Analysis of Tweets in Arabic Language for Detection of Road Traffic Conditions

Ebtesam Alomari[1(✉)] and Rashid Mehmood[2]

[1] Faculty of Computing and Information Technology, King AbdulAziz University, Jeddah, Saudi Arabia
EAlomari0011@stu.kau.edu.sa

[2] High Performance Computing Center, King AbdulAziz University, Jeddah, Saudi Arabia
RMehmood@kau.edu.sa

Abstract. Traffic congestion is a worldwide problem, resulting in massive delays, increased fuel wastage, and damages to human wealth, health, and lives. Various social media e.g. Twitter have emerged as an important source of information on various topics including real-time road traffic. Particularly, social media can provide information about certain future events, the causes behind the certain behavior, anomalies, and accidents, as well as the public feelings on a matter. In this paper, we aim to analyze tweets (in the Arabic language) related to the road traffic in Jeddah city to detect the most congested roads. Using the SAP HANA platform for Twitter data extraction, storage, and analysis, we discover that Al-Madinah, King AbdulAziz, and Alharamain are the most congested roads in the city, the tweets related to the road traffic are posted mostly in the rush hours, and the highest traffic tweeting time is 1 pm.

Keywords: Twitter analysis · Traffic congestion · Arabic language
Big data analytics · SAP HANA · Smart cities

1 Introduction

Traffic congestion is a worldwide problem, resulting in massive delays, increased fuel wastage, monetary losses, and damages to human health and lives. Traditional approaches for traffic measurement have relied on road sensors. Video analysis, mobile data, and vehicular ad hoc networks (VANETs) are being used also for traffic monitoring purposes. More recently, various social media such as Twitter have emerged as an important source of information on various topics including real-time road traffic. Particularly, social media can provide information about certain future events, the causes behind certain behavior, anomalies, and accidents, as well as the public feelings on a matter.

Twitter is widely used for communication, and sharing personal status, events, news, etc. It was originally introduced in 2006, and the number of active users is growing every year. According to the sixth edition of Arab Social Media Report in 2014 [1], the number of the active Twitter users in Saudi Arabia is 2.4 million, which represent the highest number of Twitter users in the Arab region. Furthermore, the

R. Mehmood et al. (Eds.): SCITA 2017, LNICST 224, pp. 98–110, 2018.
https://doi.org/10.1007/978-3-319-94180-6_12

Kingdom accounts for over 40% of all active Twitter users in the Arab region. By 2016, the number of Twitter users in Saudi Arabia had reached 4.99 million [2]. Saudi Arabia, hence, presents an excellent opportunity for extracting useful information from Twitter media.

Twitter provides easy-to-access APIs to enable collecting large and diverse data for analysis of valuable information such as road traffic information. Currently, road traffic congestion is one of the biggest problems in Saudi Arabia especially in large cities like Jeddah. Jeddah city is the second largest city in Saudi Arabia and arguably the most congested one. The lack of public transportation, the increasing number of vehicles, and an enormous number of pilgrim visitors all year round, have increased accidents and traffic jams in many major roads in the city.

A Large number of tweets are posted every day, by users who wish to inform their followers about current road traffic congestion condition, either to share information or to complain about the traffic problems. Also, there are official Twitter accounts, which have thousands of followers and tweets about the current road traffic conditions in Jeddah, such as @Jed_Rd1. Thus, analyzing these tweets might be useful to predict traffic congestion.

In this paper, we collect tweets in Arabic language related to traffic in Jeddah city and then analyze the data to find the most congested streets and roads. The data will be stored in *SAP HANA* database. *SAP HANA workbench* will be used to create the tables, execute the queries and analyze the data. Further, *SAP Lumira* will be used to visualize the results.

The standard text Analysis in *SAP HANA* using the *VOICEOFCUSTOMER-configuration* does not suffice. There is a need to create custom dictionaries for unknown terms in the SAP HANA system. Therefore, we will create dictionaries for Jeddah streets and roads names. Additionally, we will create dictionaries for the most popular Arabic words related to traffic jam, transportation, and the causes of traffic congestion. Finally, analysis results will be visualized using the word cloud and charts.

The main objectives and contributions of this paper can be summarized as follows:

- To discover the most congested streets and roads in Jeddah.
- To find the time periods when the number of tweets about traffic are at the highest.
- To detect the major causes of traffic congestion (accident, rains, etc.)
- To find the Arabic words that are used while tweeting about the traffic.

