CS306 – Introduction to Perl
Summer 2005
Homework Assignment #4
Due: June 29th, 2005, 5pm

The homework should be submitted in the form of a zip file which contains one program for each question. Your zip file should be named *firstname-lastname*-hw4.zip and should contain files like hw4q1, hw4q2, hw4q3, etc... Email this zip file to cs306@cis.uab.edu.

1. (20 pts) Write a mailfile analyzer. This program should take a standard mbox-formatted file of email and produce the following statistics:

- The total number of email messages in the file
- A list in descending order of the number of emails received from each sender
- A list in descending order of the number of emails sent to each recipient
- Average number of lines in the body of the email messages
- A chart showing how many email messages were received on each day of the week
- A chart showing how many email messages were received in each hour of the day
- A list of the unique subject lines (i.e. threads) and the number of messages in each thread
- A distribution chart of hops – intermediate mail servers – the messages passed through

In addition, your program must practice good Perl design principles. Specifically, you should have thorough commenting, well-aligned and well-indented code, subroutines as organizational blocks, no global variables, sensible data structures, meaningful variable names, appropriate error handling and be operating under 'use strict' and 'use warnings'.

Some useful features of mbox files:

- Each new message begins with a From line
- A blank line will separate the header from the body
- Each Received header represents one hop in the path from origin to destination

A sample mbox file will be available for download on the course website. However, a different file will be used to test your script. The output should be neat and easy to read. Please see the sample input on the next page for an example. Your output does not need to match the example, but it must be neat, well-aligned, easy to read and include all of the elements outlined above.

Hint: You may want to structure this as a finite state machine. In other words, you may want some sort of $state variable to keep track of "what type of content" you are in within the mail file.

4 POINT BONUS. Add ASCII bar charts to your output for the chart items (time of day, day of week, number of hops). They can be horizontal (bars grow from left to right) but must fit within 60 columns total width of output no matter how many messages are in the mail file. You should still indicate absolute message/hop counts somewhere on your chart.

Sample Output

```
File Analyzed: /home/fran/mail/mbox

--------------------
| Summary Statistics|
--------------------

Total No. of Messages:    12
Total No. of Threads:      7
Unique Senders:            8
Unique Recipients:         2
Average Message Size:     34 lines

-----------------------
| Detailed Statistics |
-----------------------

Thread List
-----------
Msgs      Subject
----      -------
   4      Pizza Party Friday
   3      Perl Is Fun!
   2      High School Programming Contest results
   1      CS306 HW3 question
   1      Toner low in ugrad lab
   1      [SPAM] 1StopShop: Cheap Rx, hot dates, get rich quick!


Day of Week Distribution
------------------------
Msgs      Day
----      ---
   0      Sunday
   3      Monday
   2      Tuesday
   2      Wednesday
   3      Thursday
   1      Friday
   1      Saturday


Hour of Day Distribution
------------------------
Msgs      Hour
----      ----
   0      12am
   0       1am
   0       2am
```

```
   0        3am
   1        4am
   0        5am
   0        6am
   0        7am
   1        8am
   0        9am
   2       10am
   2       11am
   0       12pm
   3        1pm
   0        2pm
   2        3pm
   1        4pm
   0        5pm
   0        6pm
   0        7pm
   0        8pm
   0        9pm
   0       10pm
   0       11pm

Sender Report
-------------
Msgs       Sender
----       ------
   5       tony@cis.uab.edu
   3       bryant@cis.uab.edu
   2       joe@acme.com
   1       student@uab.edu
   1       bob@spamworld.net

Recipient Report
----------------
Msgs       Recipient
----       ---------
  10       fran@cis.uab.edu
   1       sysadm@cis.uab.edu
   1       cs306@cis.uab.edu

Hop Report
----------
Msgs       Hops
----       ----
   1       1
   1       2
   3       3
   5       4
   0       5
   2       6
```