Lecture 2 commands:

----

head – prints first lines / characters in file

usage: **head** [OPTIONS]... [FILE]…

if no file is specified, then operate on standard input. Otherwise operate file by file.

Common options:

- -n<X> – print first X lines in file
- -n-<X> – print all but the last X lines
- -c<X> – print first X characters in file
- -c-<X> – print all but the last X characters

----

tail – prints last lines / characters in file

usage: **tail** [OPTIONS]... [FILE]…

if no file is specified, then operate on standard input. Otherwise operate file by file.

Common options:

- -n<X> – print last X lines in file
- -n+<X> – print all lines starting from line X
- -c<X> – print last X characters in file
- -c+<X> – print file starting from character X
- -f – print and wait; if file is appended, then keeps printing (useful for logging)

----

wc – counts bytes / words / lines

usage: **wc** [OPTIONS]... [FILE]…

if no file is specified, then operate on standard input. Otherwise operate file by file.

Common options:

- -l – print number of lines
- -c – prints number of bytes
- -w – prints number of words

----

cut – print a subset of characters/fields in every line in a file

usage: **cut** OPTIONS... [FILE]…

if no file is specified, then operate on standard input. Otherwise operate file by file. User must specify at least one option

Common options:

- -c<RANGE> – output characters in RANGE in every line
- -f<X>-<Y> – output fields in RANGE in every line
- -d<C> – use character C as the field delimiter

Notes: Exactly one of –c or –f has to be specified. Ranges can be specified by single number (e.g., -c4), comma-separated list (e.g., -f1,4,5), closed rage (e.g., -f4-11), open ended range (e.g., -c-6 or –f3-) (indicating that other bound is at the edge of the line), or any combination of those (e.g., -c3,4-6,9,12-).

----

paste – print a subset of characters/fields in every line in a file

usage: **paste** [OPTIONS] ... [FILE]…

if no file is specified, then operate on standard input. Otherwise operate file by file.

Common options:

- -d<SEPARATORS> – list of characters to act as field delimiters by order
- -s – converts each file to a single line, where lines of the original file are fields in the new line

Notes: If more than one separator is given, then they are used in order. For example –d" :;," will use ' ' between the first and second fields, ':' between the second and third, etc.

----

tr (translate) – substitute characters in a given file

usage: **tr** [OPTIONS]... SET1 [SET2]

outputs to stdout the stream in stdin after elimination / substitution of characters from SET1 to characters in SET2

Common options:

- -d – delete characters from SET1 (no need for SET2)
- -s – shrinks sequences of characters from SET1 to a single character (ne need for SET2)

Notes: SET1 and SET2 are given as strings of the same length. The i'th character in SET1 gets substituted by the i'th character in SET2.

----

sort – sort lines of file

usage: **sort** [OPTIONS]... [FILE]…

Sort line of file in lexicographical order (characters are ordered by ASCII value), unless otherwise stated. If no file is specified, then operate on standard input. Otherwise operate file by file.

Common options:

- -f – ignore case
- -r – sort in reverse order (large to small)
- -n – sort based on numeric value (assuming each line contains a number)
- -k<N>,<M> - sort tab-delimited table based on string that starts in the Nth field and ends at the end of the Mth field.

Notes: One can sort a tab-delimited file in hierarchical ordering of columns. For example sort -k1,1 -k2,2 will sort based on 1st field (primary) and 2nd field (secondary).

----

uniq – merges consecutive unique lines

usage: **uniq** [OPTIONS]... [FILE]…

Sort line of file in lexicographical order (characters are ordered by ASCII value), unless otherwise stated. If no file is specified, then operate on standard input. Otherwise operate file by file.

Common options:

- -i – ignore case
- -c – for each line print number of occurrences (before line)
- -d – only print duplicate lines
- -u – only print unique lines

Notes: uniq is often applied after sort, so that identical lines appear one after the other.