CSCE 4523 Database Management Systems
Homework 1
By: Joshua Rose

Objectives

The objectives of this homework are to implement a random-access file database. By doing so, I will gain a better understanding of the data management system that was popular before relational databases became ubiquitous. This will help demonstrate the benefits and limitations of this older style of data management and how it compares to the strengths and weaknesses of a relational database. As I understand from the lecture and personal experience, relational databases are more flexible and easier to deal with at scale.

Approach

I started off with the sample code provided to get a general understanding of how the program works, modifying it as needed to complete the requirements for Part 1 of the assignment. I opted to use C++ and enjoyed learning some new strategies for writing to a file from the sample code. Previously, I had not used the '<<' and '>>' operators for anything other than 'cin' or 'cout/cerr' but I had seen the sample code methods used in textbooks, so it was beneficial to learn this technique for simple file operations using an FStream object.

My strategy was to attempt to write only one method at a time before doing some sort of testing to verify that my intent was being carried out before moving forward. I only felt a need to use two classes, one for the database and one for the main method and user menu functions. It might be beneficial to break out my user menu into a helper class to keep my main class tidier, but I think it is reasonable without doing so. Whenever possible, I passed by reference to avoid copying objects and strings unnecessarily to optimize program size.

Where I deviated from the sample code the most is how I handle intake and modification of individual record columns as I wanted my design to be able to accommodate a varying number of columns per record (assuming all records were equal in column quantity). For that reason, I chose to use a single vector of strings object and associated pointer for record retrieval and input versus hardcoded strings for the columns. This way I could use loops based on vector size to create a more dynamic set of functions.

Benefits of this are easier iteration for printing the elements for testing and display and passing one pointer argument instead of multiple string pointer arguments to methods. Columns of one record are related to the index of the vector. Additionally, a matching sized vector was used in the database to hold default field names and max widths, programmatically set by the .config file contents. This made it much more convenient to handle truncation of excessive user inputs through trivial vector access versus reopening the .config file each time.

The format of the record itself is determined by the .config file contents. The first line of the .config file tracks state of sorted record quantity versus overflow record quantity which lets my program know how many entries the database will contain. The remaining lines of the .config

each contain one default column value which doubles as the header value for displaying records, comma separated from the second value per line which is the max width of that column. Conveniently, I was able to add all the latter together after reading the .config file and increment by one to determine the record size (using non-Windows OS). Here are the contents of my .config file after the tests add two overflow records:

```
500,2
NAME,38
RANK,3
CITY,19
STATE,2
ZIP,5
EMPLOYEES,7
```

My record entries contain no padding, spaces are converted to underscores, and the order of columns remains the same as the intake .csv. My program assumes that the first column is the unique, sorted key for the record. There is no delimiter between column entries, only the linefeed character at the end of each record entry. For this dataset including the linefeed character, total record size is 75. Here is a sample entry:

```
FIDELITY_NATIONAL_INFORMATION_SERVICES301JACKSONVILLE      FL3220455000
```

To delete records without breaking binary search, soft deletes were used. I would simply change all fields beyond the key to '-1' and modify 'readRecord()' to check that the second column of a successful binary search did not contain '-1' before displaying a successful search to the user. One obvious flaw of this logic is that re-adding a record of the same key name will put it in the overflow area but binary search will not reach it afterward as it will succeed in finding the deleted record first (but will not show it as successfully found to the user). This could be easily fixed with some additional logic, but the requirements did not state anything about re-adding deleted records, so I did not spend time on it.

Overflow was handled without too much trouble. Essentially, it required adding logic to binary search that would be executed if the number of overflow records was greater than zero and that binary search was exhausted based on the number of sorted records. Then it simply calls 'readRecord()' for each overflow record, checking for a match of the search key each time. Finally, my 'db.close()' method was modified to reopen the .config file and update the first line with the new number of overflow records. This allows the new records to persist between sessions.

Results

Error handling was implemented through checking that the database was opened before accessing most method logic and input checking for valid inputs from the menu logic as opposed to handling bad inputs in private methods. My menu will reject empty string inputs for record columns, requiring at least a space character and my 'writeRecord()' method uses the data from .config to truncate any excessively long input to its max allowable width. This does not throw any error though, so the user will not know until reading the record.

