Software Design Document for ESCORT: Visual Analytics of Trajectories for Selecting Billboard Locations

Loai Hamdy, Mina Kamal, Ahmed Wael, Amr Ali Farrag

February 14, 2018

1 Introduction

1.1 Purpose

The document is MIU university, business owners (marketing companies), data analysts, software developers that might upgrade the system in the future, This document is to present a detailed description of the ESCORT system: a web-enabled system for the analysis of traffic trajectories and Escort will be built on HTML, PHP and Python. This document will explain the purpose and features of the system, and describe its interfaces.

1.2 Scope

Understanding traffic patterns and crowd motion is a challenging task for urban developers and law enforcement agencies. Spatial-temporal data collected through positioning systems and smart phones offer great opportunities for analyzing motion trajectories and identifying patterns. However, a challenge remains due to the large size and dimensionality of the collected data. This project aims to develop a web-based tool for the visualization and analysis of motion trajectories and the speed of the car in a specific time. Specifically, the idea of the project is to study the traffic patterns in areas,Detected patterns along with raw data will be fed to the visualization system to support high level inference.

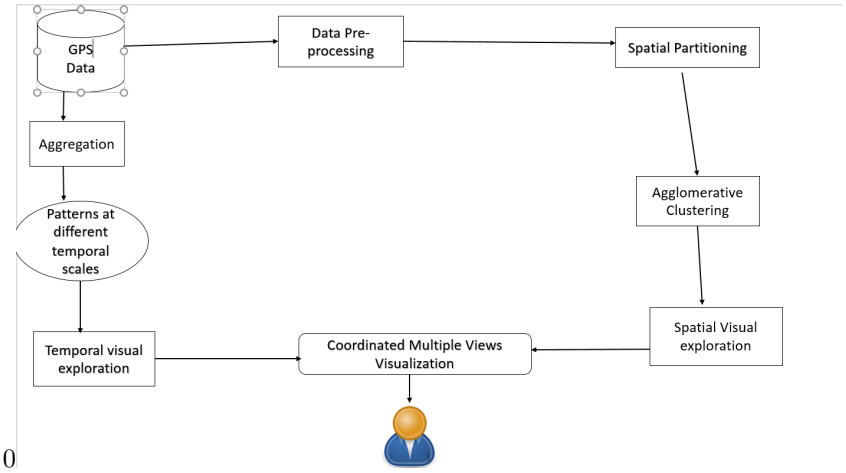
1.3 Overview

This document targets the front end users like marketing companies that will use Escort system to satisfy their needs. It will also be beneficial and helpful for data analysts and developers that may work on the Escort system in the future.Also provides an overview of the system that the system some functions is working to show the clustering and visualizing of trajectory.

1.4 Definitions and Acronyms

Term	Definition
1. Google Firebase	A mobile and web application platform.
2. Google Maps API	and API that allows using the google maps functionalities elsewhere.
3. SVM	Support Vector Machine.
4. ANN	Artificial Neural Network.
5. CMV	Coordinated Multiple views.
6. RNN	Recurrent Neural Network.
7. KLQ	k- location Query.
8. KNN	k- Nearest neighbour.
9. MMA	Map Matching Algorithm.

2 System Overview



0

Figure 1: .

The system overview is divided into two ways first way is talking about the spatial view that take the GPS data and make data pre-processing to separate the data for making partition and clustering and classification to get the output on the CMV screen.The second way is temporal view that take the GPS data for making a report statistics for the data we have in different ways and to get the time on the CMV screen.

