CS 246 Fall 2014 - Tutorial 1

September 17, 2014

1 Summary

- General Administration Stuff
- Shell Command Review
- I/O Redirection and Pipelining
- Regular Expressions and `grep`

2 General Administration Stuff

- Course E-mail: `cs246@uwaterloo.ca`
- Kirsten's Office Hours: WF 1:30-2:20, Th 2:00-4
- Use `Piazza` for most questions
  - Questions with potential answers should be private or asked in office hours
  - If your question is made private by an instructor - keep it that way
- E-mail the course account or post on Piazza about topics you would like to see in upcoming tutorials

3 Shell Review

- Commands you should definitely know
- **cd** - change the current directory
  - With no directory or ~ returns you to your home directory
  - With - will return you to previous current directory
- **ls** - view files in the current/specified directory
  - With -l returns long form list of directory
  - With -a returns all (including hidden) files
  - You can look into a directory using `ls dir-name`
  - Can combine multiple options, e.g. ls -al
- **pwd** - prints the current directory
  - Same as $PWD
- **uniq** - removes consecutive duplicates (removes all duplicates if sorted)
  - -c option will print counts of consecutive duplicates
- **sort** - sort lines of a file/standard in
  - -n option will sort strings of digits in numeric order
- **tail** - print last 10 lines of file/standard in

4 Output Redirection and Piping

4.1 Basic Examples

- Suppose we have a program (printer) that prints to standard output and standard error. Give the redirection to redirect stdout to print.out and stderr to print.err.
  - `./printer > print.out 2> print.err`
- What if we want to redirect standard output and standard error to the same file?
  - Will `./printer > out 2> out` work?
- To print to standard out and standard error to the same file we need to tie them together.
  - For example, `./printer > out 2>&1`
  - Or `./printer 2> out 1>&2`
- What would be the purpose of redirecting output to `/dev/null`?
  - When we do not care about the actual output of the program but want it to perform some operation (e.g. checking if files are the same, executed correctly).
- What is the difference between `./printer 2>&1 > out` and `./printer > out 2>&1`
  - The first prints all odd numbers to stdout and even numbers to the file out
  - The second prints all numbers to the file out
  - Order of redirection matters!

4.2 More Complex Example

Suppose we want to determine the 10 most commonly occurring words in a collection of words (see wordCollection file) and output it to file top10. How might we accomplish this?

Idea: Use some combination of sort/uniq/tail. But how? Probably need -c option with uniq and sort -n.

Okay. `uniq -c wordCollection | sort -n`

But what's the problem? wordCollection isn't sorted!

So now: `sort wordCollection | uniq -c | sort -n`

So this gives us counts in least to most. How do we get the top 10 and output it to the file top10?

Let's try tail now. `sort wordCollection | uniq -c | sort -n | tail > top10`

For fun (not actual material): wordCollection was created using the command:

```
for i in ‘seq 1 10‘;
do
num=$((RANDOM % 10000));
echo $num;
sort -R /usr/share/dict/words | head -n $num >> wordCollection;
done
```

5 grep and Regular Expression

- Recall that **grep** allows us to find lines that match patterns in files
- To get the full power of regular expressions supported, we must use **egrep** or **grep -E**
- Some useful regular expression operators are:

  - `^` - the pattern following must be at the beginning of the line
  - `$` - the pattern preceeding must be at the end of the line
  - `.` - matches any single character
  - `?` - the preceding item can be matched 0 or 1 times
  - `*` - the preceding item can be matched 0 or more times
  - `+` - one or more times
  - `[...]` - match one of the characters in the set
  - `[^..]` - match a character not the set
  - `expr1|expr2` - match expr1 or expr2
- Also, recall that concatenation is implicit
- Note that parentheses can be used to group expressions
- egrep can be especially useful for finding occurrences of names in source files
  - The option -n will print line numbers
- The following are some examples:

```
> egrep "count" file.c
> egrep "count" *.c
> egrep -n "count" *.c

> # Give a regular expression to find all lines starting with 'a' and ending with 'z'.
> egrep "^a.*z$" /usr/share/dict/words

> # Give a regular expression to find lines starting with 'a' or lines ending with 'z'.
> egrep "^a|z$" /usr/share/dict/words

> # Give a regular expression to find lines with one or more occurrences of the characters a,e,i,o,u
> egrep "[aeiou]" /usr/share/dict/words

> # Give a regular expression to find lines with more than one occurrence of the characters a,e,i,o,u
> egrep "([aeiou].*)([aeiou].*)+" /usr/share/dict/words

> # Want all lines in all .c files that modify count by assigning either 0 or 1 aside from initialization
> # Let's try the obvious thing first
> egrep "^ *count *= *0|1;$" *.c
> # Doesn't work. Why?
> # Let's use parenthesis
> egrep "^ *count *= *(0|1) *; *$" *.c
> # Excellent, this works.
```

6 Vi Tip of the Week

- Hitting the '/' key, while in command mode, will allow you to search the text of an open file for the specified regular expression
  - 'n' navigates to the next match
  - 'N' navigates to the previous match
- Entering ':N', while in command mode where N is a number will take you to the N'th line of the currently open file