I believe my program is running at optimum efficiency given the constraints of using binary search with an appended linear search for overflow. If we denote number of sorted records as N and unsorted as M, then efficiency should $\log(N) + M \Rightarrow$ so $O(\log(N))$ overall, not including small amounts of overhead per query for checking fields widths, headers, etc.

Testing

I made sure to include all sample test cases and added a few that I deemed necessary. These included invalid menu inputs, excessively long field inputs, empty field inputs, and attempting to use menu options in an illogical order to see if anything would break. One bug I discovered that I am unable to figure out is that creating the same database during the same session will cause an empty database to be written. This happens even when you close the database but don't close the program, then recreate the database.

This is quite baffling to me because starting the program with an existing database is nearly identical, but the new database will be written properly. This does not cause any requirements of the assignment to be missed, but it is something I wish to correct in the future when I have more time. My intuition is that is has something to do with the reinitialization of the vector containing the .config information that occurs when restarting the program but not when closing the database and recreating it.

During test I also discovered another bug. Since I was storing my entries with underscores instead of spaces for the keys, String's .compare() method would value an underscore as higher than a letter of the alphabet versus a space which is normally valued lower. This caused an issue with the search for "FIRST AMERICAN FINANCIAL" when it was compared against "FIRSTENERGY" because the space was replaced by an underscore, compared to 'E' and sent binary search in the wrong direction. I barely caught this while writing this report and just used two temporary strings to replace underscores with spaces before calling '.compare().'

