# Algorithms Report

Raziq Khan - A00990021
Pedram Nazari - A00931203



## Contributions:

Throughout the project, we worked on every file alongside each other and were in a VOIP call working on the same files and functions. Usually around 4-5 times a week, 4 hours each day.

The contributions would be 50/50.

We spent most of our time researching and implementing the QuadTree class. After we had a working version of the QuadTree we started to implement the debugQuad and debugWorld logging functions. Through doing this we found a bunch of errors in our QuadTree and started fixing those issues. Then we started to implement the generic features for the BufferPool. Then the HashTable and NameIndex. After implementing the HashTable/NameIndex. We realized we needed to write to the database file and started implementing that as well inside the Logger class. Finally after that we created the BufferPool - GISRecord GetRecord() function since we needed to retrieve from the buffer pool first. Finally we went through the logging functions one by one and fixed our implementation if it wasn't matching the example log file.

Raziq:
On Day 1 I created a base for the GISRecord so we could focus on the quadtree and hashtable. This is basically properly reading the lines for future usage. After day 1, we worked together on everything else in a voice call.

Pedram:
On Day 1, I started off creating the CommandProcessor. more specifically parsing all the arguments correctly to start working on the algorithmic modules. Then Raziq would “yo” me every morning until we got it done.

## How to Run:

**When reading text files, the application expects the text files to have LF line endings.**
When writing to files, it will write in LF line endings regardless of OS platform.
To run the project, run the Makefile given in the project folder using “make” in the command prompt. This project uses g++ and c++11 to compile. To run the exe, an example would be **./GIS.exe databaseFile scriptFile logFile.**

## Self Project Grading

| | |
|---|---|
| Report explains different parts of the source code | 1 |
| Report provides a UML diagram depicting the relation of different classes of the project | 1 |
| Report provides resulting log file from running Script01.txt | 1 |
| Main GIS.cpp file is provided | 1 |
| Project compiles using g++ with no compilation error | 1 |
| Project Invocation schema is as required | 1? UML diagram? |
| File names are not hard coded | 1 |
| The database file creation/recreation is as required | 1 |
| Program acts sainly when script file does not exist (shows an error and does not throw any error after getting closed) | 1 |
| Program creates/recreates the log file as required | 1 |
| Log file initial lines are filled as instructed | 1 |
| No library other than STL is used | 1 |
| NameIndex is implemented | 1 |
| A proper hash table is implemented for NameIndex (not | 1 |

| | |
|---|---|
| std::map) | |
| Hash table uses array structure | 1 |
| Hash table resizes itself automatically when 70% full | 1 |
| Quadratic probing function resolution implemented | 1 |
| Proper hash function implemented for the hash table | 1 |
| hash table entries store a feature name and state abbreviation and the file offset(s) of the matching record(s) | 1 |
| Debug command displays the content of nameIndex | 1 |
| CoordinateIndex is implemented | 1 |
| Implementation of bucket PRQuadtree | 1 |
| Insertion to already full leaf nodes results in node separation | 1 |
| Value of K is a parameter and can be properly modified | 1 |
| Quadtree entries store a geographic coordinate and a collection of the file offsets of the matching GIS records in the database file | 1 |
| Each quadtree node index can contain references to an unlimited number of different GIS records | 1 |
| Debug command displays the content of the coordinateIndex | 1 |
| BufferPool is implemented | 1 |
| During an import operation the buffer pool is bypassed | 1 |
| When performing searches, retrieving GIS records from the db file is managed through the buffer pool | 1 |
| Pool is capable of buffer up to 15 records | 1 |
| Pool implements LRU replacement | 1 |
| Debug command displays content of the buffer pool listed from MRU to LRU entry | 1 |
| Lines beginning with a semicolon are ignored but are correctly copied to log file | 1 |
| World command is parsed and run properly | 1 |

| | |
|---|---|
| After import command, right log values are reported | 1 |
| Debug world command displays world properly | 1 |
| What_is command runs properly | 1 |
| What_is_at command runs properly | 1 |
| What_is_in command without filter runs properly | 1 |
| What_is_in command with long tag runs properly | 1 |
| Quit command prevents execution of commands after its line | 1 |
| A Descriptive message is logged if a search yields no matching records | 1 |
| | |
| Script Testing | |
| (S01) Basic testing of simple search commands with a tiny DB | 3 |
| (S02) Basic testing of the region search commands with a tiny DB | |
| (S03) Test of multiple import commands with a couple of tiny DB files | |
| (S04) General test of searches with a larger database file | |
| (S05) Test of extreme region searches with a small database file | |
| (S06) Test of searches with a large database file | |
| (S07) Test of searches and the buffer pool with another large database file | |
| (S08) General test with a large dB and multiple imports | |
| (S09) Test of search failures with a large database file | |
| (S10) Test of searches and the buffer pool with another large database file | |
| | |

# Describe Solution:

We use a Makefile to build the project that runs under g++ c++11. This can be executed by running “make” in the command prompt where the Makefile is located.