3 System Architecture

3.1 Architectural Design

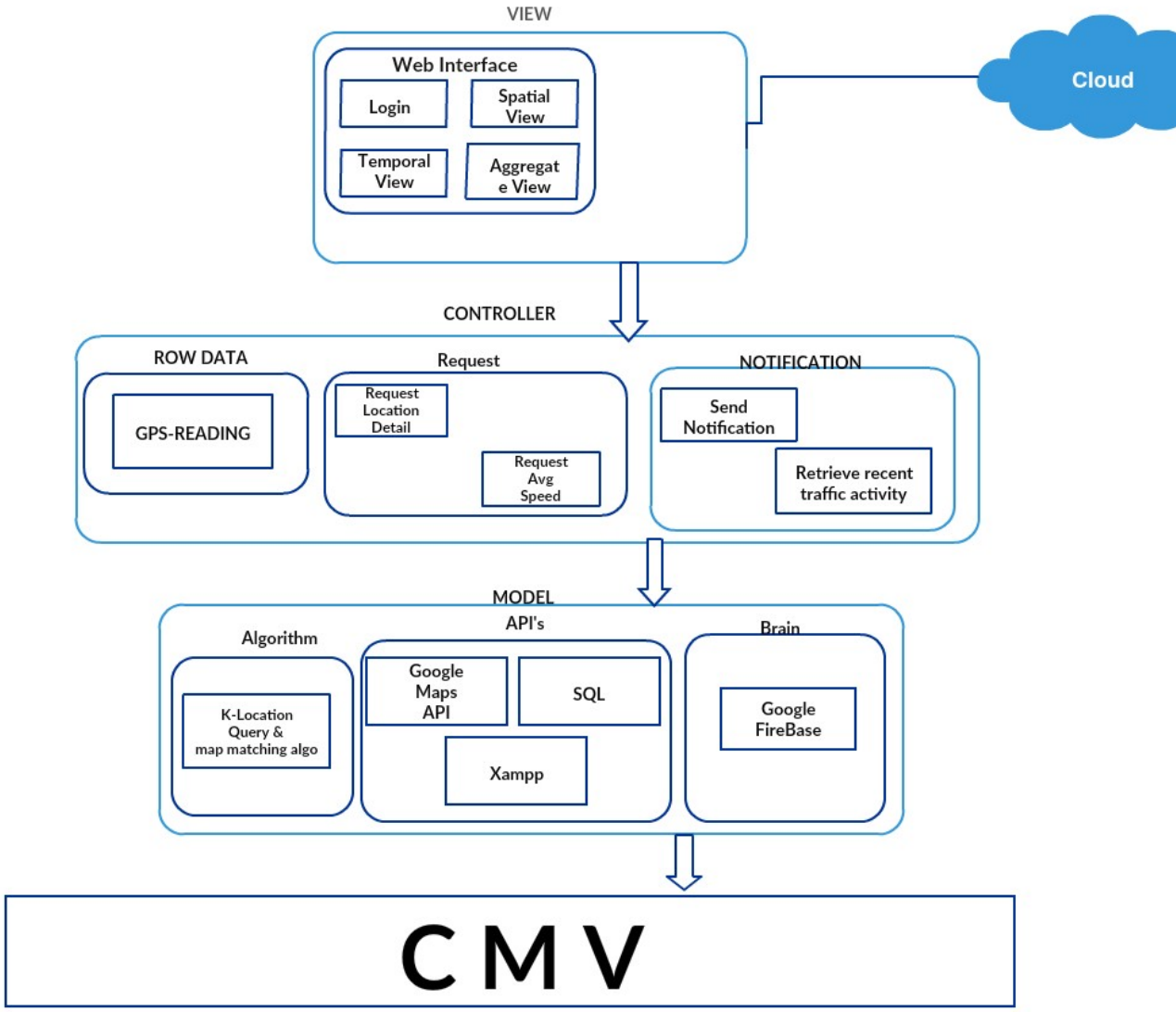


Figure 2: .

3.1.1 View

This represents the user interface which is divided into two interfaces.

1- BI Analyst interface: consists of a work to help the user through the functionality provided by the web interface. Each page represents a certain functionality, such as logging in, viewing the time and clustering on maps.

2- Customer interface:contains a lot of screens to see all statistics and view trajectories.

3.1.2 Controller

This is a class that match between view and models.Controller take the input from the view and send the data to database.The coming input of the data is getting from GPS trajectories and see the request of location details and speed limit, also the BI get notifications by sending and receiving data.r, it is responsible for distribution of the raw data between the View and the Model to be processed and analyzed.

3.1.3 Model

3.1.4 Algorithms

KLQ: short for K Location Query. It is a classification algorithm where a set of examples is collected and each one has a known class that Contains KNN (K-nearst neighbour. This set is used as a training dataset for the algorithm so when a new data is fed to it a comparison takes place between the training dataset and the new data. K-Location Query is a classification algorithm where a set of examples is collected and each one has a known class. KLQ(Location Query) is found based on the value of K. This is an effective and simple approach in the case of large training datasets.

