

# Onto-Traffic: A Semantic Traffic Analysis Tool based on GPS Data for Smart-Cities

Juan J. Sánchez-Escobar<sup>‡</sup>, Liliana I. Barbosa-Santillán\*,  
Luis F. Barbosa-Santillán\*, Fatemeh G. Nasri-Shandiz\* and Gabriel A. León-Paredes\* <sup>†</sup>

\*University of Guadalajara

<sup>†</sup>Politécnica Salesiana University

<sup>‡</sup>Technical and Industrial Teaching Center

email: jjsanchez@ceti.mx, ibarbosa@cucea.udg.mx,

francisco.barbosa@alumno.udg.mx, ghazalnasri@hotmail.com and gleon@ups.edu.ec

**Abstract**—Nowadays by growing usage of vehicles, control and monitor the traffic is a promise for smart cities. Recommendation Systems are one of the solutions provided by smart cities for such issues. These kind of systems are trend in Information Technologies and they are acquiring acceptance among scientific communities and enterprises as well. Such systems feed users with valuable information about their environment and help them to take better decisions related to traffic jam. This paper presents OntoTraffic, a recommendation system based on ontologies, which supports the user to obtain information related to urban traffic in the city of Beijing. The information is provided according to the user's requests. This recommendation system works with Global Positioning System (GPS) traces taken from the Beijing data set, which has twenty one millions traces. In addition, 5888 traces were gathered through smartphones of Guadalajara citizens. In OntoTraffic system, data are collected by certain algorithms that extract the data from a reliable databases. Produced information from such system, which depends on the other recommendation, allows users being more productive and efficient in urban issues. This information also improves user's life quality by spending less time and money on their daily activities. OntoTraffic also provides recommendations to obtain: (1) average traffic on a specific street, (2) find rush hours for each street, (3) discover the most crowded streets in a certain time, (4) demonstrate the distance between two points on the map, and (5) show the bustle streets between 7:00 pm to 3:00 am. Finally this paper suggests some activities for future work such as: (a) build new models to support smart cities, (b) supporting recommended system developments for decision makers, (c) valuing data analysis and ontologies for future High Performance Computing (HPC) center.

**Keywords**—Recommendation systems; urban traffic; smart cities; OntoTraffic system; smart applications; GPS data analysis.

## I. INTRODUCTION

Over the last five years, urban density in big cities has been increased constantly and in some cases, such as less-developed regions it has expanded dramatically. As a result, the increase of urban density generated many issues like healthcare, public security, pollution, etc. Moreover, urban mobility has been affected drastically by the increase of population in major cities. This situation has a tendency to get worse due to increase of migration from rural areas to urban areas in the coming years [1]. These issues can be more remarkable on travel seasons, because there are more people and more vehicles traveling at the same time to the same place. In addition, road infrastructures cannot expand at the same pace to solve the problem. As a consequence, the inhabitants of big cities are the ones who suffer from daily troubles such as traffic jam, delays and high fuel consumption [2]; in addition, there are some other intangible problems such

as fatigue, stress, air pollution, higher rate of accidents and diseases, that are caused by urban growth and traffic. The aforementioned issues are very complex and involve multiple factors. Hence, these significant problems have been analysed and treated from different perspectives and via different entities. Finding a solution for cited problems is not easy because there are also some governmental issues involved, such as politics, guidelines, infrastructure investments and lack of using intelligent applications. However, systems like mobile sensors that employ Global Positioning System (GPS) are growing so fast and almost all of the smartphones are using GPS benefits, but still there are issues to provide the accurate users location. Calabrese et al., [3] mention in their work that use of smartphones for geo-location is only useful when traffic routes have been defined and identified. Herring et al., [4] have utilized gadgets with GPS to improve the accuracy of user's location. Disadvantage of this phenomena is that data acquisition depends on people participation by downloading and executing the application on their smartphones. Incurion of social networks and massive spread of smartphones have contribute to the develop of various projects where GPS are utilized as data collectors systems, like tracking GPS trajectories projects. For example, the GeoLife project [5] identifies the relationship between people and places, enabling people to share life experiences and build connections among each other using location history. Users share travel experience using GPS trajectories. Also, they obtain information from the GPS traces with the objective to recommend sites where other users and travellers visited. In other project presented by Microsoft Research Asia [5], they have focused on characterizing and comprehending people behavior, based on supervised learning through GPS traces to infer and predict their mobility behavior. Another ambitious project called T-Drive [6] Yuan et al., have utilized taxis GPS as mobile sensor to register the trajectory. Hence, they use the experience and ability of taxi drivers to locate the best route for the user's destiny. In this project, a database has been created with most used routes trajectories of Beijing city using GPS traces of 33,000 taxis during 3 months. The authors could identify each segment (every block in a trajectory) and time estimation between each node (an intersection between streets). Thus, the fastest route from one point to another could be calculate and demonstrate. Subsequently, users can choose the best route to their selected destination [7]. Many developing countries have witnessed an explosive vehicular growth. Our motivation is to improve people's lives through the knowledge of a recommender, by using an open source standard-based learning management system, that allows them to choose the better available option.

