Journal of Information Technology Management
ISSN-OnLine: 2980-7972
University of Tehran

Twitt Ray: Sentiment Visualization of Arabic Tweets

Nadia Al-Ghreimil

Assistant Prof., Department of Information Technology, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. E-mail: ghreimil@ksu.edu.sa

Manal Alhassoun*

Corresponding author, Data Scientist and Research Associate, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia. E-mail: smalhassoun@kacst.edu.sa

Nora Al-Twairesh

Associate Professor, Department of Information Technology, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. E-mail: twairesh@ksu.edu.sa

Duaa AlSaeed

Prof., College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. E-mail: dalsaeed@ksu.edu.sa

Journal of Information Technology Management, 2026, Vol. 18, Issue 3, pp. 155-180
Published by the University of Tehran, College of Management
doi: https://doi.org/10.22059/jitm.2026.108009
Article Type: Research Paper

Received: January 14, 2026
Received in revised form: March 23, 2026
Accepted: May 09, 2026
Published online: July 01, 2026

Abstract

Sentiment Analysis of Twitter offers valuable insight into public sentiment. However, humans absorb most information through visual perception. Therefore, visual representation of data enables them to comprehend data more quickly and effectively than in text form. Unfortunately, the majority of ex-isting work mainly focuses on sentiment analysis and pays less attention to sentiment visualization; this potentially reduces the real value of sentiment analysis results. In this paper, we present an easy-to-use sentiment visualization tool for Arabic tweets called Twitt Ray. The goal of Twitt Ray is to serve anyone, from individuals to companies and social scientists, who are interested in knowing, analyzing, and visualizing the current sentiments that people express on Twitter. Moreover, Twitt Ray provides seven different visual representations of sentimentally analyzed data to achieve a better understanding of different aspects of the data. Also, a usability study was performed to evaluate the usability of the tool for end users.

Keywords: Sentiment Visualization, Social media analytics, information visualization, sentiment analysis

Introduction

Previously, the term visualization meant forming a visual image in the human mind. Modern use of this term now implies the act or process of representing data, information, and knowledge in a visual form, such as in pie charts, scatter charts, and histograms, to gain a better insight into the data and to communicate information adequately (Chow, 2018). The concept of Information Visualization (IV) evolved because of the way our brains function. Cognition occurs slowly compared to visual perception, which occurs immediately and with little effort (Ware, 2012). As the old adage goes, "A picture is worth a thousand words." With the high rate of data being produced everywhere, IV becomes essential to aid and speed up interpretation and make sense of data (Wong, 2010). IV is recognized as an essential component of the process of analyzing unstructured text data with the goal of deriving meaningful information, e.g., in text analytics (Keim et al., 2013). One of the sub-disciplines of text analytics is sentiment analysis (SA), which is the process of classifying the polarity of a given text as positive, negative, or neutral using natural language processing (NLP) techniques (Pang & Lee, 2008; Liu, 2012). SA offers many benefits (e.g., analyzing trends, determining ideological bias, determining opinions, and gauging reactions) in different areas such as marketing, politics, public relations, social sciences, and psychology. For example, analyzing trends in consumer sentiment helps brands gauge the impact of their public relations and marketing campaigns.

Currently, the Web is the main source of information, and social networking and microblogging sites are rapidly growing. As a result, in recent years, social media and microblogging sites, where millions of users share their opinions on various aspects of everyday life, have become a gold mine for public SA. Twitter is one of the world’s most well-known microblogging sites; hence, SA on Twitter has become an active research area, and many tools have been built and developed for SA of tweets. However, as with any other text analysis tool, an important issue in developing an SA tool is representing analysis results in a way that facilitates an easy understanding of the sentiment of tweets. Unfortunately, existing tools often focus on the SA aspect and disregard the sentiment visualization part, despite it being equally important and essential to show the real value of SA. There exist some SA tools and applications that provide sentiment visualization, but the majority of these tools are designed for English text. Many low-resourced languages lack sentiment visualization tools, such as the Arabic language. To address this issue, this paper presents Twitt Ray (ray is an Arabic word that means sentiment), which is a web application that collects, analyzes, and visualizes the current sentiments of Arabic tweets using a proper assortment of visualization techniques in order to leverage the real value of Twitter SA. To the best of our knowledge, this is the first online sentiment visualization tool for Arabic tweets. As stated in a recent survey on sentiment analysis in Arabic (Al-Twairesh, 2018), there is no previous work on sentiment visualization for the Arabic language. Twitt Ray utilizes a previously developed classifier for SA of Arabic tweets, which was presented in Al-

Ghreimil et al. (2017). According to Healey and Enns (2012), usability evaluation is critical for IV, and most of the existing work on sentiment visualization do not consider the end-user. Therefore, we opted to perform a usability evaluation study to determine how usable, satisfactory, efficient, and effective Twitt Ray is for users.