Algorithm 1 Bloom filter k-mer counting algorithm

```
1: B ← empty Bloom filter of size m
2: T ← hash table
3: for all reads s do
4:    for all k-mers x in s do
5:       x_rep ← min(x, revcomp(x)) //x_rep is the canonical k-mer for x
6:       if x_rep ∈ B then
7:          if x_rep ∉ T then
8:             T[x_rep] ← 0
9:       else
10:         add x_rep to B
11: for all reads s do
12:    for all k-mers x in s do
13:       x_rep ← min(x, revcomp(x))
14:       if x_rep ∈ T then
15:          T[x_rep] ← T[x_rep] + 1
16: for all x ∈ T do
17:    if T[x] = 1 then
18:       remove x from T
```

Figure 3: .

MAP Matching Algorithm: GPS receivers, capable of measuring location instantaneously and accurately, are also ubiquitous in most mobile platforms, enabling applications relating and combining GPS-measured locations and routes with digital maps. The Map Matching Algorithm was observed that shortest path codes are much more compact than greedy path codes, justifying the larger time complexity.

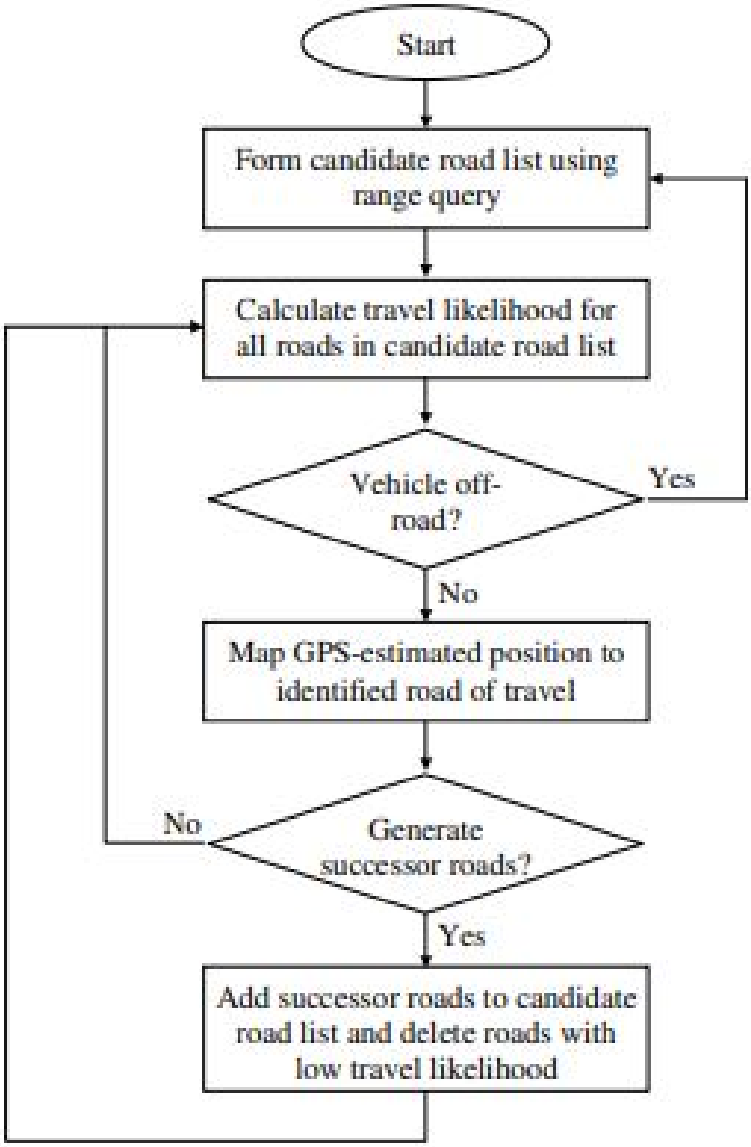

Start
Form candidate road list using range query
Calculate travel likelihood for all roads in candidate road list
Vehicle off-road?
Yes
No
Map GPS-estimated position to identified road of travel
Generate successor roads?
No
Yes
Add successor roads to candidate road list and delete roads with low travel likelihood

Figure 4: .

3.1.5 API's

** Google Maps API: this is the API that is being used to provide the BI Analyst monitor with a map display on the web interface and use its functionalities.

