Subject: RegExp: newlines in replacement string
Posted by norbert1968 on Thu, 22 Feb 2018 09:40:08 GMT

View Forum Message <> Reply to Message

I have a question regarding Upp::Regex. As I am not a very experienced programmer, my apologies if this is a stupid question.

In RegExp::ReplaceGlobal(String& t, const String& r, bool backref=false), replacements containing newlines are not processed. However, when using a Vector<String> as argument, this problem does not arise. Thus,

```
#include <Core/Core.h>
#include <plugin/pcre/Pcre.h>
using namespace Upp;

String s1 = "Too \n\n many \n\n\n   newlines!", s2 = s1;
RegExp reg;
reg.SetPattern("(\\s*\\n\\s*)"); // find (multiple) newlines and adjacent white space
String repl = "(\n)";           // trim to single newline
reg.ReplaceGlobal(s1, repl);    // doesn't work as expected, but deletes any match
DUMP (s1);

reg.ReplaceGlobal(s2, Vector<String>({"\n"}));    // does work
DUMP (s2);
```

yields the following output:
s1 = Toomanynewlines!
s2 = Too
many
newlines!

The reason seems to be that RegExp::Make_rv(const String&) uses a regex itself to identify text within parenthesis, and this regex apparently does not use the "dot matches newline" option. Is this behaviour by design? I would prefer the s1 example to work as expected.