The paper is organized as follows: Section 2 reviews related work. In Section 3, the methodology used to develop Twitt Ray is presented. In Section 4, the details of the user interface for the tool are illustrated. The usability evaluation study is discussed in Section 5. Results and discussions are presented in Section 6. Finally, section 7 concludes the paper.

Literature Review

A comprehensive survey of sentiment visualization techniques was presented in Wang and Lin (2014), where the authors discussed and summarized over 130 techniques found in different peer-reviewed publications. They also developed an interactive survey that can be found online. In this study, the publications were categorized into different groups according to data domain, data source, data properties, analytic tasks, visualization tasks, visual variables, and visual representation. Since text visualization is a subfield of information visualization and visual analytics (Healey & Enns, 2012), sentiment visualization is part of text visualization. The authors in Healey and Enns (2012) survey the surveys on text visualization and identify the challenges in the field. One of the reported challenges is the lack of quantitative or qualitative evaluation for text visualization techniques and the lack of user interactivity that can provision the analysis tasks. We tried to overcome these by involving the users in the usability study.

Also, according to Shammim et al. (2018), who performed a usability study on several sentiment visualization systems, the most important metrics are eye-pleasing, easy to understand, user-friendly, informative design, usefulness, and representation style. They also showed that the top five visualizations used in most systems are bar charts, glowing bars, tree maps, line graphs, and pie charts. In the design of Twitt Ray, we attempt to follow the results of this study.

Another survey on visual text analytics was presented in Choudhury and Khosla (2020), where the authors attempt to identify the relationships between visual analytics and text mining techniques by analyzing over 4,600 papers in both areas. They highlight the research trends and gaps in the relationship between the two fields of visual text analytics and text mining, including sentiment analysis. These surveys review work that is on sentiment visualization of text in different genres. However, in this section, we focus on work that is on sentiment visualization of Twitter data.

Early work on Sentiment visualization of Twitter data streams can be found in Diakopoulos and Shamma (2010), where they present two visualization approaches: (1) a pixel sentiment calendar that shows the topics in the rows and the time intervals in the columns then the cell color represents the sentiment (green: positive, gray: neutral, red: negative); (2) a pixel sentiment geography map that visualizes the sentiment of tweets according to the geolocations on the map. In Dastgheib and Vazirani (2017), the sentiment of a topic of interest is visualized in a Sentiment Card. The color of the card represents the polarity (green for positive, red for negative), a word cloud, and a map visualization. The card also has a trend opinion component that shows the temporal sentiment of the topic. The sentiment percentage is represented in a bar chart or line chart. To evaluate the system, a usability test was conducted on six participants, and the system scored 75 % using a system usability score (SUS). SentiCompass was presented in Lu and Li (2013), where the temporal information of tweets is combined with the valence and arousal of the sentiment of the tweet. They were used for the visualization of a 2D cyclic representation with color coding and the time tunnel representation. Social Helix is a visual analysis system proposed by Guo et al. (2017) to visualize sentiment divergence in social media. The system aims to help users to understand the complex phenomenon of sentiment divergence by visualizing three aspects: when the sentiment divergence started and ended, why it occurred, and who is involved in it (the community or opposing groups involved).

They utilized a unique visual metaphor, which is the DNA molecule that is made up of two twisting helices (hence the name Social Helix). The system was evaluated by three domain experts in computational social science, human-computer interaction, and social media analyzer. The uVSAT tool (Chung et al., 2018) targets linguistics researchers and NLP experts who are working on stance analysis, including sentiment analysis. It is a web-based visual analytics application for detecting sentiment in social media documents. It provides interactive history diagrams and aggregation charts for the analyzed documents. The basic charts used are line plots for time series and bubble charts for aggregation charts. In Zhao et al. (2019), StanceViz Prime was introduced, which is a tool for visual analytics of the sentiment and stance of temporal text extracted from different social media outlets (Twitter and Reddit). The main visualization types used are a stacked graph to represent the counts of documents processed over time and bar charts for the different sentiment polarity and stance data series.