The Twitter data analysis carried out in this paper reveals that the most congested roads are Al-Madinah Rd, King AbdulAziz Rd, and Alharamain Rd. We also found that the most tweets related to the road traffic are posted in the rush hours and the highest traffic tweets time is 1 pm.

The rest of the paper is organized as follows. Section 2 presents brief information about SAP HANA. Section 3 reviews the related works. Section 4 illustrates the methodology. Section 5 explains the analysis process, followed by the results in Sect. 6. Finally, we draw our conclusions in Sect. 7.

2 Background: SAP HANA

SAP HANA is a relational database management system (RDBMS) developed by SAP SE. Additionally, it is an in-memory columnar database offering groundbreaking performance. It is the integration of transactional and analytical workload within the same database management system [3]. Further, SAP HANA Extended Application Services (SAP HANA XS) provides the *SAP HANA Web-based Development Workbench* that supports developing entire applications in a Web browser without the need to install any development tools.

SAP HANA Web-based Development Workbench includes a *Catalog* and *Editor* tools [4].

- Catalog: enables developing and maintaining SQL catalog objects in the SAP HANA database. It also supports creating tables, executing SQL queries and creating a remote source to collect data. Furthermore, catalog supports text analysis and text mining.
- Editor: supports running design-time objects in the SAP HANA Repository. It supports a great information view, which is a calculation view. The data foundation of the calculation view can include tables, column views, analytic views and calculation views. Also, it enables creating Joins, Unions, Aggregation, and Projections on data sources.

Moreover, SAP offers a data visualization tool for reporting on top of SAP HANA, named *SAP Lumira*[1]. It can be used to investigate GBI data stored in SAP HANA. It is a straightforward and user-friendly interface which allows users to swiftly analyze data without the need for scripting. To analyze the text in SAP HANA, there is a need to create *full-text indexing* on the text column and this results in the new table '$TA_table.' It supports tokenization, which means it decomposes word phrase or sentence into tokens. After that, it automatically specifies a type of each token such as persons, products or places.

Moreover, *catalog* supports text analysis for Arabic text. Words can be extracted and linked to the corresponding topics. Additionally, it can detect if any emoticons are used in the text. To use the default configuration, developers simply need to include *VOICEOFCUSTOMER* parameter in a query. However, if the standard configuration doesn't suffice to the requirement, developers need to customize keywords in new dictionaries.

3 Literature Review

Many traffic monitoring systems have been proposed to detect road congestion using video and image processing technologies. Wei and Dai suggested a real-time traffic congestion estimation approach based on image texture feature extraction and texture analysis [5]. A congestion detection approach using video processing has been

[1] http://saplumira.com/.

proposed in [6]. A congestion model based on speed and density of vehicles was proposed in [7]. Traditional approaches for traffic measurement have relied on sensors that are buried under the road (such as inductive loops) or installed on roadside [8]. However, these approaches require sensors and other equipment such as cameras and thus the deployment and maintenance are costly.

Several approaches have been proposed, particularly during the last decade, to use vehicular ad hoc networks (VANETs) for monitoring traffic [8–10], in general, and for specific purposes, such as for traffic coordination and disaster management [11–13]. Simulations have also been playing a key role in transportation planning and control [14]. A number of works on operations research related to transportation in smart cities have also been proposed, see e.g. car-free cities [15], intelligent mobility [16], big data in transport operations [17, 18], prototyping in urban logistics [19], and autonomic transportation systems [20–22],

More recently, Twitter has become a popular social platform to share real-time traffic information since numerous users tweet to report about problems that may affect traffic such as traffic accidents. Furthermore, there are specific and official Twitter accounts created to report on traffic conditions and events in particular cities. These accounts generate useful sources of information for the followers. Therefore, there is an enormous amount of traffic updates and information available in different Twitter accounts and can be freely obtained via the easy-to-access APIs [23].

Recently, some methodologies, techniques, and tools have been proposed to analyze traffic tweets. Gu *et al.* [24] proposed a method to collect, process and filter public tweets about traffic in Pittsburgh and Philadelphia Metropolitan. They have used Twitter REST API to collect real-time tweets using a dictionary of relevant keywords and their combinations that can indicate traffic condition. After that, the collected tweets are geocoded to define their locations. Then, the tweets are categorized into one of the five event classifications, which are Accidents, Roadwork, Hazards & Weather, Events, and Marathon. However, they mainly focused on identifying the incident categories, and they did not identify the most congested roads.