To run the exe, you will need a database file, script file and log file. These can be found in the Data folder. If all 3 parameters are not given, the exe will not run. It will also check if the script file has proper arguments and will not run if it is incorrect.

The project will create a world boundary (Longitudes and Latitudes) given by the script file in DMS format. The valid records inside the world boundary from the import command will be added to a database file. In some areas, the DMS format is converted into decimal values for proper functionality.

Everything ends up traversing through the GIS record file as a middleman. The GIS record goes towards the coordinate index and name index based on the commands inputted from the command processor. The buffer, coordinate index, and name index are held in the GIS record. All offsets saved are in character offsets in not line offset.

The Coordinate Index uses a bucket PR Quadtree to hold its data. The data in the quad tree can be searched and inserted using the database file. The PR Quadtree holds the row and offset value to properly obtain the record from the database.

The Name Index uses a hash table as a way to hold values from the database. The hash table is using elfhash as its hash function and quadratic probing as its collision resolution. Keys are using the feature name and state abbreviation.

The Logger is a class by itself that can be called anywhere to perform a write to the log file. The Logger class uses System Manager to write and read files. The System Manager class can use the logger to write to the log file. The system manager copies comments from the script to the log file when read.

The System Manager is general system operations based on the commands given from the script file. Such as reading the database log, reading the script file, and also finding the offset and parsing the line individually from the database log.

The Command Processor is where the commands are actually put through to achieve its final output. It uses the command given from the script file to locate what requirements are necessary and sends it through the GIS Record for its proper output.

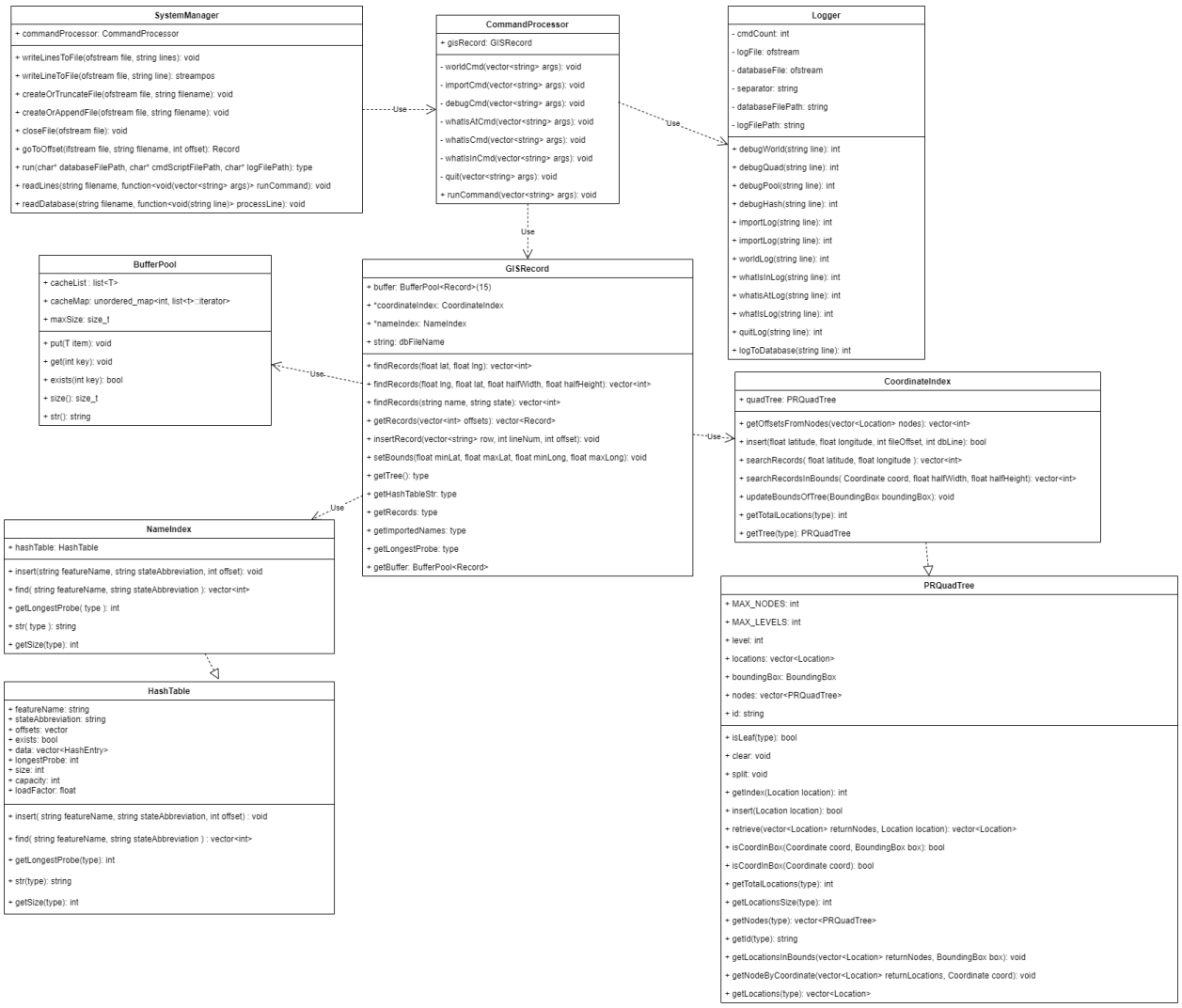
The Buffer Pool displays the most recently used record to the least recently used record when commands are giving an output. Whenever an output is made on commands such as what_is, what_is_at and what_is_in, it will add the record to the buffer pool. The buffer pool uses a

