

# Twitter-based reporting system for public infrastructure in Saudi Arabia

Abdullah Alsaedi  
University of Southampton

Roobaea Alroobaea  
Taif University, Taif, Saudi Arabia

Sepi Chakaveh  
University of Southampton

## ABSTRACT

In recent years, many governments have adopted the open government concept to serve several purposes such as increasing transparency and facilitating communication with their citizens. One of the aims of open government is to help society by engaging the citizens in the country's development processes. Citizens' participation is one of the most important factors in improving countries today, and this motivates governments to focus on encouraging their citizens to play a role in the improvement process. In this regard, the citizens can participate in several areas, and problem reporting is one of the most common ways of citizen participation. There are many applications that can perform these tasks in different countries; however the Saudi Arabian government does not currently employ these applications. This paper aims to propose a model and application that allow citizens' participation in reporting problems with public infrastructure. The application can inform the Saudi Arabian government about the problems facing public infrastructure that are encountered everyday by the citizens, such as potholes, graffiti, damaged lights and any broken or dirty streets that need repairing or cleaning. Because of the shortage of classic methods of reporting, this project depends on Twitter as an intermediate platform to collect the citizens' reports and deliver them to the relevant government agencies, due to the large number of Twitter users and problems with communication between government and citizens. The result of the study shows that the proposed application yields significant results with regard to improving services in Saudi Arabia. It helps to improve the quality of public infrastructure, facilitate the communication between government and citizens, and increase transparency by publishing reports and enabling citizens to view them and express their opinions.

**Keywords:** Open government, Public infrastructure, Twitter-based reporting system, Sentiment analysis, linguistic resources

Copyright statement: Authors retain the copyright to the manuscripts published in AABRI journals. Please see the AABRI Copyright Policy at <http://www.aabri.com/copyright.html>

## INTRODUCTION

Many governments in the world depend on their citizens to perform some tasks that facilitate problem solving; the role of the citizens might be different, depending on the problem type. For example, some governments provide open data that can be accessed by their citizens to develop innovative applications that help to make their life easier, or valuable statistics may be published that help to increase transparency. In the UK, for instance, there is an independent application developed by a non-profit organisation [1] called FixMyStreet [2], which enables citizens to report problems they have found on the streets, such as potholes, graffiti, or damaged lights. The public can easily use this application to send a report that will be forwarded to the city council. The reports sent through this application will also be available for public view. In this regard, public infrastructure in a country needs to be constantly observed so that its problems and flaws can be reported. It could be difficult for the government to know about the problems in different places immediately, especially in large countries. In such cases, the role that citizens play is necessary to communicate problems to the government quickly.

Statistics in [3] showed that Saudi Arabia has the highest percentage of active Twitter users in the Arab World. Thousands of Tweets are posted every day and some of them contain complaints about public services and their quality. Based on these interesting statistics, the social network is deemed to perform the same task as other reporting applications. During the initial phase of this project, a quick test was conducted to identify the efficiency of this idea by searching for complaints about public infrastructure posted by Saudi Arabia's citizens. The test found that there are many complaints containing valuable information. Therefore, this idea was developed further for the collection and processing of the citizen's reports. This paper aims to use Twitter to find, collect, analyse and classify these complaints, to deliver them so that they can be dealt with by the relevant government agencies. In this regard, the use of a social network enables access to the largest possible number of complaints from the biggest possible number of users, instead of using an independent, isolated application. Thus, this project focuses on complaints about public infrastructure and public services sent by Saudi Arabian citizens.

This paper is developed as follows. After the introduction, work related to open government and reported systems will be reviewed to identify the research gap. Next, the proposed system will be explained. Then, the proposed system design will be presented, and finally, the future work and conclusion will be discussed.

## LITERATURE REVIEW

Open government is defined as “the transparency of government actions, the accessibility of government services and information, and the responsiveness of government to new ideas, demands and needs.” [4]. The open government has certain characteristics that could provide a detailed picture of the concept [5]. These characteristics are: 1) information transparency, to enable the citizens to know how the government works by publishing government data (e.g. budgets, contracts, voting records, and political finance) to be accessible in the open data portal; 2) citizens' participation, which engages citizens in the improvement of the government by encouraging them to contribute and use their expertise to benefit the government in building projects that use open government data; 3) accountability, which means the ability of citizens hold the government to account for its actions and behaviours, in order to reduce corruption.

In this regard, the US government started working on their own open government in January 2009 through the Open Government Data (OGD) Portal ([Data.gov](http://Data.gov)) [6]. It is claimed that

openness would strengthen their democracy through three main points: transparency, participation and collaboration. A year later in 2010, the UK government publicly launched its OGD Portal ([Data.gov.uk](http://Data.gov.uk)) [7]. This portal contains many datasets from different resources and enables citizens to participate by innovative applications. Consequently, many other countries became interested in the open government concept and built their own OGD Portals to benefit from the portals' advantages.

In terms of the government of Saudi Arabia, it realised the importance of e-government in 2005 when it established the “Yesser” programme. Its aim was to transform Saudi society into an information society, and to facilitate communication between the citizens and the government in order to increase the efficiency of public services [8] [9]. Moreover, in order to improve the e-government platform in Saudi Arabia, the Yesser program introduced the Saudi OGD Portal ([Data.gov.sa](http://Data.gov.sa)) in 2015 [10]. It contains different datasets from several institutes and government ministries. However, the OGD Portal was not able to achieve its established goals as there was a lack of interaction and participation from the citizens, and a lack of real transparency. One of the main reasons is that most datasets in the portal are not actually open, and do not follow the principles of publishing that enable people to benefit from the data.

To publish open datasets, certain criteria must be followed in the publishing process. The e-Government Interest Group sets eight principles for the data to be considered as open [11]; these principles briefly stated that the data should be: 1) completely published without limitations; 2) in its original form and not modified; 3) available as quickly as possible to save its value; 4) accessible to the widest range of people; 5) structured in machine-readable format; 6) available to everyone without registration; 7) available without any exclusive control for entities; and 8) not subject to any control through copyrights, trademarks, or patents. Therefore, any open data publisher must follow the above principles to provide valuable open data. However, the OGD Portal of Saudi Arabia was analysed and many of these principles were not applied. Firstly, most datasets had been modified or aggregated and were not in their original forms.

Secondly, the latest added datasets were published a long time ago, which means that no timely data were provided. Lastly, most datasets are not in a machine-readable format. Therefore, there is a lack of apparent interaction within the OGD Portal and there is an absence of innovative applications that can solve the government's problems, as the data are unusable. In addition, several types of reporting applications have been used (e.g. FixMyStreet and SeeClickFix), whether by the governments or private organisations. Most of these applications work in similar ways, where the application is used by the user to write a description of the problem, and then to send the report to the responsible parties. Consequently, these parties will review the complaints and finally fix the problem and provide a reply to the user.

Furthermore, reporting applications in Saudi Arabia that report public infrastructure problems are not very common. The public usually uses traditional communication channels to report such problems – nevertheless, not all authorities depend on these channels as recently some authorities have adopted new technologies to facilitate the reporting process. With the lack of centralisation, each municipality in Saudi Arabia has different ways of reporting. Common reporting methods include: 1) a phone call to an official number (e.g., almost in all municipalities); 2) an email message (e.g., most municipalities adopt this method); 3) web and mobile reporting applications (e.g., Riyadh Municipality); 4) communicating through social networks such as Twitter and Facebook (e.g., Almadinah Municipality); and 5) WhatsApp messages (e.g., Makkah Municipality).

