

Java – Database Assignment Report

STAGE 1 – RECORDS

1.1 DESIGN DECISIONS

- The first decision was to use an ArrayList for storing the strings for each Record. Since this program will not be used for very large databases, we don't need to worry about the minutiae of how the data is stored for optimal space/speed efficiency. Using the built in java utilities will save a lot of time, since all of the resizing, setting, getting, length etc. functions are already available to us.
- Using an ArrayList is the most suitable option for data storage since we need to be able to retrieve data from individual cells and store the strings in a set order which will not change.
- Almost every function in Record needs to be public, since all the functions are currently to do with retrieving data from the record. However, the ArrayList object can be private, meaning it can only be accessed or modified through the provided functions.

1.2 NEW FILES AND RESTRUCTURING

- Two files have been created for this section, Database.java and Record.java
- Database is the controlling class, currently this has little purpose other than for testing the creation of new records. Later it will be used to set up the databases by using all the other classes.
- Record is the class that acts as a single 'row' of a table. Currently there are no tables, so the records have no restrictions on size. The constructor builds the record by taking an string array as input, then adding all the strings to the list in order.

2 TABLES

2.1 DESIGN DECISIONS

- Database.java has been left unchanged and currently serves no purpose other than to test Table and Record. When the database software is more functional, all the other classes will be used by the Database class.
- Changed max line/comment length to 100 chars, 80 was becoming too restrictive and causing print statements to roll over multiple lines.
- Initially I wrote the Table class with a separate List for storing the attribute names. Later it became clear that it would be easier to write DRY code if I was able to iterate simply over one list rather than two separately, so the code was modified so that attribute names for a table are stored in index 0 of the records List.
- Success and Failure print statements were added to all the public functions, to make it clearer what happens when the functions are used.

- Initially the select function was written to just print out the selected record, with tabs between each attribute. To help make it clearer which attribute was which, this function was modified to also print the attribute names above the record attributes. This required several additional functions to properly align the names with the values, but the much nicer output is worth it, and the table formatting code will be highly reusable in future iterations of the database.

```

+-----+-----+-----+
| Id | Name | Kind | Owner |
+-----+-----+-----+
| 1 | Fido | dog | ab123 |
+-----+-----+-----+
| Id | Name | Kind | Owner |
+-----+-----+-----+
| 2 | Wanda | hippopotamus | ef789 |
+-----+-----+-----+

```

2.2 NEW FILES AND RESTRUCTURING

- Some small changes were made to the Record class:
 - The return value of `getElement()` when given an invalid index was changed to a null character. Previously `getElement()` would return a string, "ERROR", which could potentially cause issues if the user created a table with the word ERROR in it.
 - The `printRecord()` function has been removed. Now that the Table class has been added, this function is obsolete since all console output will be controlled by Table.
- One new file has been added, Table.java. This class acts as a single table from a database. The constructor builds the table using a String array of the attribute names. The table then keeps an ArrayList of records, where the record at index 0 stores the table attribute names and the remaining records in the list are the rows of the table. Several public functions are available:
 - `insert(Record newRecord)` – inserts a new record into the table.
 - `delete(int index)` – removes a record from the table at the given index.
 - `update(int index, Record newRecord)` – updates the record at the given index to the given new record.
 - `select(int index)` – prints out the record at the given index, correctly formatted into a table with the attribute names above.
- 4 other functions, `printSelected()`, `setPrintwidth()`, `printRow()` and `printDivider()` are all supporting functions for `select()`, used to properly format the table.

3 FILE INPUT/OUTPUT

3.1 DESIGN DECISIONS

- Originally I planned to write FileIO as a separate class called independently from the Table class to save and load files, which seemed like a logical layout. However, it became clear that this would make it difficult to avoid cyclic dependencies. The controlling class would need to use functions from Table and FileIO, and FileIO would need to send and receive data from Table in order to function. I decided to instead have the `saveTable()` and `loadTable()` functions inside Table.java, which can be called from the controlling classes. Since these functions are so directly related to tables, it is reasonable to keep them in this class.