** XAMPP: free and open source cross-platform web server solution stack package developed by Apache Friends, consisting mainly of the Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages. XAMPP stands for Cross-Platform (X), Apache (A), MariaDB (M), PHP (P) and Perl (P)

3.1.6 Core

** Google FireBase DB maneger: is a mobile platform from Google offering a number of different features. Specifically, these features revolve around cloud services, allowing users to save and retrieve data to be accessed from any device or browser. This can be useful for such things as cloud messaging, hosting, crash reporting, notifications.

3.2 Accuracy

Term	Accuracy
1. KNN	92.40
2. MMA	95
3. SVM	97-100

3.3 Decomposition Description

3.3.1 Class Diagram

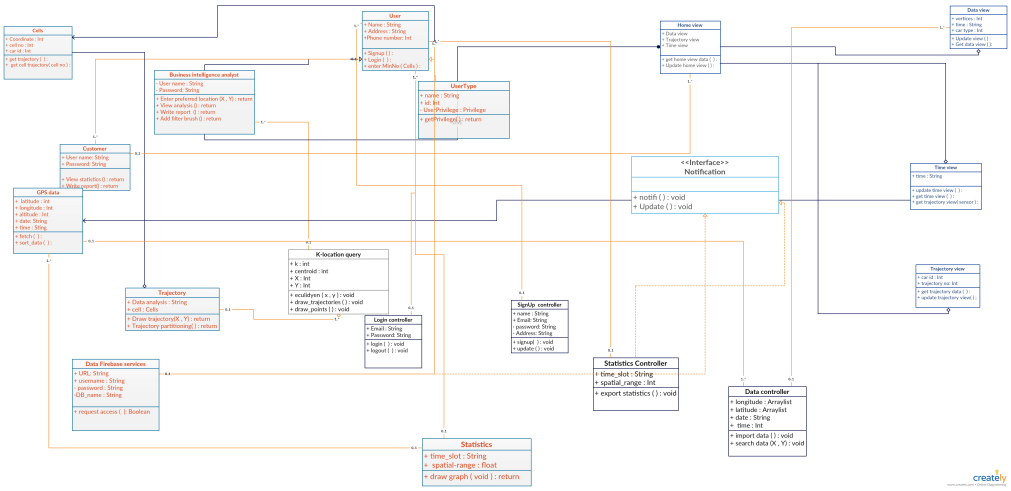


Figure 5: .

3.3.2 Sequence Diagram

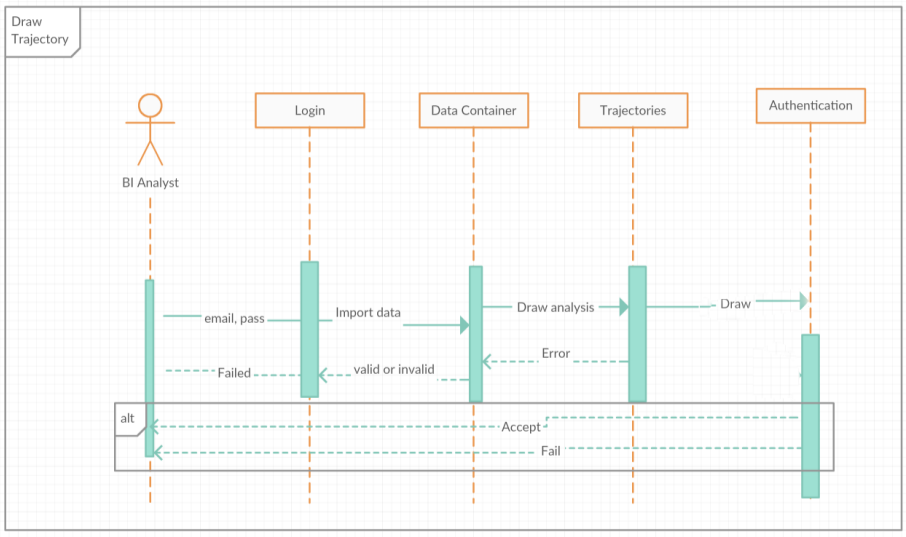


Figure 6: .