Moreover, techniques have also been proposed to analyze tweets using data mining. Kurniawan *et al.* [25] conducted experiments to classify real-time road traffic tweets using data mining. They collected real-time data about Yogyakarta Province, Indonesia using Twitter Streaming API. Further, they specified search parameters such as follow and track parameters. In follow parameter, they defined a list of the twitter account that reports about traffic, to get tweets from these accounts only. Furthermore, track parameter was used to identify the keywords that should be included in the retrieved data. Additionally, they compared classification performance of three machine learning algorithms, namely Naive Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT). However, they only classified tweets into the traffic or non_traffic categories.

Similar work is proposed by D'Andrea *et al.* [26]. They suggested an intelligent system, based on text mining and machine learning algorithms. They collected real-time tweets of several regions of the Italian road networks and then assigned the appropriate class label to each tweet, as to whether the tweet is related to a traffic event or not. Ribeiro *et al.* [27] analyzed tweets to detect traffic events. They manually listed the most frequent types of events that were used to indicate traffic conditions. The dataset is collected from ten twitter accounts whose primary purpose is to report traffic conditions in Belo Horizonte, Brazil. After that, they created a set of place names, called GEODICT. Subsequently, they detected the locations and streets names by using string matching technique by searching for substrings from the tweet that can be detected in GEODICT. Wongcharoen and Senivongse [28] built a congestion severity prediction model to predict traffic congestion severity level. The collected tweets are geotagged and contain traffic-related keywords. However, like previous approaches [25, 27], the tweets are fetched only from particular accounts. Finally, Suma et al. [29] have analyzed Twitter data to detect events related to road traffic and other topics for smart cities planning purposes. Their focus is on the use of big data platforms including Spark and Hadoop to analyse large amounts of data. The results are presented by analyzing 500,000 tweets about the London city.

In this paper, we execute several queries using the Arabic keywords and synonyms to fetch tweets that are posted about Jeddah traffic, and then we analyzed the tweets to identify the most congested roads. To the best of our knowledge, most papers on traffic detection using Twitter data analysis focus on languages other than Arabic, and none of them have analyzed Arabic language tweets in Jeddah for traffic related purposes. Twitter offers two categories of search APIs [30]; (a) Streaming API that gives Twitter's global stream of Tweet data, and (b) Twitter REST API or batch query API that supports access to read and write Twitter data. In this work, we use the REST API because it supports writing search queries to retrieve tweets that include the specified keywords or tweets from specific Twitter accounts. Additionally, there is a possibility to use 'OR' and 'AND' logical operations between keywords to design good quality queries. However, the search index has a 7-day limit. So, there is a need to re-execute the queries to retrieve new tweets.

Furthermore, REST API supports *geocode* parameter to restrict query by a given location using "latitude, longitude, radius". Thus, when executing the queries, the search API will first attempt to search for Tweets which have lat/long within the queried geocode. If there are no results, it will attempt to detect Tweet's location information from the location data in user's profile.

4 Methodology

We generated a list of Arabic keywords related to road traffic, transportation, and traffic reasons. In addition, we collected a list of streets and roads names in Jeddah using OpenStreetMap[2]. We searched for the most popular Twitter accounts that tweet about Jeddah and traffic conditions in the city.

As shown in Fig. 1, the main implementation steps can be summarized as follows:

1. Search queries will be executed in *SAP HANA Web-based Development Workbench Catalog* to collect tweets using twitter REST search API. The retrieved tweets will be stored in the created table '`TrafficJed`'.
2. The duplicated tweets (retweets) will be deleted from the table.

[2] https://www.openstreetmap.org/.