- For now I will avoid separate directories for files and store all csv files in the tableFiles folder. Since I intend to store all tables from a database in the same directory, the database class will deal with setting up folders once it has been implemented.
- For now in the CSV files commas are not allowed in strings. Any records with additional commas will fail to insert into the table. Generally for storage of simple words and names commas will be unnecessary. However, if I add specific data types for attributes at a later point then storage of strings will need to change and this feature could be added.
- An example csv file (test.csv) is created during testing to test the loading and saving features. It is then deleted at the end of testing.
- A return type was added to select for testing purposes. The number of records printed is returned as an integer value.

3.2 NEW FILES AND RESTRUCTURING

3.2.1 Table.java Changes:

- A `tableName` variable was added (used for the filename when saving/loading tables) .
 - `getName()` function, getter for the table name.
- A second Table constructor for use when loading a table from file rather than initialising an empty table.
 - `Table(String fileName)`, loads table from the input file name.
- `saveTable()` function, saves table to CSV file, using the table name ("tableName".csv)
- `loadTable(String filename)`, called by the second Table constructor.
- The `select()` function and the related functions `printSelected()` & `setPrintwidth()` were all modified to make select more versatile:
 - Some error checks were removed, under the assumption that the user will not directly call 'select' etc., so higher levels of the database can ensure that the input indexes will be valid beforehand.
 - `select()` and `printSelected()` were merged, since select without the error checks was significantly shorter.
 - The remaining `select()` and `setPrintwidth()` functions were modified to take an index array instead of single index, so multiple records can be printed together in the same table.
 - The input `int[] {0}` is treated as a 'select all records' statement (SELECT *).
- The table printing from select statements was improved to look nicer, now uses the box drawing Unicode characters.
- `insert()` was causing a `NullPointerException` when the first record (attribute names) was added to the table via the `loadTable()` function. The insert function checks that the new record is the same size as the first record in the table, but before the attribute names are added the first record is NULL. The insert function was modified to accept an record if the records ArrayList size = 0.
- The testing function was modified & extended to reflect the changes made above.
 - Equals functions were added to Table (& record) in order to check if two tables are identical.

3.2.2 Other Changes:

- Renamed database.java to controller.java

- The name database was misleading, since a database is technically a collection of tables (which does not yet exist in my program).
- Controller is more appropriate since it functions by making calls to the supporting classes, table and record.
- The Controller now loads a table from file and prints out all the records, as an example use case.

Example run of the program loading and printing sample CSV election data:

