

Common Sed Commands

Sunday, June 11, 2017

8:46 PM

`tr -c 'A-Za-z' '[\n*]' < assign2.html` : After running this, all non letter characters were replaced with a new line. So, if there are multiple non letter characters in a row then there will be multiple new lines.

`tr -cs 'A-Za-z' '[\n*]' < assign2.html` : After running this command, it outputs a similar to the previous command except that if there are multiple non letter characters in a row, they are replaced with only one new line. There is no blank lines between lines with text.

`tr -cs 'A-Za-z' '[\n*]' < assign2.html | sort` : The output of this command has the same characteristics of the previous command, but now the output is sorted alphabetically. I initially put `< assign2.html` at the end, but the terminal somewhat froze and I had to use C-c to exit. Putting it before piping to sort fixed the issue.

`tr -cs 'A-Za-z' '[\n*]' < assign2.html | sort -u` : This outputted the same result as the previous command, but using the -u option removed any duplicate words that are spelled using the same characters. If, there are different characters used to spell the same word they are kept. For example, for all, All, and ALL, none of the words are deleted, but if it was all, all, and all, only one of the all words is kept.

`tr -cs 'A-Za-z' '[\n*]' < assign2.html | sort -u | comm - words` : This command outputs 3 columns.

The first column outputs the words that only appear in assign2.html. The second outputs the words that only appear in words, and the third column outputs the words that are in both files.

$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

[illegible][illegible]

$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$\frac{1}{2} \times \frac{1}{2}$ $\frac{1}{2} \times \frac{1}{2}$ $\frac{1}{2} \times \frac{1}{2}$ $\frac{1}{2} \times \frac{1}{2}$

`sed '/^\s$/d' |` : This command gets rid of all the empty lines.
Initially, I used the command, `sed 's/^\s*$//g'` and nothing happened. It seemed to replace the empty line with another empty line. By using the delete function with `d`, the empty lines were replaced.

`tr '[:upper:]' '[:lower:]' |` : This changes the uppercase letters into lowercase letters

`sed "/[^pk'mnwlhaeiou]/d" |` : This deletes any line that does not contain the Hawaiian letters. The `^` inside the brackets makes it so the `sed` command deletes the characters that are not in the set listed.

`sort -u` : This sorts the word list alphabetically and removes duplicates. Then, it saves the output into `hwords`.

<code>sed 's/[^0-9]*//g'</code>	delete non numbers
<code>sed -i 's/ //g'</code>	delete leading blank space
<code>sed 's/[:blank:]*\$//'</code>	delete trailing whitespace
• <code>sed -r 's/\s+//g'</code>	remove all white space
<code>tr -d '\n'</code>	remove new lines
<code>sed '/^\$/d'</code>	Remove empty lines
<code>sed 's/[^0-9]*//g'</code>	remove non numbers