cachemap for efficient removal and insertions of items. We use the iterator to access or remove an item, in this case we are sorting from MRU to LRU.

The debug functions are fully functional and able to be viewed to properly see the hashtable, quad tree , and buffer pool.

# General Architecture:

### UML Diagram




First, files are checked for validity in the SystemManager. Once valid, world bounds are created in the CommandProcessor. After the world bounds are created, it will import a file which will read line by line into a record struct which will then insert a valid record that fits the world bounds into a database file. These valid records will then be passed into the NameIndex and CoordinateIndex to create a HashTable and PRQuadTree. Then based on the command file, the

GISRecord will either call the BufferPool, NameIndex, or CoordinateIndex. The GISRecord will check the BufferPool first for some commands. Once an output is created, it will go through the Logger for better formatting to be written to the Log file.

## Structs Created

**BoundingBox**: This creates the rectangle that can be used to search for location, defined by a center point, halfWidth, and halfHeight. The center point is in the middle of the rectangle, and halfWidth/halfHeight are used to create the width and height of the rectangle.

**Coordinate**: This is defined / used to store longitude and latitude.

**DMS**: This is used to convert DMS format strings to decimal format, total seconds, and to log strings.

**HashEntry**: Hash Entry stores the feature name, state abbreviation and offsets. These will be passed into a hash function.

**Location**: This stores the point on a map using its coordinate and offset. It can refer to multiple different database lines if they are the same coordinate. It can also return the total seconds of either the latitude or the longitude.

**Record**: This represents a line in the database file. Contains the offset of the line in the database file. Also contains the row as a vector whereby each column of the line can be accessed when specified.

## STL Libraries Used:

- Fstream:
- Sstream
- Iostream
- Iterator
- Algorithm
- **Map**: We used map for the filtering of records. We map the feature class to a filter option so that when we are given a list of records, we only get records that are of the selected filter option.
- Stdexcept
- String
- **Vector**: We used vector in many places, including the QuadTree and HashTable. This is because we can dynamically resize it instead of deleting and creating a new array.
- **Unordered_map**: We implemented Buffer Pool using unordered map because an unordered map gives an efficient way for accessing or removing elements. We use the

iterator in the cacheMap for being able to search and remove directly in constant time O(1) instead of searching through the whole list O(n).

- **List**: We use list in the Buffer Pool. We store the index in the list into the unordered map so we can properly search and remove using the iterator.