Entries loaded: 142
Failed entries: 0

date	ward	electorate	candidate	party	votes	percent
2014-05-22	Avonmouth	9185	Patrick Hulme	Trade Unionists and Socialists Against Cuts	16	0.47
2014-05-22	Avonmouth	9185	Ian Humfrey Campion-Smith	Liberal Democrat	108	3.17
2014-05-22	Avonmouth	9185	Stephen Kenneth James Norman	Independent	135	3.97
2014-05-22	Avonmouth	9185	Justin Michael Quinnell	Green	149	4.38
2014-05-22	Avonmouth	9185	Spud Murphy	UKIP	878	25.79
2014-05-22	Avonmouth	9185	John Thomas Bees	Labour	1051	30.88
2014-05-22	Avonmouth	9185	Matthew Simon Melias	Conservative	1067	31.35
2014-05-22	Bedminster	9951	Robin Victor Clapp	Trade Unionists and Socialists Against Cuts	204	5.47
2014-05-22	Bedminster	9951	Thom Oliver	Liberal Democrat	264	7.08
2014-05-22	Bedminster	9951	Sarah Helen Cleave	Conservative	680	18.23
2014-05-22	Bedminster	9951	Catherine Slade	Green	838	22.46
2014-05-22	Bedminster	9951	Mark Bradshaw	Labour	1745	46.77
2014-05-22	Bishopston	10303	Martin James Saddington	Trade Unionists and Socialists Against Cuts	65	1.37
2014-05-22	Bishopston	10303	Owen James Evans	Conservative	511	10.76
2014-05-22	Bishopston	10303	Barry John Cash	Liberal Democrat	757	15.95
2014-05-22	Bishopston	10303	Eileen Means	Labour	1168	24.61
2014-05-22	Bishopston	10303	Tim Malnick	Green	2246	47.31
2014-05-22	Bishopsworth	8857	Joseph John Etherington	Trade Unionists and Socialists Against Cuts	88	3.28
2014-05-22	Bishopsworth	8857	Gareth Owen	Liberal Democrat	92	3.43
2014-05-22	Bishopsworth	8857	Alan Wilson Baker	Green	230	8.58
2014-05-22	Bishopsworth	8857	Jon Craig	Independents for Bristol	431	16.07
2014-05-22	Bishopsworth	8857	Kye Daniel Dudd	Labour	783	29.19
2014-05-22	Bishopsworth	8857	Kevin Michael Quartley	Conservative	1058	39.45
2014-05-22	Brislington East	8893	Matthew Gordon	Trade Unionists and Socialists Against Cuts	35	1.11
2014-05-22	Brislington East	8893	Philip Collins	Independent	138	4.39
2014-05-22	Brislington East	8893	Pauline Mary Allen	Liberal Democrat	199	6.33
2014-05-22	Brislington East	8893	Peter Antony Goodwin	Green	241	7.66
2014-05-22	Brislington East	8893	Perry Hicks	Conservative	645	20.5
2014-05-22	Brislington East	8893	John Langley	UKIP	886	28.16
2014-05-22	Brislington East	8893	Mike Wollacott	Labour	1002	31.85
2014-05-22	Brislington West	8439	Ibado Ali Mahamoud	Trade Unionists and Socialists Against Cuts	63	2
2014-05-22	Brislington West	8439	Roy Towler	Conservative	501	15.91
2014-05-22	Brislington West	8439	Peter Henry Main	Liberal Democrat	800	25.4
2014-05-22	Brislington West	8439	Christopher James Robinson	UKIP	813	25.82
2014-05-22	Brislington West	8439	Rhian Elena Greaves	Labour	972	30.87
2014-05-22	Filwood	8299	Crispin Allard	Liberal Democrat	102	4.54
2014-05-22	Filwood	8299	Marion Jackson	Trade Unionists and Socialists Against Cuts	174	7.75
2014-05-22	Filwood	8299	Sylvia Christine Windows	Conservative	301	13.4
2014-05-22	Filwood	8299	Ryan Brinkley	Green	335	14.92
2014-05-22	Filwood	8299	Christopher David Jackson	Labour	1334	59.39
2014-05-22	Hartcliffe	8760	Paul Elvin	Liberal Democrat	122	5.29
2014-05-22	Hartcliffe	8760	Robert Nash	Trade Unionists and Socialists Against Cuts	223	9.67
2014-05-22	Hartcliffe	8760	Patrick Charles Gordon Slade	Green	283	12.27
2014-05-22	Hartcliffe	8760	Jonathan Robert Hucker	Conservative	648	28.09
2014-05-22	Hartcliffe	8760	Naomi Grace Rylatt	Labour	1031	44.69
2014-05-22	Henbury	7781	David Rawlings	Trade Unionists and Socialists Against Cuts	93	3.24
2014-05-22	Henbury	7781	Thomas Stephens	Liberal Democrat	139	4.85
2014-05-22	Henbury	7781	Ruby Tucker	Green	236	8.23
2014-05-22	Henbury	7781	Rosalie Walker	Labour	907	31.65
2014-05-22	Henbury	7781	Mark Weston	Conservative	1491	52.02
2014-05-22	Hengrove	9125	Mark Baker	Trade Unionists and Socialists Against Cuts	28	0.95
2014-05-22	Hengrove	9125	Neil Oliver Maggs	Respect	114	3.88
2014-05-22	Hengrove	9125	Graham Hugh Davey	Green	123	4.18

4 KEYS

4.1 DESIGN DECISIONS

- For the implementation of keys, I considered a few options:
 - MySQL style – user defines which column (of any datatype) in the table is the ‘primary key’. Each entry in this column must be unique.
 - Simple style - Every record has an integer ID, which autoincrements when records are added.

