

Forum:[Hack Forums](#)

Thread: [Pattern Matching with PCRE and SLRE](#)

geekmorpheus403 - November 26, 2014, 3:25 am

Ok,

I have been recently trying to do a pattern matching with regular expressions. I have to do this in C (No other language). I have used slre library and wrote this:

```
Code #include <stdio.h>
```

```
#include <stdlib.h>
```

```
#include <string.h>
```

```
#include "slre.h"
```

```
#include "slre.c"
```

```
int main(void)
```

```
{
```

```
    char* mystring = "potty <body class=\"admin-yokohoma admin-235\"> lovely baby :*";
```

```
    struct slre_cap caps[2];
```

```
    static const char* regex = "<body class=\".*(admin-).*(admin-).*\>";
```

```
    if(slre_match(regex, mystring, strlen(mystring), caps, 2, 1) > 0)
```

```
    {
```

```
        char buffer[100];
```

```
        printf("We have found it! Yay!\n");
```

```
        char* ptr = caps[0].ptr + 6;
```

```
        char* loc = strchr(ptr, ' ');
```

```
        int ttlsz = (int) (loc - ptr);
```

```
        printf("%d\n", ttlsz);
```

```
        strncpy(buffer, ptr, ttlsz);
```

```
        buffer[ttlsz] = '\0';
```

```
        printf("Matched:%s", buffer);
```

```
    }
```

```
}
```

This works fine for short string. But my problem is that I will have to check for this pattern in a whole web page source code about 60000 characters but unfortunately slre is not able to handle it for some reason. Whenever I use this code in that program to look for this pattern in webpage, it never ends. There is no loop. The program doesn't terminate. And with smaller strings it is ok. But slow.

So I wanted to use PCRE. But I am not getting it how to use it. I tried downloading the library but pcre.h is not there. Can somebody show me how to use PCRE like this code. Starting from the download. Please I really need it. Or show me some other effective way to handle huge strings and matching patterns please.

Best Regards.

-ceramic- - November 26, 2014, 6:44 am

Regex is not suited to parse SGML.

Try this XML parser, it's really lightweight and will run a LOT faster

hxxp://tibleiz.net/asm-xml/tutorial.html

Also see

[hxxp://stackoverflow.com/questions/6751105/why-its-not-possible-to-use-regex-to-parse-html-xml-a-formal-explanation-in-la](https://stackoverflow.com/questions/6751105/why-its-not-possible-to-use-regex-to-parse-html-xml-a-formal-explanation-in-la)

Mouse? - November 26, 2014, 6:48 am

<https://hackforums.net/showthread.php?tid=4553947> Don't make a new thread for the same problem.

geekmorpheus403 - November 26, 2014, 7:21 pm

(Nov 26, 2014 07:44 AM)-ceramic- Wrote: [\[CITE\]](#) Regex is not suited to parse SGML. Try this XML parser, it's really lightweight and will run a LOT faster

[hxxp://tibleiz.net/asm-xml/tutorial.html](http://tibleiz.net/asm-xml/tutorial.html)

Also see

[hxxp://stackoverflow.com/questions/6751105/why-its-not-possible-to-use-regex-to-parse-html-xml-a-formal-explanation-in-la](https://stackoverflow.com/questions/6751105/why-its-not-possible-to-use-regex-to-parse-html-xml-a-formal-explanation-in-la)

Thanks Bro for clearing all the confusions,

I have looked it up. But can you give me a sample code how to use it. I have a huge string which contains HTML source. I want to look for that specific string pattern and return its matches. Can you show me with Asm-XML. I googled it I couldn't find helpful examples.

Best Regards.