Subject: Stregex question - extracting substring
Posted by PMan on Thu, 19 Dec 2013 20:44:06 GMT
View Forum Message <> Reply to Message

I am trying to extract a substring after a colon. My string looks like:

xxxxxx:yyyyyyyyy

where x's can be spaces, upper and lower case letters

then there's the colon

and finally the y's, which can be pretty much anything and will often include colons (which is why strsplit("xxxxxx:yyyyyyyyy", ":") won't work for me here)

I just want the yyyyyyyyy part and have been trying to extract it with stregex, but no luck. But before I give up and will try a different approach, does anyone know how to construct a regular expression for IDL that would extract just the yyyy.. part with stregex?

Thanks for your time.

Subject: Re: Stregex question - extracting substring
Posted by John Correira on Thu, 19 Dec 2013 20:53:08 GMT
View Forum Message <> Reply to Message

On 12/19/2013 03:44 PM, Paul Mallas wrote:
> I am trying to extract a substring after a colon. My string looks
> like:
>
> xxxxxx:yyyyyyyyy
>
> where x's can be spaces, upper and lower case letters
>
> then there's the colon
>
> and finally the y's, which can be pretty much anything and will often
> include colons (which is why strsplit("xxxxxx:yyyyyyyyy", ":") won't
> work for me here)
>
> I just want the yyyyyyyyy part and have been trying to extract it
> with stregex, but no luck. But before I give up and will try a
> different approach, does anyone know how to construct a regular
> expression for IDL that would extract just the yyyy.. part with
> stregex?
>
> Thanks for your time.
>

Not a stregex solution, but I think something like

strjoin((strsplit(string,':',/extract))[1:*],':')

would do it.

John

Subject: Re: Stregex question - extracting substring
Posted by Helder Marchetto on Thu, 19 Dec 2013 21:09:41 GMT
View Forum Message <> Reply to Message

On Thursday, December 19, 2013 9:53:08 PM UTC+1, John Correira wrote:
> On 12/19/2013 03:44 PM, Paul Mallas wrote:
>
>> I am trying to extract a substring after a colon. My string looks
>
>> like:
>
>>
>
>> xxxxxx:yyyyyyyyy
>
>>
>
>> where x's can be spaces, upper and lower case letters
>
>>
>
>> then there's the colon
>
>>
>
>> and finally the y's, which can be pretty much anything and will often
>
>> include colons (which is why strsplit("xxxxxx:yyyyyyyyy", ":") won't
>
>> work for me here)
>
>>
>
>> I just want the yyyyyyyyyy part and have been trying to extract it
>
>> with stregex, but no luck. But before I give up and will try a
>

>> different approach, does anyone know how to construct a regular
>
>> expression for IDL that would extract just the yyyy.. part with
>
>> stregex?
>
>>
>
>> Thanks for your time.
>
>>
>
>
>
>
>
>
> Not a stregex solution, but I think something like
>
>
>
> strjoin((strsplit(string,':',/extract))[1:*],':')
>
>
>
> would do it.
>
>
>
> John

Hi John
I just answered and the original post was deleted, so my answer was dumped.
My solution was: strmid(str,strpos(str,':')+1)
It's good as long as there is at least one ':'. (if necessary one can check for it...).

Cheers,
h

Subject: Re: Stregex question - extracting substring
Posted by Phillip Bitzer on Thu, 19 Dec 2013 22:31:40 GMT
View Forum Message <> Reply to Message

OK, how about a STREGEX solution:

str = ['xxxxxxx:yyy', 'xxxxxxx:yyyyyyy', 'x:yyyy'] ;define the string array

yStrColon = STREGEX(str, ':.+$', /EXTRACT) ;get everything past the colon, including the colon

yStr = STRMID(yStrColon, 1) ;strip the colon

About that regular expression:

: get the substring starting with the colon
.+ get one or more instances of the "dot" (so, any character)
$ anchor at the end of the string

Subject: Re: Stregex question - extracting substring
Posted by PMan on Fri, 20 Dec 2013 15:53:32 GMT
View Forum Message <> Reply to Message

On Thursday, December 19, 2013 5:31:40 PM UTC-5, Phillip Bitzer wrote:
> OK, how about a STREGEX solution:
>
>
>
> str = ['xxxxxxx:yyy', 'xxxxxxx:yyyyyyy', 'x:yyyy'] ;define the string array
>
>
>
> yStrColon = STREGEX(str, ':.+$', /EXTRACT) ;get everything past the colon, including the colon
>
>
>
> yStr = STRMID(yStrColon, 1) ;strip the colon
>
>
>
> About that regular expression:
>
>
>
> : get the substring starting with the colon
>
> .+ get one or more instances of the "dot" (so, any character)
>
> $ anchor at the end of the string

I figured out my solution, just moments after my post (isn't that always the case??) Here is what I did, which is similar to Philip's approach:

I was starting with an array of strings, hence the indices at the end:
x = (stregex(splLines, ':(.*)$' , /extract, /sub))[1, *]

(x = (stregex(splLines, ':(.*)$' , /extract, /sub))[1] for a single string)

The () encapsulates what the /sub extracts and [1] index gets the subexpression only, the 0th element starts with the colon, which I did not want.

Thanks for the reply to my short-lived post!
