CSE 391
Regular Expressions
sed

ROADMAP

- Introduction to the command line
- Input/output redirection, pipes
- More input/output redirection, tee, xargs
- Git: Fundamentals
- Git: Branches, merging, and remote repositories
- Regular expressions
- More regular expressions, sed
- Users and permissions
- Bash scripting

AGENDA

- More regular expressions
  - Negative matches
- sed

THINK:

classes.txt

CSE142
CSE143
CSE311
CSE391
CSE416
CSE446
CSE507
CHEM142
MATH324
MATH409
MATH461
PHIL322
HIST210

Suppose we have the file classes.txt on the left. Each line contains the name for a single class at UW. What is the full grep command to print all CSE classes which are not at the 300 level?

Syntax	Functionality
[0-9]	Any digit
[^0-9]	*Not* any digit
^	Beginning of line
$	End of line

1:00

SED

- sed: a *stream editor* for filtering and transforming text
  - Similar to tools such as "search and replace" in most word processors such as Google Docs or Microsoft Word
- General format
  - `sed -r 's/REGEX/TEXT/' file.txt`
    - This will replace all matches of *REGEX* in the given file with *TEXT* and output it to the console.

SED

```
sed -r 's/Kanye/Josh/' names.txt
```

names.txt

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Kanye
Jo, Jo
Lamar, Kendrick
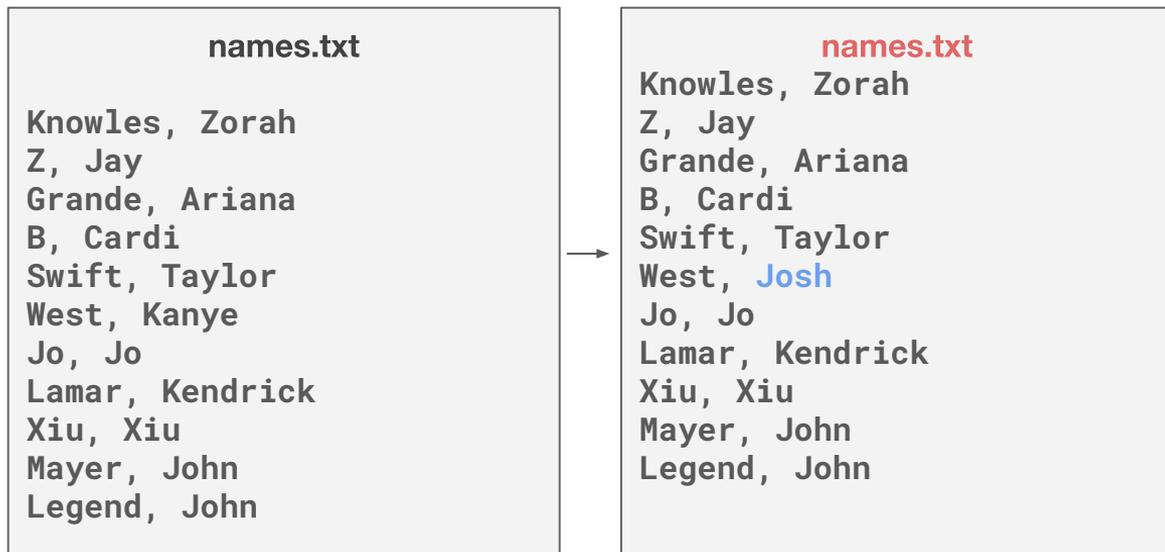Xiu, Xiu
Mayer, John
Legend, John

→

Console output

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Josh
Jo, Jo
Lamar, Kendrick
Xiu, Xiu
Mayer, John
Legend, John

Notes:

- The thing we are "replacing" (i.e. Kanye), is a regular expression
- **By default, sed outputs to the console and does not modify the original file**

SED

```
sed -ri.bak 's/Kanye/Josh/' names.txt
```

names.txt

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Kanye
Jo, Jo
Lamar, Kendrick
Xiu, Xiu
Mayer, John
Legend, John

→

names.txt

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Josh
Jo, Jo
Lamar, Kendrick
Xiu, Xiu
Mayer, John
Legend, John

Notes:

- **If you want sed to change the contents of your file, use the -i (in-place) flag to do so. You may add an extension to create a backup file.**

SED

```
sed -r 's/Jo/Josh/' names.txt
```

names.txt

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Kanye
Jo, Jo
Lamar, Kendrick
Xiu, Xiu
Mayer, John
Legend, John

Console output

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Kanye
Josh, Jo
Lamar, Kendrick
Xiu, Xiu
Mayer, John
Legend, John

Notes:

- By default, sed will only replace the first occurrence, **per line**, of the regex

SED

```
sed -r 's/Jo/Josh/g' names.txt
```

names.txt

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Kanye
Jo, Jo
Lamar, Kendrick
Xiu, Xiu
Mayer, John
Legend, John

Console output

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Kanye
Josh, Josh
Lamar, Kendrick
Xiu, Xiu
Mayer, Joshhn
Legend, Joshhn

Notes:

- If the g option is included, then we perform a **global** match and replace all matches on each line

SED

```
sed -r 's/\<Jo\>/Josh/g' names.txt
```

names.txt

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Kanye
Jo, Jo
Lamar, Kendrick
Xiu, Xiu
Mayer, John
Legend, John

Console output

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Kanye
Josh, Josh
Lamar, Kendrick
Xiu, Xiu
Mayer, John
Legend, John

Notes:

- Recall the left-most string is a regular expression and can be used for more complex matchings.
- Also recall that \< and \> match the beginning and ending of words.

THINK:

emails.txt

```
hschafers@uw.edu
djdan@uw.edu
zorahlf@uw.edu
lruzzo@uw.edu
```

Console Output

```
hschafers@cs.washington.edu
djdan@cs.washington.edu
zorahlf@cs.washington.edu
lruzzo@cs.washington.edu
```

Suppose we have the file emails.txt on the left. Each line contains the email address of a CS faculty member. What is the full sed command to output to the console all uw.edu addresses replaced with cs.washington.edu addresses?

1:00

SED

```
sed -r 's/^(.*),(.*)$/\2, \1/g' names.txt
```

names.txt

Knowles, Zorah
Z, Jay
Grande, Ariana
B, Cardi
Swift, Taylor
West, Kanye
Jo, Jo
Lamar, Kendrick
Xiu, Xiu
Mayer, John
Legend, John

Console output

Zorah, Knowles
Jay, Z
Ariana, Grande
Cardi, B
Taylor, Swift
Kanye, West
Jo, Jo
Kendrick, Lamar
Xiu, Xiu
John, Mayer
John, Legen

Notes:

- We can match first/last names on the left and right side and then flip them on the right side.

THINK:

emails.txt

```
hschafers@uw.edu
djdan@cs.washington.edu
zorahlf@cs.washington.edu
lruzzo@uw.edu
```

Suppose we have the file emails.txt on the left. Each line contains the email address of a CS faculty member. What is the full sed command to output just the basename of their email address (i.e. before the @ symbol)?

Console Output

```
hschafers
djdan
zorahlf
lruzzo
```

2:00