Typescript

```
Script started on Sun Feb 12 10:52:58 2023

g++ -I include/ -c src/rafdb.cpp -o objs/rafdb.o
g++ -I include/ -c src/main.cpp -o objs/main.o
g++ objs/*.o -o bin/program
Dir contents before creating database:
total 96
drwxr-xr-x  3 admin  staff      96 Feb 12 10:53 bin
drwxr-xr-x  4 admin  staff     128 Feb 12 10:53 objs
-rw-r--r--  1 admin  staff      43 Feb 12 10:52 typescript
-rw-r--r--  1 admin  staff      54 Feb 12 10:49 Fortune500.config
-rw-r--r--  1 admin  staff     836 Feb 11 20:49 input2
-rwxr--r--  1 admin  staff     203 Feb 11 14:00 tests.sh
-rw-r--r--  1 admin  staff      35 Feb 11 13:59 input1
drwxr-xr-x  4 admin  staff     128 Feb  1 20:40 src
drwxr-xr-x  3 admin  staff      96 Feb  1 20:40 include
-rw-r--r--  1 admin  staff     211 Feb  1 20:40 Makefile
-rw-r--r--@ 1 admin  staff   21691 Jan 30 10:25 Fortune500.csv
>> [testing no choice followed by 'enter' shows menu]
************* MENU **************
*                              *
*    #1) Create new database    *
*    #2) Open database          *
*    #3) Close database         *
*    #4) Display record         *
```

```
*    #5) Update record         *
*    #6) Create report         *
*    #7) Add record            *
*    #8) Delete record         *
*    #9) Quit                  *
*                              *
********************************
>> [testing wrong name for create database]
Please enter the input .csv name without file extension:
CREATE DATABASE "WrongCSVName500"
Error opening 'WrongCSVName500.config'
        Hint: does file exist?
>> [testing correct name for create database]
Please enter the input .csv name without file extension:
CREATE DATABASE "Fortune500"
Successfully created Fortune500.data from Fortune500.csv
[menu option 9 to quit]
>> exiting...
Dir contents after creating database:
total 176
-rw-r--r--  1 admin  staff  37500 Feb 12 10:53 Fortune500.data
drwxr-xr-x  3 admin  staff     96 Feb 12 10:53 bin
drwxr-xr-x  4 admin  staff    128 Feb 12 10:53 objs
-rw-r--r--  1 admin  staff     43 Feb 12 10:52 typescript
-rw-r--r--  1 admin  staff     54 Feb 12 10:49 Fortune500.config
-rw-r--r--  1 admin  staff    836 Feb 11 20:49 input2
-rwxr--r--  1 admin  staff    203 Feb 11 14:00 tests.sh
-rw-r--r--  1 admin  staff     35 Feb 11 13:59 input1
drwxr-xr-x  4 admin  staff    128 Feb  1 20:40 src
drwxr-xr-x  3 admin  staff     96 Feb  1 20:40 include
-rw-r--r--  1 admin  staff    211 Feb  1 20:40 Makefile
-rw-r--r--@ 1 admin  staff  21691 Jan 30 10:25 Fortune500.csv
Running program with remaining tests:
>> [enter to show menu]
************* MENU **************
*                              *
*    #1) Create new database   *
*    #2) Open database         *
*    #3) Close database        *
*    #4) Display record        *
*    #5) Update record         *
*    #6) Create report         *
*    #7) Add record            *
*    #8) Delete record         *
*    #9) Quit                  *
*                              *
********************************
>> [test attempting display record before opening database]
No database is open yet!
>> [testing attempt to open wrong name database]
Please enter the name of a database to open:
OPEN DATABASE "WrongDataBaseName500"
Error opening 'WrongDataBaseName500.config'
        Hint: does file exist?
>> [testing opening database correct name]
Please enter the name of a database to open:
OPEN DATABASE "Fortune500"
Succesfully opened Fortune500.data
>> [testing display record]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "FIRST AMERICAN FINANCIAL"
--------------------------------------------------------------------------------------
|NAME                          |RANK|CITY            |STATE|ZIP  |EMPLOYEES|
--------------------------------------------------------------------------------------
|FIRST AMERICAN FINANCIAL      |464 |SANTA ANA       |CA   |92707|19531    |
--------------------------------------------------------------------------------------
>> [testing display record]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "INTEL"
--------------------------------------------------------------------------------------
|NAME                          |RANK|CITY            |STATE|ZIP  |EMPLOYEES|
```

```
-----------------------------------------------------------------------------
|INTEL                            |47  |SANTA CLARA        |CA  |95054|106000   |
-----------------------------------------------------------------------------
```
>> [testing display record]
```
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "TEXAS INSTRUMENTS"
-----------------------------------------------------------------------------
|NAME                             |RANK|CITY               |STATE|ZIP  |EMPLOYEES|
-----------------------------------------------------------------------------
|TEXAS INSTRUMENTS                |206 |DALLAS             |TX  |75243|29865    |
-----------------------------------------------------------------------------
```
>> [testing display record]
```
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "CONSOLIDATED EDISON"
-----------------------------------------------------------------------------
|NAME                             |RANK|CITY               |STATE|ZIP  |EMPLOYEES|
-----------------------------------------------------------------------------
|CONSOLIDATED EDISON              |234 |NEW YORK           |NY  |10003|14960    |
-----------------------------------------------------------------------------
```
>> [testing display record doesn't exist]
```
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "ABC SYSTEMS"
No results found.
```
>> [testing display record]
```
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "CONSOLIDATED EDISON"
-----------------------------------------------------------------------------
|NAME                             |RANK|CITY               |STATE|ZIP  |EMPLOYEES|
-----------------------------------------------------------------------------
|CONSOLIDATED EDISON              |234 |NEW YORK           |NY  |10003|14960    |
-----------------------------------------------------------------------------
```
>> [testing edit record rank]
```
