Amanuel Shiferaw
CS 4395.001
Professor Karen Mazidi
September 02, 2022

Homework 1

Describe what the program does

The program reads text from an employee file in order to parse and process the text. The program first splits the text into sentences and processes each sentence of the text omitting the first line of the text as it only describes the formatting of the data and is not the attribute of a specific person. For each person, we delimit the sentences using a comma and process each attribute of a person to fit the requirements provided (i.e Capital case first and last name, capitalize middle initial). The program then adds every person object to the dictionary where the person's id is the key in the dictionary. Finally, the program saves the dictionary as a pickle file, opens/reads the pickle file, and displays all the person objects in the dictionary using the display function defined in the person class.

Explain how to run it

In order to run this program, a prerequisite is to have python3 installed. Clone the git repository and run the following command to run the python script via the terminal.

python3 Homework1_aks210003.py <relative_location_of_data>

Note: Please make sure to provide the system argument to the system.

Write a couple of sentences describing the strengths/weaknesses of Python for text processing, in your opinion

Python has the natural language took kit NLTK which is a group of libraries that are used to make text processing easier. This is complemented by the simple-to-understand syntax and object-oriented programming nature of python. This way we are able to parse large texts and are able to model them as objects to be processed.

Write a couple of sentences describing what you learned in this assignment, or what was a review for you

This is the first python program that I have written and it has allowed me to better understand the syntax of python in detail. In addition, I also learned how to use regular expressions and initially thought that the program needed to convert the phone number from the format they were into the desired 123-456-7890 format and got more in-depth with regular expressions to understand how grouping works in regex. Moreover, I also learned about taking input from the user and outputting to a user. Lastly, I also learned how to pickle/serialize an object and read from a pickle file in python.