



Introduction to Bash video lecture

12 -Text processing commands



grep

- `grep pattern file` - search for lines containing pattern (even substrings)
- `grep -i pattern file` - ignore case
- `grep -w pattern file` - search for lines with words matching pattern
- `grep -v pattern file` - search for lines that don't have pattern
- `grep -i wv pattern file` - ignore case, lines without word matching pattern
- `grep -r pattern directory` - search whole directory tree for files with lines matching pattern

Regular expressions intro

- `.` - any one character
- `\.` - an actual dot
- `pattern$` - line ends with pattern
- `^pattern` - line starts with pattern
- `[a-z]` - any one character between a and z in the ascii table
- `[^a-z]` - any one character not among a-z
- `pattern?` - one or zero occurrences of the character expressed by pattern
- `pattern*` - zero or more occurrences ...

Printing parts of files, etc

- `head` - print the ten first lines
 - `head -5` - print the five first lines
- `tail` - print the ten last lines
 - `tail -7` - print the seven last lines
- `cat` - print all lines from file or std in
- `tac` - print all lines in reverse order from file or std in
- `rev` - print all lines backwards
- `sort` - print all lines sorted by the ASCII table or as specified

Cutting up URLs

```
rikard@newdelli:~/bash-intro/text-files$ cat a_few_urls.txt
http://www.gu.se/bazinga
https://ait.gu.se/forskning/journalister
http://www.elsewhere.org/pomo/
http://snarxiv.org/vs-arxiv/
http://www.physics.nyu.edu/faculty/sokal/afterword_v1a/after
word_v1a_singlefile.html
```

How can we get rid of the protocols?

Cutting up lines using cut

```
# Cut lines using / as delimiter, show fields 3 and up
rikard@newdelli:~/bash-intro/text-files$ cut -d '/' -f3-
a_few_urls.txt
www.gu.se/bazinga
ait.gu.se/forskning/journalister
www.elsewhere.org/pomo/
snarxiv.org/vs-arxiv/
www.physics.nyu.edu/faculty/sokal/afterword_v1a/afterword_v1
a_singlefile.html
```

Cutting up lines using cut

```
# Cut lines using / as delimiter, show field 3 only
rikard@newdelli:~/bash-intro/text-files$ cut -d '/' -f3
a_few_urls.txt
www.gu.se
ait.gu.se
www.elsewhere.org
snarxiv.org
www.physics.nyu.edu
```

Translating characters

- `tr 'a' 'e'` - translate all 'a' to 'e' from standard in
- `tr -d 's'` - remove all 's' from standard in
- `tr 'ng' 'ss'` - translate all 'ng' to 'ss'
- `tr '[a-z]' '[A-Z]'` - translate any lowercase to its uppercase version

```
rikard@newdelli:~/bash-intro/text-files$ tr 'a-zA-Z' 'n-za-mN-ZA-M'  
Top secret message!  
Gbc frperg zrffntr!          <- tr replies  
Gbc frperg zrffntr!  
Top secret message!          <- tr replies  
(user presses Ctrl-D)  
rikard@newdelli:~/bash-intro/text-files$
```


Uniquify sorted consecutive lines

- `uniq some_sorted_text.txt resulting_file.txt` - remove duplicate lines and save the result in `resulting_file.txt`
- `uniq -c sorted.txt result.txt` - remove duplicate lines and report how many duplicate each line had in the resulting `result.txt` file

More about sort

- `sort` - sort lexicographically
- `sort -r` - sort descending (reverse)
- `sort -n` - sort numerically
- `sort -k1` - sort using column 1