CS 3030 Project Proposal: Tool for Manipulating Excel Files
28 September 2019

Team Members: Sydney Fowler, Priyanka Philip, and Matt Hileman
Github Link: https://github.com/sydneyfowler/CS3030

NOTE: In the original project proposal, I, Sydney, accidentally left off Priyanka's name. As discussed after class on Wednesday, please give her full credit for this assignment.

Project Description:

The purpose of this project is to create a command-line tool of various functions used to work with Excel files. These functions will include:
- Importing / exporting Excel files from various other file types (e.g. csv, txt, etc.)
- Extracting and exporting a set of rows based on a search criteria
- Data validation and cleanup features
- Compressing files
- Sending a copy of the file to a list of email addresses
- Removing duplicate rows based on a set of criteria
- Data Analysis: Sum, Count, Max, Min, all unique, Average, etc.

The tool will contain a main menu, allowing you to choose from the various functionalities it provides. Once you have selected a function, you will drag and drop your file into the tool (or manually provide the path), and the tool will perform the chosen function. Once it finishes, it will return to the main menu, where you can either exit the tool or choose another function.

For example, the data validation and cleanup tool would take a file as input, produce a list of headers, and allow the user to apply cleanup/validation rules to each column of data in the file.

Some potential fields for cleanup:
- Phone numbers
- Email addresses
- Mailing addresses
- Dates
- Web addresses
- Social media information

The tool could also look for invalid pattens (e.g. finding two people listed as one record by looking for " and " or " & " in a name field and then splitting the row into two records). It could take a column and create a unique list of entries, or verify that a column's entries are contained within a unique list inputted by the user. It could truncate fields based on a character limit, or verify the entries in a field are the correct data type (whole number, decimal value, currency, text, datetime stamp, etc.). The idea behind this feature is to automate simple cleanup procedures and help eliminate human error in validation.