Practice exercises with the shell

Srikumar K. S.

18 July 2024

Contents

1   Counting the number of PNG images in the current directory

The current directory has a number of PNG format image files (extension '.png') mixed with JPG format image files (extensions '.jpg' and '.jpeg'). Your task is to determine the number of PNG format files in the current directory and print it on the console/terminal.

Hint: Use `ls` with wild cards and then count the number of files. Note that `ls` will output one file per line, so you'll basically have to count the number of lines it outputs.

Hint: If you want to mock a directory with such files, you can use the `touch` command to create empty files.

2   Combining files

Your friend has downloaded a large PNG file in small chunks split up at 10KB boundaries. These files are all in your current directory and named `chunk01`, `chunk02`, …, `chunk50`. Use the command line to combine all these chunks into a single file named `combined.png`. Note that the files must be combined (i.e. concatenated) in the order of their numerical suffixes.

Hint: Use wild card characters.

Point to ponder: What would've happened had the files been named `chunk1`, `chunk2`, … `chunk9`, `chunk10`, `chunk11`, …, `chunk50` instead? Can you modify your command line program to work for this case?

3   Sorting a file listing

The output of `ls -l` in one of my folders is as shown below –

```
total 68728
drwx------@  33 sriku  staff      1056 Oct 30  2023 Complete interpreter
-rw-r--r--@   1 sriku  staff     50846 Oct 29  2023 Complete interpreter.zip
drwx------@  46 sriku  staff      1472 Nov  1  2023 Generators Microkanren
-rw-r--r--@   1 sriku  staff     29516 Oct 29  2023 Generators Microkanren.zip
drwx------@  30 sriku  staff       960 Oct 31  2023 Mid term submissions
-rw-r--r--@   1 sriku  staff     46658 Oct 29  2023 Mid term submissions.zip
drwx------@ 148 sriku  staff      4736 Oct 29  2023 SMoL Tutor Submissions
-rw-r--r--@   1 sriku  staff  35048545 Oct 29  2023 SMoL Tutor Submissions.zip
-rw-r--r--    1 sriku  staff        77 Aug  4  2023 comp308-assignment3-interpreter.rst
```

I want the dates to be formatted like 2023-10-30 (instead of "Oct 30 2023"), 2023-10-29 and so on instead of the way it is shown, and I also want the output to be sorted by that date column in ascending order. Use `man ls` to learn how to format the date column and process the output of `ls -l` using the `sort` program to sort by the date column. Hint: The man page for `ls` might point you to another place that you then have to read. And, if you haven't realized yet, use `man sort` to learn how to use the `sort` program.

4   Filtering a CSV file

You're given the following CSV ("comma separated values") file contents in the file named `data.csv` -

```
number,english,hindi
1,one,ek
2,two,do
3,three,teen
4,four,chaar
5,five,paanch
6,six,chey
7,seven,saath
8,eight,aat
9,nine,now
10,ten,das
```

Your task is to device a command line program combination to output all the Hindi number words in the file which contain the character `e`. So the expected output is –

```
ek
teen
chey
```

5   Listing the URLs in a web page

(This will involve some hunting around manuals and learning a number of things to solve.)

HTML pages may feature links given in "anchor" tags that look like this –

```
<a href="https://somewebsite.com/somewhere/somepage.html">Hyper text</a>
```

The `href=` is called an "attribute" of the `<a></a>` tag. Your task is to construct a command line process using the above known commands and piping, to take in a URL and output a list of URLs found as such "href" attributes in the HTML page retrieved from the URL. You must output one such URL found per line (you have some flexibility with that format if you want).

For example, if the given URL were `https://www.google.com`, then the first 5 lines of your output would be something like –

```
https://www.google.com/imghp?hl=en&tab=wi
https://maps.google.co.in/maps?hl=en&tab=wl
https://play.google.com/?hl=en&tab=w8
https://www.youtube.com/?tab=w1
https://news.google.com/?tab=wn
```

Hint: You'll need to figure out how to construct the "regular expression" that will detect the 'href="…."' part. See the "see also" section of `man grep` for where to learn that. Note that you won't need any commands beyond those listed in this document.