

## Lab 1: Why Databases? Part II

**Due date:** Saturday, October 1, 8:00pm.

**Note:** Lab 2 will be distributed on September 28.

## Honor System

**This part of the assignment is handed to each group separately,  
upon submission of the first part of the lab.**

While each group will receive the same assignment, I ask you to use the honor system and not directly convey the contents of the assignment to students from the teams that have not yet completed part I of the assignment. Please let other teams concentrate on finishing the first part, without worrying about the content of the second part.

## Lab Assignment

### The Task

In the second part of the assignment, you have to change your program to accommodate for the changes in the input data format.

Recall that the original `students.txt` file had the following format:

```
StLastName, StFirstName, Grade, Classroom, Bus, GPA, TLastName, TFirstName
```

Generally speaking, if each teacher teaches in exactly one classroom, it is a bit of a waste of space to put teacher's name against the name of every student. Therefore, we now revise the format of the input file. Now, the original `students.txt` is replaced with a pair of files: `list.txt` and `teachers.txt`.

The format of `list.txt` is

`StLastName, StFirstName, Grade, Classroom, Bus, GPA`

A sample line from `list.txt` is

`DROP, SHERMAN, 0, 104, 53, 2.67`

(“Sherman Drop is a kindergarten student assigned to classroom 104 who takes bus number 53 and whose GPA is 2.67.”)

The format of `teachers.txt` is

`TLastName, TFirstName, Classroom`

A sample line from `teachers.txt` is

`NIBLER, JERLENE, 104`

(“Jerlene Nibler teaches in classroom 104.”)

The data files are available on-line. The direct URLs for the files (they won’t be linked to the web page until later) are:

<http://www.csc.calpoly.edu/~dekhtyar/365-Fall2016/labs/lab1/list.txt>

<http://www.csc.calpoly.edu/~dekhtyar/365-Fall2016/labs/lab1/teachers.txt>

The information in these files is exactly the same as in the original `students.txt` file, so answers to test searches in both versions of your program should be the same.

Your goal is to change the `schoolsearch` program you have developed for Part I of this assignment to handle the new input data formats. In addition, you are asked to design, and implement in the new program extra search facilities.

## Specs

With the exception of the new functionality, described below, the new `schoolsearch` program shall have the same functionality as the `schoolsearch` program, working exactly the same way, i.e., it shall accept the same search commands and produce the same results.

The key difference is that now, instead of obtaining the data from a single file called `students.txt`, your program will read data from two files: `list.txt` and `teachers.txt`. Same assumptions about the location of the files (current directory) and error handling (minimal) apply.

## Additional Functionality: Extended Search

You are asked to design and implement additional search functionality. In particular, current language of search instructions for the program allows

one to search for information about a student, given student's last name; find the bus route of a student, given student's last name, list all students attending a class taught by a specified teacher; list all students in a specific grade and all students taking a specific bus route and find students with the highest and the lowest GPA in a given grade.

Three new searches need to be added to the program:

- Given a classroom number, list all students assigned to it.
- Given a classroom number, find the teacher (or teachers) teaching in it<sup>1</sup>.
- Given a grade, find all teachers who teach it.

## Additional Functionality: Analytics

The final portion of the functionality for your new program is a bit of analytics.

We are interested in determining whether student performance somehow is in a relationship with some other factors. We use student GPA as a proxy for overall student performance. There are three observable factors that can possibly influence student performance, that are available to us in this data set. These factors are:

- The grade level of the student
- The teacher teaching the student
- The bus route the student is on<sup>2</sup>

Add to your program the commands that allow a data analyst to extract appropriate data to be able to analyze whether student GPAs are affected by the student's grades, student's teachers or the bus routes the students are on.

**Note:** I am being **vague** here **on purpose**. Part of the assignment is to figure out not just what the commands should look like, but also, what needs to be computed and how it needs to be reported.

The program you are creating is not supposed to perform any data analysis by itself, but the analytical commands you develop shall output the information that would allow a data analyst to study the questions of the relationship between a specific factor (from the list of three factors mentioned above) and the GPAs of students.

---

<sup>1</sup>In our test file, there is one teacher per classroom. However, your search algorithm cannot assume that it will always be the case.

<sup>2</sup>You might find it silly to consider that a bus route may somehow be a predictor of student performance in school, but in some places bus routes may be viewed as proxies for socio-economic status of the families of the students, which is a characteristic that is known to be in a relationship with student performance.

For both types of new commands (extended searches and analytics) you shall

- extend the language of search instructions to allow for these three searches to be specified by users;
- implement the functionality supporting each command.

## Implementation notes

Obviously, there are two ways to do this task. One way is to start from scratch and build a brand new program. The other way is to adapt the first program to the new data format. The decision of what to do is left up to individual teams.

## Testing and Deliverables

You have to submit your code, a README file **describing the syntax of the new search commands** and instructions on how to compile (if needed) and run your program, and a write-up.

The write-up shall contain the same information as the write-up for the first part of the assignment. We specifically are interested in how you elected to change the design of your Part I submission in order to address the modifications in the requirements. In addition, I encourage each member of each team, write some additional comments about the assignment: what do you think is the purpose of the assignment? What have you learned?

**Submission.** Use handin as follows:

```
$ handin dekhtyar lab01-2 <Files>
```

**Good Luck!**