To fix and clean public infrastructures, the public can make reports to many authorities but each city's municipality is responsible for cleaning up garbage, graffiti etc. from the streets. Moreover, the authority responds to reports about dangerous roads that contain potholes or are in the need of speed bumps. However, these types of reports can also be sent to the traffic police who will arrange a solution. Therefore, the public has the option to send the reports to many authorities, and social network applications could be used for this purpose. There are many reasons to adopt social network applications as they have a powerful impact on our world. For example, they were used (Twitter especially) earlier in the Arab World and the Middle East, in the incident known as the "Arab Spring". Twitter has played a significant role in some of the biggest social and political events in different countries such as Tunisia, Egypt, Libya, Syria and Yemen. The people of these countries used Twitter to communicate with each other and to post pictures and videos encouraging other citizens to join the protest [12] [13]. Because of the great impact that Twitter had on these revolutions, they have been called "Twitter Revolutions" [14]. Regardless of the future of these revolutions, they have clearly shown the important role that social media could play.

In this paper, Twitter was used because of a report published in 2014 [3] which stated that Saudi Arabian users on Twitter represent 40% of all active users in the Arab World. Moreover, a recent survey by Saudi Aramco found that the number of Tweets from Saudi Arabia exceeded 500,000 Tweets daily [15]. Based on these reports, we know that Twitter could be effective if the government used it to analyse citizens' opinions and to study their needs [16].

## **THE PROPOSED SYSTEM ARCHITECTURE**

The proposed system is founded on a Twitter-based reporting system. It can be divided into several parts, depending on the stages that each report goes through, from the beginning when it is extracted, until its publishing. Each stage has its rules and implementations that are applied to the data before they proceed to the next stage. These stages are as follows:

### **EXTRACTING REPORTS**

In this stage, the system extracts all of the data related to reports on public infrastructure in any way. It uses the Twitter search feature to find the Tweets that contain complaints on public services or public infrastructure.

Many tests have been done during this stage to identify the most effective keywords and hashtags that provide the most valuable data. These tests have been applied manually using the commonly used words in reports. The search focused on formal and informal words that indicate semantically reporting issues, such as words that are related to the subject of the problems and negative words that people usually use to express their dissatisfaction.

The system uses the Twitter API to perform the search requests and to obtain the results. To obtain more precise and accurate data, the system refines the result using the advanced search feature, which provides more options and specifications. In advanced search, Twitter enables the client to use some operators and special characters that have special meanings, such as [17]:

- 1) OR operator: to search for Tweets that contain any word (e.g., "potholes" OR "garbage")
- 2) AND operator: to search for Tweets that contain both words (e.g., "potholes" AND "garbage")
- 3) quotation marks: to search for Tweets that contain the exact phrase (e.g., "potholes garbage")

- 4) minus sign: to exclude Tweets that contain some words (e.g., "potholes - garbage" would exclude "garbage" words)
- 5) since and until operators: to search only for Tweets that were sent in a specific period (e.g., "potholes since: 2016-06-01 until: 2016-06-24")
- 6) filter: to filter the result to show only Tweets that contain the filtered option (e.g., "potholes filter: images", to show only Tweets that have images)

The extractions of Tweets are divided based on problem type, and for each type, there are many requests that use special keywords, hashtags and accounts to identify the reports. The main types of problem the system focuses on are damaged roads, where the local authorities need to fix potholes and unsafe roads, install speed bumps, and anything related to transport. The second problem is cleanliness issues such as garbage and graffiti. Lastly, the system looks for other types of reports in an extra part that focuses on general complaints. In this part, it searches for reports that have no specific type or are difficult to classify, such as complaints about corruption or complaints that do not contain enough information for classification.

Therefore, the abbreviations for this analysis are named based on these three types of problems – Fixing, Cleaning and General. In this light, when the system looks for reports from Saudi people, the search keywords are in Arabic, and those keywords are grouped for each type of problem, which is divided into many requests to obtain as many reports as possible. The system also considers common spelling mistakes and commonly used words in both formal and informal reporting as shown at Table 1.

Problem Type	Keyword (English)	Keyword (Arabic)
Fixing	Pothole	حفرة
	Potholes	حفر
	Bump	مطب
	Bumps	مطبات
	Asphalt	أسفلت
	Asphalting	سفلتة
	Sewerage	مجاري
	Light	إنارة
	Road	ريق
	Street	شارع
	Neighborhood	حي
Cleaning	Garbage, Rubbish, Waste	قمامة
	Trash	نفايات
	Municipality	بلدية
	Cleanness	نظافة
	Cleaning	تنظيف
	Road	ريق
	Street	شارع
Neighborhood	حي	
General	Derelection	قصير
	Disregard, Overlooking	إهمال
	Municipality	بلدية

Table 1: Keywords used.

Twitter search provides good support for the Arabic language; it ignores many spelling mistakes and returns the correct result even if the user wrote the words incorrectly. This is

because in the Arabic language there are similar letters that people often write incorrectly, as shown in Table 2.

Similar Letters	أ	ا	آ	إ
	ي	ى	ئ	
	و	ؤ		
	ه	ة		

Table 2: Similar Arabic Letters.

If the user replaces any of these letters with other, Twitter would assume that he means the right one. In addition, it automatically corrects mistakes in Arabic diacritics, which usually represent the word in many forms, as shown in Table 3.

Arabic Diacritics	◌َ	◌ِ	◌ُ	◌ْ
	◌ِ	◌ِ	◌ِ	◌ِ

Table 3: Arabic Diacritics[18].

This support helps to increase the number of Tweets in the search result by covering all possible forms of letters and words. Before making the request through the API, the system prepares the final form of the search phrases, using keywords with special search characters and operators to improve the result. The phrases are grouped depending on the number of results that they have. Table 4 illustrates the groups of phrases that are used in requests:

Problem Type	Request No.	Search Phrase (English)	Search Phrase (Arabic)
Fixing	#1	Bump OR Bumps OR Asphalting OR Asphalt Street OR Road OR Neighborhood - filter: retweets	مطب OR مطبات OR سفلتة OR أسفلت شارع OR ريق □ OR حي - filter: retweets
	#2	Potholes OR Pothole OR Sewerage OR Light Street OR Road OR Neighborhood - filter: retweets	إنارة OR مجاري OR حفر OR حفرة OR ريق □ OR حي - filter:retweets
Cleaning	#1	Garbage OR Trash Street OR Road OR Neighborhood -filter: retweets	ريق □ OR نفايات شارع OR قمامة OR حي - filter: retweets
	#2	Municipality Cleanness OR Cleaning -filter: retweets	تظيف □ بلدية نظافة filter: retweets
General	#1	Dereliction OR Disregard Municipality - filter: retweets	إهمال بلدية OR قصير - أمانة filter: retweets

Table 4: Search Phrases.

The system classifies the problem types and arranges the API requests for each of the search phrases. When the system is used for the first time, the API fetches the most recent 100 reports from Twitter for each request, without specifying the dates when the Tweets were posted.

The maximum limit specified by Twitter as a result for each request is 100 Tweets. After fetching the first set of Tweets, the system saves and records the date in the database to be used in the next extracting process.

In Twitter, there are three types of search results [19]: 1) popular: returns only the most popular results in the response; 2) recent: returns only the most recent and real-time results in the response; 3) mixed: returns both popular and recent results in the response. To return the largest possible number of reports, the system requests the search using the recent result type. The popular search results only present Tweets that show these characteristics: having many retweets and many likes, and sent by a highly regarded user. However, in this case, the researcher needs to obtain reports from all people, even those who have no followers or are not famous. This is because many reports received by the system are usually made by ordinary people who complain about the problems of their daily lives. At this stage, not all of the Tweets collected by the system contain real reports, as many of them could be normal Tweets that contain only some of the words in the search phrase. Therefore, the system would save all of the reports that have been extracted in this stage into the database, to analyse them later and to exclude Tweets not related to the reporting of issues. This will be explained in the next section.

