Sami Sainio
UNIX Operating Systems TX00BW20-3004
Metropolia University of Applied Sciences

9.2.2016
Lab 3

Help can be found from page 2 onwards!!!

Fetch from Wikipedia a list of city populations (http://en.wikipedia.org/wiki/List_of_cities_proper_by_population) and copy it to a file. It is known now, that there is bunch of weirdness in the file (such as text ending with .jpg and so on.

1. Write a grep command, that lists all cities from a country you specify.

2. Write a grep command, that finds a city by specific name.

3. Do number 2, but the other way around (exclude the given city name).

Hint: man grep.

4. List all cities with a population above 10 million

5. How many cities have more than 10 million people? (Print with a command)

6. By using grep command, find all processes, whose parent process ID (PPID) is 1, including all kernel threads from the output of command “ps –ef”.

7. Do your best to use sed command to remove *.jpg, *.JPG and other image filytype extensions.

8. Filter the citylist so that you only print one city per country.

Hint: sort, uniq.

Sami Sainio
UNIX Operating Systems TX00BW20-3004
Metropolia University of Applied Sciences

About the command ’grep’

Search, that will find lines starting with one (1)

• grep '^1' (^ searches from the beginning of the line)

Search, that will find lines that end to a six (6)

• grep '6$' ($ searchers from the end of the line)

Search, that will find lines starting with a number

• grep '^[0-9]' (number range defined between the square brackets [XX]

Search, that will find lines where there is one capital letter A-C or X-Z

• grep '^[A-C,X-Z]$'

Search, that will find lines there there is a single capital letter A-C or X-Z that is followed by three characters

• grep '^[A-C,X-Z]...$' (.. matches to any character)

Search, that will find a line that starts with a number and a dot, followed by a space

• grep '^[0-9]\. ' (\ removes the special function of the following character)

Search, that will find same as above, except that the line must end to a small letter

• grep '^[0-9]\. [a-z]$'

Search, that will find a line where the number is not from the range 0-9

• grep '^[^0-9]'

Search, that find a line that starts with one and continues with one or more zeros

• grep '^10*$' (* repeats the character 0 zero or more times)

Search, that finds a line that starts with a number, followed by a dot and a space

• grep '^[1-9][0-9]*\. ' (* repeats the previous parts zero or more times)

Search, that finds a line like above, but the line contains only small characters.

• grep '^[1-9][0-9]*\.[a-z]*$'

More complicated search can be done using egrep, example:

egrep '^[0-9]{7}:'

doable with grep still:

grep '^[0-9]\{7\}:'

Above searches will find lines that starts with 7 numbers and a semicolon

Shortly about sed:

• sed 's/mio/moi/g' (replaces the string ’mio’ with moi)

• sed 's/[0-9]//g' (removes all numbers from the input)

• sed 's/[1-9][0-9]*\./REF/g' (replaces all numbers that are followed by dot with a string REF)