

Forum: [Hack Forums](#)

Thread: [Regex from website source](#)

Soldiercrimes - January 5, 2018, 5:28 am

Code

6.1

i have this in a website source and im using this to capture the 6.1:

Code	Dim r As New
------	--------------

```
System.Text.RegularExpressions.Regex("&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&~([\"'\""]*)</td>")  
Dim matches As MatchCollection = r.Matches(response)
```

For Each match As Match In matches

```
captureee = match.Value.ToString()
Next
```

[illegible]

```
Dim finalcap As String = Regex.Replace(cap, "^\s+$[\r\n]*", "",  
RegexOptions.Multiline)
```

But this returns the 6.1 With another blank line.. and i cant use the blank line...

Dialog - January 5, 2018, 6:14 am

You can of course [Trim](#) the result. But eh... you really shouldn't be using regex to "parse" HTML.

Soldiercrimes - January 5, 2018, 7:23 am

(Jan 5, 2018 07:14 AM)Dialog Wrote: [\[CITE\]](#) You can of course [Trim](#) the result. But eh... you really shouldn't be using regex to "parse" HTML.

What else then?

Dialog - January 5, 2018, 7:26 am

(Jan 5, 2018 08:23 AM)Soldiercrimes Wrote: [\[CITE\]](#) What else then?
HTML parsers sound like they're decent at parsing HTML. Consider HTML Agility Pack:
<http://html-agility-pack.net/>.

Soldiercrimes - January 5, 2018, 7:35 am

(Jan 5, 2018 08:26 AM)Dialog Wrote: [\[CITE\]](#) HTML parsers sound like they're decent at parsing HTML. Consider HTML Agility Pack: <http://html-agility-pack.net/>.
Hmm regex and Trim() do the job for now :p thanks dude

Barret - January 5, 2018, 8:04 am

(Jan 5, 2018 07:14 AM)Dialog Wrote: [\[CITE\]](#) You can of course [Trim](#) the result. But eh... you really shouldn't be using regex to "parse" HTML.
Dialog is right, regex is for "regular text" which html is not, regex is not self correcting or forgiven for mistakes in the markup, it is not recommended.

System[X] - January 5, 2018, 1:28 pm

Quickly done this but not tested.

Regex pattern
Code " (\n+\s+)(\d\.\d)(\n+\s+)</td>"
Then use match.Value(2)

Thanks

Edit: depending on the rest of the html code you could just use (\d\.\d) but if there are more number.number sequence it wont work

Soldiercrimes - January 5, 2018, 2:13 pm

(Jan 5, 2018 02:28 PM)System[X] Wrote: [\[CITE\]](#) Quickly done this but not tested.

Regex pattern
Code " (\n+\s+)(\d\.\d)(\n+\s+)</td>"
Then use match.Value(2)

Thanks

Edit: depending on the rest of the html code you could just use (\d\.\d) but if there are more number.number sequence it wont work
Thanks for the tip man, it's already working with the Trim function but I will keep this in mind.

System[X] - January 5, 2018, 2:35 pm

(Jan 5, 2018 03:13 PM)Soldiercrimes Wrote: [\[CITE\]](#) Thanks for the tip man, it's already working with the Trim function but I will keep this in mind.
Using a better regex pattern removes the need to use Trim or Split but then again Trim and Split is only an extra 2 lines of code :)

Thanks

Dialog - January 5, 2018, 3:55 pm

(Jan 5, 2018 03:35 PM)System[X] Wrote: [\[CITE\]](#) Using a better regex pattern removes the need to use Trim or Split but then again Trim and Split is only an extra 2 lines of code :)

Thanks

Using regex to parse HTML is for retards.

Ninja - January 5, 2018, 4:20 pm

(Jan 5, 2018 04:55 PM)Dialog Wrote: [\[CITE\]](#) Using regex to parse HTML is for retards.
If it is a long document and you need a single string parsed then it is not always bad to use.

Dialog - January 5, 2018, 4:38 pm

(Jan 5, 2018 05:20 PM)Ninja Wrote: [\[CITE\]](#) If it is a long document and you need a single string parsed then it is not always bad to use.
Yes it is always bad to use.

Fire - January 5, 2018, 5:40 pm

Learn how to use xpath instead.

Ninja - January 5, 2018, 7:29 pm

(Jan 5, 2018 06:40 PM)90210 Wrote: [\[CITE\]](#) Learn how to use xpath instead.
HtmlAgilityPack has XPath, and Chrome has an option where you can right click an element and select "copy as XPath" or something like that.

System[X] - January 6, 2018, 1:28 am

(Jan 5, 2018 04:55 PM)Dialog Wrote: [\[CITE\]](#) Using regex to parse HTML is for retards.
How is it not good to use regex for a simple task like this just to grab "6.1" ?

Yes lets make it better by using the agility pack but over complicates things when a single regex pattern can do the job just as good.

Il it comes down to is what you prefer and we all prefer different approaches to solve tasks.

Thanks

Soldiercrimes - January 6, 2018, 2:59 am

(Jan 6, 2018 02:28 AM)System[X] Wrote: [\[CITE\]](#) How is it not good to use regex for a simple task like this just to grab "6.1" ?

Yes lets make it better by using the agility pack but over complicates things when a single regex pattern can do the job just as good.

Il it comes down to is what you prefer and we all prefer different approaches to solve tasks.

Thanks

Yeah its to only grab the 6.1 in a huge source file so regex is the best option in my opinion.

Sn0wX - January 14, 2018, 12:24 pm

(Jan 5, 2018 05:38 PM)Dialog Wrote: [\[CITE\]](#) Yes it is always bad to use.
That's not stricly true. The post from stackforums that you all clearly go by, since your own lack of experience with regex is nothing at best. For Once i'l