Homework 3

Scripted Log Parsing

# Outcomes

* Learn to work with regular expressions.
* Learn Perl implementations of command line argument parsing, file I/O, and useful data structures.

Rubric

* 2 points each for 5 scripts.

# Description

Given the supplied logfile, write 5 perl (or python, make sure to use python3) scripts named one.pl(py), two.pl(py), etc. which are capable of performing the following tasks on the supplied logs. You should use regular expressions for this homework and may find this [Perl regular expression cheat sheet](https://umanitoba.ca/computing/ist/internal/admin_sys/project_review/media/Perl_Regular_Expressions_Quick_Reference_Card.pdf) and the [Python documentation on regular expressions](https://docs.python.org/3/library/re.html) helpful for this assignment.

## one.pl(py)

Takes a logfile as a command line argument and prints to a file named hourlyInfo which is not a command line argument the number of postfix rejects every minute and the numbers of amavis quarantines every minute.

Example output should contain one line per minute, so you will have 60 lines (because the logfile we supplied only covers an hour) that have the following format.

Mar 1 hh:mm [postfix rejects:27] [amavis quarantines:30]

Mar 1 hh:mm [postfix rejects:17] [amavis quarantines:33]

## two.pl(py)

Opens a file in the current directory called log2 which is not a command line argument (you should make a copy of the log file with this name for testing) and print out the number of ips postfix was connected to/from by both known and unknown IPs and list the IP with the most known connections and the one with the most unknown connections.

Example output

Total Known connection: 99 – [x.x.x.x] accounts for 23 connections

Total Unknown connections 32 – [x.x.x.x] accounts for 21 connections

## three.pl(py)

Takes a logfile as a command line argument and prints out the top 5 From address and the top 5 To addresses. Note you can include both mail that was sent and received both successfully and unsuccessfully.

Example output (I am only showing 2 from and 1 to, you will need 5 of each)

From bill@linux.org 42

From snoopy@freethedogs.net 32

To bill@microsoft.com 21

…

Note you should list them with the top five From first, ranked highest to lowest of the top 5, followed by To, ranked highest to lowest of the top 5.

## four.pl(py)

Opens a file in the current dir called log4 which is not a command line argument (you should make a copy of the log file with this name for testing) and print out the number of messages rejected based on dnsbl.sorbs.net and how many unique ips, unique from addresses and unique to addresses associated with those rejected messages.

Example output

400 messages rejected

23 unique IP’s

27 unique from addresses

12 unique to address

## five.pl(py)

Takes a logfile as a command line argument alters it to cover your tracks. It will remove any line containing any word supplied on the command line as an argument following the file name

The following execution would remove any line that contained reject or bill@microsoft.com or dnsbl.

./five.pl logfile reject bill@microsoft.com dnsbl

# Submission

.tgz (tarred and gzipped) file containing five Perl scripts: one.pl(py), two.pl(py), three.pl(py), four.pl(py), five.pl(py).