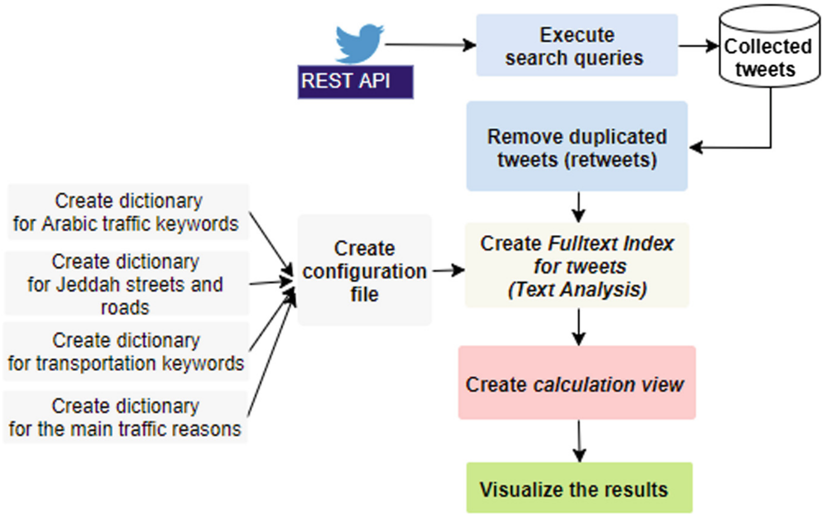

Fig. 1. Implementation Flowchart

3. The created lists of custom dictionaries will be used to create a new configuration file for analysis using *SAP HANA Web-based Development Workbench Editor*.
4. The created configuration file will be used to create the *fulltext index* on 'Tweets' column to split the text into tokens and specify the token type based on the created dictionaries.
5. Calculation view will be created.
6. In SAP Lumira, the result will be virtualized using charts and word clouds.

4.1 Search Queries

We tried to design high-quality search queries to retrieve a large number of highly related tweets about traffic in Jeddah. We have used the collected list of Arabic keywords about traffic and transportation to write a large number of queries. In addition, we have used the following accounts that tweets about Jeddah (@Jed_Rd1, @eMoroor, @SukkanJeddah, @jedgovsa, @JeddahNow and @jeddah_ar) to fetch their tweets about traffic.

Generally, there are two types of location information:

1. Tweets from geotagging enabled smartphones that carry latitude/longitude coordinates of the locations where users posted the tweets.
2. Location name referred in tweet texts.

In this work, this two location information is used to collect the tweets related to traffic and transportation in Jeddah city. To handle the problem of non-geotagged tweets, we re-execute all queries after adding the keyword 'Jeddah' and modifying the locations parameters to 'null'. However, there are still some tweets that are not included in our analysis because they do not carry any location information.

Query Example:

```
UPSERT "GBI_001"."Trafficjed" SELECT * FROM "GBI_001"." Tweets"
('@Jed_Rd1 OR @eMoroor OR @SukkanJeddah OR @jedgovsa OR @JeddahNow OR
@Jeddah_ar AND ( زحام OR زحمه OR يزدحم OR مزدحم OR مزدحمة )',
1500,null, null, '21.3891, 39.8579, 35mi' , null , null , null , null );
```

4.2 Tweets Collection

We created a table to store the retrieved tweets. The created table includes several attributes such as 'UserId', 'Tweet', ' UserName', 'CreatedAt', 'Latitude', 'Longitude', 'Country', 'Place_name'. All the collected tweets will be stored in 'Tweet' column. Location fields will be important in the analysis process. However, if the user did not add information about the city and county in his/her profile, 'Country' and 'Place_name' fields would be empty. In addition, if they disable location service in their smartphones 'Latitude', 'Longitude' will be empty. 'CreatedAt' attribute refers to the time and date of posting the tweet. It will be helpful to find the top tweet time. Moreover, after we created the table, we executed all search queries. We collected tweets in the period between 13/4/2017 and 19/5/2017. Before analyzing the data, we delete all duplicated tweets in the table.

5 Tweets Analysis

5.1 Custom Dictionaries

The standard text Analysis in SAP HANA using the VOICEOFCUSTOMER-configuration does not suffice where not all Arabic tokens are classified under the right token type. Therefore, we need to add a custom dictionary for unknown terms in the SAP HANA system and then create a new configuration file. We created four dictionaries, which are:

- JeddahStreets: includes the main streets and roads in Jeddah city.
- TrafficJam: includes the collected list of Arabic traffic congestion synonyms (such as زحام, يزدحم, تزدحم, مزدحمة, مزدحمه).
- Transportation: includes the collected Arabic keywords about transportation (such as طريق, الطريق, طرق, الطرق, سيارة, سياره, السيارة).
- TrafficReasons: include the collected Arabic words about traffic congestion reasons (such as حادث,حريق, مطر).

