Overview of Assignment 1

Assignment 1 delves into text processing with Python. Given a file path, the program opens the employee file; extracts textual information; and performs text processing. Specifically, it parses the information and stores it inside a dictionary. Regarding incorrect data, the program uses regular expressions to detect invalid phone numbers and ID's. It then prompts the user to enter a valid version of the data. The program also exits if it detects duplicate ID keys. After processing the data, the program saves the dictionary to a pickle file. The program finally loads the file and displays the employees' information.

This script can be runned in a Linux terminal. If you already installed Python 3, you can execute the script with the line "python3 Assignment-1.py [file_path]". For this scenario, the file path would be "data/data.csv".

Regarding text processing, Python offers different strengths and weaknesses. Python supports multiple libraries and built-in functions. With these NLP-friendly tools, programmers can quickly perform tasks such as file I/O and string manipulation. By emphasizing simplicity, Python also supports code readability. On the other hand, Python is an interpreted and dynamically-typed language. This leads to slow code execution. Since NLP interacts with large collections of text, Python scripts might endure longer execution time.

This assignment provided a thorough review on Python programming. I had the opportunity to write code for file I/O, text parsing, and regular expression. I enjoyed learning how Python supports NLP tasks. Specifically, I learnt how to perform string manipulation with parsed text. I also learnt how to use regular expressions for filtering information.