## **ANALYZING REPORTS**

Before the aspects of report analysis are discussed, it is important to clarify some issues. First, it is important to note that the system deals with numerous reports that contain texts and in the case of this study, these texts consist of many sentences and words written in the Arabic language. Deep searching in these data to extract a valuable result that talks about a certain topic is not a simple process, as to analyse these results, the system needs to use many NLP (Natural Languages Processing) related tools and techniques such as data mining, artificial intelligence and others. These need to be considered as future work. However, there is a special field of study that is dedicated to such subjects, called Text Mining (also known as Text Analytics).

## **TEXT MINING**

Text mining is a field of research that analyses unstructured text to identify patterns. Unlike search engines, which aim to facilitate information retrieval by looking for keywords, text mining aims to extract values from a text through operations. These operations include lexical analysis, sentence restructuring, parsing and deriving linguistic characteristics [20].

Technological advances in NLP and Machine Learning in recent decades have contributed to building the foundations of text mining. Previously, it was expected that this topic would play a significant role in solving business problems; however, the significant increase of textual contents on different platforms such as Wikipedia, Forums, Blogs, and Social networks like Facebook and Twitter, has provided an interesting opportunity to investigate many other areas where text mining could be applied. These include article summarising, security, auto-replying to messages and marketing. Article summarising refers to reading an article thoroughly and summarising it into fewer words or sentences. The final result depends on the technique that is used to summarise the article. In this light, extractive summarisation extracts parts from the original text, while abstractive summarisation includes new parts that were not in the original text. With regard to security, it is possible to monitor social network activities and inform about suspicious messages. Messages can be responded to automatically, depending on the

understanding of their content, and text mining can also be used in marketing as commercial companies focus on social networks to influence people – they need to know what people like and what they need.

There is a feature in text mining that is related to the concept of this proposed system. This feature is called ‘Sentiment Analysis’. The system searches a massive number of Tweets to identify citizens' complaints about public infrastructure. In this regard, the classic search method is insufficient as it depends on finding keywords and hashtags, and the result contains Tweets with the specified keywords that are not complaints, such as news reports. Thus, the system should find a method to differentiate between complaints and other types of reports. In this regard, it is important to note that making a complaint is an action that expresses a negative attitude of dissatisfaction. Hence, to know whether the Tweet is a complaint, the system needs to use the sentiment analysis techniques.

### SENTIMENT ANALYSIS

Sentiment analysis, or also known as opinion mining, is the process of using NLP, text analysis and machine learning to determine the feelings and attitudes of the writer [21]. In general, sentiment analysis applications can adopt one of two approaches. The first one is most used approach, which depends on linguistic resources. The other one works through machine learning techniques. The idea of machine learning approach is to provide a set of pre-defined texts (called corpus) for the computer to use as models to predict new text. Consequently, the more data that computer should learn from, the more accurate result it gives.

Due to the limitation of the pre-defined corpora for the Arabic language [22], the system would use the first approach, which depends on linguistic resources. This approach works by manually labelling words as positive or negative. Consequently, sentences that contain many positive words are considered as positive sentences, and those containing negative words are considered as negative. Sentences that do not contain positive or negative words are considered as neutral. Advanced algorithms set the values for words (e.g. "bad" has less negative value than "horrible"), and work with probabilities in that the sentence composed of combination of positive and negative words, as shown an example in Table 5. Moreover, Emoji plays a significant role in the sentiment analysis as many used them to express their feelings. Therefore, even Emoji and Emoticons signify positive or negative (see Table 6). However, some sentences are difficult to classify, as shown an example in Table 7.

Sentence	Words	Sentiment analysis
I like the Department of Electronics and Computer Science.	Like	Positive
The work was too hard.	Hard	Negative
This is my book.	-	Neutral

Table 5: Examples of Sentiment Analysis[20].

Type	Emoji
	<3
	:)



Positive	:~)
	(:
	(~:
	:-D
	:D
	:p
	:-p
	:>
	^_^
Negative	:S
	:(
	:~(
	:!(
	:!~(
	:~<
	:<
	:/
	>:(
	:S
	:s
	:!s

Table 6: List of Emoji[23].

Type	Sentence
Mixed Opinions	The new iPad is beautiful but expensive.
Metaphor	I want a coffee so bad.
Sarcasm	"Some cause happiness wherever they go; others whenever they go." - Oscar Wilde

Table 7: Examples of sentences that difficult to classify[20].

From Table 7, it is apparent that there are some types of sentences that are not easy to determine whether they represent positive or negative opinions. In the meantime, people share their actual beliefs and opinions clearly, and engage in discussions about different topics freely in social networks. Therefore, governments, organisations and companies will benefit from analysing texts people sent in social networks. However, there are some challenges in analysing social networks, first, its dynamic content as well as the fact that users of social networks often use informal terms and expressions which creates new difficulties for the implementations and create inaccurate results.

Arabic language is the largest member of the Semitic family, and it is usually classified into three forms, MSA (Modern Standard Arabic), CA (Classic Arabic), and DA (Dialectical Arabic). Some researchers claimed that Arabic textual content on the internet can be divided into a few categories [22]; contents are classified based on its resources to chat, Twitter, Web forums, and Wikipedia Talk. It is agreed that chat contents are usually conveyed in DA, while

other resources are presented using a mix of MSA and DA. Both two categories have their own features, which make computations harder. DA also lacks a standard form there are a few different dialects used. Moreover, MSA and DA are used ambiguously, such as for sarcasm, or composed of multi-lingual content (e.g. Arabic and English). As a result, since the system deals with Twitter, the study considered both MSA and DA.

Consequently, the task is harder in this system, because Arabic language is morphologically and grammatically complicated [24]. Moreover, the number of studies on algorithms and applications for sentiment analysis in Arabic are limited, compared to those in English. Furthermore, as mentioned, NLP's operations such as parsing and lexical analysis are customised to be for a certain language, through the understanding of its characteristics, which might affect the process in Arabic.

## **THE ADOPTED APPROACH FOR TEXT ANALYSIS**

At this stage, the system database collected many Tweets during the data extracting process, however, these tweets looked messy and difficult to classify. Therefore, some analysis and ranking process would be conducted to help identify the Tweets that system needs.

To do this, the system uses an algorithm that works on each Tweet to rank them according to the probability of the Tweets being a real report. This algorithm applies a lexical analysis on the Tweets by looking for a set of words in it, hence, when a Tweet contains some of these words, its rank value will increase. In this regard, the ranking keywords consist of Arabic words usually used for complaining about something, or to express a dissatisfaction.

The ranking keywords contain both formal and informal words. These keywords are focused on common words that Saudi people use. However, to obtain better results, the system should use many more keywords to help to increase accuracy. On the other hand, in the Arabic language, it is difficult to define all forms of words because they are too many words and each word can have many forms, for example, in English, the word "report" could be written in many forms such as "report", "reports", "reporting" and "reported". Therefore, the system uses the lexical roots and stems (the primary part in the words that always remains after removing all affixes) for these words. In addition, for some keywords the root word is used to reduce the number of keywords. Consequently, this guarantees that the system covers almost all forms of words that people use in their reporting. Table 8 shows samples of the ranking keywords and its translations after self some analysis on Twitter's search result.

Keyword (English)	Keyword (Arabic)
Need For: We need, I need, they need, etc...	حِتاَج For: يحتاج، نحتاج، محتاج، محتاجة، محتاجين
Hope For: We hope, I hope, etc...	مَنى For: نتمنى، يتمنى، أتمنى
Wish For: We wish, I wish, etc...	رَجو For: نرجو، أرجو، يرجو
Reporting (v) For: We reported, I reported, He reported, etc...	بَلغ For: نبليغ، يبلغ، أبلغ
Report	بلاغ
Complaint	شكوى
We demand	نطالب
They demand	يطالبون
We want	نريد
We have	عندنا، لدينا
We (it is not like the part of previous phrases in Arabic)	نحن
When	متى
Since	منذ
Disregard	إهمال
Not done	لم يتم
There is, there are	يوجد
Dangerous	خطير، يهدد
Safe, Safety	سلامة

Table 8: Examples of English and Arabic ranking keywords.

