Homework #4

Due Thursday, March 26, 2009

1. Write an AWK program to convert an Excel style CSV (comma separated) file to pipe delimited file.

   Input files for this program will contain lines with comma-separated data. If a field on a line contains an embedded comma, then that field will be enclosed in double quotes (“).

   An example of an input file could be:

   ```
   Tom Jones,08/15/2007,”Huntsville, AL”,158
   Brad Thompson,09/11/2004,”Nashville, TN”,197
   Sue Moore,03/05/2001,”Atlanta, GA”,594
   ```

   Output from this file would be:

   ```
   Tom Jones|08/15/2007|Huntsville, AL|158|
   Brad Thompson|09/11/2004|Nashville, TN|197|
   Sue Moore|03/05/2001|Atlanta, GA|594|
   ```

   In addition, this program should print the following statistics at the end of the output:

   1. Total number of char in the input file (ignoring field delim – comma and quote)
   2. Total number of fields in the input file
   3. Total number of lines in the input file
   4. Mean (average) line length in number of fields
   5. The number of fields in the longest line
   6. The number of fields in the shortest line
   7. Mean (average) field length in number of characters

   You will be provided a test file shortly before the due date. You must submit the output of your program using this file.

2. Write a AWK program to print a statistical analysis of the words from its input file:

   This program should print a list of words in the input file along with the number of times each word appears in the input file, and the percentage of that number to the total.

   This program should suppress printing words that exists in the input file less than 3 times.

   You should ignore case, and punctuation when evaluating the input file.

   Note: You will be provided a test file shortly before the due date. You must submit the output of your program using this file.