- I chose to go with the simpler style for my database. From my experience with database software, there are very few situations where you wouldn't want to add an ID column to every table, and essentially no situations where it would be detrimental to have one.
- Certainly, whenever foreign keys are involved it is much more sensible to use an integer ID which is stored across the two tables:
 - Uses less space than a string or other data type.
 - Since ID cannot be modified by the user, it is much simpler to keep the foreign key matched between the two tables.
- The only major downside is storage space, but one additional attribute per record is unlikely to make a significant difference.
- Primary key implementation decisions:
 - The ID value will auto-increment as records are added.
 - It will not reassign IDs of deleted rows (similar to MySQL).
 - The ID value cannot be user modified, but is visible when printing.
 - ID will be of type integer rather than type long. 2 billion+ IDs should be more than enough, unless a table has records deleted and added a huge number of times over a long period of time. In the future changes could be made so the user can choose to have int or long IDs, depending on the application.
 - Each record will store the ID as a separate int, rather than as an additional attribute in the elements ArrayList. This will be beneficial for two reasons:
 - Will require less refactoring of the table class.
 - The record elements are currently all stored as strings, so it would be inefficient to keep converting back and forth from int/string.
 - The attribute names record (index 0 of records) will have the ID 0. I considered making this -1 so that it would be differentiated from the actual data records, but this means that every record's index and ID is off by a factor of one (record ID = 0, index in records list = 1) which is likely to cause mistakes in the future.
- I have also decided at this point to use custom file extensions to distinguish between importable CSV files and program table files.
 - The program will still be able to import files in .CSV format
 - These files will need to be processed and have an ID column generated
 - The program will use files in .TAB format
 - These files will still essentially be in CSV format and will be readable in any text editor
 - However, the .TAB extension signifies there is an ID column already present in the file, so it is compatible with my database software.

4.2 NEW FILES AND RESTRUCTURING

- Added `private int ID` and a `getID()` function to Record.
- Added `nextID` to Table, stores the ID value that will be assigned to the next record inserted into the table
- Added a `getID()` method in table, which returns the next record ID and increments the `nextID` value
- Added `ID` comparison to the `Record.equals()` method and `nextID` comparison to `Table.equals()`

- Changed Table.insert() to take a string array as input rather than a Record, so the record is created only when it is being added to the table. This is to prevent the situation where a record is created with a new ID, so the nextID increments, but then the record fails to be added to the table (so an ID is wasted/lost).
 - The same changes were made to Table.update().
- The test records in the test function were changed to String arrays to work properly with insert() & update(), and the ID attribute was removed (since all records now have an auto ID by default).
- Load table will have 2 options, one for importing external data (no ID column in file) and one for loading (with ID column). Save table will always save the ID column to file (to maintain foreign keys, when they are added).
- Save table now saves with .TAB file extension.
- When loading a table which already has the ID column, the Table.nextID value needs to be set to one higher than the highest ID from the file.