It is essential to have a recommendation system that could predict the behaviour of certain routes based on user’s previous queries. Therefore, the use of recommendation systems could reduce the travel in time and costs, and by this way users can reach their destination with less stress.

One of the challenges is the architecture proposal, which transforms the input using a data structure (ontologies) with a huge datasets.

A. Contributions

Our contributions are as follows: 1) automatically identifying traffic interests from a user’s history, 2) alert techniques that can be used to better determine whether a new recommendation is interesting, and 3) present a sample of specific SPARQL queries in order to evaluate system’s answers.

B. Structure of the paper

This paper is organized as following: in Section 2, methodology is introduced. In Section 3, experiment results are presented. Finally, Section 4, presents conclusions.

II. METHODOLOGY

A description of our recommender system based on Ontologies is shown in Figure 1. In the first level, dataset are grouped into two sets (a) 21 millions of Beijing GPS records and (b) 5888 Guadalajara GPS records. On the second level, the GPS traces are transformed into Ontology Web Language (OWL) Lite. Then, OntoTraffic is populated and its consistency is verified with Protege. In the fourth level, valid questions for OntoTraffic were identified and agreed. Next, competence questions are translated to SPARQL language. We use ARQ and JENA engines in order to run the SPARQL expressions into OntoTraffic. Results are validated by a focus group, and translated to XML format to integrate with the project TUI-TRAFFIC. Finally, all files transformed to CAP-XML file format.

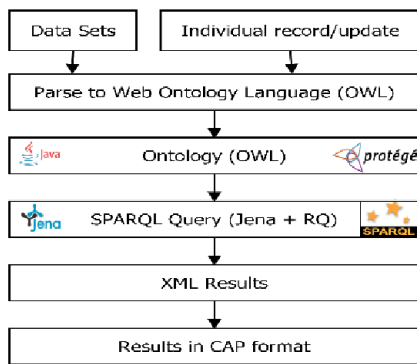


Figure 1. Recommender system architecture based on OntoTraffic

Some of the objectives that OntoTraffic covers are:

- Average traffic flow on a specific street.
- Obtain street rush hours.
- Identify the most crowded streets in a certain time.

- Measure distance between two particular points on the map.
- Recognize the most crowded streets between 7:00 pm to 3:00 A.M.

On the other hand, GPS Taxi Cabs files contain data collected by monitoring local taxicabs GPS in the city of Beijing, China. This dataset was obtained from [5] where 10383 files have been collected. Each file contains the GPS lecture of a single taxicab. This means that 10383 taxicabs were tracked. Figure 2 shows a fragment of a GPS trace file content. Each file has a ".txt" extension and the information stored in all of them is plain text.

```
1131,2008-02-02 13:30:59,116.45847,39.86964
1131,2008-02-02 13:31:04,116.45852,39.86954
1131,2008-02-02 13:31:09,116.45851,39.86955
1131,2008-02-02 13:31:14,116.45851,39.86956
1131,2008-02-02 13:31:24,116.45859,39.86954
1131,2008-02-02 13:31:29,116.45867,39.8695
1131,2008-02-02 13:31:34,116.45889,39.86946
1131,2008-02-02 13:31:39,116.45879,39.86946
1131,2008-02-02 13:31:44,116.4588,39.86943
1131,2008-02-02 13:31:49,116.45897,39.86941
1131,2008-02-02 13:31:54,116.45925,39.8694
1131,2008-02-02 13:31:59,116.45954,39.86939
1131,2008-02-02 13:32:04,116.46003,39.86935
1131,2008-02-02 13:32:09,116.46044,39.86934
1131,2008-02-02 13:32:14,116.46081,39.86941
1131,2008-02-02 13:32:19,116.46122,39.86951
1131,2008-02-02 13:32:24,116.46152,39.86951
1131,2008-02-02 13:32:29,116.46157,39.86959
1131,2008-02-02 13:32:34,116.46136,39.87011
1131,2008-02-02 13:32:39,116.46127,39.8703
1131,2008-02-02 13:32:44,116.46127,39.8703
1131,2008-02-02 13:32:49,116.46127,39.8703
1131,2008-02-02 13:32:54,116.46127,39.87029
1131,2008-02-02 13:32:59,116.4615,39.87025
```