As observed, there is no need to use a complicated stemming algorithm to perform this task, as the use of stems of common words are sufficient. These keywords were selected based on an analysis on the Tweets collected. These keywords are those which appeared frequently on numerous Tweets that are actual reports. Since all of the collected Tweets contain words related to public infrastructures problems, such as 'Potholes' 'Bumps', and 'Trash', the system was able to analyse them and extract real complaints.

After making a list of these words, the system searches for each keyword separately in the Tweets, whenever a similarity is found on a certain Tweet, this gradually increase its rank value. At the end of this process, all Tweets will have rank values, starting from zero, which represents the lowest probability of a Tweet being a real report, to the maximum number. This value equal the number of keywords in the list, and this indicates to the highest probability of the Tweet being a real report, as shown in Figure 1 and Table 9.

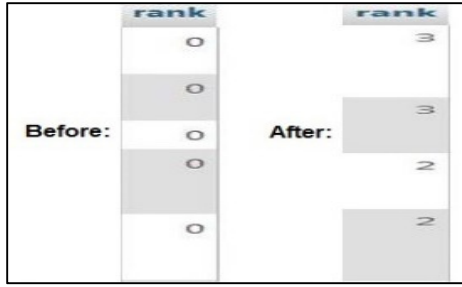


Figure 1: A sample from the database for reports' ranks before the analyzing process and after.

Tweet (English)	Tweet (Arabic)	Keywords (English)	Keywords (Arabic)	Rank
We are the people of Alrawdah neighbourhood; we need bumps at the entrance of the road because there is a dangerous slop.	نحن سكان حي روضه يا حلقة شرفيه نحتاج اي مطبات صناعيه يوجد منحدر ريق داخل اي سكني يهدد حياة ساكنيه	We. Need. There is. Dangerous.	نحن. نحتاج. يوجد. يهدد.	4
We have a big street and there people drifting in the neighbourhood, accidents happened many times. We really need a bump, we reported and no one respond	و سمحتم عندنا شارع كبير يا حي فحيطات دايم واكثر من مره حوادث وسيارة اخوي واقفه صدموها نحتاج مطب ضروري بلغنا مارود علينا	We have. Need. Reported	عندنا. نحتاج. بلغ.	3
There is no Asphalt in the road of Madinah towards the Municipality street since more than 15 days; in addition, there are potholes.	ريق يا مدينة خدمات يا جاه شارع بلدية لايوجد فيه زفت منذ اكثر من ١٥ يوم وزحمة خانقة وفوق ك حفرة وار فاع فتحات تصريف عن الارض	There is. Since.	يوجد. منذ.	2

Table 9 : Samples of real Tweets and their rank value.

After finishing the analysis and ranking, the system can use the rank values to show Tweets that more likely to be real complaints at the beginning, to facilitate manual classification in the next stage. To illustrate the difficulty of classifying Tweets without the ranking process,

the below tables show Tweets found in the first result of the ranking process, where Table 10 presents Tweets that were extracted without considering their rank values, while Table 11 presents Tweets that were extracted by considering their rank values.

#	Tweet (English)	Tweet (Arabic)	Real Complaint?
1	In a quick response to what we have published, Tabuk's municipality removes the trash in Farhodiah road...	في فاعل سريع مع ما نشرناه.. أمانة بوك زيل نفايات ريق فرهودية: أزلات أمانة بوك نفايات...	No (News)
2	Tomorrow Mohammed Abdu says that the road of Talal Maddah is full of potholes and bumps. #street_of_Talal_Maddah	بكره يطلع محمد عبده ويقول شارع لال ما يصلح كله حفر ومطبات. شارع لال مداح#	No (Joke)
3	Do we wait for a gruesome accident to do something about Alrwaidah road?	هل ننتظر حادثا شنيعا يتحرك مسؤول عن #حفر موت في #ريق رويضة سنام	Yes

Table 10 : Samples of Tweets that appeared at the beginning without considering the ranking.

#	Tweet (English)	Tweet (Arabic)	Real Complaint?
1	I hope that you add more waste containers and pay attention to the cleanliness of Alrayan neighborhood in Barqa street, thank you.	ارجو من منكم زيدون في حاوية نفايات والاهتمام في نظافة شارع.. مع جزير شكر.. حي ريان شارع برقة.	Yes
2	The street of Khalid bin Alwaleed in Riyadh need re-asphalting in both sides. The street is terrible and full of potholes...	شارع خالد بن ابيد بالرياض يحتاج إعادة سفلتة في الإجهين.. وضع شارع مزري كله حفر. نرجو الاهتمام	Yes
3	We are the people of a neighborhood next to street no 24, we are suffering from over speed and the lack of bumps...	نحن سكان حي بجوار شارع ٢٤ ونعاني سرعة وعدم وجود مطبات صناعيه وهو ملتقى شارعين ٢٤ و ٢٠	Yes

Table 11: Samples of tweets that appeared at the beginning of the process when the ranking value was considered.

The analysis and ranking processes have shown significant results where clear majority of high ranked Tweets were identified as real complaints from citizens. Using these results, the system would be able to easily display the Tweets collected and sort them according to the rank value, in order to facilitate its classification.

## CLASSIFYING REPORTS

Prior to this process, the system had obtained a set of reports related to different problem types from several locations. Consequently, there is a need to these reports depending on the responsible parties so that it can be successfully delivered to the right responsible. As previously

mentioned about the reporting methods in Saudi Arabia, there are many authorities that can deal with the reports. However, the system should select the most relevant authority to forward the reports to. In this regard, there are two specific types of problems that the system deal with, which are problems regarding roads such as potholes, bumps and road safety, as well as issues on cleanliness, such as garbage, graffiti, etc.

The general department of traffic in Saudi Arabia, which is a part of the ministry of interior, deals with reports related to traffic. Among these problems are potholes and roads safety. The municipalities of each area also receive this kind of reports. In the meantime, Saudi Arabia is divided into 13 regions, where there are municipalities that deal with the complaints about problems that occur in each area. Table 12 shows the regions of Saudi Arabia and their municipalities.

Regions' Municipalities	Riyadh Municipality
	Makkah Municipality
	Madinah Municipality
	Qassim Municipality
	Tabuk Municipality
	Northern Borders Municipality
	Jawf Municipality
	Eastern Province Municipality
	Hail Municipality
	Jizan Municipality
	Najran Municipality
	Asir Municipality
	Bahah Municipality

Table 12: Saudi Arabia's Municipalities[25].

Local municipalities are considered better than other authorities because of their many features such as: 1) The ability to deal with many types of problems: this minimises the number of authorities that the system reports to, and will overcome the confusion of which authority should be selected; 2) Depends on the location: which could be known from the report content, to use it to report to the correct municipality; 3) More Responsive: municipalities receive a big amount of reports from different sources, and deal with it in a short time. In this light, the local municipality is deemed as the best authority to forward the reports to. Furthermore, all Saudi Arabia's municipalities provide an email address dedicated to receiving different types of reports. Therefore, the system uses these emails to send the report that containing all the problem's details in the next stage.

The big challenge is how to know the location of problem in the report. The system management team would be able to recognise the right municipality from the information in the reports. Usually, reports contain names of regions, cities, places or streets. It is also possible that some Tweets have not enough information that tell where it should be send. However, the clear majority of the available reports contain detail about problem's location, and the analysis of the collected reports showed that there are no reports that provide no information about the location of the problem. It is also possible that some reports contain names of locations that are not well known. In this case, if it is possible to identify the name of the district or regions of this location,

the team can forward the report to the municipality of this area who has more knowledge about its places and will be able to recognise it.

