



Lab 3/CIS*2250

Pair Programming 2



courselink.uoguelph.ca



A. Hamilton-Wright
K. Raymond &
D.A. Stacey

Overview

This lab is about integrating data between multiple files.

Learning objectives: ○ Learning to manage multiple open files

● Selecting a strategy to deal with file iteration ● Constructing strategies to calculate values based on multiple pieces of data within a file

Skills

coordination + communication (3/6)

organization + planning (3/6)

teamwork (3/6)

programming + tools (5/6)

strategy (1/6)

visualization (0/6)

(*)[The skill scale is from 0 (Fundamental Awareness) to 6 (Main Focus).]

Image description

A pair of sea otters. Image courtesy of US National Oceanic and Atmospheric Administration, who placed this image into the public domain.

Pair Organization

For this lab you will be pair programming with a new partner. This new partner is assigned in the CourseLink group for this lab. You will work with this partner for this lab. Next lab you will have a new partner again. If your assigned partner for this lab is absent, tell the TA and you will be assigned a new partner.

What you learned from the last lab:

1. How to read through a CSV file and select the items that you want.
2. How to read parameters off of the command line, such as the name files that we used in Lab 2 containing the names of people in the United States born in a particular year.

In this lab we will do two things:

1. We will look for the names of people in the class in those American name files.
2. We will trace popular names across a number of years.

Be sure to *read the entire lab description* before beginning the lab, as some important tips are included at the end. (Always a good strategy for any assignment, actually!)

Lab Task Description

Open the CourseLink page for our course (link is in the sidebar) and collect the materials for Lab Assignment 3. You will need the files `cis2250Names.csv` and `usaSSnames.zip`. Unzip these to get the input data for the tasks below.

Retrieve a copy of your Perl script from *last week's lab*. Decide with your new partner which of your implementations to use. If you did not get the previous assignment working, you can download working code from the CourseLink site for but you will lose 10% of your grade for this lab. You will see this available as "Emergency Kit: Solution for Lab 2."

Copy your code into a new file called `findFirstNames.pl` and change it according to the following description:

- the command line should list two files:
 1. the name of one of the "yob" name data files stored within `usaSSnames.zip` (these files contain data about a given "Year Of Birth" in the US)
 2. the name of a file containing first names only, one per line
 - The `cis2250Names.csv` file you have downloaded from CourseLink is in this form
- Your code should read in both files and output the following:
 - Print out each of the names in the `cis2250Names.csv` file (one name per line) and next to each name also print out in brackets its ranking. The ranking should appear in one of the following three forms, depending on the data:
 1. if the name is used for both women and men, then the name should have the female then the male ranking, e.g., "Andrew (3070,7)"
 2. if the name is used *exclusively* for only women *or* for men, then only the appropriate number should appear, e.g., "Kassy (4475)"
 3. if the name does not appear in the name data file at all, it should have ranking zero e.g., "Otter (0)"
- At the end it will print out both:
 - the number of names from the second file that were found in the first file, and
 - the number of names from the second file that *were not* in the first file.



Lab 3/CIS*2250

Pair Programming 2



courselink.uoguelph.ca



A. Hamilton-Wright
K. Raymond &
D.A. Stacey

Overview

This lab is about integrating data between multiple files.

Learning objectives: ○ Learning to manage multiple open files

● Selecting a strategy to deal with file iteration ● Constructing strategies to calculate values based on multiple pieces of data within a file

Skills

coordination + communication (3/6)

organization + planning (3/6)

teamwork (3/6)

programming + tools (5/6)

strategy (1/6)

visualization (0/6)

(*)[The skill scale is from 0 (Fundamental Awareness) to 6 (Main Focus).]

Image description

A pair of sea otters. Image courtesy of US National Oceanic and Atmospheric Administration, who placed this image into the public domain.

Strategies for This Lab

Think about your loop structure: you want to compare each value from one file against every value from the other file. Do you need:

- two loops, one after another?
- two nested loops (one inside the other)?
- three loops (two nested and one later)?

How can you decide this?

Remember that the female names come first and the male names follow so you will have to make sure that the “ranking” for the male names reflects this.

For example, if we search for “Andrew” in the file `yob2000.txt` we would find `Andrew,F,45` on line 3070 of the file, and `Andrew,M,23641` on line 17662.

The first male name, `Jacob,M,34477` is on line 17656. Thus to mark “Jacob” as the male name ranked as number 1 we need to subtract 17655 from the line count in the file. To calculate the rank of the male name Andrew correctly, we also have to subtract 17655 from its line count.

An example output looks like the following:

```
$ perl findFirstNames.pl yob2000.txt cis2250Names.csv
Aakil (0)
Aasim (6313)
Abdul (1061)
Abdullah (869)
Adarsh (4065)
.....
Umut (0)
Wesley (6468,171)
William (3521,11)
Zayn (3920)
Zufar (0)
Number of found names: 90
Number of missing names: 28
```

Overall Programming Strategies

You will likely run into some debugging problems in this lab. Debugging, when you don’t know what is going on, is made much harder if you only have large files to work with.

Ask yourself this question: “*Where can I get smaller data files to help me debug this lab?*”

Another issue in this lab is ensuring that you have exercised each of the important cases. *How many different ranking situations are there? How can you ensure that you have each situation covered, and that you know what the correct answer is for your code to be printing?*

For this lab, it may be the case that Perl::Critic complains about the complexity of your code. We will address this next week.

Once you have completed this lab, go to the CIS*2250 CourseLink site and upload the `findFirstNames.pl` file for grading.