Figure 2. GPS Traces located in the txt files.

It can be noticed that every taking reading is represented as a single line. The name of the fields is as follows: id\_taxi, date, hour, latitude and longitude.

Additionally, the first column of Table 1 shows the name of the fields from the collected data of GPS lectures, while on the second column we present an example of values. This was done for clarification purposes and all parameters are self-explanatory.

TABLE I. SCHEMA AND EXAMPLE OF A GPS TRACE FILE

avgflow	
Field	Example
id_taxi	9999
date	2008-02-02
hour	13:39:30
latitude	40.03238
longitude	116.28176

In order to process all the instances of OntoTraffic we convert the CSV files to turtle files as shown in Figure 3.

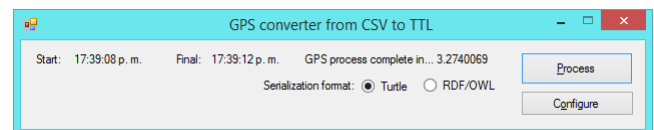


Figure 3. Convert the CSV files to turtle files.

### III. RESULTS

This system can also give recommendations on (1) average traffic on a specific street, (2) rush hours of streets, (3) most crowded streets in a certain time, (4) distance between two points on the map, and (5) most crowded streets between 7:00 pm to 3:00 am, such as explained from A to E.

#### A. Average traffic flow on a specific street

The first test aim was to find the average traffic flow on a specific street. Our dataset has been used for this test. Figure 4 shows results of the SPARQL query:

```
SELECT (AVG(?Flujo) AS ?Avg_flow)
WHERE {
  ?IDStreet s:flujo?Flujo;
  s:name "DKMLTADKHFYLRKHQRMPQ";
  s:col "SZMAXRLJMHPKURUJYYOG".
}
```

Avg_flow
98.190623122084040777777777

Figure 4. Result of average traffic flow on a specific street by using Beijing’s dataset

#### B. Streets rush hours

Second test purpose was to find the busiest points in a certain time (rush hours). To fulfill this test aim, we used Beijing’s dataset. Figure 5 shows results in a statistical graph and Figure 6 shows the result of the SPARQL query:

```
SELECT ?latitude ?longitude ?Hour
(COUNT(?gps) AS ?Amount)
WHERE {
  ?gps gps:latitude ?latitude;
  gps:longitude ?longitude;
  gps:hour?Hour
  FILTER(?latitude != 0.0).
  FILTER(?longitude != 0.0).
}
GROUP BY ?latitude ?longitude ?Hour
HAVING(COUNT(?gps) > 1)
ORDER BY DESC (?Amount)
LIMIT 20
```

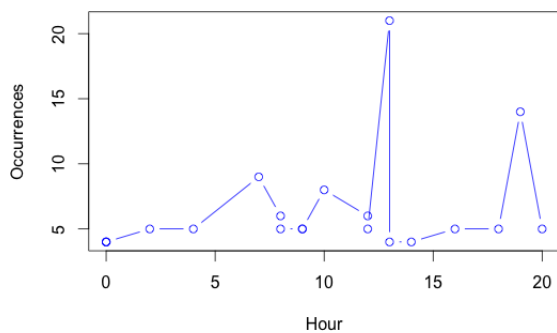


Figure 5. Results of busiest points in a certain time on Beijing

latitude	longitude	Hour	Amount
116.39696	39.81439	13	21
116.49653	39.9659	19	14
116.49711	39.95284	07	9
116.4436	39.82555	10	8
116.01101	40.35158	08	6
119.11908	40.06314	12	6
116.50432	40.00928	02	5
116.3726	39.96738	04	5
116.2737	39.9397	08	5
115.46215	41.23102	09	5
117.0063	40.14755	09	5
117.18312	40.18602	12	5
116.16002	39.67661	16	5
116.48792	40.01089	18	5
116.31802	39.93755	20	5
116.39075	39.84218	00	4
116.41175	39.93947	00	4
116.51686	40.23957	00	4
116.49892	39.723	13	4
116.46863	39.97601	14	4

