

Re: stripping out ASCII chars using regexp?

Source: <http://coding.derkeiler.com/Archive/Perl/comp.lang.perl.misc/2004-03/1649.html>

From: Anno Siegel (anno4000_at_lublin.zrz.tu-berlin.de)

Date: 03/18/04

Date: 18 Mar 2004 11:55:28 GMT

Greg <djbitchpimp@snowboard.com> wrote in comp.lang.perl.misc:

> *I am trying to get rid of the ASCII chars from the end of a string*
> *that I download from a webpage using LWP::Simple. The script downloads*
> *the HTML from a webpage and then uses HTML::TableExtract to extract*
> *the information from specific tables on the page.*
>
> *This basically gives me a string like this:*
>
> *���,Temper Tantrum�,Take Care Comb Your*
> *Hair�,;,CD�,23.56��,;�,0 Days*
> *Ago�,Scotla�*
>
> *which I then split into an array using:*
>
> *my @line = split (',', \$line) ;*
>
> *I then do a comparison on \$line [6]:*
>
> *if (\$line [6] >= 75) {*
>
> *... do something*
>
> *When I run this using -w, I get the following error:*
>
> *Argument "15.99M- M- " isn't numeric in numeric ge (>=) at*
> *./parse_wants.pl line 49.*

You seem to be confused about what's in your string. With your data, \$line[6] is "23.56��", not "15.99M- M- ".

> *This is because somehow some extended ASCII chars got in the end of*
> *the string. If I do:*

Somehow? So they shouldn't be there and you don't know how they get there?

That's a reason to check the logic where these elements are produced. Fixing this by deleting the unwanted characters is nothing but band aid.

comp.lang.perl.misc: Re: stripping out ASCII chars using regexp?

The bug remains.

[snip attempts]

> *Can anyone figure out a way to get rid of these trailing ASCII
> characters using a regular expression?*

The only correct way is not to pick them up from wherever they
come from.

Anno