Program Summary

The program I developed for this project is intended as a means to support database designers construct commands to input information into the database. This project and program do not use or support interaction with a database, but is an automatic way to construct the SQL commands to support input of large amounts of data. If this were done manually it would be a very labour intensive process. My program takes input information from text files (input files supplied), processes the information and returns string values that are SQL commands. The python program stores the strings in output files that it creates within the program.

The input information is derived from references of a peer reviewed journal publication. specifically the authors, publication name, publication year, journal, number of citations, DOI and link to publication. Official APA citation is typically used, as follows:

Dale, A.M., Fischl, B., Sereno, M.I. (1999).Cortical surface-based analysis: I. Segmentation and surface reconstruction. NeuroImage,9(2) 179-194.

The program takes as input, **1)** a text file that contains the names of Journals corresponding to papers that will be used as input (**FischlPubJournals.csv**). This input file has repeat names of journals and the program will take this input and process only a list of unique journal names. **2)** The next input file contains researcher names who are authors on the papers that will be processed by the program (**ResearcherNameData.csv**). This file contains only unique researcher names. **3)** The final input file into the program is an excel sheet that contains the remaining aspects of a paper reference, with each attribute in a separate column (**FischlPubsSamp.xlsx**). The program will output three files that contain SQL commands as text: **ComdSqlJournal.csv**, **ComdSqlPaper.csv** and **ComdSqlResearcher.csv**.

The main program that the user should interact with is **auto_construct_command.py** and user defined classes can be found in **class_resercher.py**. The class module is imported into the main program as: **import class_researcher as cr**.

Program Instructions

1) Open auto_construct_command.py and check input files, read comments to ensure the correct files are assigned to the correct variable.

2) Check output files are assigned and will be output in the local directory i.e. no direct paths.

3) Lines 54-64 can be used as a unit test for the user defined function.

4) The journal section takes the journal input file variable pub_journal_data and creates a list. Each item in that list is then passed to the user defined function journal_insert_func().

5) The SQL commands created are stored in the journal SQL output file.

6) The paper section takes the input excel file in the variable pub_all_data and created a pandas dataframe.

7) Each column of the data frame is passed to a variable as a list.

8) The lists are all read as once using zip() in a for loop, the paper insert command is created and saved in the output file.

9) The researcher section takes the text in the researcher input file and creates a list using the variable researcher_lst.

10) Additionally there are four lists of names that specify what type of researcher the person is.

11) Using a for loop and if statements each item in researcher_lst is compared to the specification lists to determine the type of researcher.

12) The following user defined classes are then used to construct the SQL command and output a string: Researcher() - main class, LcnFaculty() - inherited class, LcnMember() - inherited class, LcnCollaborator() - inherited class.

13) At the end of each if statement in the loop the output command string is saved to the researcher output file.