Figure 6. Results of busiest points in a certain time on Beijing

#### C. The most crowded streets in a certain time

The third test aim was to find certain points with more occurrences regardless of day and time. To achieve results for this test, our dataset has been used. Figure 7 shows the results of SPARQL query:

```
SELECT ?latitude ?longitude
(COUNT(?gps) AS ?Amount)
WHERE {
  ?gps m:Latitude_from ?latitude;
  m:Longitude_from ?longitude.
  FILTER(?latitude != 0.0).
  FILTER(?longitude != 0.0).
}
GROUP BY ?latitude ?longitude
HAVING(COUNT(?gps) > 1)
ORDER BY DESC (?Amount)
```

latitude	longitude	Amount
20.6724837	-103.3305215	3
20.671611	-103.3463266	2
20.6748546	-103.3568629	2
20.6750575	-103.3401175	2
20.6973759	-103.3100631	2
20.6980102	-103.3116547	2
20.6982709	-103.3121455	2
20.6992375	-103.3139986	2
20.6992935	-103.3115487	2
20.6997488	-103.3136851	2
20.6999986	-103.3154296	2
20.7001552	-103.3110331	2
20.7002576	-103.3133749	2
20.7005223	-103.3150944	2
20.7015567	-103.3181134	2
20.7025497	-103.3118661	2
20.7031258	-103.3114649	2
20.7056635	-103.3100847	2
20.7061499	-103.3116284	2
20.706695	-103.3133045	2
20.7067842	-103.3096909	2
20.7070555	-103.3144593	2
20.7072601	-103.3112226	2
20.7097507	-103.3230464	2
20.7108174	-103.3224945	2
20.7137486	-103.30201	2
20.7175645	-103.2967534	2
20.7203743	-103.2957234	2

Figure 7. Results of the most crowded streets in Beijing

#### D. Distance between two particular points on the map

Fourth test target was to measure and demonstrate the distance between two particular points on the map. For this result, Beijing dataset has been used. Figure 8 shows the result to SPARQL query:

```

SELECT (SUM(?Distancia_Seg) AS
      ?Distance_mts)
WHERE
{
  ?gps1 m:IDSegmento ?Segmento;
        m:Distancia_Seg ?Distancia_Seg .

  FILTER(?Segmento > A ).
  FILTER(?Segmento < B ).
}

```

```

=====
! Distance_mts !
=====
! 9822.914631589 !
=====

```

Figure 8. Results of distance between point A to B on Beijing dataset

E. The most crowded streets between 7:00 PM to 3:00 AM

The aim of the fifth test was to find the crowded points between 7:00 PM to 3:00 AM. For obtain results of this test, Beijin dataset has been used. The following query was used in this aim.

```

SELECT ?longitude ?latitude ?t_hour
      (COUNT(?gps) as ?Amount )
WHERE
{
  {?gps gps:hour ?hour;
    gps:longitude ?longitude;
    gps:latitude ?latitude .
  BIND (SUBSTR(?hour,1,2) as ?t_hour)
  FILTER(?latitude != 0.0).
  FILTER(?longitude != 0.0).
  FILTER(?hour>="19:00:01").
  FILTER(?hour<="23:59:59").
  FILTER(?latitude=?latitude).
  FILTER(?longitude=?longitude)
}
UNION
{?gps gps:hour ?hour;
  gps:longitude ?longitude;
  gps:latitude ?latitude .
BIND (SUBSTR(?hour,1,2) as ?t_hour)
FILTER(?latitude != 0.0).
FILTER(?longitude != 0.0).
FILTER(?hour>="00:00:01").
FILTER(?hour<="03:00:00") .
FILTER(?latitude=?latitude).
FILTER(?longitude=?longitude)
}
}
GROUP BY ?longitude ?latitude ?t_hour
HAVING (COUNT(?gps)>2)
ORDER BY DESC (?Amount)