Visual Twitter Analytics (Vista) was presented in Bögl et al. (2013) to visualize the sentiment of tweets collected through a search query. The main visual representation is a timeline coupled with color encoding to represent the three classes of sentiment (positive, negative, and neutral). A geographic map is also provided to reflect the geolocation of the tweets. The tool also provides interactive filtering, zooming, and aggregation of the collected tweets. Social Sentiment Sensor (SSS) (Zhao et al., 2016) is a system that detects real-time sentiment of hot topics on the Weibo microblogging platform. It presents the visualization through line charts that represent the temporal sentiment, and bar charts for the sentiment of the detected hot topics. It also includes a nationwide emotion index visualized through a pie chart, a radial gauge, and a choropleth map. In Yaqub et al. (2020), the authors performed SA of Twitter location data gathered during the US and UK elections of 2016 and 2017, respectively. The overall objective was to evaluate the similarity between the sentiment of location-based tweets and on-ground public opinion reflected in election results. In their approach, they used data analyses with visualizations to answer their research questions. They created visualizations by plotting public sentiment toward different candidates on a map. They concluded that Twitter location sentiment does corroborate with the election result in both cases.

Authors in Samah et al. (2022) proposed implementing a web-based dashboard to visualize the performance of the airline companies in Malaysia through Twitter SA. They designed and developed a model for bilingual (English and Malay) SA of Twitter, then deployed it in a web-based dashboard to visualize results. The visualized results can be used to increase customer satisfaction and ensure retention. In their dashboard, data was displayed in different visualizations such as bar charts, pie charts, gauge charts, and word clouds for better insights. The authors concluded that their proposed SA web-based dashboards reflect a good solution and can assist anyone in understanding the public views on airline companies in Malaysia. The study in Villavicencio et al. (2021) was about Twitter SA toward COVID-19 Vaccines in the Philippines. The authors used different visualization techniques for the visual analysis of their data. A grouped bar chart was used to display processed datasets over time; the grouping was by SA polarity (positive, negative, or neutral). A word cloud was also used to show the terms used in the dataset.

In Healey et al. (2019), a system was developed to visualize online customer service chats with a focus on three elements: topic identification, text summarization, and sentiment analysis. Using a stream graph, the conversation topics, emotions, emotion patterns, chat similarity, and semantic tags are shown in a four-layer visualization. Britzolakis et al. (2017) presented a data visualization tool for identifying political popularity over Twitter in the Greek Language. A sentiment analyzer for Greek tweets was developed, then used as input to the web-based data visualization tool. The visualized data included, in addition to the sentiment of the tweets, other factors such as likes and retweets of the collected tweets. The visualizations used included pie charts, bar charts, and line charts. Another way of using IV in connection with SA is presented in Kim et al. (2017), where Visualizations like bar plots provide interpretability and explainability of predictions, which in turn can act as debugging tools.

As we can see from the work reviewed above, most of the approaches for sentiment visualization have been developed for the English language. Moreover, in spite of the significance of the visualization of analysis results, most research on SA of Arabic tweets has been focused on the actual classification of the data (Al-Twairesh, 2018), and the visualization of sentiment data has not been given much focus. Therefore, this paper presents Twitt Ray, which is an online, publicly accessible Twitter sentiment visualization tool for Arabic tweets. Twitt Ray attempts to minimize the gap, fulfill the growing needs, and overcome some of the existing limitations. Thus, it visualizes tweets and their sentiments using several visualization techniques in order to show different aspects of the data. Moreover, it shows the sentiment of individual tweets and informs the user about the approach to sentiment classification.

Methodology

Twitt Ray was developed as a web application. It consists of five main components: data collection, data preprocessing, sentiment classification, data transformation, and data visualization. A high-level overview of the architecture of Twitt Ray is depicted in Figure 1. The subsequent subsections provide further details on each component.

