Regular expressions, command grep.

Jan Trdlička

Department of Computer Systems
Faculty of Information Technology
Czech Technical University in Prague

trdlicka@fit.cvut.cz

November 12, 2018

Contents

Discussion of homework

- Create alias `lss`, which prints names of files in the working directory sorted by file size.
  - ```
    alias lss='ls -al . | tail -n +2 | sort -k5,5n | \
    tr -s " " | cut -d" " -f9 '
    ```
- Create shell script that prints names of the 10 largest files (including their sizes), which are in your home directory and in its direct subdirectories.
    ```

Discussion of homework

```
#! /bin/bash

LIST=""
cd "$HOME"

for i in * .*
do
  if [ -f "$i" ] ; then
    LIST="$(printf "%s\n%s\n" "$LIST" "$(ls -l "$i")")"
  elif [ -d "$i" -a ! \( "$i" = "." -o "$i" = ".." \) ] ; then
    for j in "$i"/* "$i"/.*
    do
      if [ -f "$j" ] ; then
        LIST="$(printf "%s\n%s\n" "$LIST" "$(ls -l "$j")")"
      fi
    done
  fi
done

echo "$LIST"

echo "$LIST" | tail -n +2 | sort -k5,5nr | head -10 | \
     tr -s ' '| cut -d' ' -f5,9-
```

Grep – options

- What is the default behaviour of the command grep?
  - Print lines matching a pattern.

```
ls /home/* 2>/dev/null | grep novak
```

- What is the meaning of the following options?
  - -v ... invert the sense of matching, to select non-matching lines,

```
echo $PATH | tr ':' '\n' | grep -v bin
```

  - -c ... print a count of matching lines for each input file,

```
grep -c root /etc/passwd
```

  - -i ... ignore case distinctions,

```
man ls | grep -i command
```

Grep – options

- What is the meaning of the following options?
  - -l ... suppress normal output, instead print the name of each input file from which output would normally have been printed,

```
grep -l start /etc/* 2>/dev/null
```

  - -n ... prefix each line of output with the 1-based line number within its input file

```
grep -n bash /usr/bin/* 2>/dev/null
```

grep – extended regular expressions (ERE)

- What is the meaning of the following symbols in ERE?
  - `^` (caret/hat) ... begin of line,
  - `$` ... end of line,
  - `\<` ... begin of word,
  - `\>` ... end of word,
  - `.` ... any single character
  - `[a-d]`, `[abcd]`, `[[:lower:]]` ... single character from interval/list,
  - `[^a-d]`, `[^abcd]`, `[^[:lower:]]` ... single character not from interval/list,
  - `?` ... the preceding item will be matched zero or one times,
  - `+` ... the preceding item will be matched one or more times,
  - `*` ... the preceding item will be matched zero or more times,
  - `{n}` ... the preceding item is matched exactly `n` times,
  - `{n,}` ... the preceding item is matched `n` or more times,
  - `{,n}` ... the preceding item is matched at most `n` times (GNU extension),
  - `{m,n}` ... the preceding item is matched at least `m` times, but not more than `n` times.

grep – extended regular expressions (ERE)

- What is the meaning of the following symbols in ERE?
  - `|` ... the choice, that matches any string matching either alternate expression.
  - `( )` ... defines a marked subexpression. The string matched within the parentheses can be recalled later.
  - `\n` ... matches the same string of characters as was matched by an expression enclosed between `(` and `)` earlier in the same ERE.

    For example, the expression `^(.*)\1$` matches a line consisting of two repeated appearances of the same string.

grep – extended regular expressions (ERE)

- Copy file `/usr/share/lib/dict/words` from `fray1.fit.cvut.cz` to local system.

```
scp fray1.fit.cvut.cz:/usr/share/lib/dict/words .
```

- Try the following commands and explain the output.

```
grep -E 'o' words
```

```
grep -E 'oo' words
```

```
grep -E 'oo$' words
```

```
grep -E '^oo' words
```

```
grep -E 'o.o' words
```

```
grep -E 'o.*o' words
```

```
grep -E '^o.*o$' words
```

grep – extended regular expressions (ERE)

- What is the difference between the following commands?

```
grep -E 'o*' words
```

```
grep -E 'oo*' words
```

- What is the difference between the following commands?

```
grep -E '[^o]' words
```

```
grep -E -v 'o' words
```

- What is the difference between the following commands?

```
grep -E '^a.*a$' words
```

```
grep -E '^\(a\).*\1$' words
```

```
grep -E '^\(.\).*\1$' words
```

grep – extended regular expressions (ERE)

- Use the output of the command `ps -ef`. How to print the number of processes running under the identity of the user `root`?
  Hint: try to find the correct solution for both Linux and Solaris.

```
ps -ef | grep -Ec '^root '     # Linux
```

```
ps -ef | ggrep -Ec '^ *root '   # Solaris
```

- Use the output of the command `ps -eo pid,user,comm` on the server `fray1.fit.cvut.cz`. How to determine how many times the program `sshd` is running on this server?

```
ps -eo user,pid,comm | ggrep -Ec '/sshd$'
```

grep – table

- Use the output of the command `ps -eo pid,user,comm` on the server `fray1.fit.cvut.cz`. How to determine how many times the program `sshd` is running under the identity of the user `root` on this server?

```
ps -eo pid,user,comm | ggrep -Ec ' root *[^ ]*/sshd$'
```

- Use the output of the command `getent passwd` on the server `fray1.fit.cvut.cz`. How to determine how many students have the account on this server?
  Hint: student's account has the flag `student` at the end of the fifth column.

```
getent passwd | ggrep -Ec ' student:[^:]*:[^:]*$'
```

grep – file

- Assume that the variable `file` is defined as follows

```
file=/home/courses/BIPS1/public/07/NAMES.TXT
```

Use the output of the command `getent passwd` on the server `fray1.fit.cvut.cz`. How to find account info about users whose names are in the file `NAMES.TXT`?

```
getent passwd | ggrep -Ef "$file"
```

grep – extended regular expressions (ERE)

- How to print names of the days of the week from file `/usr/share/lib/dict/words` on the server `fray1.fit.cvut.cz`?

```
Mon day
Tues day
Wednes day
Thurs day
Fri day
Satur day
Sun day
```

```
grep -E '[MTWFS][ouehra][neduit][esnru]*day' words
```

```
grep -E '(Mon|Tues|Wednes|Thurs|Fri|Satur|Sun)day' words
```

Homework

- How to print lines containing the same character at the beginning and at the end of line.
- How to print all palindromes of length 2, 3, 4 and 5 characters from the file `/usr/share/lib/dict/words` on server `fray1.fit.cvut.cz`.