## Screenshot of Log File using Script01:

```
Course Project for COMP 8042
Student Name: Raziq Khan, Student Id: A00990021
Student Name: Pedram Nazari, Student Id: A00931203
Begin of GIS Program log:
dbFile: ./db_test.txt
script: ./script01.txt
log:    ./log_test.txt
Start Time: Fri Jul 07 13:58:41 PDT 2023
; Script 1
;
; Testing using a small dataset
;
; Specify boundaries of coordinate space:
;
world   0794530W    0792630W    381000N 383000N

-----------------------------------------------------------------------------------------
Latitude/longitude values in index entries are shown as signed integers, in total seconds.
-----------------------------------------------------------------------------------------
                      World boundaries are set to:
                                138600
                      -287130             -285990
                                137400
-----------------------------------------------------------------------------------------
;
; Import the dataset [the address here is a relative address]
Command 1: import   ./VA_Monterey.txt
```

```
;
; Import the dataset [the address here is a relative address]
Command 1: import   ./VA_Monterey.txt

Imported Features by name: 62
Longest probe sequence:   5
Imported Locations:        62
Average name length:       14
-------------------------------------------------------------------------------
Command 2: what_is_at   382812N 0793156W

    The following feature(s) were found in (38d 28m 12s North, 79d 31m 56s West)
    6547:  "Possum Trot"  "Highland"  "VA"
-------------------------------------------------------------------------------
Command 3: what_is  Church  VA

    No records match "Church" and "VA"
-------------------------------------------------------------------------------
Command 4: what_is  Central Church  VA

  685: Highland (38d 29m 53s North, 79d 33m 23s West)
-------------------------------------------------------------------------------
Command 5: what_is  Town of Monterey    VA

  8399: Highland (38d 24m 42s North, 79d 34m 51s West)
-------------------------------------------------------------------------------
```

```
Command 6: what_is  Smith Field VA

  7425: Highland (38d 18m 9s North, 79d 30m 29s West)
-------------------------------------------------------------------------------
Command 7: what_is  Smith Field CO

    No records match "Smith Field" and "CO"
-------------------------------------------------------------------------------
Command 8: what_is_at   382812N 0793156W

    The following feature(s) were found in (38d 28m 12s North, 79d 31m 56s West)
    6547:  "Possum Trot"  "Highland"  "VA"
-------------------------------------------------------------------------------
Command 9: what_is_at   381816N 0793700W

    Nothing was found at (38d 18m 16s North, 79d 37m 0s West)
-------------------------------------------------------------------------------
Command 10: what_is_at  381816N 0793708W

    The following feature(s) were found in (38d 18m 16s North, 79d 37m 8s West)
    7042:  "Trimble"  "Highland"  "VA"
-------------------------------------------------------------------------------
Command 11: what_is_at  381612N 0793256W

    The following feature(s) were found in (38d 16m 12s North, 79d 32m 56s West)
    6167:  "Clover Creek"  "Highland"  "VA"
```

```
--------------------------------------------------------------------------------
Command 12: what_is_at  382951N 0793238W

    The following feature(s) were found in (38d 29m 51s North, 79d 32m 38s West)
    2769:  "Key Run"  "Highland"  "VA"
--------------------------------------------------------------------------------
Command 13: what_is_at  382856N 0793031W

    The following feature(s) were found in (38d 28m 56s North, 79d 30m 31s West)
    1429:  "Forks of Waters"  "Highland"  "VA"
    4709:  "Strait Creek"  "Highland"  "VA"
--------------------------------------------------------------------------------
Command 14: what_is_in  382812N 0793156W    60  90

The following 7 feature(s) were found in (38d 28m 12s North +/- 60 , 79d 31m 56s West +/- 90)

