Data Structures and Algorithms/Big Data and Database Programming
Chulalongkorn University
Marko Niinimaki
Fall 2024

Topic 2: Reading and writing files.

Q1: Create a text file called lorem.txt that looks like the one below.

```
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Nulla justo elit, sagittis et gravida at, efficitur at nunc.
Sed leo ante, dapibus sed interdum vitae, egestas eu leo.
Ut leo ex, interdum et vulputate sit amet, euismod in ex.
Quisque feugiat tempus augue, at sollicitudin enim luctus quis.
Maecenas blandit urna malesuada, scelerisque dui sit amet, malesuada sem.
Aenean iaculis porta sem, quis pretium arcu sollicitudin nec.
In volutpat imperdiet dolor vitae bibendum.
```

Using this example https://www.w3schools.com/python/python_file_open.asp write a Python program that reads the file and prints how many lines the file contains.

Q2: Write a program that finds the most frequent **word** in the file. You can use ChatGPT, but ask it to explain the code.

Q3: Structured data. Put the data below in a file called company.csv.

```
Fname,Minit,Lname,Ssn,Bdate,Address,Gender,Salary,Super_ssn,Dept
John,B,Smith,123456789,1965-01-09,731-Fondren-Houston-TX,M,30000,333445555,5
Franklin,T,Wong,333445555,1955-12-08,638-Voss-Houston-TX,M,40000,888665555,5
Alicia,J,Zelaya,999887777,1968-01-19,3321-Castle-Spring-TX,F,25000,987654321,4
Jennifer,S,Wallace,987654321,1941-06-20,291-Berry-Bellaire-TX,F,43000,888665555,4
Ramesh,K,Narayan,666884444,1962-09-15,975-Fire-Oak-Humble-TX,M,38000,333445555,5
Joyce,A,English,453453453,1972-07-31,5631-Rice-Houston-TX,F,25000,333445555,5
Ahmad,V,Jabbar,987987987,1969-03-29,980-Dallas-Houston-TX,M,25000,987654321,4
James,E,Borg,888665555,1937-11-10,450-Stone-Houston-TX,M,55000,NULL,1
```

Using the Python CSV (comma separated values) module, write a program that reads the data and prints only the employees' last names.

Please notice that we don't want to print the first line's "Lname".

For reference:

https://www.youtube.com/watch?v=q5uM4VKywbA Python CSV.

Example: Uses csv, skips the first line, prints last names

import csv

```python
with open('company.csv') as csv_file:
    csv_reader = csv.reader(csv_file, delimiter=',')
    line_count = 0
    for row in csv_reader:
        line_count += 1
        if line_count > 1:
            print(row[2]) #lname in pos 2 (start 0)
```