Quick notes

- Check Piazza
- Reminders:
  - Weekly quiz and assign2 **DUE TODAY**
- Assignment 3 is hard
  - Due on 30 Oct
- Questions?

Feedback / Office Hours

- Tameez Latib
  - tameezlatib@gmail.com, please add “CS35L” to the subject line
  - Office Hours: Monday **4pm-6pm** (or by appointment)
  - Feedback: https://forms.gle/6kcJ2aJtzAzFMhHQ7 (anonymous google form)

Recap: I/O redirection

- >, >>, <, |, 2>, 2>>
  - You should know what these do!
  - 0 (in), 1 (out), and 2 (err)
- New: &
  - Specifies that you're talking about stdin/out/err
  - &1 -> stdout
  - 2>&1 -> redirect stderr to stdout

Regular expressions

- Recall linux wildcards (regex is very similar)
- ?: 0 or 1 of preceding
  - ab?c matches ac and abc, anytext ac anyothertext
- *: 0 or more of preceding
  - ab*c matches ac, abc, abbc, abbbbbbbc
- +: 1 or more of preceding
  - ab+c matches abc, abbc, abbbbbbc
- {n}: exactly n preceding
- {n,}: n or more
- {n,m}: n to m

Regular expressions

- ^: matches beginning of line
- $: matches end of line
- [ ] matches anything inside
  - [aeiou] matches a vowel
  - [^aeiou] matches a non-vowel
- [a-z], [A-Z], [0-9], [a-zA-Z]
- abc|xyz matches abc OR xyz
- \$ matches character &
  - \. \?, ...
  - Basic regex = \$ means end of line
  - Extended regex = \$ means match character $

Regular expressions

- Save previous matches with () and use with \1, \2, ...
  - \1 = first capture group, \2 = 2nd, ...
- ([a-z])c\1 matches aca, bcb, zcz
- https://regex101.com/ or https://regexr.com/ or ...

Back to shell commands

- wc : word count, newline, word, byte
  - -l, -w, -c
- ls -a | wc -l
- sort
  - -u for unique
- tr: translate using regex
  - tr [a-z] 0
  - -c (complement), -s (squeeze / remove duplicates), -d (delete)
- grep: find pattern using regex
- ls -l | grep \.txt
- comm: compare sorted files, show which lines unique to which file

Back to shell commands

- sed ‘s/[find pattern]/[replace pattern]/[flags]’
  - Works on strings, not character
  - echo 00000 | sed ‘s/0/1/’ --> 10000
- Lots of different ways to use command, many different flags
  - By default, uses only first match, flag g (global) matches every occurence on line

Bash scripts

- Use multiple commands together
- Create a file (script), run the script
- First line is (usually) a shebang. Which interpreter to use
  - #!/bin/bash (has more tools)
  - #!/bin/sh (simple version)
  - Doesn’t matter too much
- Variable declaration: two=2
- Variable usage: echo $two
- Script to make temp directory, cd into it, create a file, ls, and cd out
- More advanced topics on wednesday
- Loops, if, functions, etc

Lab 3, getting started

- Understand the english checker
  - What does each step do? Why is it necessary?
- Follow the instructions!
- wget to get the HTML page
- remove ?, <u>, and </u>
- Treat uppercase letters as if they were lowercase
- Treat grave accent ` as if it were apostrophe ‘
- Treat hyphens as if they were spaces
- Extract lines of the form **A**<td**X**>**W**</td>**Z**
  - **A** is zero or more spaces
  - **X** is anything *except* the > character
  - **W** is any sequence of one or more Hawaiian characters and spaces.
  - **Z** is zero or more spaces

Lab 3, getting started

- Extract lines of the form **A**<td**X**>**W**</td>**Z**
  - **A** is zero or more spaces
  - **X** is anything *except* the > character
  - **W** is any sequence of one or more Hawaiian characters and spaces.
  - **Z** is zero or more spaces
- We only want W
- Sort the remaining list of words, remove duplicates, and print to stdout
- Also check the hints slide