
Subject: RegExp support

Posted by [hojtsy](#) on Mon, 03 Apr 2006 22:05:13 GMT

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

I was playing around with pcre library, and found out that it provides two matching algorithms:

The NFA algorithm (pcre_exec) doesn't find the longest possible match

The DFA algorithm (pcre_dfa_exec) doesn't support backreferences.

I see it as a drawback that pcre can not simultaneously provide these two features. For example boost regex library is able to do that.

I think this aspect should be taken into account before selecting pcre for integration into U++ library.

Subject: Re: RegExp support

Posted by [mirek](#) on Mon, 03 Apr 2006 22:35:00 GMT

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

hojtsy wrote on Mon, 03 April 2006 18:05I was playing around with pcre library, and found out that it provides two matching algorithms:

The NFA algorithm (pcre_exec) doesn't find the longest possible match

The DFA algorithm (pcre_dfa_exec) doesn't support backreferences.

I see it as a drawback that pcre can not simultaneously provide these two features. For example boost regex library is able to do that.

I think this aspect should be taken into account before selecting pcre for integration into U++ library.

Thanks!

Mirek

Subject: Re: RegExp support

Posted by [mirek](#) on Mon, 10 Apr 2006 09:02:25 GMT

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

BTW, right at the moment, RegExes are quite enemy territory for me.

Perhaps you or somebody else should step in and tried to create the right "plugin" package, ready for subsequent integration with String (WString, maybe even Stream?).

BTW, Stream variant would be usable for String as well via addapting it using StringStream - performance wise, it is quite cheap. But there then remains the unicode problem... maybe use utf-8 only? E.g. in CodeEditor (in TheIDE), all search operations are performed in utf-8 anyway, ditto for files in most cases.

Mirek

Subject: Re: RegExp support
Posted by [c420](#) on Sun, 05 Nov 2006 18:35:25 GMT
[View Forum Message](#) <> [Reply to Message](#)

This is simple class that use libpcre, to support Regular Expresion in Upp.
I wrote him to validate some EditFields in my application.

File Attachments

1) [RegExp.tar.gz](#), downloaded 1798 times

Subject: Re: RegExp support
Posted by [unodgs](#) on Sun, 05 Nov 2006 19:57:10 GMT
[View Forum Message](#) <> [Reply to Message](#)

I wrote something similar and it's available in uppbbox. If you have time please check it. Maybe there is something in my wrapper that isn't in yours (but I doubt it). Anyway If you could continue your work we could include it in upp as a plugin (pcre is quite good, although some of upp users claim that it lacks some very advanced fetures that are important to them).

Subject: Re: RegExp support
Posted by [mirek](#) on Sun, 05 Nov 2006 22:05:44 GMT
[View Forum Message](#) <> [Reply to Message](#)

Actually, recently I was thinking about this issue - maybe we can use PCRE now and switch to something better later, with the same interface?

Mirek

Subject: Re: RegExp support
Posted by [agent86](#) on Fri, 19 Jan 2007 06:18:41 GMT
[View Forum Message](#) <> [Reply to Message](#)

It seems I need/want regular expression capability in theIDE, and offer the following for consideration. I have a public domain implementation which I have been using in my work for awhile now. (excerpt from comments from the file)

- * These routines are the PUBLIC DOMAIN equivalents of regex
- * routines as found in 4.nBSD UN*X, with minor extensions.
- *

and I include the entire set of comments for your review. Note that I made some mildly important additions. I am not an expert in the legalities, etc. but these routines do work. They find the longest match and have back references. See what you think, If you want to take a better look, I'd be happy to contribute them. Just let me know how.

chuck

File Attachments

1) [regex_comments.txt](#), downloaded 529 times

Subject: Re: RegExp support

Posted by [mirek](#) on Fri, 19 Jan 2007 09:04:58 GMT

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

PCRE RegExp support was added to the library month ago.

BTW, about TheIDE: wildcards are nowhere as powerful as full regular expressions, but for 99% of cases, they are quite handy (and perhaps easier to use).

Mirek
