TTCS Assignment 2

Truman Ellis

September 17, 2010

Exercise 1

The entries in /usr/share/dict/words are formatted one dictionary word per line in the order A-Z, a-z, and special characters.

Exercise 2

Part a

```
for LETTER in {a..z}; do echo -n $LETTER:; grep -i "^$LETTER" words
| wc -l; done
```

```
a:5902
b:6095
c:9507
d:5811
e:3813
f:4075
g:3452
h:3856
i:3613
j:1259
k:1247
l:3437
m:5952
n:2015
o:2240
p:7539
q:464
```

```
r:5285
s:11072
t:5002
u:1899
v:1584
w:2711
x:56
y:380
z:287
```

Part b

```
grep -i "s$" words | wc -l
```

46104 words end in `s`

```
grep -i "'s$" words | wc -l
```

24431 words end in `'s`

Part c

```
grep -i "foo[^tdl]" words | wc -l
```

7 words contain *foo*, but not *food, fool,* or *foot.*

Part d

Find the top 7 longest words in `words`.

```
cat words | awk '{print length, $0}' | sort -nr | head -7
```

```
23 electroencephalograph's
22 electroencephalographs
22 electroencephalogram's
22 counterrevolutionary's
22 counterrevolutionaries
22 Andrianampoinimerina's
21 electroencephalograph
```

Exercise 3

/h1 and /h2 Home directory, for all code developement, reports, and personal data. Network filesystem, acked up daily.

/org Appears to be where specific applications go that are used by various groups and centers. Network filesystem, backed up daily.

/opt Provides applications / compilers. Read only.

/work For generating shared data files. Network filesystem, not backed up.

/workspace For generating data files on the local machine, much faster. Local filesystem, not backed up.

Exercise 4

It took 0.093 seconds to write 10MB and 0.556 seconds to write 100MB to `/workspace`, while it takes 0.914 seconds to write 10MB and 9.094 seconds to write 100MB to `/h2/truman`. It is faster to write to `/workspace` because it is a local disk, writing to my home directory has to go over the network.

Exercise 5

Part a

```
du -hs
```

```
find . -type f | wc -l
```

```
find . -type f | grep '.c$' | wc -l
```

The linux-2.6.24.1 directory is 304MB with 23062 files, of which 9748 are ".c" files.

Part b

```
find . -type d | wc -l
```

```
ls -FR | grep '/$' | wc -l
```

```
tree -d
```

According to `ls` and `tree`, there are 1355 directories. `find` counts 1356 because it also counts the current directory, "`.`" in it's list.

Part c

```
wc -l `find . -printf '%s %p\n' | sort -nr | head -1 | cut -f2 -d' '`
```

```
13947 ./fs/nls/nls_cp949.c
```

The file `fs/nls/nls_cp949.c` is the largest with 13947 newlines.

Part d

```
wc -l `find . -type f | grep '.[ch]$'`
```

counts 7803521 newlines in all `.c` and `.h` files.

```
awk 'NF {x++}END{print x}' `find . -type f | grep '.[ch]$'`
```

counts 6737091 non-blank lines.

Part e

How many lines in the Linux kernel make reference to "happymeal"

```
egrep -ir "(happymeal)" * | wc -l
```

There are 152 lines that have the upper, or lower case string "happymeal"

Exercise 6

My answer from last time was:

```
getent passwd | cut -f5 -d: | awk 'BEGIN{x="wwwwwwwwwwwww"}/[[:alpha:]]+\
[[:alpha:]]+/ {if (length < x && length >1){x=length;name=$0}}END{print name}'
```

The shortest name of an ICES user is "Jun Li"

But now I understand the use of backticks and your solution which found all the shortest user names:

```
getent passwd | cut -f5 -d: | awk '{ if (NF==2 && length==''`getent passwd | cut
-f5 -d: | awk '{if(NF==2) {print 1+length($1)+length($2)} }' | sort -n | head
-1`') {print $1, $2}}'
```

which returns

```
Jun Li
Na Sai
Na Lei
Eh Tan
Ju Liu
J Shim
```