    1429:  "Forks of Waters"  "VA" (38d 28m 56s North, 79d 30m 31s West)
    1707:  "Ginseng Mountain"  "VA" (38d 28m 50s North, 79d 31m 39s West)
    3045:  "Laurel Run"  "VA" (38d 27m 25s North, 79d 31m 59s West)
    3863:  "Peck Run"  "VA" (38d 28m 6s North, 79d 31m 9s West)
    4709:  "Strait Creek"  "VA" (38d 28m 56s North, 79d 30m 31s West)
    5749:  "Wooden Run"  "VA" (38d 27m 18s North, 79d 32m 1s West)
    6547:  "Possum Trot"  "VA" (38d 28m 12s North, 79d 31m 56s West)
--------------------------------------------------------------------------------
```

```
Command 15: what_is_in  382012N 0792330W    60  90

    Nothing was found in (38d 20m 12s North +/- 60 , 79d 23m 30s West +/- 90)

--------------------------------------------------------------------------------
Command 16: what_is_in  382148N 0793109W    15  15

The following 1 feature(s) were found in (38d 21m 48s North +/- 15 , 79d 31m 9s West +/- 15)

    3200:  "Little Doe Hill"  "VA" (38d 21m 48s North, 79d 31m 9s West)
--------------------------------------------------------------------------------
Command 17: what_is_in  -long   382148N 0793109W    15  15

The following 1 feature(s) were found in (79d 31m 9s West +/- 15, 38d 21m 48s North +/- 15)

     Feature ID   : 1484896
     Feature Name : Little Doe Hill
     Feature Cat  : Summit
     State        : VA
     County       : Highland
     Longitude    : 79d 31m 9s West
     Latitude     : 38d 21m 48s North
     Elev in ft   : 3241
     USGS Quad    : Monterey SE
     Date created : 09/28/1979

--------------------------------------------------------------------------------
```

```
Command 18: what_is_in  382148N 0793109W    60  60

The following 2 feature(s) were found in (38d 21m 48s North +/- 60 , 79d 31m 9s West +/- 60)

    3200:  "Little Doe Hill"  "VA" (38d 21m 48s North, 79d 31m 9s West)
    4137:  "Seldom Seen Hollow"  "VA" (38d 21m 45s North, 79d 30m 31s West)
--------------------------------------------------------------------------------
Command 19: what_is_in  382148N 0793109W    120 120

The following 4 feature(s) were found in (38d 21m 48s North +/- 120 , 79d 31m 9s West +/- 120)

    3200:  "Little Doe Hill"  "VA" (38d 21m 48s North, 79d 31m 9s West)
    4137:  "Seldom Seen Hollow"  "VA" (38d 21m 45s North, 79d 30m 31s West)
    6041:  "Bear Mountain"  "VA" (38d 20m 12s North, 79d 32m 54s West)
    6300:  "Doe Hill"  "VA" (38d 23m 13s North, 79d 31m 13s West)
--------------------------------------------------------------------------------
Command 20: what_is_in  382148N 0793109W    180 180

The following 6 feature(s) were found in (38d 21m 48s North +/- 180 , 79d 31m 9s West +/- 180)

    282:  "Buck Hill"  "VA" (38d 19m 2s North, 79d 33m 58s West)
    3200:  "Little Doe Hill"  "VA" (38d 21m 48s North, 79d 31m 9s West)
    4137:  "Seldom Seen Hollow"  "VA" (38d 21m 45s North, 79d 30m 31s West)
    6041:  "Bear Mountain"  "VA" (38d 20m 12s North, 79d 32m 54s West)
    6300:  "Doe Hill"  "VA" (38d 23m 13s North, 79d 31m 13s West)
    8120:  "Strait Creek School (historical)"  "VA" (38d 24m 47s North, 79d 32m 17s West)
--------------------------------------------------------------------------------
```

```
Command 21: what_is_in  -long   382148N 0793109W    180 180

The following 6 feature(s) were found in (79d 31m 9s West +/- 180, 38d 21m 48s North +/- 180)

     Feature ID   : 1482110
     Feature Name : Buck Hill
     Feature Cat  : Summit
     State        : VA
     County       : Highland
     Longitude    : 79d 33m 58s West
     Latitude     : 38d 19m 2s North
     Elev in ft   : 3291
     USGS Quad    : Monterey SE
     Date created : 09/28/1979

     Feature ID   : 1484896
     Feature Name : Little Doe Hill
     Feature Cat  : Summit
     State        : VA
     County       : Highland
     Longitude    : 79d 31m 9s West
     Latitude     : 38d 21m 48s North
     Elev in ft   : 3241
     USGS Quad    : Monterey SE
     Date created : 09/28/1979
```

```
Feature ID   : 1486995
