
Subject: What's wrong in this regexp ? [SOLVED, FIXED]

Posted by [awlee](#) on Thu, 17 May 2012 14:08:58 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi,

I have some problem with this code and I don't understand what's wrong

```
#include <CtrlLib/CtrlLib.h>
#include <plugin/pcr/pcr.h>
#include <Web/Web.h>
```

```
using namespace Upp;
```

```
GUI_APP_MAIN
```

```
{
    String content;
    HttpClient web;
    web.URL("http://ultimatepp.org");
    content = web.ExecuteRedirect(3,3); // read html page
```

```
String regexp = "http-equiv=\"content-type\"[^\"]+charset=([^\s\"]+).*?<title>([^\"]+)\"";
```

```
RegExp r0(regexp, PCRE_CASELESS|PCRE_MULTILINE|PCRE_DOTALL);
```

```
PromptOK("Test1");
if (r0.Match(content))
    if (r0.GetCount() > 0)
        for (int i=0; i<r0.GetCount(); ++i)
            PromptOK(r0[i]);
```

```
PromptOK("Test2");
content = "<meta http-equiv=\"content-type\" content=\"text/html; charset=utf-8\" /><meta
name=\"generator\" content=\"U++ HTML Package\"><title>Wooooow!</title>";
if (r0.Match(content))
    if (r0.GetCount() > 0)
        for (int i=0; i<r0.GetCount(); ++i)
            PromptOK(r0[i]);
```

```
PromptOK("End");
```

```
}
```

At "Test1" it won't show anything

At "Test2" all is ok

P.S. Windows 7 Home Basic x64, MSSDK 7.1 x64, U++ b4193

Subject: Re: What's wrong in this regexp ?
Posted by [sergeynikitin](#) on Thu, 17 May 2012 16:42:30 GMT
[View Forum Message](#) <> [Reply to Message](#)

I recommend free open-source software KIKI for testing regular expression. There is in standard repositories under linux. For Windows use this link
<http://code.google.com/p/kiki-re/downloads/list>

(I use ver.0.5.6)
Very simple interface.

Subject: Re: What's wrong in this regexp ?
Posted by [awlee](#) on Thu, 17 May 2012 16:52:48 GMT
[View Forum Message](#) <> [Reply to Message](#)

I think I was wrong with description of my problem...

- regular expression is the same in both cases
- results are different
- we can save HttpClient response and we'll see html page

why this regexp works so different?

both strings without specific UTF8 chars (I think)

Subject: Re: What's wrong in this regexp ?
Posted by [dolik.rce](#) on Thu, 17 May 2012 17:05:11 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi awlee!

It seems like U++ doesn't support PCRE_DOTALL flag correctly at this moment. For a quick workaround, add '(?s)' switch at the beginning of your regexp: String regexp =
"(?s)http-equiv=\"content-type\"[^\"]+charset=([^\s\"]+).*?<title>([^\"]+)\"";

I'll try to put together a patch to support this (and other) missing feature soon.

Best regards,
Honza

Subject: Re: What's wrong in this regexp ?
Posted by [sergeynikitin](#) on Thu, 17 May 2012 17:06:33 GMT
[View Forum Message](#) <> [Reply to Message](#)

I suggest to say what you want to do with this expression.

Subject: Re: What's wrong in this regexp ?
Posted by [awlee](#) on Thu, 17 May 2012 17:09:15 GMT
[View Forum Message](#) <> [Reply to Message](#)

dolik.rce wrote on Thu, 17 May 2012 21:05Hi awlee!

It seems like U++ doesn't support PCRE_DOTALL flag correctly at this moment. For a quick workaround, add '(?s)' switch at the beginning of your regexp: String regexp = "(?s)http-equiv=\"content-type\"[>]+charset=([^\s\"]+).*?<title>([<]+)";

I'll try to put together a patch to support this (and other) missing feature soon.

Best regards,
Honza

Yep.. Thanks a lot for workaround. It works for me

Question resolved.
=)

Subject: Re: What's wrong in this regexp ?
Posted by [dolik.rce](#) on Thu, 17 May 2012 18:16:09 GMT
[View Forum Message](#) <> [Reply to Message](#)

Issue reported and patch submitted to Mirek for check

You can see it all here: <http://www.ultimatepp.org/redmine/issues/286>

Honza

Subject: Re: What's wrong in this regexp ?
Posted by [dolik.rce](#) on Sun, 20 May 2012 09:42:10 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi awlee,

The issue should be fixed in version 4970. All the PCRE_* options should work now, even without the (?s) workaround, see plugin/pcre/RegExp.h for details (such as list of all recognized options).

Best regards,
Honza

Subject: Re: What's wrong in this regexp ?
Posted by [awlee](#) on Tue, 22 May 2012 22:33:09 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks a lot for quick answer and fix!

patch applied by hands)