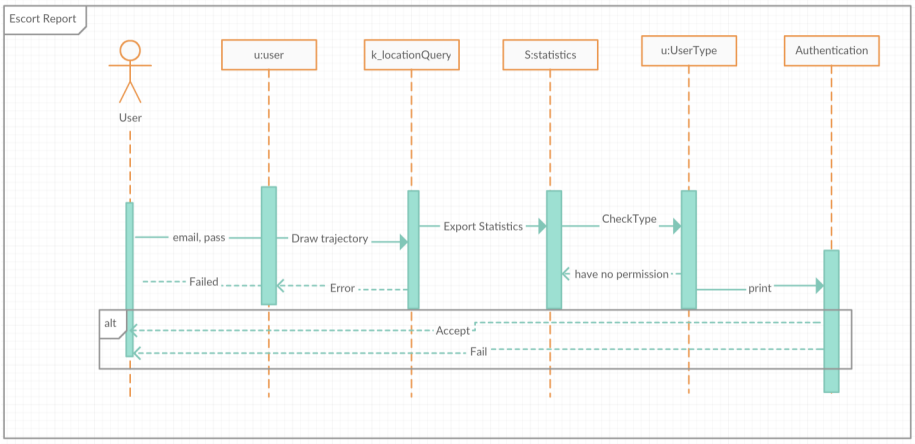


Figure 7: .

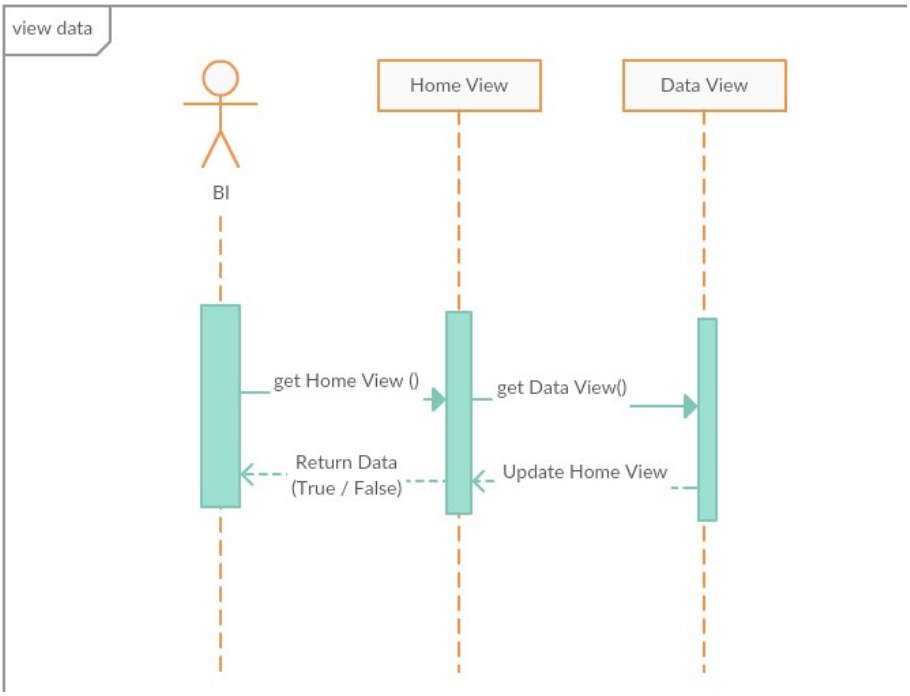


Figure 8: .

3.3.3 ER-Diagram

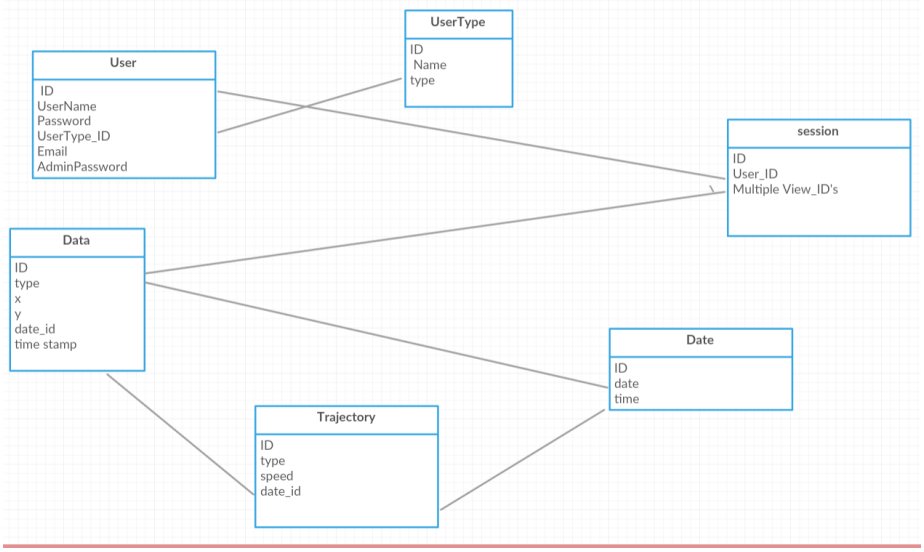


Figure 9: .

