

## Re: Perl script to clean up file -- Dont know if it can be done

**Source:** <http://coding.derkeiler.com/Archive/Perl/comp.lang.perl.misc/2004-09/1721.html>

---

**From:** thundergnat ([thundergnat\\_at\\_hotmail.com](mailto:thundergnat_at_hotmail.com))

**Date:** 09/22/04

Date: Wed, 22 Sep 2004 11:15:00 -0400

LHradowy wrote:

> *I thought I would throw this out there, I think it can not be done, but I am  
> not a guru.*

>

> *This is the problem I get a file that I must pull the pertinent data out. I  
> has a header and footer, as well as page breaks, this is all in ASCII  
> format. I need to pull out just the columns.*

> *I do this all manually (delete the header and footer, and well as all the  
> page breaks) there are also at times a 0 at the beginning of a record that I  
> do not want there as well.*

>

> *This is what the file looks like...*

>

[snip]

> *0 1555200 00 0 12 02 CUSTOMER HAS  
> 2555206 00 0 05 01 CUSTOMER HAS  
> 4555208 00 0 03 06 TELN NOT BILL*

[snip]

> *0 1555200 00 0 12 02 CUSTOMER HAS  
> 2555206 00 0 05 01 CUSTOMER HAS  
> 4555208 00 0 03 06 TELN NOT BILL*

[snip]

What tranforms are you applying to change the numbers?  
Where did the extra CUSTOMER HAS line come from?

>

> *I manually disect this file to make it look like this...*  
> *1555002 00 0 04 27 TELN NOT BILL  
> 3555007 00 0 06 00 CUSTOMER HAS*

comp.lang.perl.misc: Re: Perl script to clean up file -- Dont know if it can be done

> 5555410 00 0 12 10 CUSTOMER HAS  
> 6755012 00 0 12 06 CUSTOMER HAS  
>

[snip]

What tranforms are you applying to change the numbers?  
What happened to the other CUSTOMER HAS line?

>  
> *And I get this: WHAT I NEED!*  
> 5555002,00 0 04 27,TELN NOT BILL  
> 1555007,00 0 06 00,CUSTOMER HAS  
> 2555010,00 0 12 10,CUSTOMER HAS  
>  
> *I want to try to eliminate as much manual intervention as I can.*  
>  
>

Is there more than one different TELN NOT BILL lines in any one file?  
If so, how do you tell which CUSTOMER HAS goes with which?

I have no idea how you got "WHAT YOU NEED" from the example data.  
None of the lines in the final can be directly derive without applying  
some unknown transform.

If you JUST want the lines containing TELN NOT BILL or CUSTOMER HAS  
then sift them out and reformat the lines.