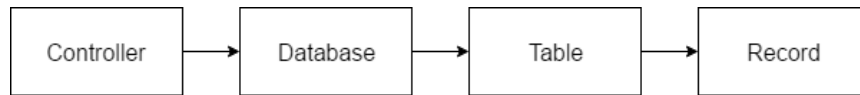
4.3 THIS SECTION IN HINDSIGHT

- I spent significantly longer on this section as I anticipated. After looking back, I decided that two decisions in particular cost me a lot of time:
 - The decision to allow importing files at this stage.
 - Although it is nice to be able to import any CSV file into the database, this massively overcomplicated the loading function
 - Two different cases were required for valid files with IDs present and invalid files where the ID must be generated for each record.
 - In hindsight it would make a lot more sense to have importing as a completely separate feature, added further down the line when the base software functionality has been established. Trying to add this early created a lot of code that needed refactoring multiple times, and was too interwoven with the Table class.
 - Including the record ID as a separate integer rather than stored as an additional attribute string.
 - This meant a lot of code had to be repeated for dealing with the ID separately
 - The printing functions also needed to be overhauled, whereas simply including another attribute would have required very little refactoring.
- Since the program is currently functioning as expected I will leave it in it's current state. At a later point I may return for a second pass to fix some of the issues, such as separating importing out into another class and making ID an attribute rather than a separate variable.

5 DATABASES

5.1 DESIGN DECISIONS

- Now that the program is made up of several interconnected classes, it's worth deciding on a class dependency diagram. This should help prevent accidentally creating cyclic dependencies and other potential issues.



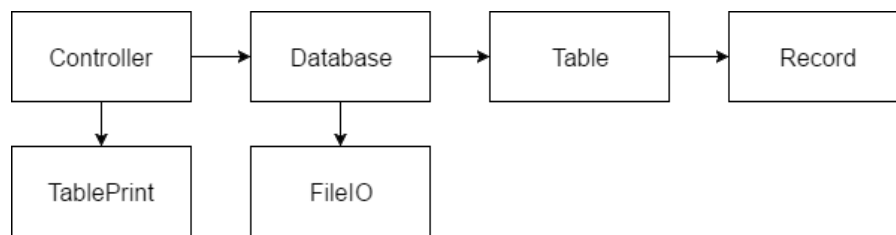
- Potential problems with the current classes:
 - Currently printing to console occurs from multiple classes. It will be worth bringing this all together into a separate class
 - File IO may fit better outside of Table.java
 - In general Table is much too large and performs too many tasks.
- Folder hierarchy for databases:
 - All database files are in the “databases” folder.
 - Each database has a folder with its name (i.e. elections database folder = “elections”).
 - All the table files for each database are stored within the relevant database folder
- The controller class will load the names of all the database folders on start-up.
- When a database is selected all the tables from that database are loaded.
- When exiting a database all the tables from that database are saved.

5.2 NEW FILES AND RESTRUCTURING

5.2.1 Major Refactoring

- For this section I decided to undergo some significant refactoring to my code, due to some of the problems mentioned above.
- This involved creating some new classes (FileIO & Display) and cutting down some old ones (Record, Table)
 - Record no longer deals with IDs separately, so the ID must be included as an input String.
 - Table is now significantly shorter (by almost 200 lines). All printing and file save/load functions have been moved to other classes.
 - Importing external CSV files is no longer possible. This can be re-added later in a more appropriate class. The program has returned to using standard CSV files for storing data.
 - Record & Table no longer prints any messages, this is left to Database & Display.

- The new class dependency diagram looks like so:



5.3 CURRENT FUNCTIONALITY OF THE SOFTWARE

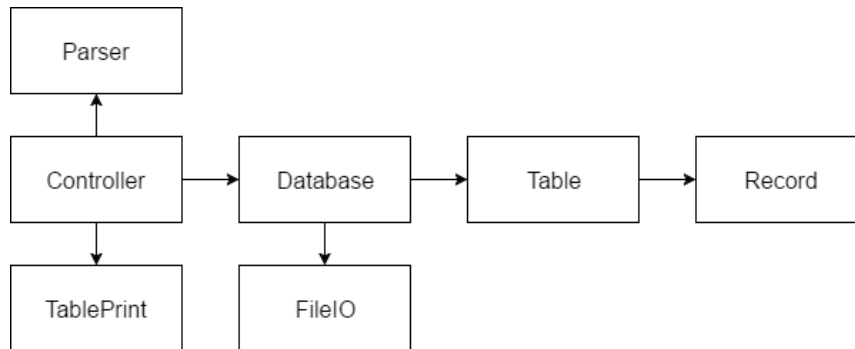
- On start-up, Controller prints a list of the databases (folders in the databases directory)
- A database can be loaded with `new Database("Example");`
 - Controller can only have one database active at one time.
 - Loading a database will load all of its tables into memory ready to be used.
- The `listTables()` function will print out a list of the tables in the active database.
- The `showTable(String tName)` function will print out the specified table name in a formatted table.
- More functionality exists below the Controller class, such as adding individual records to the table, deleting records, updating records etc.
- When a user interface is added next, all these features will be connected back up to user commands.