4 Data Design

4.1 Data Description

The data is GPS trajectory dataset was collected in (Microsoft Research Asia) Geolife project by 182 users in a period of over five years (from April 2007 to August 2012). A GPS trajectory of this dataset is represented by a sequence of time-stamped points, each of which contains the information of latitude, longitude and altitude. This dataset contains 17,621 trajectories with a total distance of 1,292,951kilometers and a total duration of 50,176 hours. These trajectories were recorded by different GPS loggers and GPS-phones, and have a variety of sampling rates. 91.5 percent of the trajectories are logged in a dense representation, e.g. every 1 5 seconds or every 5 10 meters per point.

4.2 Data Dictionary

The database recoded a broad range of users outdoor movements, including not only life routines like go home and go to work but also some entertainments and sports activities, such as shopping, sightseeing, dining, hiking, and cycling. The trajectory dataset can be used in many research fields, such as mobility pattern mining, user activity recognition, location-based social networks, location privacy, and location recommendation.

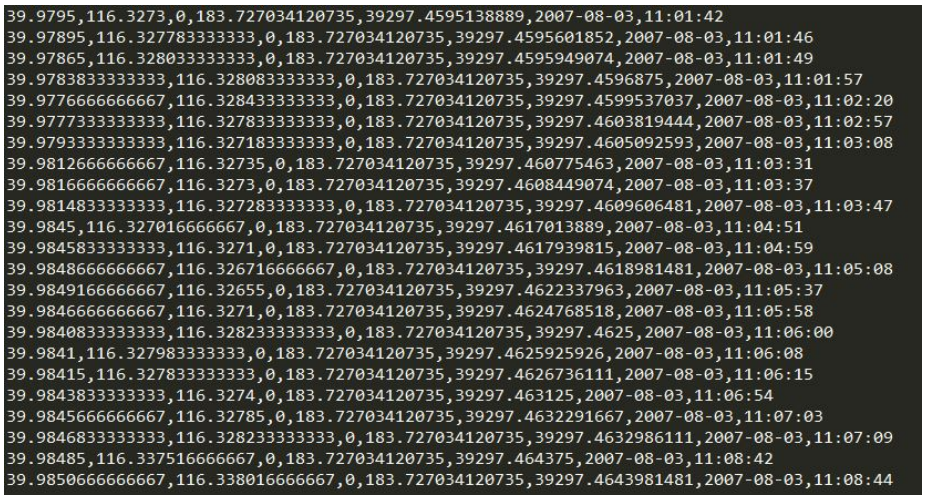


Figure 10: .

5 Component Design

Every single folder of Trajectory files stores a users GPS log files, which were converted to PLT format. Each PLT file contains a single trajectory and is named by its starting time. To avoid potential confusion of time zone, we use GMT in the date/time property of each point, which is different from the dataset previous release. PLT format:
Line 16 are useless in this dataset, and can be ignored. Points are described in following lines, one for each line.
Field 1: Latitude in decimal degrees.
Field 2: Longitude in decimal degrees.
Field 3: All set to 0 for this dataset.
Field 4: Altitude in feet (-777 if not valid).
Field 5: Date - number of days (with fractional part) that have passed since 12/30/1899.
Field 6: Date as a string. Field 7: Time as a string.