One of the suggested ideas to identify the correct location of the problem is to extract the geographical location of the Tweet (report). However, this idea has many disadvantages, and would not be effective to know the problem's location. Moreover, many people who sent the reports did not send them at the same time and right at the time they encounter the problem; it is common that people take photos of the problem and then post a report after they got home or reached their workplace. In this case, Twitter will geotag the post based on the user's location when they post the Tweets. Therefore, using Tweet's location is not a good idea. Nevertheless, the amount of reports that do not have a location information are very few and not worthy for analysis.

## SENDING REPORTS

The system sends the report to the most relevant authority identified. The system has the authorities' email addresses in its database that are officially obtained from each authority. Therefore, the email addresses are used to forward the reports to them. The email messages sent by the system contain the report content obtained from Twitter, in addition to some other information such as the date of reporting, pictures attached, etc. For each report, the system provides each municipality a secret key that enables them to access and manage the report. They can update report's status whenever they fix it. Otherwise, they can close the report if they think it is not under their jurisdictions, or forward it to the relevant authorities.

In the popular reporting systems, such as FixMyStreet and SeeClickFix, there are three statuses for reports sent as shown in Figure 2. These statuses are: 1) Open; 2) Fixed; 3) Closed. The 'open' status indicates that the responsible still have not provide any response to the report, reports with this status might need a long time be processed by the relevant authority. At the same time, the system displays the date when the report was sent to the municipality, hence, public will know whether that the municipality received the reports and time needed to process each report. In the meantime, the second status, which is 'fixed', indicates reports that have been processed and that the problems had been fixed. The relevant municipality can provide more details about the processing operation and the citizens would be able to view it in the system homepage. Lastly, the 'closed' status refer to reports closed by the authority. These reports were processed but not fixed because of reasons such as an unclear description for the problem, or the problem is not under the jurisdiction of the authority.

The authorities can describe the status that they selected by adding a comment that can be viewed by users. The municipality can forward the report to the correct authority if they think are not responsible for the problem. Consequently, they can inform citizens about what actions taken and who they have forwarded the report to in the comment area. This report status will be recorded as 'open' until the other responsible authority take actions and change the status. The municipality can easily provide the other party the secret key to access the report's management page. Otherwise, they can provide a link that enables the other party to directly control the report. Lastly, any report that contains inappropriate content such as bad words, denigration or privacy violation of people, house addresses, or car plates will not be accepted.

Figure 2: Report's Management Page.

## PUBLISHING REPORTS

Publishing process works in three different areas: 1) Main page: as a text for the content of the report to enable citizens to view, comment, and vote; 2) Visualized data: as a visualization on the map to represent all reports in Saudi Arabia classified by the municipality; 3) Open Data: as an open format file for all data in the system (see Appendix A and Appendix B).

## MAIN PAGE

After sending a report to the municipalities, it would be published directly on the main page of the system so that it is available for public views. Such display will help to achieve transparency, which is one of main goals of the open government. By default, the status of all reports is 'open' when they posted, and then, the responsible authority can change it and add new information depending on what they did with the report. The status of the report will not change until the responsible authority changes it.

For all the published report, the system would show responsible authority and update the status of the report. In addition, there is an optional comment section where the responsible authority can add to the report's status to explain the actions taken. Therefore, through these reports, people can monitor how authorities work and which of them are active. Moreover, the published report contains other details such as images attached to the Tweet, as shown in Figure 3.

Figure 3: One of published reports on the main page.



To increase public contribution, the system enables them to make comments to discuss about the problem and vote on the published reports to express their satisfaction as shown in Figure 4. This voting feature was inspired by SeeClickFix, which enables its users to vote on the reports, unlike FixMyStreet that does not. To protect the privacy of posters, the system would not show their details (names and emails) on their comments. However, the manager of the system can delete any inappropriate comments.

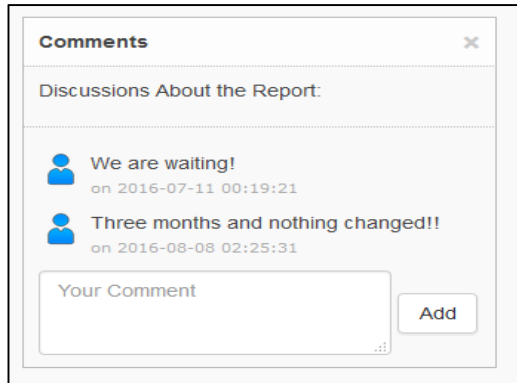


Figure 4: Report's Comment Box.

## VISUALIZED DATA

Visualisation delivers the data easily and effectively through representing the data on a picture. The system visualises all reports and represent it on the map of Saudi Arabia, after collecting all reports for each municipality. It uses Google map to generate a dynamic and clear map of Saudi Arabia, then retrieve the data from the database and perform some computations using D3.js to prepare for the representations on the map.

To provide the users an overview about each region in Saudi Arabia and the amounts of reports sent from each area the system displays the number of reports on the location of the municipality of each region, as shown in Figure 5. This number of reports includes reports in all statuses, whether open, fixed, or closed. This is the first picture that user sees, then he can select any of these municipalities to view its details and reports' statuses. Through visualisation, the users will be able to know about public infrastructures reports in Saudi Arabia simply and quickly. Furthermore, the users would be informed about municipalities' interaction and which of them work well.

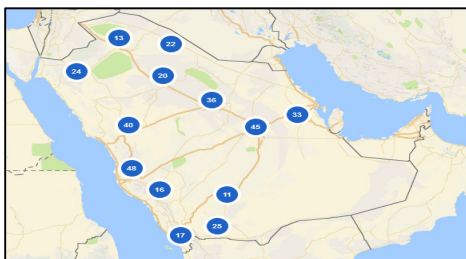


Figure 5: Map visualization for all reports in Saudi Arabia.

**OPEN DATA**

Publishing open data empowers a country by giving its citizens the ability to use the data for researches, statistics, and developing innovative application that help to make their lives easier. Moreover, the use of data and representing it on other applications increases the transparency in the country.

The system has adopted the eight principles of open data publishing, as previously mentioned [11]: 1) Complete: All reports that available in the database are published; 2) Primary: no aggregations or modifications on the published reports; 3)Timely: The system publishes a real-time data; 4) Accessible: Available for all people; 5) Machine-readable: The system publishes the data on CSV format, which is a machine-readable; 6) Non-discriminatory: Available on the main page and no need for registration; 7) Non-proprietary: no exclusive control for any report; 8)License-free: no copyright. Thus, people can benefit from these open data in general and from the proposed system in two points. The first point is innovative applications which provide a weekly, monthly or yearly report about best and worst municipalities in Saudi Arabia in term of performance. Also, they provides a list of unfixed problems for each municipality or for their achievements. The second point is in researches manner which find out what people want, explore what types of problems are more in a specific region, study what are the challenges for municipalities, and detect what problems are difficult to solve.

**THE PROPOSED SYSTEM DESIGN**

**Use Case Diagram**

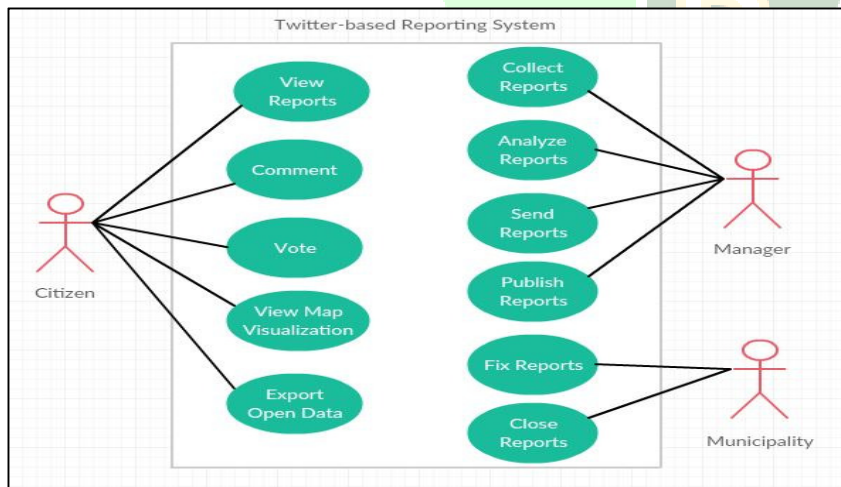


Figure 6: Use Case Diagram.