6 BASIC TEXTUAL INTERFACE

6.1 DESIGN DECISIONS

- The aim for this section was to create a simple textual interface for the functionality available in the previous section, to enable adding more complex table querying later on.
- A Parser class was written to read in a line from console, then check it against a list of pre-set valid keywords.
- I decided to use the alternative querying language suggested in the coursework. However for now only 4 queries have been added:
 - "list" – prints a list of the tables in the current database.
 - "load x" – prints the formatted table 'x'.
 - "quit" – saves the tables and returns to the database selection.
 - "basicprint (true/false)" – changes the table printing to use simple characters or box drawing characters.
- The Parser has a number of other query keywords already set up, but these have been commented out until the features are added.

- Class Dependency diagram for this stage:



6.2 CURRENT FUNCTIONALITY

- The Controller run function has two nested while loops:
 - The outer loop is for selecting a database.
 - This will print a list of the available databases, then request a choice from the user. If a database is selected then all its tables are loaded.
 - The program can be exited if the user enters “quit”.
 - The inner loop is for querying a database.
 - This will wait for a valid query from the user
 - If an invalid query is entered it will print an error message
 - If a valid query is entered, then it is performed (“list” and “load”)
 - If “quit” is entered, the program returns to the outer database selection loop

Example of current program output:

```

Welcome! These are the available databases:
[Empty, Pets]

Choose a Database to load:
Pets
Table loaded: animals (3 Entries)
Table loaded: owners (4 Entries)

list
[animals, owners]

load animals

```

ID	name	species	ownerId
0	Fido	dog	0
1	Wanda	fish	2
2	Garfield	cat	0

```

basicprint true

load owners
+-----+
| ID | name |
+-----+
| 0 | Jo |
| 1 | Sam |
| 2 | Amy |
| 3 | Pete |
+-----+

quit
Table saved: animals
Table saved: owners

Welcome! These are the available databases:
[Empty, Pets]

Choose a Database to load:
Empty
Database is empty, no tables to load!

list
[]

load table
Table "table" does not exist.

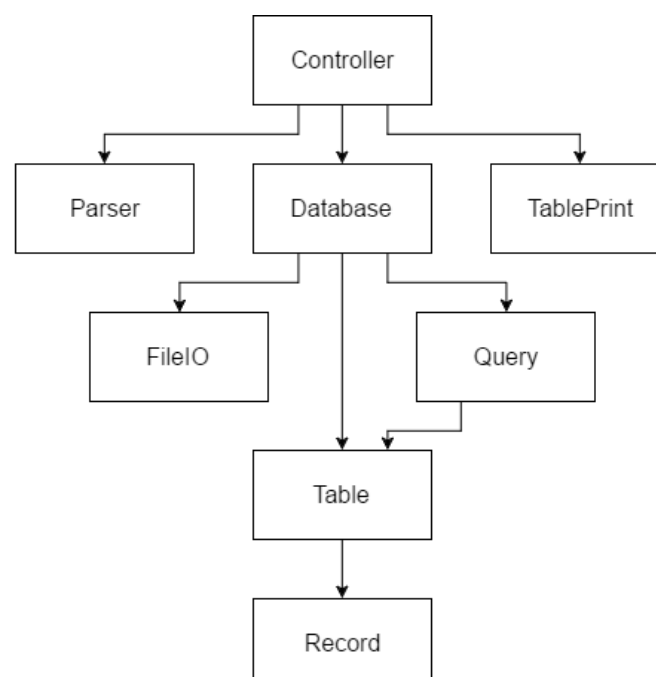
```

