**Asset Cleanser**

Asset Cleanser is a utility tool to redact sensitive information from office documents (Word, Excel, PowerPoint, and Visio).

The User interface is built using java swing and it runs a PowerShell script in background to redact data defined by the user as sensitive.

It works on the basis of an Excel file containing data about the string to be masked. Its functionalities are:

1. Changes all the string provided in the Excel file with the masked string in Word, Excel, PowerPoint and Visio files and updates a csv log file with Project Name (Picked up from the Excel file name), the file changes are made in, the data changed, and the number of changes made.
2. Removes all meta data i.e. the file properties such as author name from the aforesaid documents and updates the log file with the file name, and in changes made column it writes “FILE PROPERTY”.
3. Count the no. of images in the files and update the details in the log file.
4. Changes all the string provided in the Excel file with the masked string in Word Headers and Footers and updates the log file.
5. Redact sensitive strings as defined by the used from the File Name as per the Excel masking file and updates the log file mentioning “FILE NAME CHANGED” in the file name column.
6. Project name is extracted from the Excel masking file and is mentioned in the log file while updating each files.
7. Deletes all the images in headers and footers of the Word file and mentions Image in the “change” column of the log file.
8. It redacts the Email ID’s present in the documents by using regular expressions, and changes the string with “Email” in those documents and updates the log file accordingly.
9. It redacts the Phone Numbers present in the documents by using regular expressions, and changes the string with “PhoneNo” in those documents and updates the log file accordingly.
10. It also handles the zip files, by extracting them using 7-zip, processing the all the files and then it compresses them back into the zip version. When a zip file is extracted and compressed is mentioned in the log file so that the user knows which files processed were in the zip file (All rows mentioning the files names in between the Un-zip and zip information are inside the zip file)
11. All the above mentioned functionalities are processed for each and every documents present in the Working directory and its sub-folders and also for zip files and its sub zip files.

The user interface is built for easy running of the PowerShell script which on clicking on the “Run Script” button runs the PowerShell script in background, and keep on displaying the updates done so that the user knows that things are working in the back. The user interface also mentions about the success and failure of the script.