

## Re: Parsing 'dirty/corrupt data'. Advice wanted

**Source:** <http://coding.derkeiler.com/Archive/Perl/comp.lang.perl.misc/2004-10/3084.html>

---

**From:** James Willmore (*jwillmore\_at\_adelphia.net*)

**Date:** 10/29/04

Date: Fri, 29 Oct 2004 08:01:59 -0400

burlo\_stumproot@yahoo.se wrote:

<snip>

> <example> # Block is 9 lines, line nr of data added, the rest is junk  
> 1: 030 RAN  
> 2:  
> 3: 00002 00002  
> BUG440  
> BUG440 : 00AC76B2 00001002 00008018 00004913 0000 19 0001 001  
> 000 0 73168 000020A5 00006137 00000008 00000000 0000 0001 000  
> BUG440 + 0471C390 044C8418 044C5340 044C5016 04366226  
> <<<< Here there can be many more lines like these >>>>  
> BUG440 + 04365EB2 04365E10 0435E0A8 04B486AA 04B4837A  
> BUG440 + 04B48306  
>  
> 4:  
> 5: 0000000 00000  
> 6: 0000000 00003  
> 7: 00000 00000  
> 8: 00000  
> 9: 0000000 00000  
> </example>  
>  
>  
> In one file I found what appears to be a login session complete with  
> commands and output. \*sigh\*  
>  
>  
> Any help, pointers, reading suggestions???

Know your data. Know why one line is valid and another isn't. The data may appear to have no "logic" or "pattern" to it, but it's there somewhere.

First place I might start is either split the line on whitespace or use `unpack` to get at least the first column. Then start testing for what is requires for a valid line. That's at first glance and without having any clue as to what the data is supposed to be/represent.

comp.lang.perl.misc: Re: Parsing 'dirty/corrupt data'. Advice wanted

HTH

Jim