7 QUERY LANGUAGE

7.1 DESIGN DECISIONS

- The aim for this section is to add a functional query language, based upon the simple style suggested in the coursework description.
- This is an interactive querying style, so the current modified table is printed to console after each query.
 - Loading a table places it at the top of a stack
 - Tables on the stack can be queried and modified
 - To save changes or make new tables the save query must be used
- I decided to make a new class Query, which stores the stack of tables and performs the query on the stack.
- Current database queries:
 - “basicprint true/false” switches basic table character printing on/off.
 - “list” – list the names of all tables in the current database.
 - “quit” – return to the database selection.
 - “show x” – prints table x without adding it to the stack.
 - “clear” – clears the stack of tables.
 - “load x” – loads table x onto the stack.
 - “save name” – saves the current table to the file: “name.csv”.
 - “project x y z” – removes all columns that aren’t x/y/z from the current table.
 - “delete name” – deletes the file “name” from the database.
 - “add e1 e2 e3” – adds the entry to the current table. Must include attributes for all columns except ID.
 - “create c1 c2 c3” – creates a new empty table with the given attribute names.
 - “rename c1 c2” – renames the column c1 to c2.
 - “join x y” – join the current table1 to the table2 below where table1.x = table2.y

Final Class dependency diagram:



8 FINAL PROGRAM

8.1 POINTS OF DISCUSSION

- Unfortunately I ran out of time to complete the join query. It more or less works, but some combinations cause the program to crash. Also, not all required columns are copied over to the join table.
- The textual interface has not been fully bug tested for unusual inputs. Misspelling certain queries may cause the program to crash. However, with correct inputs all of the queries function correctly except for join.
- Testing has not been added for some of the higher classes, such as Database. It is not feasible to properly test this class since it would require keeping dummy test databases to load in, modify and complete numerous queries on.
- However, many different combinations of program functions have been tested to ensure everything works as expected.

8.2 EXAMPLE PROGRAM USAGE

- Loading the Pets database, loading animals & owners onto the stack, then joining them together.

```
Welcome! These are the available databases:
[Census, Empty, Pets]
```

```
Choose a Database to load:
```

```
Pets
```

```
Table loaded (4 Entries)      animals
```

```
Table loaded (4 Entries)      owners
```

```
load animals
```

ID	name	species	ownerId
0	Fido	dog	0
1	Wanda	fish	2
2	Garfield	cat	0
3	Jeffrey	hippopotamus	3

```
load owners
```

ID	name
0	Jo
1	Sam
2	Amy
3	Pete

```
join ID ownerId
```

```
Specified column names not found in tables.
```

```
join ID ownerId
```

```
New table has been created.
```

ID	name	animals.name	animals.species
0	Jo	Fido	dog
1	Jo	Garfield	cat
2	Amy	Wanda	fish
3	Pete	Jeffrey	hippopotamus

- Renaming two column names, saving the new table and listing the tables in the current database again.

```
rename animals.name pet
Column renamed successfully.
```

ID	name	pet	animals.species
0	Jo	Fido	dog
1	Jo	Garfield	cat
2	Amy	Wanda	fish
3	Pete	Jeffrey	hippopotamus

```
rename animals.species species
Column renamed successfully.
```

ID	name	pet	species
0	Jo	Fido	dog
1	Jo	Garfield	cat
2	Amy	Wanda	fish
3	Pete	Jeffrey	hippopotamus

```
save petOwners
Table saved: petOwners

list
[animals, owners, petOwners]
```

- Quitting out of the database, then quitting out of the program.

```
quit
Table saved: animals
Table saved: owners
Table saved: petOwners

Welcome! These are the available databases:
[Census, Empty, Pets]

Choose a Database to load:
quit
Program closing. Bye!
```