

Mission Name: Text Processing

- * CLI is faster than python as commands interact more intimately with the filesystems.
- * We can concat two or more text file using "cat" command.

Eg:- cat east west
 cat *
 ↖ Concat all files

- * We can sort the content of a file using sort command. It will sort content of file in Lexicographical Order.

Eg:- sort east
 sort east west

- * Reverse sorting can be done using "-r" option.

Eg:- sort -r east
 sort -r east west

- * To keep only unique elements, we can use "-u" option.

Eg:- sort -u east west

- * If we want dataset based on a particular column:-
 col. index (start)

sort -t"," -k 1:1 example.csv
 ↖ delimiter ↖ Index specification ↖ same if 1 column (stop)

- * Other example:- sort -t":" -k 3,3 -k 4,4gr example
 ↖ sort as numeric

- * to select certain columns, we can use "cut" command.

Eg:- `cut -d"," -f 2,4,5 example.csv`

delimiter

fetch

2nd, 4th & 5th cols.

`cut -d"," -f 2-5 example.csv`

Select columns 2 to 5.

- * `grep` → global regular expression print
- * `grep` can be used to apply regex on a dataset and extract columns/rows data.

* `grep -n [aeiou].[aeiou] eg.csv`

show line number to which the regex match corresponds to.

invert-match

* `grep -v [aeiou].[aeiou] eg.csv`

show rows that doesn't matches the regex

- * `-i` option can be used for ignore case.