Generally, there are three types of users for the system, manager, municipalities and citizens, as shown in Figure 6. Each of these three has its own features that the system provides for, and all of them can use the public functions.

## Manager

The manager is responsible for using the system's control panel to perform tasks like:

- Collect reports from Twitter.
- Analyse and rank the collected reports.
- Send or forward reports to the municipality.
- Publish reports to enable public to view it in the main page of the system.

In addition, the manager can use features that available for citizens (the public). It is not necessarily that the manager is one person, the management might be controlled by a team who works permanently on its tasks.

## Municipalities

As mentioned before, the selected municipalities are those responsible in dealing with the problem. Other agencies, such as organizations, companies or institutes could be added whenever the needed by the system

The municipalities are responsible for handling reports that the managers send to them. They can view reports containing the complaint about public infrastructure that happened somewhere in the locality of each municipality. These problems are divided into three types, either something to fix, something to clean, or general problems that usually about neglects such as abandoned projects and those which take a long time to finish.

Whenever a system manager sends a report to the municipality, the report is published for the viewing of all citizens. At the same time, the municipality receives a key with the report, which grants them a permission to access and control the report.

By using the key or by just by following the private link that they receive in the email, they have access to a private control panel in the system, which enables the municipality to:

- View the report
- View picture of the problem that user's sent.
- View citizens' comments and dissections about the problem.
- Forward report to another responsible.
- Change report status to "Fixed".
- Change report status to "Closed".
- Add more details about the report status, to inform citizens why it was handled like this.
- Like other users, municipality does not need special permissions to access the public services such as to view all the reports.

## Citizens

Citizens make up the majority of users who are able to use public services of the system without any special permissions or registration. Most of system's services are catered to fulfil the citizens' needs, and they are the most important users in the system. In this regard, the citizens can interact with the system and use features, such as:

- View published reports, its status, and to which responsible this report was sent.

- View reports' pictures.
- Comment and view others' comments.
- Vote and view votes on all reports.
- View a map visualisation for the problem in Saudi Arabia's regions.
- Export an open data for all reports without any modification or aggregation.

## DATA MODEL DIAGRAM

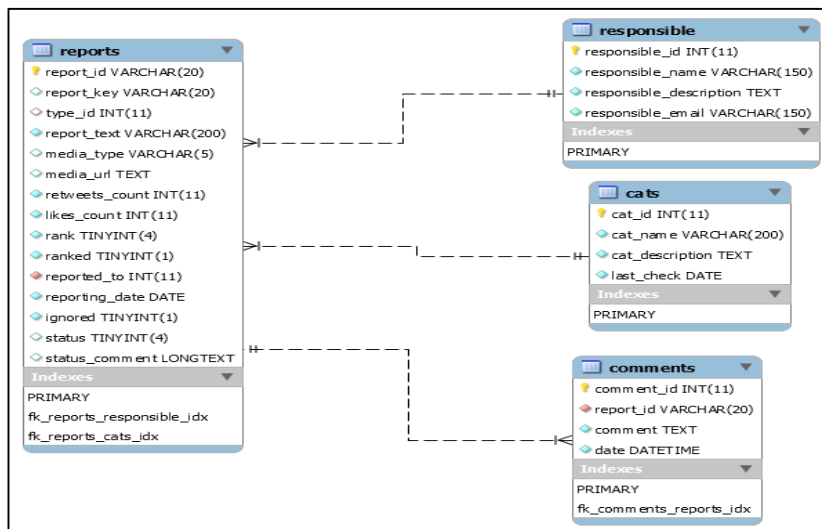


Figure 7: Database structure.

### Reports' table

Reports' table contains the details of reports collected from Twitter. It is the main table that is related with all the other tables in the system. The fields for the reports' table are:  
**Report ID:** a unique value originally brought from Tweet's object. It holds an integer value that represents Tweet's identifier in Twitter.

**Report Key:** the secret key given to the municipalities to grant them an access to view and modify reports that sent to them. The value of this field generated when the report send to the municipality, and it contains 20 random digits consist of a mixture of letters and numbers.

**Type ID:** a foreign key related to cats' table, and it has a value equals to one of categories' IDs which represent the type of problem in the report, such as 1 for fixing, 2 for cleaning, and so on.

**Report Text:** the actual content of the report sent by Twitter users. Usually, it consists of sentences and hashtags that written in Arabic.

**Media Type:** the media type is a variable in the Tweet's object, and it has some optional values, such as pictures and videos. This field helps the system to recognize the type of the media, to show it to the users. Not all reports contain media; however, this field would be empty if the report does not contain any media.

**Media URL:** the actual link to the media's resource, the system uses this URL to present the media whether it is a picture or a video. The resource file usually retrieved from Twitter's server.

**Retweets Count:** a variable in the Tweet's object and it represents the number of users who 'retweeted' this Tweet (report). The number of retweets imply the importance of the report based on how many other users are interested in it.

**Likes Count:** a variable in the Tweet's object represents the number of users who liked the Tweet (report). The number of likes imply the importance of the report where there are many users interested in it.

**Rank:** contains the result of the ranking process. The value is different for each report. The minimum value is zero which indicates that the content is not likely to be a real report.

**Ranked:** holds a Boolean value for each report, which states whether the report has been ranked (the ranking process applied on it).

**Report to:** a foreign key related to the responsible table, its value is empty by default and placed when the report sent to the municipality. The value could be one of "responsible\_id" values in responsible' table.

**Reporting Date:** is the date when the report was sent to the municipality. It holds a value of type 'date' which is filled after forwarding the report.

**Ignored:** a Boolean value that determines if the report should be displayed in the control panel for the manager. By default, no report is ignored; however, the manager can ignore certain reports and prevent them from appearing, which helps to bring focus to other more significant reports.

**Status:** determines the status of the report, either open, fixed, or close. All reports are open by default and it changed whenever the municipality modified it. The data type of the field is an integer, and it represents the status as follows: open is zero, fixed is one, and closed is two. **Status Comment:** this is for the details that municipality provides to explain the action they made. Its content could be a long text to enable the municipality to write as much as they need. However, it is optional for the municipality to add a comment reports they process.

### **Categories' table**

The categories' table contains all categories that report could be classified in, such as fixing and cleaning. The table contains three main categories, which is fixing, cleaning and general. Every report must belong to only one category. The table's fields are:

**Category ID:** a unique integer identifier for the category that generated incrementally by the DBMS. It also related to the 'type\_id' field in reports' table.

**Category name:** a label for the category to be displayed for users, such as "Fixing Problems".

**Category Description:** an optional field that could be used to describe the category and its contents.

**Last Check:** the date that represents the last time the system collected the reports for this category from Twitter. The system uses this value each time he receives an order to collect new reports. It adds this date to the API's request to ask Twitter not to return old reports.

### **Responsible authorities' table**

Responsible authorities' table contains the details of all responsible authorities, such as government agencies, companies or any group that deal with public infrastructures' problems. The system uses this table to provide the manager with the list of options of responsible

authorities to select where report should be sent to. Initially, the system has a list of 13 municipalities across Saudi Arabia. Responsible authority fields include:

**Responsible ID:** a unique integer identifier for the responsible that generated incrementally by the DBMS. It also related to the 'reported\_to' field in reports' table.

**Responsible name:** a label for the responsible to be displayed to the manager when sending the report, such as "The Municipality of Riyadh".

