

What this program does:

This program takes the data from a CSV file created in an obsolete system and processes the text data inside to be compatible with the new format. The program checks to make sure all of the values conform to the new system, and prompts the user to input new data should it be incompatible. It then saves this data into instances of the Person class, which are inserted into a dictionary, pickled, unpickled, and then displayed for the user.

How to run it:

Navigate to the folder containing the script in your terminal (ensure that the data folder containing data.csv is in the same folder) , and run it using `"python3 nlw180001RegEx.py data/data.csv"`. You will then be prompted to fill in new data should the currently existing data not fit with the new system. After this is done the program will output the data in the new format.

Python for text processing:

One of the major strengths Python has with text processing is that many NLP tools such as the NLTK (natural language toolkit) were written in python and are thus easily usable by the language. However, one drawback to using Python for text processing is the unfortunate fact that the language is oftentimes slower than its counterparts, which can be counterproductive when working with large bodies of text or when real time processing is needed.

What I learned:

While this did allow me to refresh my general programming skills (my recent classes did not require much if any coding) and practice with regular expressions, when all is said and done, I think what I learned the most while doing this assignment was how to use Python at a more advanced level than the small programs I used it for in high school. I have not had to use Python at all during my time at UTD, so it is good that I am able to have an opportunity to relearn and practice with it.

Example run:

Opening file at data/data.csv ...

Raw CSV data:

Last,First,Middle Initial,ID,Office phone

Smith,Smitty,S,WH1234,5557771212

WILLIAMS,WITTY,W,S4454,555-877.4321

Luka,Luka,L,OF4321,555.888.3456

jason,jake,,WH409,555 777 2094

Krishna,krishna,k,SA9384,555 888 0093

Phone number invalid for line: 1

Phone numbers should be in the form ###-###-####

Enter a new number:

111-111-1111

ID invalid for line: 2

IDs should be 2 capital letters followed by 4 digits.

Enter a new ID:

AA1111

Phone number invalid for line: 2

Phone numbers should be in the form ###-###-####

Enter a new number:

222-222-2222

Phone number invalid for line: 3

Phone numbers should be in the form ###-###-####

Enter a new number:

333-333-3333

ID invalid for line: 4

IDs should be 2 capital letters followed by 4 digits.

Enter a new ID:

AA1112

Phone number invalid for line: 4

Phone numbers should be in the form ###-###-####

Enter a new number:

444-444-4444

Phone number invalid for line: 5

Phone numbers should be in the form ###-###-####

Enter a new number:

555-555-5555

Employee list:

ID: WH1234

Name: Smitty S Smith

Phone Number: 111-111-1111

ID: AA1111

Name: Witty W Williams

Phone Number: 222-222-2222

ID: OF4321

Name: Luka L Luka

Phone Number: 333-333-3333

ID: AA1112

Name: Jake X Jason

Phone Number: 444-444-4444

ID: SA9384

Name: Krishna K Krishna

Phone Number: 555-555-5555