Finding the percent of colleges in the ranklist (wc, grep)

For this step, we need to introduce the concept of pipes, which is one of the most powerful feature in Bash shell. Pipes are represented by the vertical bar symbol (|) . Essentially, a pipe allows you to forward the output of one command to the input of another command, without having to save the intermediate output to a file. The following videocast explains the lesson tasks:
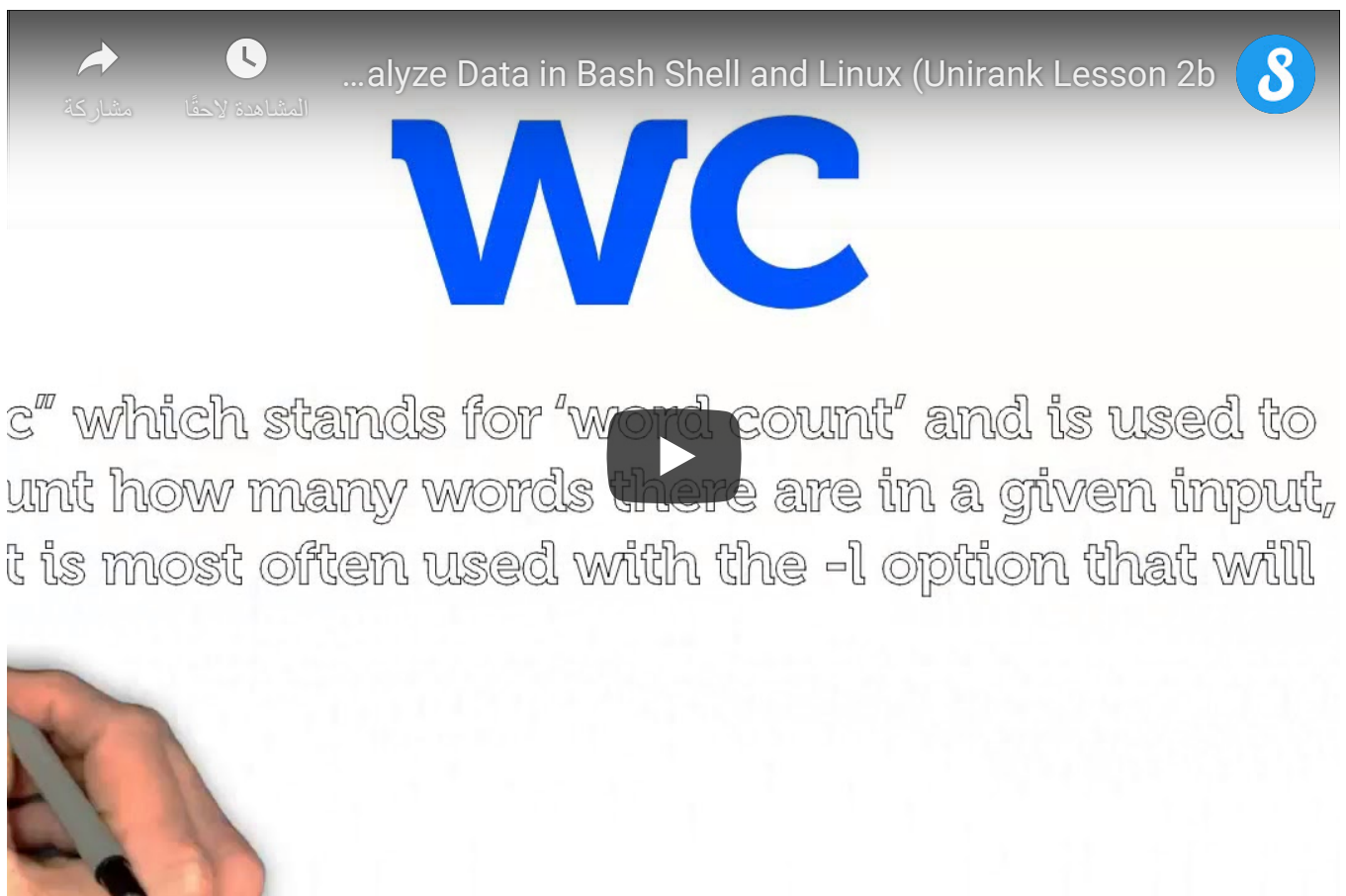
Video lecture: Finding the percentage of colleges

We would like to give the output of grep to another command which can tell

us how many lines that output has. As it turns out, there is a command that will let us do just that: `wc` which stands for word count and is used to count how many words there are in a given input, but is most often used with the `-l` option that will instead count the number of lines in our input. So let's start by counting how many lines total there are in the `unirank.csv` file:

```bash
#!/bin/bash
wc -l unirank.csv
```

Now we count the number of colleges, let's now use pipes and bring everything we learned together:

```
grep -i "college" unirank.csv | wc -l
```

The output of grep—the lines in the file that contain "college"—is passed on to `wc -l`, which will count the number of lines in the file. Notice that we didn't specify a filename in the `wc -l` command; we didn't have to because we piped our data into `wc -l`, so it didn't need to read the data from a file. Running this command should give you the output of 7. So in this dataset, `((8-2)/232) x 100% = 2.5%` institutes comes from the US colleges. We deducted the 2 wrongly identified colleges.

Do you want to know more?

📎 'wc' man page ⬇ ↗