5.2 Fulltext Index

To analyze the tweets in SAP HANA, we need to create a fulltext index for 'Tweet' column. Creating the index required executing the following SQL statement, which will lead to creating a new table containing analysis results.

```
CREATE FULLTEXT INDEX"GBI_001"."Trafficjed_Sentiment" ON
"GBI_001"."Trafficjed_001" (TWEET) TEXT ANALYSIS ON LANGUAGE COLUMN
"ISOLANGUAGECODE" LANGUAGE DETECTION ('AR') CONFIGURATION 'gbi-student-
001.CustomDictionaries:: StreetConfig_001';
```

The created table ‘`Trafficjed_Sentiment`’ will include:

- TA_token: the list of keywords extracted for the tweets.
- TA_type (token type): It will be one of the types specified in our newly created dictionaries, i.e., TrafficJam or TrafficReason).

5.3 Calculation View

To get the result of the text analysis, we created calculation view. We need two columns from the table ‘`TrafficJed`’, which are ‘`ID`’ the id of the tweets and ‘`CreatedAt`’, which represents the date and time of posting a tweet by the user. We also need the ‘`ID`’, ‘`TA_Token`’ and ‘`TA_Type`’ from ‘`Trafficjed_Sentiment`’.

Therefore, we added two projections the first one for the table that includes the tweets (`TrafficJed`) and the second one for the table created for analysis (`Trafficjed_Sentiment`). Further, we need the month and hour of posting the tweets, so we calculate these data from ‘`CreatedAt`’ attribute. When we executed the created calculation view, we got result table. Figure 2 shows a sample of the results.

6 Results

To achieve the main objective of this paper, we created a chart to represent the most congested streets and roads in Jeddah, as shown in Fig. 3. However, in analysis results, we noticed that SAP HANA represents the same Arabic words that end with ‘ـه’ or ‘ـة’ as different words. For instance, ‘طريق المدينة’ and ‘طريق المدينه’ are the same roads, which is ‘Al-Madinah Rd’. In addition, ‘شارع التحلية’ and ‘شارع التحليه’ are the same street ‘Tahlia St’. Moreover, ‘طريق الملك’ is mostly used as an abbreviation for ‘طريق الملك عبدالعزيز’ ‘King Abdulaziz Rd’. To handle this issue, we calculated the total tweets frequency for each street or road that appear in analysis results as different words, and then we draw a chart to represent the most congested roads and streets. As shown in Fig. 4, the top 5 congested roads are *Al-Madinah* Rd, *King Abdulaziz* Rd, *Alharamain* Rd, *Tahlia* St. and *Makkah-Jeddah* highway.

We draw a chart to find the peak time. We noticed that users post tweets about traffic at 7 am, and 1, 3, 4 pm more than any other time in the day. As shown in Fig. 5, the highest tweets time is at 1 pm. These results are expected because they represent rush hours (time to get to *work*/school or to get home from *work/school*).

We discovered the top traffic reasons that mentioned in the tweets. The results indicated that the synonyms of the word ‘حادث’ (accidents) are the most traffic reasons mentioned in the collected tweets in May. On the other side, the synonyms of the word

	HOUR	ID	MONTH	TA_TOKEN	TA_TYPE
9	13	1	4	تقاطع	transportation
10	13	1	4	شارع فلسطين	JedStreets
11	11	1	4	زحمه	trafficJam
12	14	1	4	زحمه	trafficJam
13	16	1	4	الزحمه	trafficJam
14	16	4	4	زحمه	trafficJam
15	16	2	4	الشارع	transportation
16	19	1	4	زحمه	trafficJam
17	4	1	4	الحريق	TrafficReason
18	4	1	4	المزدحمه	trafficJam
19	4	1	4	طريق مكه	JedStreets
20	4	1	4	مزدحم	trafficJam
21	4	1	4	شاحنات	transportation

