CSc 352
Text Processing and Regex

Benjamin Dicken

Announcements

- Exam 2 on Thursday
- ILC 130 for review session, 5pm on Wednesday
- PA 6 extra credit

Capture Groups

Use the ( ) to specify a capture group

Can use to capture a subset of the full regular expression

For example:

```
sed -n 's/\([aeiou][aeiou]\)/\1\1/p' sample.txt
```

What does this regular expression do?

file.txt:

```
The great large ant
Should be over those hills to the left
```

Command:

```
sed -n 's/\b\(...\)\b/"\1"/p' file.txt
```

Man page for sed, grep

Let’s take a look!

Common sed patterns (learn these!)

```
# Open in.txt and replace every match of search_regex with to_replace and send to out.txt
sed -E 's/search_regex/to_replace/g' in.txt > out.txt

# Open in.txt and print the lines that have a match for search_regex
sed -n -r '/search_regex/p' in.txt

# Open in.txt and replace search_regex with to_replace in a case insensitive manner and send to out.txt
sed -r 's/search_regex/to_replace/i' in.txt > out.txt

# More!
```

https://www.gnu.org/software/sed/manual/sed.html

Download resources

Look at the man pages for **`curl`** and **`wget`**

What do these do?

NBA Rosters

https://www.nba.com/suns/roster

Write the command (or sequence of commands)

- Fetch the roster for the Phoenix Suns (or your favorite team)
- Extract the first / last names of each player on the team and print one per line
- The URL: https://www.nba.com/suns/roster

For example:

```
$ your_command . . .
Deandre Ayton
Bismack Biyombo
Devin Booker
. . . .
$
```

NBA Rosters

```
#!/bin/bash
wget https://www.nba.com/suns/roster 2> /dev/null
cat roster | sed 's/[{}]/\n/g' > roster2
sed -n -E 's/Person","name":"([A-Za-z ]+)"/\1/p' roster2 | cut -d '"' -f 4
```

Icelandic currency exchange

https://apis.is/currency/m5

Write the command (or sequence of commands)

- Fetch the Icelandic currency data
- Extract the currency conversion info and print it out as shown
- The URL: https://apis.is/currency/m5

For example:

```
$ your_command . . .
1 USD = 128.54 Icelandic krona
1 DKK = 18.771 Icelandic krona
1 EUR = 139.6 Icelandic krona
1 NOK = 14.682 Icelandic krona
1 GBP = 167.47 Icelandic krona
1 CHF = 137.47 Icelandic krona
1 SEK = 13.586 Icelandic krona
1 TWI = 177.12 Icelandic krona
$
```

Icelandic currency exchange

```
$ wget https://apis.is/currency/m5 -O currency.json
$ cat currency.json | sed 's/},{/\n/g' > out.txt
$ cat out.txt | sed -rn
's/.*"shortName":"([A-Z]{3})".*"value":([1-9.]+),"ask.*/1 \1 = \2 Icelandic
krona/p'
$ echo ""
```

Further Reading

https://www.digitalocean.com/community/tutorials/the-basics-of-using-the-sed-stream-editor-to-manipulate-text-in-linux

https://www.digitalocean.com/community/tutorials/using-grep-regular-expressions-to-search-for-text-patterns-in-linux