

Re: ascii nulls, awk, and tr

Source: <http://unix.derkeiler.com/Newsgroups/comp.unix.solaris/2004-08/0289.html>

From: Richard L. Hamilton (*Richard.L.Hamilton_at_mindwarp.smart.net*)

Date: 08/03/04

Date: Tue, 03 Aug 2004 21:22:36 -0000

In article <jppvg0lugiqg7csdiqvtfcqm6fe605j5un@4ax.com>, Gerard C Blais <gerard.blais@mci.com> writes:

>
>
> *I have text files that may contain ascii nulls.*
>
> *I'd like to delete the null characters.*
>
> *I also have a lot of other things to do to the rows, and I'm doing*
> *them in awk already, so adding something to an awk script would be a*
> *nice way to go.*
>
> *None of the awks on my Solaris (/usr/bin/awk, /usr/xpg4/bin/awk,*
> */usr/bin/nawk) recognize ascii nulls. Someone suggested mawk, and I'm*
> *trying to get that installed. If anyone knows if mawk will or won't*
> *work, I'd be interested.*

GNU awk (gawk) seems to handle null bytes. If you have the Sun freeware installed for Solaris 9, it's in /opt/sfw/bin/gawk. If not, depending on the version of Solaris you have, you might find pre-built binaries at www.blastwave.org or www.sunfreeware.com. I haven't tried mawk, so I don't know how well it compares or whether it can handle nulls. But what I particularly like about gawk is that it seems to handle just about anything: arbitrarily long lines that neither the supplied awk nor nawk can handle, nulls just like other characters, etc.

> *Meanwhile, it seemed like a perfect fit for tr.*
>
> *BUT, if I try `tr -d \000 null.txt > null.nonnull`,*
>
> *tr just hangs, and there is no output. null.txt is a 3 line file*
> *with nulls in the middle of line 2.*
>
> *Any ideas?*

tr uses standard input, not filenames on the command line (so it was reading the terminal after you typed that command). Also, older versions (/usr/bin/tr, /usr/ucb/tr) always strip null bytes, while /usr/xpg4/bin/tr

comp.unix.solaris: Re: ascii nulls, awk, and tr

will handle them just like anything else.

So, either

changing a space to a space looks like it does nothing, but older
versions of tr always strip null bytes, so they do the job as a