Feature Name : Seldom Seen Hollow
Feature Cat  : Valley
State        : VA
County       : Highland
Longitude    : 79d 30m 31s West
Latitude     : 38d 21m 45s North
Elev in ft   : 2461
USGS Quad    : Monterey SE
Date created : 09/28/1979

Feature ID   : 1495244
Feature Name : Bear Mountain
Feature Cat  : Summit
State        : VA
County       : Highland
Longitude    : 79d 32m 54s West
Latitude     : 38d 20m 12s North
Elev in ft   : 3530
USGS Quad    : Monterey SE
Date created : 09/28/1979
```

```
Feature ID   : 1495470
Feature Name : Doe Hill
Feature Cat  : Summit
State        : VA
County       : Highland
Longitude    : 79d 31m 13s West
Latitude     : 38d 23m 13s North
Elev in ft   : 3970
USGS Quad    : Monterey
Date created : 09/28/1979

Feature ID   : 1673781
Feature Name : Strait Creek School (historical)
Feature Cat  : School
State        : VA
County       : Highland
Longitude    : 79d 32m 17s West
Latitude     : 38d 24m 47s North
Elev in ft   : 3068
USGS Quad    : Monterey
Date created : 11/13/1995
```

```
Command 22: what_is_in  -filter structure   382600N 0793310W    120 120

The following 7 feature(s) were found in (38d 26m 0s North +/- 120, 79d 33m 10s West +/- 120

    0:  "Asbury Church"  "VA" (38d 26m 7s North, 79d 33m 12s West)
    2080:  "Highland High School"  "VA" (38d 24m 26s North, 79d 34m 44s West)
    4303:  "Seybert Chapel"  "VA" (38d 25m 12s North, 79d 32m 25s West)
    5026:  "Thorny Bottom Church"  "VA" (38d 27m 4s North, 79d 31m 41s West)
    7832:  "Highland Elementary School"  "VA" (38d 24m 27s North, 79d 34m 46s West)
    7976:  "Monterey Methodist Episcopal Church"  "VA" (38d 24m 42s North, 79d 34m 46s West)
    8120:  "Strait Creek School (historical)"  "VA" (38d 24m 47s North, 79d 32m 17s West)
--------------------------------------------------------------------------------
Command 23: what_is_in  -filter water   382850N 0793030W    120 240

The following 9 feature(s) were found in (38d 28m 50s North +/- 120, 79d 30m 30s West +/- 240

    1277:  "Elk Run"  "VA" (38d 29m 36s North, 79d 31m 53s West)
    1553:  "Frank Run"  "VA" (38d 29m 53s North, 79d 33m 10s West)
    2769:  "Key Run"  "VA" (38d 29m 51s North, 79d 32m 38s West)
    3045:  "Laurel Run"  "VA" (38d 27m 25s North, 79d 31m 59s West)
    3863:  "Peck Run"  "VA" (38d 28m 6s North, 79d 31m 9s West)
    4426:  "Simmons Run"  "VA" (38d 26m 54s North, 79d 32m 9s West)
    4709:  "Strait Creek"  "VA" (38d 28m 56s North, 79d 30m 31s West)
    5436:  "West Strait Creek"  "VA" (38d 26m 53s North, 79d 32m 4s West)
    5749:  "Wooden Run"  "VA" (38d 27m 18s North, 79d 32m 1s West)
--------------------------------------------------------------------------------
```

```
Command 24: what_is_in  -filter pop 382000N 0793530W    3600    3600

The following 5 feature(s) were found in (38d 20m 0s North +/- 3600, 79d 35m 30s West +/- 3600

    6167:  "Clover Creek"  "VA" (38d 16m 12s North, 79d 32m 56s West)
    6418:  "New Hampden"  "VA" (38d 29m 34s North, 79d 33m 48s West)
    6547:  "Possum Trot"  "VA" (38d 28m 12s North, 79d 31m 56s West)
    7042:  "Trimble"  "VA" (38d 18m 16s North, 79d 37m 8s West)
    7289:  "Monterey"  "VA" (38d 24m 44s North, 79d 34m 50s West)
--------------------------------------------------------------------------------
Command 25: import  ./VA_Bath.txt

Imported Features by name: 41
Longest probe sequence:   5
Imported Locations:        41
Average name length:       16
--------------------------------------------------------------------------------
Command 26: quit

Terminating execution of commands.

--------------------------------------------------------------------------------
End time: Fri Jul 07 13:58:41 PDT 2023
```

# Commands Implemented:

- **Everything that was asked so you don't have to read the rest :)**
- NameIndex Module and its Required HashTable
- CoordinateIndex Module and Bucket PR Quad Tree
- BufferPool Module
- World command
- Import Command
- Debug world
- Debug hash
- Debug pool
- Debug quad
- What_is
- What_is_at
- What_is_in, both with -long and -filter, and without
- Quit