Fig. 2. Sample of analysis results

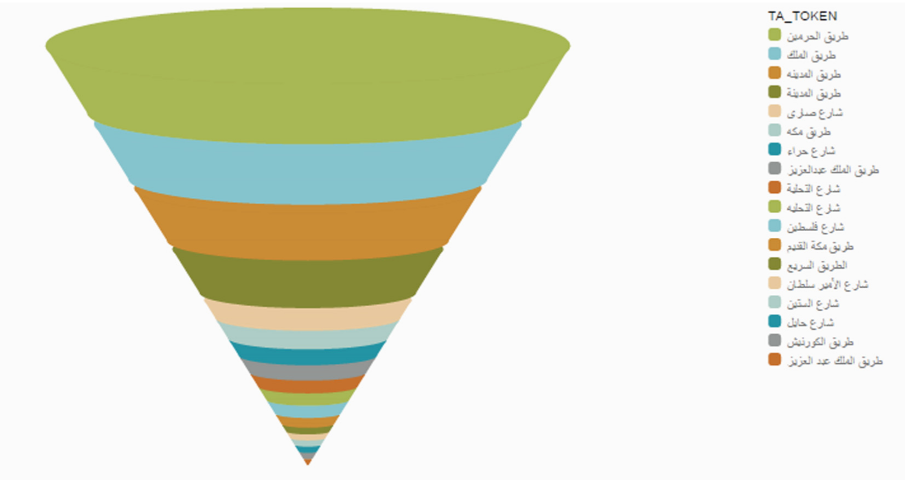

Fig. 3. Congested streets and roads

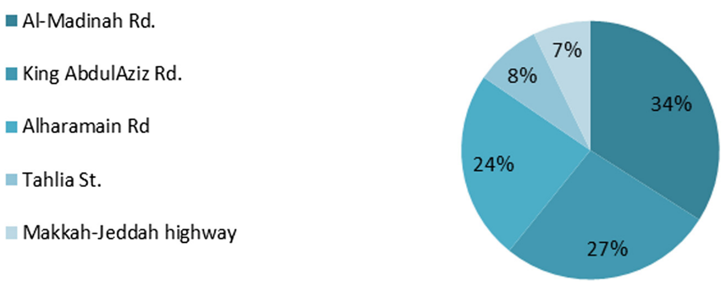

Fig. 4. Top 5 congested streets and roads in Jeddah

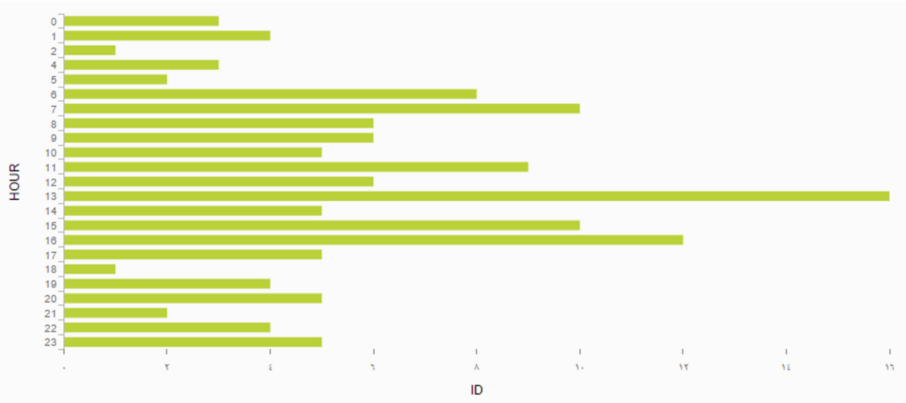

Fig. 5. Number of the posted tweets about traffic per hour

‘حريق’ (fire) are the most reason that mentioned in the collected tweets in April because fuel truck caught fire on the main road between Makkah and Jeddah city.

Furthermore, Fig. 6 shows the word cloud for the Arabic traffic jam, traffic reasons and transportation words and synonyms. We can notice that the most used words are the synonyms of the following three keywords ‘طريق’ (road), ‘شارع’ (street), and ‘زحمة’ (crowding).

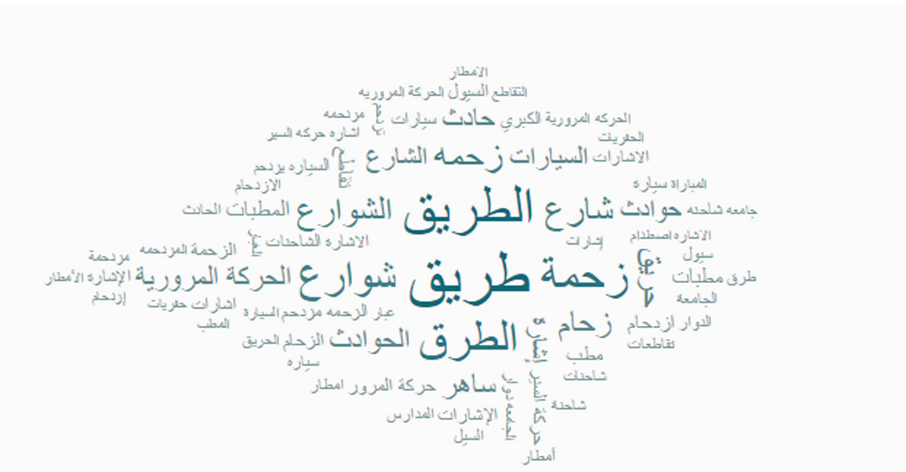

Fig. 6. Traffic and transportation keywords

7 Conclusions

Twitter becomes one of the top social platforms for sharing user-generated content. Several users post tweets about current traffic conditions. Additionally, there are official accounts specialized on this subject. In this paper, we focused on analyzing tweets about traffic in Jeddah to find the most congested streets and roads. We collected tweets using Twitter REST API and stored the collected data in SAP HANA database.

