|  |  |
| --- | --- |
| Applicant Name | Prajwal waykos |
| Applicant Email | Pwaykos1@gmail.com |
| Applicant Phone | +917249542810 |
| Applicant Gateway | https://prajtech.xyz |
|  |  |

Report – Document Automation

Problem Statement – To automate the replace of variables within given documents and arrange them in the desired format.

Note – You can directly run the code by doing the following steps:

1 .Install necessary packages.

pip install pandas python-docx shutil

2. Run the file automate.py

3. To go back to the normal file arrangements (no variables replaced or no master folder) comment the arrange\_files() functiion and run the restore\_base() function

**Algorithm (Approach)–**

**Step -1 -** Replace All the variables to <<variable>> format in the given documents manually

**Step -2** Read the given data using pandas and create a dictionary of variable-value pairs

**Step-3 –** Parse the Docx one by one and search for the desired variables.

**Step-4** – If a Match is Found, Replace the variable with it’s value from the dictionary formed in step-2

**Step -5**- Create the required folder hierarchy and move the files into their corresponding location as described in the readme.docx

**Step-6** – If anything goes Wrong, Rollback to the original state.

**Detailed Description:**

1. I have created 7 Functions that help me do the desired task.
2. Create\_variables(data): - This function takes in a pandas data frame and outputs a dictionary called variables.

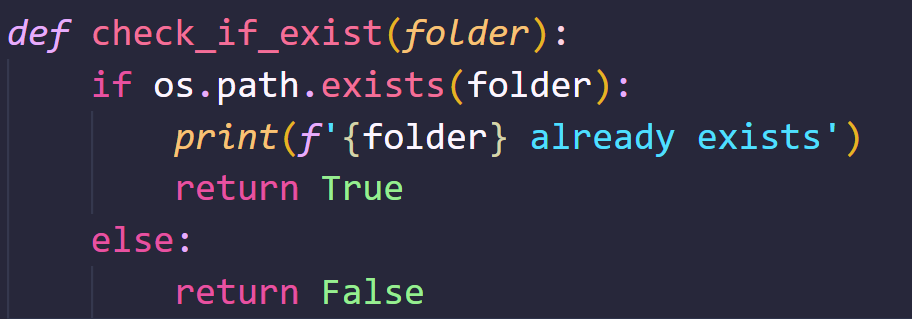
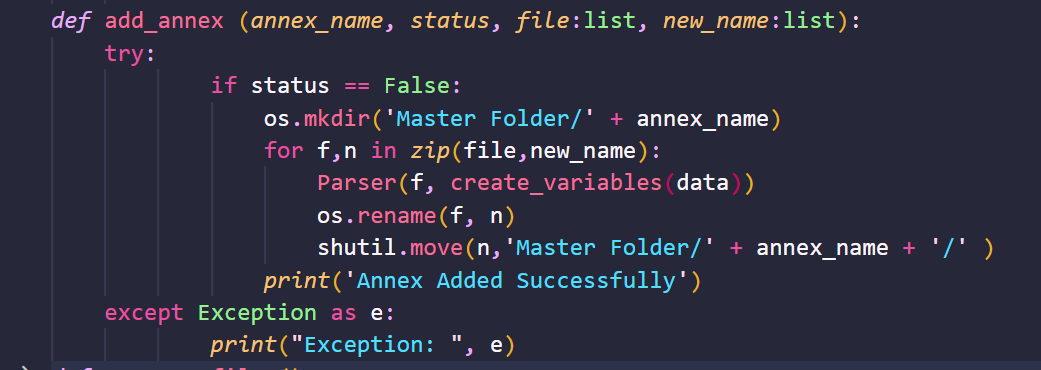
Text

Description automatically generated

1. Parser(file, variables): - This is the main function that parses the given file (parameter) and replaces the variables (<<variable>>) to their value using the values stored in the variables dictionary (parameter)

Text

Description automatically generated

1. Check\_if\_exist(folder): This is a simple one liner function that helps in checking the condition weather a particular folder (parameter) Already exist or not. 
2. Add\_annex (annex\_name, status, file:list, new\_name:list): this is one of the most important functions in the code. it is this function’s responsibility to create the annexes and arrange the files in the given format, it also renames the files as necessary. It takes the parameters Annex\_name (parameter) which is name of the folder (eg. Annex A,B). The status (weather the folder already exist?), file\_list a list of files that are to be placed in it. I had to give a list here since there were multiple files in the Annex C . And new\_name this is also a list due to multiple files in Annex C and it contains the names specified in the Readme.Docx. 
3. Arrange\_files(): This function is a wrapper over all the functions discussed above. It arranges all files in the desired locations and also, by creating the main Master Folder. Text

   Description automatically generated
4. Restore\_base() – This is an additional function that I have introduced , it helps me to get back to the previous state in case I want to undo my process. Text

   Description automatically generated
5. Replace\_with\_formatting() – This is a function that would have replaced text with its formattings but it was having constant failures and many methods didn’t work. Hence, this function Is currently not In use.

A screenshot of a computer

Description automatically generated with medium confidence

Strengths :

1. Robust in changing the variables
2. Can tackle the variables even if they occur multiple times.
3. Less tedious, everything happens in a single click.
4. Simplistic structure.
5. Hardcoding limited to file names (number of files is automated)
6. Additional Feature to rollback to the original state (in-case of failure )
7. Self – sustaining (handles/reports minor issues)
8. Secure file I.o (no miss-replacement or miscalculation)
9. Can handle variations in file structure upto a limit.
10. Can also, change variables within table structures.
11. Optimized code.

weeknesses :

1. Unable to preserve the formatting of the variables.
2. Error handling is not done, due to lack of time
3. Can only handle variations in the file structure upto a limit.

Thanks and Regards,

Prajwal waykos

[Pwaykos1@gmail.com](mailto:Pwaykos1@gmail.com)

<https://prajtech.xyz>