6 Human Interface Design

6.1 Overview of User Interface

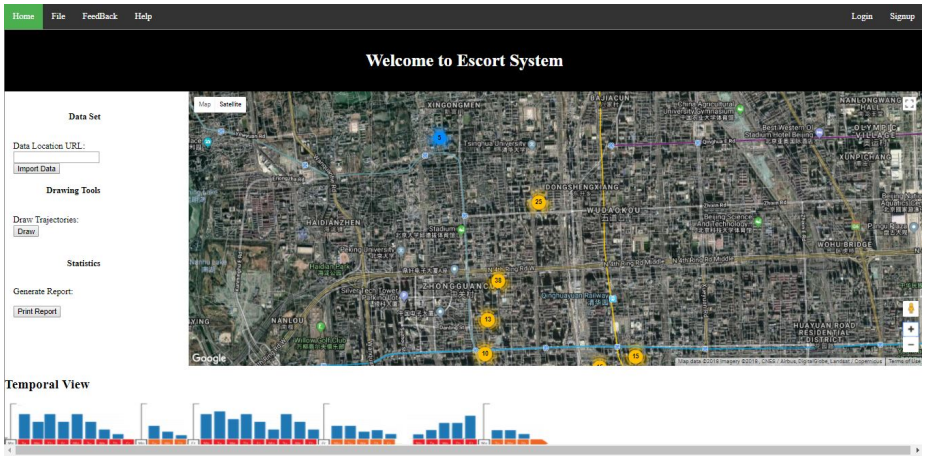


Figure 11: .

6.2 Screen Images

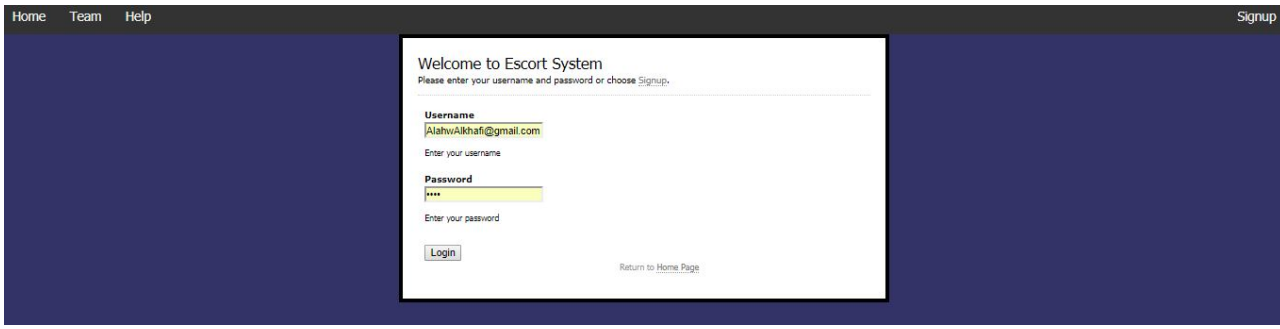


Figure 12: .

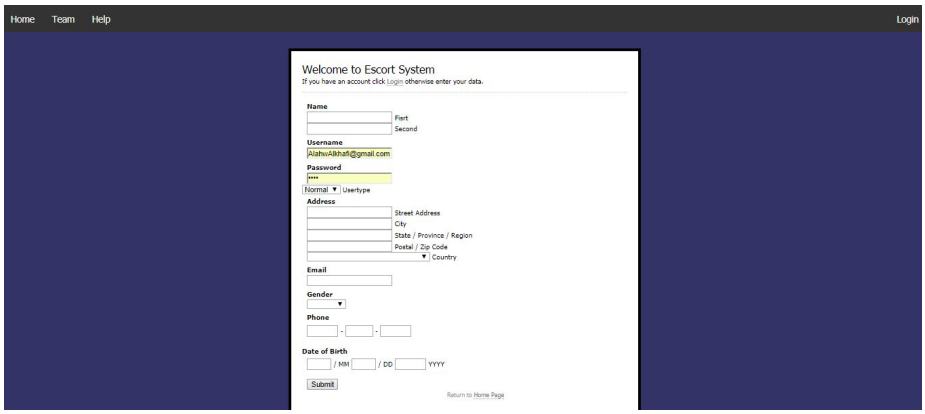


Figure 13: .

6.3 Screen Objects and Actions

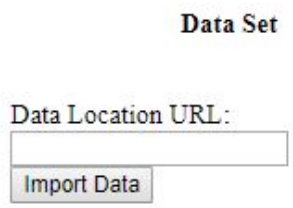


Figure 14: .

This fields is to Upload data to Fire-base to start using it from the Fire-base directly

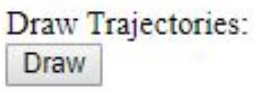


Figure 15: .

Figure 16: .

Figure 17: .

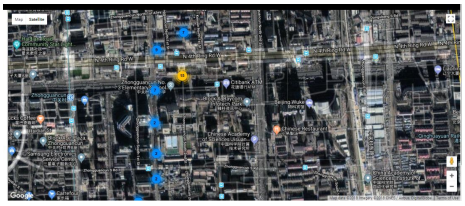


Figure 18: .

