT

View Forum Message <> Reply to Message

My application has a parsing system for retrieving URL's from a string of HTML. I am trying to port this from a C# application which uses a regular expression in multiple places. Now I noticed that the RegExp in U++ is not the same as a regular one like in C# or equal. I have a few regexes from which I know that they don't work in U++ or might not work. Could you help me out with finding a solution to them?

This regex is working.

(<p><b>.*?)\n

I found out that this regular expression is not working at all. The purpose of this Regex is retrieving multiple URL's from <a href...> blocks in HTML text. The input string for this regex is a block of HTML text in which 0 or more <a href...> blocks are located.

(<a href=.*?)</td></tr>

I also doubt the functionality of this regex. I didn't test it yet, since I have to match the regex above to have the correct data available. This regex should retrieve the exact URL in the href="..." part of the output of the above regex. Is this a correct regular expression for the U++ RegExp class?

<a href=\"\"(.{17})

Thanks in advance!

Subject: Re: Regexp is not working
Posted by mirek on Tue, 23 Oct 2012 06:58:33 GMT

View Forum Message <> Reply to Message

U++ RegExp is based on PCRE, perhaps PCRE docs will help:

http://perldoc.perl.org/perlre.html

Subject: Re: Regexp is not working
Posted by crydev on Sun, 11 Nov 2012 19:22:53 GMT

View Forum Message <> Reply to Message

Sorry for such a late response. I fixed my problem by using the C++ version 11 regex in the STD library.

http://www.cplusplus.com/reference/std/regex/