

Subject: find, cut and sort... they can be your best friends...
From: Joseph Picone <joseph.picone@gmail.com>
Date: 6/10/23, 11:46 AM
To: nedc_research <nedc_research@googlegroups.com>
CC: ECE 1111 <temple_engineering_ece1111@googlegroups.com>

When you have a database of 15,000 directories or files of this form:

```
foo/00064/000645729/001023748/urinrpr  
foo/00068/000685156/001023753/urinrpr  
foo/00068/000685158/001023666/urinrpr  
foo/00068/000687605/001023754/ukown  
foo/00068/000683798/001023733/unknown
```

How do you find incorrect entries, misspellings, etc.?

Well, you could read all 15,000 entries because you are a hard-working diligent person... But most likely in this case you would still miss many things, and that is obviously too inefficient.

You could write a Python script, and spend time debugging it. But that is also too slow.

A better way is to do something like this:

```
find foo -maxdepth 4 -mindepth 4 | cut -d"/" -f5 | sort -u
```

This produces this result:

```
nedc027_[1]: find foo -maxdepth 4 -mindepth 4 | cut -d"/" -f5 | sort -u  
ukown  
unknown  
urinrpr  
urinrpr
```

You can examine the list and determine if there are things on it that don't belong. Then you can use find to find those specific anomalies.

The above command is a good, simple, example of something I have talked about for a while - how you can build interesting capabilities by piecing together command line programs. We call this command line programming and it is a very important skill to have in Unix.

There are lots of variants of this command:

```
nedc027_[1]: find foo -maxdepth 3 -mindepth 3 | cut -d"/" -f4 | sort -u  
001023666  
001023733  
001023748  
001023753  
001023754
```

and many ways to do these kinds of things using awk, sed, etc. You can do quite a lot from the command line in Unix.

-Joe