Please enter the name of record to edit (case-sensitive):
-----------------------------------------------------------------------------
|0                                |1   |2                  |3   |4    |5        |
-----------------------------------------------------------------------------
|CONSOLIDATED EDISON              |234 |NEW YORK           |NY  |10003|14960    |
-----------------------------------------------------------------------------
Index of element to edit (cannot edit index 0, key):
Enter new value for element:
UPDATE "CONSOLIDATED_EDISON" RECORD RANK FROM "234" TO "250"
Success updating entry with key: CONSOLIDATED EDISON
```
>> [testing edit record rank excessive length input]
```
Please enter the name of record to edit (case-sensitive):
-----------------------------------------------------------------------------
|0                                |1   |2                  |3   |4    |5        |
-----------------------------------------------------------------------------
|CONSOLIDATED EDISON              |250 |NEW YORK           |NY  |10003|14960    |
-----------------------------------------------------------------------------
Index of element to edit (cannot edit index 0, key):
Enter new value for element:
UPDATE "CONSOLIDATED_EDISON" RECORD RANK FROM "250" TO "250TESTINGMAXWIDTH"
Success updating entry with key: CONSOLIDATED EDISON
```
>> [testing edit record change city]
```
Please enter the name of record to edit (case-sensitive):
-----------------------------------------------------------------------------
|0                                |1   |2                  |3   |4    |5        |
-----------------------------------------------------------------------------
|CONSOLIDATED EDISON              |250 |NEW YORK           |NY  |10003|14960    |
-----------------------------------------------------------------------------
Index of element to edit (cannot edit index 0, key):
Enter new value for element:
UPDATE "CONSOLIDATED_EDISON" RECORD CITY FROM "NEW_YORK" TO "NewYork"
Success updating entry with key: CONSOLIDATED EDISON
```
>> [testing display edited record]
```
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "CONSOLIDATED EDISON"
-----------------------------------------------------------------------------
|NAME                             |RANK|CITY               |STATE|ZIP  |EMPLOYEES|
-----------------------------------------------------------------------------
|CONSOLIDATED EDISON              |250 |NewYork            |NY  |10003|14960    |
```

```
--------------------------------------------------------------------------------
>> [testing display record before edit]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "R.R. DONNELLEY & SONS"
--------------------------------------------------------------------------------
|NAME                                 |RANK|CITY            |STATE|ZIP  |EMPLOYEES|
--------------------------------------------------------------------------------
|R.R. DONNELLEY & SONS                |388 |CHICAGO         |IL   |60606|44360    |
--------------------------------------------------------------------------------
>> [testing edit record attempt to edit key/name]
Please enter the name of record to edit (case-sensitive):
--------------------------------------------------------------------------------
|0                                    |1   |2               |3    |4    |5        |
--------------------------------------------------------------------------------
|R.R. DONNELLEY & SONS                |388 |CHICAGO         |IL   |60606|44360    |
--------------------------------------------------------------------------------
Index of element to edit (cannot edit index 0, key):
UPDATE "" RECORD
Error: cannot edit index 0, key!
>> [testing display record after attempted key/name edit]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "R.R. DONNELLEY & SONS"
--------------------------------------------------------------------------------
|NAME                                 |RANK|CITY            |STATE|ZIP  |EMPLOYEES|
--------------------------------------------------------------------------------
|R.R. DONNELLEY & SONS                |388 |CHICAGO         |IL   |60606|44360    |
--------------------------------------------------------------------------------
>> [testing display earlier edited record]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "CONSOLIDATED EDISON"
--------------------------------------------------------------------------------
|NAME                                 |RANK|CITY            |STATE|ZIP  |EMPLOYEES|
--------------------------------------------------------------------------------
|CONSOLIDATED EDISON                  |250 |NewYork         |NY   |10003|14960    |
--------------------------------------------------------------------------------
>> [testing delete record]
Please enter the name of record to delete (case-sensitive):
Success deleting record: PHILLIPS 66
>> [testing display deleted record]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "PHILLIPS 66"
No results found.
>> [testing delete record]
Please enter the name of record to delete (case-sensitive):
Success deleting record: J.B. HUNT TRANSPORT SERVICES
>> [testing display deleted record]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "J.B. HUNT TRANSPORT SERVICES"
No results found.
>> [testing delete record]
Please enter the name of record to delete (case-sensitive):
Success deleting record: ACTIVISION BLIZZARD
>> [testing display deleted record]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "ACTIVISION BLIZZARD"
No results found.
>> [testing generate report]

CREATE REPORT FIRST 10 RECORDS
--------------------------------------------------------------------------------
|NAME                                 |RANK|CITY            |STATE|ZIP  |EMPLOYEES|
--------------------------------------------------------------------------------
|3M                                   |94  |MAPLEWOOD       |MN   |55144|91584    |
|A-MARK PRECIOUS METALS               |395 |PHILADELPHIA    |PA   |19107|83000    |
|ABBOTT LABORATORIES                  |135 |SANTA MONICA    |CA   |90401|75000    |
