

Report:

1st Problem:

I have solved the 2 problems using Python. I wrote Regular expressions and called the Python function for finding all the strings that matches with my given regular expression. The emails, phones, and names extracted using pattern matching are stored in separate lists. There was some garbage collected along with the names of the faculty members which was removed using the word.replace function given by Python for a list. I replaced those garbage words with nothing or with just a space.

Then, once I have extracted and stored all the emails, phones, and names of the faculty members. I created a plain text file named 'facultyRecords.txt' and wrote a header first which contains Name, Email and Phone on the first line.

Then, the information of each faculty member is written using a for loop onto the file one by one. And at the end I closed the file safely. The 'facultyRecords.txt' file is attached in the folder.

2nd Problem:

In the second problem, I opened the pdf file by using Python library for handling pdf files. I created an object for reading pdf files. Then I extracted text from each page of the pdf file using for loop until the pages end.

Then I wrote my Regular Expressions for pattern matching. First, I removed the 'AM and PM' from the schedule.pdf and saved the remaining text in a new string variable. After that, I stored the schedule times into lists. And used that list to convert the time from 12 hour format to 24 hour format using type casting of string into integer. And I stored new converted time into another list.

And at the end I replaced the unconverted time with the converted time using replace function provided by Python.

I created a pdf file 'new-schedule.pdf' and wrote all the text into it with the new time 24-hour format. The 'new-schedule.pdf' file is attached along.