Subject: Re: again strsplit
Posted by Paolo Grigis on Wed, 21 Jan 2004 14:45:56 GMT
View Forum Message <> Reply to Message

Hallo Reimar,

is not $ a special command character for regular expressions?
If you enclose it in [] braces, then it is not interpreted, I
think.

IDL> print,a
ABC$DEF
IDL> print,strsplit(a,'[$]',/extr,/regex)
ABC DEF

Cheers,
Paolo

Reimar Bauer wrote:
> Hi all
>
> I have learned the /regex keyword but who could explain this:
>
> a='ABC$DEF'
> print,strsplit(a,'$',/extr)
>
> ABC DEF
>
> and
>
> a='ABC$DEF'
> print,strsplit(a,'$',/extr,/regex)
>
> ABC$DEF
>
>
> Why is this differnt?
>
>
> regards
>
> Reimar

Reimar Bauer wrote:
> Hi all
>
> I have learned the /regex keyword but who could explain this:
>
> a='ABC$DEF'
> print,strsplit(a,'$',/extr)
>
> ABC DEF
>
> and
>
> a='ABC$DEF'
> print,strsplit(a,'$',/extr,/regex)
>
> ABC$DEF
>
>
> Why is this differnt?
>

Hi Reimar,

Funny, I was just fiddling with this, too.

I assume that you want to split the string into substrings using the '$'
character. To do so, you need to use the escape character
(backslash)before the '$' since '$' is also a REGEX control character.

IDL> a = 'ABC$DEF'
IDL> print,strsplit(a,'\$',/extr,/regex)
ABC DEF

Cheers,
Ben

Subject: Re: again strsplit
Posted by tam on Wed, 21 Jan 2004 14:48:08 GMT

Reimar Bauer wrote:

> Hi all
>
> I have learned the /regex keyword but who could explain this:

>
> a='ABC$DEF'
> print,strsplit(a,'$',/extr)
>
> ABC DEF
>
> and
>
> a='ABC$DEF'
> print,strsplit(a,'$',/extr,/regex)
>
> ABC$DEF
>
>
> Why is this differnt?
>
>
> regards
>
> Reimar

Presuming IDL uses the same conventions for regular
expressions that have been popularized in Perl
I'd assume that if you've turned on regular expressions
that an unescaped $ has the special meaning of matching
the end of the string (and ^ the beginning).

So
    print,strsplit(a,'f$', /extr, /regex)
will try to split on a terminal f and you should get

    ABC$DE

as the result.

If you want to match with the character '$' then just escape
it with a backslash.

    print,strsplit(a,'\$',/extr,/regex)

 Regards,
 Tom McGlynn

Subject: Re: again strsplit
Posted by savoieNoSpam on Wed, 21 Jan 2004 16:42:41 GMT

Reimar Bauer <R.Bauer@fz-juelich.de> writes:

> Hi all
>
> I have learned the /regex keyword but who could explain this:
>
> a='ABC$DEF'
> print,strsplit(a,'$',/extr)
>
> ABC DEF
>
> and
>
> a='ABC$DEF'
> print,strsplit(a,'$',/extr,/regex)
>
> ABC$DEF
>
>
> Why is this differnt?

Because '$' is the regular expression for end of line. If you want to have
the actual dollar sign, you need to escape it with a back slash first.

IDL> print,strsplit(a,'\$',/extr,/regex)
ABC DEF

Helps?

Matt

--
Matthew Savoie - Scientific Programmer
National Snow and Ice Data Center
(303) 735-0785   http://nsidc.org

Subject: Re: again strsplit
Posted by R.Bauer on Thu, 22 Jan 2004 14:59:00 GMT

savoieNoSpam@nsidc.org wrote:
> Reimar Bauer <R.Bauer@fz-juelich.de> writes:
>
>
>> Hi all

>>
>> I have learned the /regex keyword but who could explain this:
>>
>> a='ABC$DEF'
>> print,strsplit(a,'$',/extr)
>>
>> ABC DEF
>>
>> and
>>
>> a='ABC$DEF'
>> print,strsplit(a,'$',/extr,/regex)
>>
>> ABC$DEF
>>
>>
>> Why is this differnt?
>
>
> Because '$' is the regular expression for end of line. If you want to have
> the actual dollar sign, you need to escape it with a back slash first.
>
> IDL> print,strsplit(a,'\$',/extr,/regex)
> ABC DEF
>
> Helps?
>
> Matt
>

Dear all

Thanks,

Is there a function available which returns the key codes of regular
expression signs.

This is now a bit complicated. I have to use /regex because the
separation could be done by more as one sign. But if it is a $ or
something else of a regex code I have to add a \ sign in front.

If a variable is used for the separator have I always to test on this?

The old str_sep routine without regex was much clearer.

regards

Reimar

Subject: Re: again strsplit
Posted by tam on Thu, 22 Jan 2004 16:11:06 GMT
View Forum Message <> Reply to Message

Reimar Bauer wrote:

> savoieNoSpam@nsidc.org wrote:
>
>> Reimar Bauer <R.Bauer@fz-juelich.de> writes:
>>
>>
>>> Hi all
>>>
>>> I have learned the /regex keyword but who could explain this:
>>>
>>> a='ABC$DEF'
>>> print,strsplit(a,'$',/extr)
>>>
>>> ABC DEF
>>>
>>> and
>>>
>>> a='ABC$DEF'
>>> print,strsplit(a,'$',/extr,/regex)
>>>
>>> ABC$DEF
>>>
>>>
>>> Why is this differnt?
>>
>>
>>
>> Because '$' is the regular expression for end of line. If you want to
>> have
>> the actual dollar sign, you need to escape it with a back slash first.
>> IDL> print,strsplit(a,'\$',/extr,/regex)
>> ABC DEF
>>
>> Helps?
>>
>> Matt
>>
>
>
> Dear all

>
> Thanks,
>
>
> Is there a function available which returns the key codes of regular
> expression signs.
>
> This is now a bit complicated. I have to use /regex because the
> separation could be done by more as one sign. But if it is a $ or
> something else of a regex code I have to add a \ sign in front.
>
> If a variable is used for the separator have I always to test on this?
>
> The old str_sep routine without regex was much clearer.
>
> regards
> Reimar
>
>

When dealing with regular expressions I normally escape all
non-alphanumeric characters. Not all of them are special, but it's
probably easiest to assume they are. It doesn't hurt to
escape a character that is not special (at least in
other languages).

Of course if you don't have a need for regular expressions, there
is no need to use that keyword.

 Tom