Basic UNIX Commands: Reference

1 clear.rice.edu: Rice Linux

1. Access instructions: `http://www.clear.rice.edu/`
2. ssh to ssh.clear.rice.edu with netid and password
3. e.g.: `ssh km21@ssh.clear.rice.edu`

2 Useful commands

wget url	simple web download tool
ls	list files and directories
ls -a	include (hidden) dot files and directories
mkdir	make a directory
cd directory	change to named directory
cd or cd ~	change to home-directory
cd ..	change to parent directory
pwd	display the path of the current directory
cp file1 file2	copy file1 and call it file2
mv file1 file2	move or rename file1 to file2
rm file	remove a file
rmdir directory	remove a directory
cat file	display a file
more file	display a file a page at a time
head file	display the first few lines of a file
tail file	display the last few lines of a file
grep 'regex' file	search a file with regexp
egrep 'regex'	grep with extended regexp features
wc file	count number of lines/words/characters in file
cat file1 file2 > file0	concatenate file1 and file2 to file0
sort	sort data
sort -n	sort data numerically
uniq	omit repeated lines
uniq -c	count and omit repeated lines
command > file	overwrite file with standard output
command >> file	append standard output to a file
command < file	redirect standard input from a file
command1 | command2	pipe the output of command1 to the input of command2
sed	stream editor: apply regexps to data
history	show commands you have executed

3 Some more complicated examples

```
sed -e 's/\s\+/\n/g' file.txt > wordlist.txt % convert spaces to new lines
```

```
sed -e 's/\s\+/\n/g' file.txt | sort | uniq -c % count words in text
```

```
% count words in text and sort the results numerically
sed -e 's/\s\+/\n/g' file.txt | sort | uniq -c | sort -n
```

```
% count words in text, sort the results numerically, and save the result
sed -e 's/\s\+/\n/g' file.txt | sort | uniq -c | sort -n > wordcounts.txt
```

Question 1: What happens if you leave out that first `sort` command? Try it and see!

Question 2: Look at the output, how might we remove some of the non-word 'words' from our counts?

4 Checking for Zipf's Law

man command	read the help documentation for the command
head file	view first N lines of file
tail file	view last N lines of file
paste file1 file2	combine file1 and file2 as columns
nl	number lines of a text file (or stream)
awk 'print $2' file	print column 2 of file
nano	edit a text file (see also vi, emacs, pico, etc.)

```
% sort the words in our text numerically in reverse, add line numbers
% so words are ranked by frequency, extract only the ranking and frequency
% columns, and save it all to a data file
sort -rn wordcounts.txt | nl | awk '{print $1, $2}' > ernest.dat
```

```
gnuplot> set out "file.png"
gnuplot> set terminal png
gnuplot> set xlabel "rank"
gnuplot> set ylabel "frequency"
gnuplot> plot "ernest.dat"
```

5 Language Knowledge

Question 3: How could we find (and maybe count) only the pronouns in this document?
Question 4: Now how about the nouns?
Question 5: What if we wanted to find only the plural nouns? Or progressive verbs?