Introduction to AWK

CS-210: Introduction to Unix

Robert Bruce

awk: What is it?

- awk is powerful tool for processing text file data with a built-in extended regular expression matching capability.
- awk is another standard tool installed on Unix and Unix-like operating systems.

Example dataset: test.csv

FIRST NAME, LAST NAME, STUDENT ID NUMBER

Bill, Blass, 12312942

Shirley, Jones, 21951234

Quimby, Blastonovich, 10121296

Example: extracting comma separated values: column 1

```
cat test.csv | awk 'BEGIN {FS=","} {print $1}'
```

FIRST NAME

Bill

Shirley

Quimby

Example: extracting comma separated values: column 2

```
cat test.csv | awk 'BEGIN {FS=","} {print $2}'
```

LAST NAME

Blass

Jones

Blastonovich

Example: extracting comma separated values: column 3

```
cat test.csv | awk 'BEGIN {FS=","} {print $3}'
```

```
STUDENT ID NUMBER
12312942
21951234
10121296
```

Example: matching text on column 2 (last name)

```
cat test.csv | awk 'BEGIN {FS=","} {if ($2 ~ /Jones/) print $1,$2,$3}'
Shirley  Jones  21951234
```

Example: matching text on column 2 (last name)

```
cat test.csv | awk 'BEGIN {FS=","} {if ($2 ~ /Jones|Blass/) print $1,$2,$3}'
Bill  Blass  12312942
Shirley  Jones  21951234
```

Example: matching text on column 2 (last name) using cat

```
cat test.csv | awk 'BEGIN {FS=","} {if ($2 ~ /Blas/) print $1,$2,$3}'
Bill  Blass  12312942
Quimby  Blastonovich  10121296
```

Example: matching text on column 2 (last name)

```
cat test.csv | awk 'BEGIN {FS=","} {if ($2 ~ /Blas/) print $1,$2,$3}'
Bill  Blass  12312942
Quimby  Blastonovich  10121296
```

Hints

- I purposefully and strategically used the cat command with the pipe symbol to send the contents of a file to awk to be processed. I did this to suggest you use Unix pipes.
- Sending awk output through a pipe is effectively using awk as a filter. That is exactly what you want to achieve when filtering the results of our in-class exercise today.
- Using the Unix pipe command in conjunction with regular expressions means you can filter content. Then you can *sort* (hint) or *count* (hint) the content (when applicable).