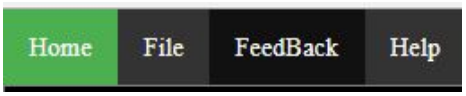


Figure 19: .

7 APPENDICES

Escort system is using Google Maps API, Escort added maps based on Google Maps data to Escort Project. The API automatically handles access to Google Maps servers, data downloading, map display, and response to map gestures.

8 References

1- **http://vacommunity.org/VAST+Challenge+2017+MC1**

2-**Ma J, Wang C, Shene CK. FlowGraph: A compound hierarchical graph for flow field exploration. InVisualization Symposium (PacificVis), 2013 IEEE Pacific 2013 Feb 27 (pp. 233-240). IEEE.**

3- **Ma, Jun, Chaoli Wang, and Ching-Kuang Shene. "FlowGraph: A compound hierarchical graph for flow field exploration." Visualization Symposium (PacificVis), 2013 IEEE Pacific. IEEE, 2013.**

4- **Laxhammar, R. (2011). Anomaly detection in trajectory data for surveillance applications (Doctoral dissertation, rebro universitet).**

5- **Schauer, Lorenz, and Martin Werner. "Clustering of Inertial Indoor Positioning Data." 1st GI Expert Talk on Localiza on (2015): 21.**

6- **Pahins CA, Stephens SA, Scheidegger C, Comba JL. Hashedcubes: Simple, low memory, real-time visual exploration of big data. IEEE transactions on visualization and computer graphics. 2017 Jan;23(1):671-80.**

7- **Liao, Zicheng, Yizhou Yu, and Baoquan Chen. "Anomaly detection in GPS data based on visual analytics." In Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on, pp. 51-58. IEEE, 2010.**

8- **Lee, Jae-Gil, Jiawei Han, and Kyu-Young Whang. "Trajectory clustering: a partition-and-group framework." In Proceedings of the 2007 ACM SIGMOD international conference on Management of data, pp. 593-604. ACM, 2007.**

9- **Liu Z, Wang Y, Dontcheva M, Hoffman M, Walker S, Wilson A. Patterns and sequences: Interactive exploration of clickstreams to understand common visitor paths. IEEE Transactions on Visualization and Computer Graphics. 2017 Jan;23(1):321-30.**

10- **Chen, C., Zhang, D., Castro, P.S., Li, N., Sun, L. and Li, S., 2011, December. Real-time detection of anomalous taxi trajectories from GPS traces. In International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (pp. 63-74). Springer, Berlin, Heidelberg.**

11- **Ferreira N, Poco J, Vo HT, Freire J, Silva CT. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. IEEE Transactions on Visualization and Computer Graphics. 2013 Dec;19(12):2149-58.**

12- **Monteiro, Gonalo, et al. "A framework for wrong way driver detection using optical flow." International Conference Image Analysis and Recognition. Springer, Berlin, Heidelberg, 2007.**

13- **Ferreira, Nivan, Jorge Poco, Huy T. Vo, Juliana Freire, and Cludio T. Silva. "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips." IEEE Transactions on Visualization and Computer Graphics 19, no. 12 (2013): 2149-2158.**

14- **Lee JG, Han J, Whang KY. Trajectory clustering: a partition-and-group framework. InProceedings of the 2007 ACM SIGMOD international conference on Management of data 2007 Jun 11 (pp. 593-604). ACM.**

15- **Correll M, Heer J. Surprise! Bayesian Weighting for De-Biasing Thematic Maps. IEEE transactions on visualization and computer graphics. 2017 Jan;23(1):651-60.**

16- **Ferreira N, Poco J, Vo HT, Freire J, Silva CT. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. IEEE Transactions on Visualization and Computer Graphics. 2013 Dec;19(12):2149-58.**

17- **Jagadeesh GR, Srikanthan T, Zhang XD. A map matching method for GPS based real-time vehicle location. The Journal of Navigation. 2004 Sep;57(3):429-40.**

18- **Gotsman R, . Generating Map-based Routes from GPS Trajectories and their Compact Representation. Technion-Israel Institute of Technology, Faculty of Computer Science; 2013 Jun.**

19- **UM, Jung-Ho, et al. k-Nearest neighbor query processing algorithm for cloaking regions towards user privacy protection in location-based services. Journal of Systems Architecture, 2012, 58.9: 354-371.**