|ABBVIE                               |111 |NORTH CHICAGO   |IL   |60064|30000    |
|ABM INDUSTRIES                       |500 |NEW YORK        |NY   |10176|110000   |
|ADOBE SYSTEMS                        |443 |MCLEAN          |VA   |22102|15706    |
|ADP                                  |240 |ROSELAND        |NJ   |7068 |57000    |
|ADVANCE AUTO PARTS                   |292 |ROANOKE         |VA   |24012|57500    |
|AECOM                                |161 |LOS ANGELES     |CA   |90067|87000    |
```

```
|AES                               |194 |ARLINGTON         |VA   |22203|19000   |
-----------------------------------------------------------------------------------
>> [testing add overflow record]
Enter data for field "NAME":
Enter data for field "RANK":
Enter data for field "CITY":
Enter data for field "STATE":
Enter data for field "ZIP":
Enter data for field "EMPLOYEES":
ADD RECORD "HANDMADE GRAPHICS, 501, MyCity, MyState, 12345, 100"
Success writing new record with key: HANDMADE GRAPHICS
>> [testing display added record]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "HANDMADE GRAPHICS"
-----------------------------------------------------------------------------------
|NAME                              |RANK|CITY              |STATE|ZIP  |EMPLOYEES|
-----------------------------------------------------------------------------------
|HANDMADE GRAPHICS                 |501 |MyCity            |My   |12345|100      |
-----------------------------------------------------------------------------------
>> [testing consecutive display added record]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "HANDMADE_GRAPHICS"
-----------------------------------------------------------------------------------
|NAME                              |RANK|CITY              |STATE|ZIP  |EMPLOYEES|
-----------------------------------------------------------------------------------
|HANDMADE GRAPHICS                 |501 |MyCity            |My   |12345|100      |
-----------------------------------------------------------------------------------
>> [testing display earlier record with attempted key/name edit]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "R.R. DONNELLEY & SONS"
-----------------------------------------------------------------------------------
|NAME                              |RANK|CITY              |STATE|ZIP  |EMPLOYEES|
-----------------------------------------------------------------------------------
|R.R. DONNELLEY & SONS             |388 |CHICAGO           |IL   |60606|44360    |
-----------------------------------------------------------------------------------
>> [testing edit added overflow record rank]
Please enter the name of record to edit (case-sensitive):
-----------------------------------------------------------------------------------
|0                                 |1   |2                 |3    |4    |5        |
-----------------------------------------------------------------------------------
|HANDMADE GRAPHICS                 |501 |MyCity            |My   |12345|100      |
-----------------------------------------------------------------------------------
Index of element to edit (cannot edit index 0, key):
Enter new value for element:
UPDATE "HANDMADE_GRAPHICS" RECORD RANK FROM "501" TO "510"
Success updating entry with key: HANDMADE GRAPHICS
>> [testing edit edited overflow record]
Please enter the name of record to edit (case-sensitive):
-----------------------------------------------------------------------------------
|0                                 |1   |2                 |3    |4    |5        |
-----------------------------------------------------------------------------------
|HANDMADE GRAPHICS                 |510 |MyCity            |My   |12345|100      |
-----------------------------------------------------------------------------------
Index of element to edit (cannot edit index 0, key):
Enter new value for element:
UPDATE "HANDMADE_GRAPHICS" RECORD STATE FROM "My" TO "My State"
Success updating entry with key: HANDMADE GRAPHICS
>> [testing display twice edited overflow record (STATE truncated due to limited width of 2)]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "HANDMADE_GRAPHICS"
-----------------------------------------------------------------------------------
|NAME                              |RANK|CITY              |STATE|ZIP  |EMPLOYEES|
-----------------------------------------------------------------------------------
|HANDMADE GRAPHICS                 |510 |MyCity            |My   |12345|100      |
-----------------------------------------------------------------------------------
>> [testing add record with empty fields aside from name (requires single space char)]
Enter data for field "NAME":
Enter data for field "RANK":
Enter data for field "CITY":
Enter data for field "STATE":
Enter data for field "ZIP":
```

```
Enter data for field "EMPLOYEES":
ADD RECORD "American National Insurance,  ,  ,  ,  ,  "
Success writing new record with key: American National Insurance
>> [testing display overflow record with empty fields]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "American National Insurance"
-------------------------------------------------------------------------------
|NAME                                |RANK|CITY           |STATE|ZIP  |EMPLOYEES|
-------------------------------------------------------------------------------
|American National Insurance         |    |               |     |     |         |
-------------------------------------------------------------------------------
>> [testing delete overflow record]
Please enter the name of record to delete (case-sensitive):
Success deleting record: American National Insurance
>> [testing display deleted overflow record]
Please enter the name to search (case-sensitive):
DISPLAY COMPLETE RECORD "American National Insurance"
No results found.
>> [testing menu #3, close database]
CLOSE OPEN DATABASE ANY
Closing any open database files.
>> [testing display record while database closed]
No database is open yet!
[end tests, input 9 to quit]
>> exiting...
Script done on Sun Feb 12 10:53:05 2023
```