Justin Hardy

JEH180008

CS 4395.001

Dr. Mazidi

Text Processing Assignment Overview

I. Program / Assignment Overview

This assignment provided us with an employee file created in an obsolete system, and tasked us with reading the file, processing its text, standardizing the processed text, and storing the processed files into a dictionary of Person objects containing each person's information. Therefore, the purpose of the program is to carry out such a task using regular expressions, and to serialize the standardized data into a file as a dictionary. Such a dictionary would simply map the employee's ID to that of the "Person" class; the object itself containing all relevant standardized data.

II. Program Execution

Running the program is simple, as it is merely a simple python script. First, make sure that you have Python 3 installed onto your computer. Then, use command prompt to open into the directory that contains the program. Then, use the "python" command, followed by the program name to run the program, and the file path of the file you want the program to process. Bear in mind, that running the program without inputting an argument will cause an error, as the program expects a file path to be given it as an argument.

Below is an example of the “python” command being used to execute the program. Note that for the submission on eLearning, the program name may be different as per the professor’s recommendation:

…/Assignment 1 - Text Processing with Python>***python TextProcessing.py data.csv***

III. Strengths / Weaknesses of Text Processing in Python

Python’s I/O streams are fairly easy to work with, although while working with them, I noticed a number of features often available in other languages (i.e., C#, C++, and Java), were missing in python. Most notable, an equivalent to checking when we reach the end of a file stream when reading. Additionally, using python’s read line function gave annoying results, as the newline character would be kept inside of the output, meaning that I’d have to remove it myself.

On the other hand, python’s regular expression libraries were really easy to use. Like other languages, there are functions that allow you to input a regex string and perform a number of actions or checks on another string using that regex string. The functions themselves wrote very nicely and were very easy to use, whereas with other languages I generally find the regex functions to be more of an annoyance to use.

Additionally, python makes string modification easy, as there are a number of functions available that you can use to modify strings, and you need not worry about memory management (unlike C or C++). I feel there is major benefit in that, despite python being objectively slower than the languages that make handling strings annoying.

IV. What I Learned / Reviewed

I'd used python in previous semesters, so this assignment served as a great review for python syntax as a whole, as well as working with file streams and objects. I'd never worked with serialization and regex in python specifically, so this served as a great way to get experience working with both. I also took the time with this assignment to improve my comments, as in previous python assignments I'd made, my comments were very basic and minimal.