**Responsible Description:** an optional field that could be used to describe the responsible, its responsibility, and the area that he covers.

**Responsible Email:** contains all the email addresses that the system uses to contact the responsible authorities.

### **Comments' table**

The comments' table contains all the comments that users leave on reports. Users can add comments when the report published, regardless of its status. Table's fields are:

**Comment ID:** a unique integer identifier for the comment that generated incrementally by the DBMS.

**Report ID:** related to the 'report\_id' field in reports' table, and it represents the ID of the report to which this comment belongs.

**Comment Text:** holds a value of a text type, which is the comment that user sent.

**Date:** a value of a 'DateTime' format that represents the time that the comment sent. It is displayed with all comments to show how old it is. Comment's date is important to monitor the statuses of reports, and what the municipalities did about it.

## **EVALUATION**

### **Accuracy of Reports**

Since the reports collected from a social network where people talk about different topics, it is possible that some of them are not real reports. However, the analysing process attempted to increase the accuracy of finding out the real reports through text analysis's implementations, and provided a rank for each report. To evaluate this process, Table 13 shows a statistic for 220 reports that system collected and analysed for one day. Reports of the rank zero are reports that system did not rank. Table 14 illustrates the difference in accuracy for ranked and unranked reports. As clearly appears, the accuracy of ranked reports is high comparing to the unranked, which shows the effectiveness of the analysis implementations. The accuracy of unranked reports raised because the original keywords that used in the search process to extract these reports are precise and related to reporting issues.

Rank	#of Reports	Real Reports	Unreal Reports	Accuracy
5	1	1	0	100%
4	1	1	0	100%
3	2	2	0	100%
2	14	14	0	100%
1	50	41	9	82%
0	152	80	72	52%
Total	220	139	81	63%

Table 13: Accuracy of All Reports.

	# of Reports	Real Reports	Unreal Reports	Accuracy
Ranked Reports	68	59	9	86%
Unranked Reports	152	80	72	52%

Table 14: The Difference between Accuracy of Ranked and Unranked Reports.

## IMPROVING TRANSPARENCY

One of system's aims is to improve the transparency. Thus, the system provides many services that help to achieve this task, such as:

- Publishing timely reports with the responsible name.
- Enable citizens to express their opinions and satisfaction about the reports.
- Facilitate reports' viewing through data visualisation.
- Providing an original and complete data about all reports in the system.

Through these services, citizens can monitor the work of municipalities, and deliver their messages.

## OTHER FEATURES

Comparing to other popular reporting applications such as FixMyStreet and SeeClickFix, the system provides most of features they have. Table 15 illustrates some of these features.

Feature	FixMyStreet	SeeClickFix	Twitter-based System
User Interface	Responsive Web Design	Responsive Web Design	Responsive Web Design
Reporting Method	Filling in a Form	Filling in a Form	Collecting reports from Twitter
Reporting to Authority Through	Email	Email and integration with authority's system	Email
Voting on Reports	Not available	Available	Available
Visualization on Map	Available	Available	Available
Exporting an Open Data	Not Available	Available	Available

Table 15: Features of The System and Other Reporting Applications[26].

As appears from Table 15, the system has many features that not available in other applications; also, there are some features not in the system such as the integration with authorities' systems to add the report directly, which available in SeeClickFix.

## GRAPHICAL USER INTERFACE (GUI)

### Main Page

The main page contains all published reports and enables users to access other features from the navigation bar at the top of the page. As shown in Figure 8, each report has its own details and options, such as displaying pictures, comments, and a voting button that enables the users the vote on issue. The Main page takes the user to other pages in the system, like:

Control Panel: only accessible to those with administrative/ managerial permissions.

Statistics1: A page that presents the visualisation of the data and shows the reports for each municipality on the map.

Export CSV: Provides access to the open data which enables the users to download a CSV file for all stored data in the system.



Figure 8: System's Main Page and Navigation Bar.

### Control Panel

The control panel is only available for the manager so that he can control the system and its contents. It offers many features for permitted party as shown in Figure 9. In this regard, this page is used to control each stages of the reporting which ranges from report collecting to report publishing.



Figure 9: Control Panel's Navigation Bar.

<sup>1</sup> People who do not have a technical background are more familiar with "Statistics" and "Export CSV" than "Data Visualization" and "Open Data".



## Extract Reports

As shown in Figure 10, when the manager clicks on "Extract All Reports" button, the system connects to the resources for each category and collect the reports. Subsequently, the page asks the manager to classify the collected reports. In addition, the page shows the last date of extraction for each category.



Figure 10: Reports' Extracting Page.

## Analyzing Reports

In this page, the manager can analyze and rank all reports in the system. Moreover, the manager can see the details for the reports coming into the system. Once the manager clicks on the button above, the system will perform the analyzing process and then ask the manager to move to the next step, as shown in Figure 11.



Figure 11: Reports' Analysing Page.

## Send Reports

On this page, the manager can select the reports that should be sent to the municipality and those which should be ignored. All reports will appear here in the order of their rank, where the first results will usually contain real reports that have to be sent (see Figure 12). If the manager wants to ignore all of the reports, he can easily click on the "Ignore all button" for each category. When the manager clicks on the forward icon, a new window will appear for selecting the responsible authority (see Figure 13).



Figure 12: Reports' Sending and Publishing Page.



Figure 13: Reports' Sending Window.

This window contains a dropdown menu for all authorities in the system. Immediately after sending the report, the information will be published on the main page to notify the users that the report has been sent, and it would be in "open" status (see Figure 14). The municipalities can keep tab of reports sent and read through the report management page. The municipalities can access this page through the link they received in the email.



Figure 14: Report Management Page.

## STATISTICS

This page contains a visualization for all reports on the map.

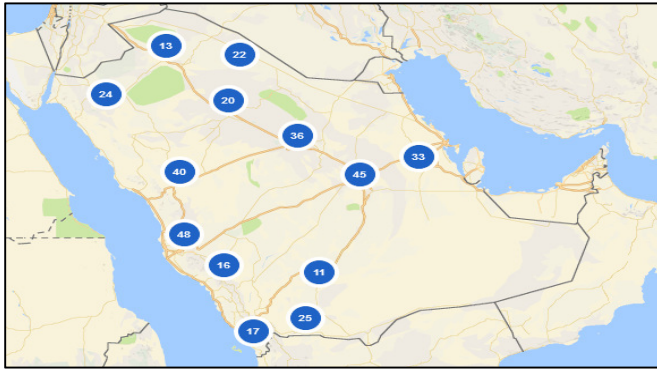


Figure 15: Statistics Page.

## CONCLUSION AND FUTURE WORK

The government needs to put constant care of public infrastructure to provide good services. Thus, reporting applications plays a significant role in engaging citizens in the development process. However, for some countries like Saudi Arabia, normal reporting applications did not work effectively in facilitating communication between the government and citizens. On the other hand, social networks contain huge amount of complaints that could be used to extract reports about public infrastructure to inform the government about such problems. In this regard, due to its openness and features such as featured search, as well as its popularity in Saudi Arabia, Twitter was selected as the main resource for these reports. The reports were collected from Twitter through pre-defined and chosen keywords that help to cover as many reports as possible. Furthermore, a sentiment analysis's process was applied on these collected data to identify real complaints. After the analysis, it was easier to find the real reports and forward them to the responsible parties. In the meantime, the system was based on local municipalities because they are the one dealing with the public infrastructure's problems.

Inherently, it forwards the reports of each region in Saudi Arabia for the municipality of this region. To increase the transparency in the country, reports were published as soon as they were sent to each municipality. This enables all citizens to view the reports, discuss about them, vote and express their satisfaction. Consequently, open government strives to achieve the goal of transparency, therefore, providing open data on collected reports will support transparency by displaying original data to the citizens so that they can be viewed or used for innovative projects. Lastly, to facilitate the exploration of reports and provide an overview for public infrastructure's reports, data visualisations were made to represent reports of each municipality on the map.

