**ASSIGNMENT**

Here is a small assignment to extract key information from an invoice. The invoice samples are given in the Images folder. The data in the invoices is to be extracted using an OCR engine(tesseract) .

This output is in unstructured format. Your task is to extract key information from these text files and create a structured output in the form of excel file (\*\*Sample output file attached\*\*)

Pls follow these steps while doing the assignment.

* Use python to do this exercise.
* Read the text files given in the folder “Plain output” ( “Contoso 1.txt”, “Contoso 2.txt”,”Contoso 3.txt”)
* Extract the fields from the invoice output“ Date”, ”Due Date”, ”Balance Due”
* Save the details to an excel file as given in the sample excel file.

Please give appropriate comments for the code you have written.

**Useful Links**:-

<https://medium.com/@marioruizgonzalez.mx/how-install-tesseract-orc-and-pytesseract-on-windows-68f011ad8b9b> (Tesseract)

<https://opencv24-python-tutorials.readthedocs.io/en/latest/> (Open cv)