To make sure that all the collected tweets are related to traffic in Jeddah, we have used the geocode parameter in REST search API and specify the lat/long of Jeddah city. But not all tweets are geotagged because some users disable location service in their smartphones. Therefore, we re-execute all queries after adding the keyword ‘Jeddah’ and without specifying a location, to collect all traffic tweets that include ‘Jeddah’ keyword. However, there are still some tweets that are not included in our analysis because they are not geotagged and not carrying location information.

Moreover, SAP HANA Web-based Development Workbench was used to create a table, execute search queries and analyze the data. While the default analysis configuration in SAP HANA is not efficient for Arabic text analysis, we created a new configuration file. We added new dictionaries for the Arabic keywords related to traffic jam, transportation, and traffic reasons. In addition, we created a dictionary for the main streets and roads in Jeddah.

Finally, we have used SAP Lumira to virtualize the results by creating charts and word cloud. We found that the most congested streets and roads in Jeddah are Al-Madinah Rd, King Abdulaziz Rd. and Alharamain Rd. We also found that the most tweets related to traffic are posted in the rush hour (at 7 am and 1, 3, 4 pm). In future, we plan to collect more tweets to analyze traffic in other cities such as Riyadh. We also plan to use other data sources to collect traffic data. Future work will also consider improving the data analyses methodology.

Acknowledgments. The work carried out in this paper is supported by the HPC Center at the King Abdulaziz University.

References

1. Mourtada, R., Salem, F., Al-Shaer, S.: Citizen Engagement and Public Services in the Arab World: The Potential of Social Media. Arab Social Media Report (2014)
2. Twitter: number of active users 2010–2016, Statista (2016). www.statista.com
3. SAP: What is SAP HANA. In Memory Computing and Real Time Analytics
4. SAP HANA Web-Based Development Workbench - Introduction to SAP HANA Development - SAP Library
5. Wei, L., Dai, H.-Y.: Real-time road congestion detection based on image texture analysis. Procedia Eng. **137**, 196–201 (2016)
6. Kanungo, A., Sharma, A., Singla, C.: Smart traffic lights switching and traffic density calculation using video processing. In: 2014 Recent Advances in Engineering and Computational Sciences (RAECS), pp. 1–6. IEEE (2014)