For future works, reports can be collected from resources other than Twitter, such as Facebook, Instagram, or even online newspapers by looking for news about public services and searching the comments section to find complaints from the public. Moreover, the system can cover more types of reports in addition reports on public infrastructure, such as complaints about education and health. Meanwhile, on the technical side, with the increase of phone users, the system should be available as a phone application for several platforms such as Android, iOS and Windows Phone. Moreover, the gain of the system's popularity might open the opportunity for

citizens to send reports through various ways such as through a form. Moreover, Natural Languages Processing, data Mining, Artificial intelligence, and Text Mining technologies should be considered to improve the proposed system during analysing the results.

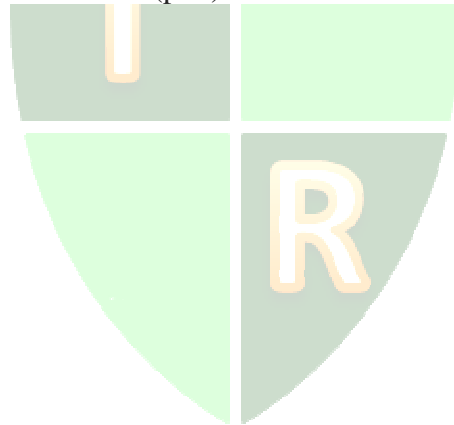
## ACKNOWLEDGEMENTS

To express my deepest gratitude to Dr. Sepi Chakaveh for her continued guidance and advices in every phase of the research. Last but not least, I am grateful to University of Southampton and Taif University.

## REFERENCES

- [1] "MySociety." [Online]. Available: <https://www.mysociety.org/about/>. [Accessed: 15-Jun-2016].
- [2] "FixMyStreet." [Online]. Available: <https://www.fixmystreet.com/>. [Accessed: 14-Jun-2016].
- [3] K. T. ABDURABB, "Saudi Arabia has highest number of active Twitter users in the Arab world," 2014. [Online]. Available: <http://www.arabnews.com/news/592901>.
- [4] K. Gavelin, S. Burall, and R. Wilson, "Open Government: beyond static measures," 2009.
- [5] N. Heller, "A Working Definition of Open Government," 2012.
- [6] B. Obama, "Transparency and Open Government," *Whitehouse.gov*, 2009. [Online]. Available: [https://www.whitehouse.gov/the\\_press\\_office/TransparencyandOpenGovernment](https://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment).
- [7] "FAQ | data.gov.uk," *Data.gov.uk*. [Online]. Available: <https://data.gov.uk/faq>. [Accessed: 05-Jun-2016].
- [8] "About e-Government," *Saudi.gov.sa*. [Online]. Available: <https://www.saudi.gov.sa/wps/portal/saudi/aboutKingdom/aboutEgovernment>. [Accessed: 07-Jun-2016].
- [9] "Overview," *Yesser.gov.sa*. [Online]. Available: <http://www.yesser.gov.sa/en/ProgramDefinition/Pages/Overview.aspx>. [Accessed: 08-Jun-2016].
- [10] I. Elbadawi, "Open data developments in Saudi Arabia, Qatar and Dubai," *Open Data Institute*, 2015. [Online]. Available: <https://theodi.org/blog/open-data-developments-in-saudi-arabia-qatar-and-dubai>. [Accessed: 09-Jun-2016].
- [11] (2007) OGDG, "Open Government Data Principles: 8 Principles of Open Government Data," *www.opengovdata.org*. [www.opengovdata.org](http://www.opengovdata.org), 2007.
- [12] Y. Ryan, "How Tunisia's revolution began," *Aljazeera*, 26-Jan-2011.
- [13] U. Friedman, "The Egyptian revolution dominated Twitter this year," *Foreign Policy*, 05-Dec-2011.
- [14] P. Beaumont, "The truth about Twitter, Facebook and the uprisings in the Arab world," *The Guardian*, 25-Feb-2011.
- [15] Arab News, "Saudi social media users ranked 7th in world." [Online]. Available: <http://www.arabnews.com/saudi-arabia/news/835236>. [Accessed: 25-Jun-2016].
- [16] "Using advanced search," *Twitter.com*, 2016. [Online]. Available: <https://support.twitter.com/articles/71577#>. [Accessed: 10-Jul-2016].
- [17] "GET search/tweets," *Twitter.com*, 2016. [Online]. Available:

- <https://dev.twitter.com/rest/reference/get/search/tweets>. [Accessed: 10-Jul-2016].
- [18] [https://en.wikipedia.org/wiki/Arabic\\_diacritics](https://en.wikipedia.org/wiki/Arabic_diacritics) [Accessed: 20-Jul-2016].
- [19] C. C. Aggarwal and C. Zhai, "An introduction to text mining," *Mining Text Data*, pp. 1–10, 2013.
- [20] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," vol. 1, no. 2, pp. 91–231, 2012.
- [21] M. Abdul-Mageed, M. Diab, and S. K??bler, "SAMAR: Subjectivity and sentiment analysis for Arabic social media," *Comput. Speech Lang.*, vol. 28, no. 1, pp. 20–37, 2014.
- [22] N. Habash and O. Rambow, "Arabic Tokenization, Part-of-Speech Tagging and Morphological Disambiguation in One Fell Swoop," *ACL '05 Proc. 43rd Annu. Meet. Assoc. Comput. Linguist.*, pp. 573–580, 2005.
- [23] Wolny, W., 2016. Sentiment analysis of Twitter data using emoticons and emoji ideograms. *Studia Ekonomiczne*, 296, pp.163-171.
- [24] "What is PHP?," *PHP.net*, 2016. [Online]. Available: <http://php.net/manual/en/intro-what-is.php>. [Accessed: 12-Jul-2016].
- [25] <http://www.mofa.gov.sa/sites/mofaen/ServicesAndInformation/aboutKingDom/SaudiGovernment/Pages/AdministrativeDivision46464.aspx>. [Accessed: 15-Jul-2016].
- [26] Al-Barrak, K., Carr, L. and Ryan, M., 2016, September. Towards a Model for Monitoring Public Services Projects in Saudi Arabia. In *10th European Conference on Information Systems Management: ECISM 2016* (p. 1). Academic Conferences and publishing limited.



APPENDIX A

Reports samples



Fixing Report Sample.



Cleaning Report Sample.

APPENDIX B

Open Data file content

reporting_c	reported_t	r	ranked	rank	favorite_c	retweet_c	media_url	media_type	report_text	type_id	report_id
23/06/16	0	1	1	1	0	0			السالم عليكم ورحمة الله بنقص نحتاج لحلوليات نفايات وريش ميديدات في حي النهضة	2	7.4E+17
23/06/16	0	1	0	0	0	0			حسينا الله ونعم الوكيل هل يعقل ما يحدث نفايات طيبة وسامة وتلوث في وسط حي سكني	2	7.4E+17
23/06/16	0	1	0	0	0	0			الحفر طريق دولي بس كله نفايات وين المراقبين اصحو فشتلونوا الحواري كلها زيبائل اعزكم	2	7.4E+17
23/06/16	5	1	0	0	0	0	<a href="http://pbs.t/photo">http://pbs.t/photo</a>	photo	ح حفر في نص طريق كنت بطيح فيها حسيننا الله ونعم الوكيل#	1	7.4E+17
23/06/16	0	1	0	3	14	14	<a href="http://pbs.t/photo">http://pbs.t/photo</a>	photo	هل ننتظر حادثا سنفيما ليحركه المسؤول عن #حفر الموت في	1	7.4E+17
11/07/16	5	1	1	1	0	0	<a href="http://pbs.t/photo">http://pbs.t/photo</a>	photo	السالم عليكم لدي بلاغ حفر وتسققات خطيرة في طريق الدمام	1	7.4E+17