```

Figure 9 shows result to the SPARQL query:

```

=====
! longitude ! latitude ! t_hour ! Amount !
=====
! 39.9659 ! 116.49653 ! "19" ! 14 !
! 39.93755 ! 116.31802 ! "20" ! 5 !
! 40.00928 ! 116.50432 ! "02" ! 5 !
! 39.84218 ! 116.39075 ! "00" ! 4 !
! 39.93947 ! 116.41175 ! "00" ! 4 !
! 39.90652 ! 116.41823 ! "20" ! 4 !
! 40.23957 ! 116.51686 ! "00" ! 4 !
! 39.96689 ! 116.2228 ! "21" ! 3 !
! 39.94916 ! 116.37407 ! "21" ! 3 !
! 39.96749 ! 116.48391 ! "21" ! 3 !
! 39.69164 ! 116.42191 ! "20" ! 3 !
! 39.93284 ! 116.45692 ! "23" ! 3 !
! 39.98753 ! 116.48178 ! "02" ! 3 !
! 40.0798 ! 116.58538 ! "22" ! 3 !
! 40.13195 ! 116.66135 ! "02" ! 3 !
! 40.11647 ! 116.89351 ! "20" ! 3 !
! 39.78471 ! 117.0719 ! "21" ! 3 !
=====

```

Figure 9. Results of the most crowded streets between 7:00PM and 3:00AM of Beijing dataset

IV. CONCLUSION

OntoTraffic was developed as a basis for creating a recommendation system. Our system will recommend the best traffic route based on five solutions. The present paper shows that, by using GPS traces from citizens mobile devices, can recommend the best route in order to travel from point A to point B in a specific time of day or under certain conditions in a moment. It also helps to know the most frequent route, the most utilized roads, time with the most heavy traffic load, city areas with the biggest traffic jam, the best places to take a Taxi, the preferred route for drivers, the different routes available to reach a place or the alternative routes according to a specific daytime, in a reasonable execution time. It was observed that the processing time increases by the quantity of GPS traces to be analysed. Even in the first experiment, the quantity of traces was relevant (more than 1,000,000). Our main contribution is to have a recommendation system that could predict the behaviour of the traffic based on five questions and answers of certain routes depending on different user’s search criteria. Besides, our methodology is simple because based on ontologies and SPARQL language is possible to discover data in a huge dataset as Beijing. Above all, it is important to highlight that this work is in development process. For future work, we propose: (a) Build new models to support smart cities. (b) Support recommender systems development for decision makers. (c) Use data analytic and ontologies for a future HPC center.

ACKNOWLEDGMENT

We thank Sciences Research Council (CONACyT) for funding this research project.

REFERENCES

- [1] UNPD, “Unpd. 2012. world urbanization prospects: The 2011 revision,” tech. rep., 2011.
- [2] D. Schrank, T. Lomax, and B. Eisele, “Urban mobility report 2011. report for the texas transportation institute,” tech. rep., 2011.
- [3] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, “Real-time urban monitoring using cell phones: A case study in rome,” *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, pp. 141–151, March 2011.
- [4] R. Herring *et al.*, “Using mobile phones to forecast arterial traffic through statistical learning,” *89th Transportation Research Board Annual Meeting, Washington DC*, 2010.
- [5] Y. Zheng, X. Xie, and W.-Y. Ma, “Geolife: A collaborative social networking service among user, location and trajectory,” *IEEE Data Eng. Bull.*, vol. 33, no. 2, pp. 32–39, 2010.
- [6] J. Yuan *et al.*, “T-drive: driving directions based on taxi trajectories,” in *GIS* (D. Agrawal, P. Zhang, A. E. Abbadi, and M. F. Mokbel, eds.), pp. 99–108, ACM, 2010.
- [7] K. Sia-Nowicka, J. Vandrol, T. Oshan, J. A. Long, U. Demar, and A. S. Fotheringham, “Analysis of human mobility patterns from gps trajectories and contextual information,” *International Journal of Geographical Information Science*, vol. 30, no. 5, pp. 881–906, 2016.