7. Widyantoro, D.H., Enjat Munajat, M.D.: Fuzzy traffic congestion model based on speed and density of vehicle. In: 2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA), pp. 321–325. IEEE (2014)
8. Mehmood, R., Nekovee, M.: Vehicular AD HOC and grid networks: discussion, design and evaluation. In: 14th World Congress on Intelligent Transport Systems, ITS 2007, pp. 1555–1562 (2007)
9. Gillani, S., Shahzad, F., Qayyum, A., Mehmood, R.: A survey on security in vehicular ad hoc networks. In: Berbineau, M., et al. (eds.) Nets4Cars/Nets4Trains 2013. LNCS, vol. 7865, pp. 59–74. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37974-1_5
10. Alvi, A., Nabi, Z., Greaves, D.J., Mehmood, R.: Intra-vehicular verification and control: a two-pronged approach. Int. J. Veh. Inf. Commun. Syst. **2**, 248–268 (2011)
11. Alazawi, Z., Altowaijri, S., Mehmood, R., Abdljabar, M.B.: Intelligent disaster management system based on cloud-enabled vehicular networks. In: Vinel, A., et al. (eds.) 2011 11th International Conference on ITS Telecommunications, ITST 2011, pp. 361–368. IEEE, St. Petersburg (2011)
12. Alazawi, Z., Abdljabar, M.B., Altowaijri, S., Vegni, A.M., Mehmood, R.: ICDMS: an intelligent cloud based disaster management system for vehicular networks. In: Vinel, A., Mehmood, R., Berbineau, M., Garcia, C.R., Huang, C.-M., Chilamkurti, N. (eds.) Nets4Cars/Nets4Trains 2012. LNCS, vol. 7266, pp. 40–56. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29667-3_4
13. Alazawi, Z., Alani, O., Abdljabar, M.B., Altowaijri, S., Mehmood, R.: A smart disaster management system for future cities. In: WiMobCity 2014. International Workshop on Wireless and Mobile Technologies for Smart Cities, pp. 1–10 (2014)
14. Ayres, G., Mehmood, R.: On discovering road traffic information using virtual reality simulations. In: 11th International Conference on Computer Modelling and Simulation, UKSim 2009, pp. 411–416 (2009)
15. Mehmood, R., Lu, J.A.: Computational Markovian analysis of large systems. J. Manuf. Technol. Manag. **22**, 804–817 (2011)
16. Büscher, M., et al.: Intelligent mobility systems: some socio-technical challenges and opportunities. In: Mehmood, R., Cerqueira, E., Piesiewicz, R., Chlamtac, I. (eds.) EuropeComm 2009. LNICST, vol. 16, pp. 140–152. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-11284-3_15
17. Mehmood, R., Meriton, R., Graham, G., Hennelly, P., Kumar, M.: Exploring the influence of big data on city transport operations: a Markovian approach. Int. J. Oper. Prod. Manag. **37**, 75–104 (2017)
18. Mehmood, R., Graham, G.: Big data logistics: a health-care transport capacity sharing model. Procedia Comput. Sci. **64**, 1107–1114 (2015)
19. Graham, G., Mehmood, R., Coles, E.: Exploring future cityscapes through urban logistics prototyping: a technical viewpoint. Supply Chain Manag. **20**, 341–352 (2015)
20. Schlingensiepen, J., Mehmood, R., Nemtanu, F.C.: Framework for an autonomic transport system in smart cities. Cybern. Inf. Technol. **15**, 50–62 (2015)
21. Schlingensiepen, J., Mehmood, R., Nemtanu, F.C., Niculescu, M.: Increasing sustainability of road transport in European cities and metropolitan areas by facilitating autonomic road transport systems (ARTS). In: Wellnitz, J., Subic, A., Trufin, R. (eds.) Sustainable Automotive Technologies, 2013 Proceedings of the 5th International Conference ICSAT 2013, pp. 201–210. Springer, Ingolstadt (2014). https://doi.org/10.1007/978-3-319-01884-3_20

22. Schlingensiepen, J., Nemtanu, F.: Autonomic transport management systems—enabler for smart cities, personalized medicine, participation and industry grid/industry 4.0. In: Sladkowski, A., Pamula, W. (eds.) Intelligent Transportation Systems – Problems and Perspectives, pp. 3–35. Springer, London (2016). https://doi.org/10.1007/978-3-319-19150-8_1
23. Wang, S., He, L., Stenneth, L., Yu, P.S., Li, Z.: Citywide Traffic Congestion Estimation with Social Media
24. Gu, Y., Qian (Sean), Z., Chen, F.: From Twitter to detector: real-time traffic incident detection using social media data. Transp. Res. Part C Emerg. Technol. **67**, 321–342 (2016)
25. Kurniawan, D.A., Wibirama, S., Setiawan, N.A.: Real-time traffic classification with Twitter data mining (2016)
26. D'Andrea, E., Ducange, P., Lazzerini, B., Marcelloni, F.: Real-time detection of traffic from Twitter stream analysis. IEEE Trans. Intell. Transp. Syst. **16**, 2269–2283 (2015)
27. Ribeiro, S.S., Davis, C.A., Oliveira, D.R.R., Meira, W., Gonçalves, T.S., Pappa, G.L.: Traffic observatory: a system to detect and locate traffic events and conditions using Twitter Sílvio. In: Proceedings of the 5th International Workshop on Location-Based Social Networks - LBSN 2012, p. 5 (2012)
28. Wongcharoen, S., Senivongse, T.: Twitter analysis of road traffic congestion severity estimation. In: 13th International Joint Conference on Computer Science and Software Engineering (2016)
29. Suma, S., Mehmood, R., Albugami, N., Katib, I., Albeshri, A.: Enabling next generation logistics and planning for smarter societies. Procedia Comput. Sci. **109**, 1122–1127 (2017)
30. Kumar, S., Morstatter, F., Liu, H.: Twitter Data Analytics. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-9372-3