Regular expressions

Part I

Grep Command

- grep - "general regular expression parser“
- A regular expression is a string that can be used to describe several sequences of characters.
- Search command for UNIX.
- Used to search for text strings and regular expressions within one or more files.
- man grep

Common grep Command Options

grep [options] pattern [files]

- -a Process a binary file as if it were text
- -b Display the block number at the beginning of each line.
- -c Display the number of matched lines.
- -i Ignore case sensitivity.
- -l Display the filenames, but do not display the matched lines.
- -n Display the matched lines and their line numbers.
- -v Display all lines that do NOT match.
- -w Match whole word.

- -o Print only matched part of the matching line
- -f Obtain patterns from *FILE*, one per line
- -H Print the filename for each match.
- -h Suppress the prefixing of filenames on output when multiple files are searched.
- -r recursive search in subdirectories.
- -m[n] return only specific number of search

How to use grep command

- Search file for a user
  - $ grep student /etc/passwd
- Search file ignoring word case
  - $ grep -i ‘STUDENT’ /etc/passwd
- Search recursively all files and directories under given directory
  - $ grep -r “student" /etc/

How to use grep command

- Search for a specific word in file
  - $ grep -w ‘welcome' filename
- Search for 2 different words in file
  - $ grep -w ‘goa|kerala' filename
- Count lines that matched in file
  - $ grep -c 'word' filename

How to use grep command

- Display lines that did not match a pattern
  - $ grep -v ‘war’ filename
- Number of lines that contain matched pattern
  - $ grep -n 'word' filename
- Display filenames that matched pattern, but not lines from the files
  - $ grep -l ‘good' *.txt
- Return only a specific number of matches
  - grep –m2 welcome file

Grep and Regular Expressions

- A "regular expression" is a pattern that describes a set of strings.
- Regular expressions are used when you want to search for specific lines of text containing a particular pattern.
- Use –E to use extended regular expressions

grep and Wildcards

^ (Caret)	match expression at the start of a line, as in ^A.
$ (dollar)	match expression at the end of a line, as in A$.
\ (Back Slash)	turn off the special meaning of the next character, as in \^.
[ ] (Brackets)	match any one of the enclosed characters, as in [aeiou]. Use Hyphen "-" for a range, as in [0-9].
[^]	match any one character except those enclosed in [ ], as in [^0-9].
. (Period)	match a single character of any value, except end of line.
* (Asterisk)	match zero or more of the preceding character or expression.
.*	match zero or any no. of characters
-e	combine multiple expressions

- print the lines that starts with the word "welcome“
  - grep –w ‘^welcome’ a
- print the lines which end with the word "university“
  - grep –w ‘university$’ a
- Print the lines which does not have any vowels
  - grep –v ‘[aeiou]’ a

Extended Regular Expressions

- x(y|z) matches xy or xz
- (x|y)z matches xz or yz
- {n} The preceding item is matched exactly n times.
- {n,} The preceding item is matched n or more times.
- {,m} The preceding item is matched at most m times.
- {n,m} The preceding item is matched at least n times, but not more than m times.
- ? The preceding item is optional and matched at most once.
- + The preceding item will be matched one or more times.

- [:upper:] uppercase letters
- [:lower:] lowercase letters
- [:alpha:] alphabets upper+lowercase letters)
- [:digit:] numbers in decimal, 0 to 9
- [:alnum:] alphanumeric meaning alpha+digits (any uppercase or lowercase letters or any decimal digits)
- [:space:] whitespace meaning spaces, tabs, newlines and similar
- [:cntrl:] control characters meaning non-printable characters

- List the lines which contain 3 a’s
  - grep –E ‘a{3}’ *
- List the lines which contains at least 3 a’s
  - grep –E ‘a{3,}’ *
- List the lines which contains at most 4 a’s
  - grep –E ‘a{,4}’ *
- List the lines which contains 3 to 5 a’s
  - grep –E ‘a{3,5}’ *
- List the lines which contains ababab
  - grep –E ‘(ab){3}’ *
- Match the word ‘time table’ or ‘timetable’
  - grep –E ‘time ?table’ *

- Assignment 3.2