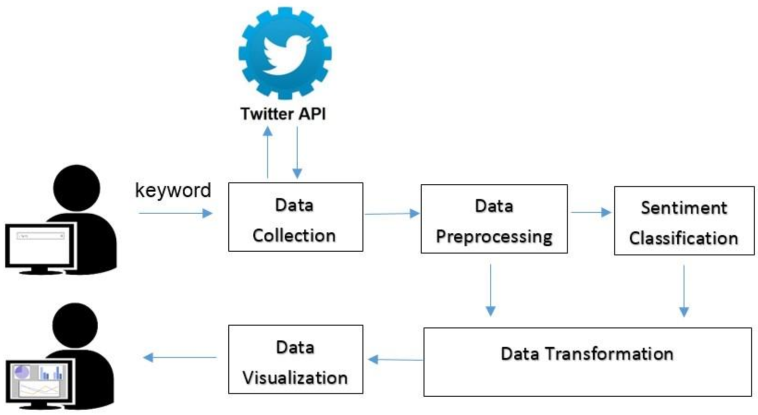

Fig. 1 System Overview of Twitt Ray

Data Collection

This component is responsible for retrieving tweets containing a given search term from Twitter. Twitter provides APIs that aid third-party applications in accessing and collecting its data (Twitter, n.d.). There are two major types of Twitter API: REST APIs and streaming APIs. In our system, we utilized the Search REST API. This API is suitable for the nature of our system because we are using a keyword-based search. In addition, we need to narrow the search results to include only tweets that were written in the Arabic language. The search API easily allows the user to filter the results by its rich set of query parameters, such as language and location (tweets' location and users' profile location). In contrast, the streaming API allows filtering of the results based on the tweet locations only. Furthermore, in general, the volume of the retrieved results from REST APIs is more than that of streaming APIs.

Data Preprocessing

The data preprocessing component is responsible for correcting inconsistencies and eliminating noise from the retrieved tweets. The preprocessing step is important in order to achieve quality results from the analysis. In our system, we perform two types of preprocessing on the text in tweets: standard (i.e., general) and Twitter-specific. For the standard preprocessing, we follow the preprocessing techniques as described in a previous study on SA of Arabic tweets (Al-Ghreimil et al., 2017):

- **Letter Normalization** is the process of unifying various forms of a letter into one form. There are four letters in Arabic that usually need to be normalized: Alif, TaMarbuta, Alif-Maqsura, and non-Alif forms of Hamza. Figure 2 shows the Arabic normalization.
- **Word Tokenization** is the process of tokenizing (i.e., segmenting) tweets into words.
- Remove non-Arabic words.

Moreover, Twitter data contains several unrelated data points; therefore, it requires further preprocessing. Thus, we considered the following tweet-specific features in the preprocessing phase (Al-Ghreimil & Al-Twairesh, 2018):

- Retweet: the retweet symbol (RT) in the tweet text is removed.
- URLs: we remove all the URLs from each tweet text.
- @username: mentions of other users (@username) in the tweets text are removed.
- #hashtag: we eliminate the hash symbol (#) in front of any word, for instance, we use upset instead of #upset.

Unnormalized		Normalization	
Letter	Shapes	Letter	Shape
Alif-Maqsura	ى	Ya'a	ي
Non-Alif forms of Hamza	ؤ	Hamza	ء
Non-Alif forms of Hamza	ئ	Hamza	ء
Ta-Marbuta	ة	Ha'a	ه
Alif	آ أ إ	Bare Alif	ا

Fig. 2. Arabic Letters Normalization

Sentiment Classification

This component is dedicated to automatically classifying the sentiment of the retrieved tweets as positive, negative, or neutral. Since our main goal is not to improve the accuracy of sentiment classification, but rather to effectively visualize SA results, we utilize the sentiment classification models developed in Al-Gremial et al. (2017). The authors utilized the Ara Senti-tweet corpus curated in Mohammad et al. (2016) to build the classification models. The corpus was constructed by manually annotating Arabic tweets according to sentiment. The size of the corpus after manual annotation was 17,573 tweets. The corpus was annotated by four sentiment labels (positive, negative, neutral, and mixed). Three sentiment classification models were developed: lexicon-based, corpus-based, and hybrid.

The lexicon-based model utilizes sentiment lexicons specifically designed to capture the nuances of Arabic tweets. These lexicons provide a prior-polarity for words, helping to identify positive and negative sentiments within tweets. The model calculates a Tweet Score by summing the sentiment intensities of words from the Ara Senti lexicon (Al-Twairesh et al., 2016), leveraging the inherent sentiment information embedded in the lexicons. The corpus-based model, on the other hand, relies on machine learning techniques to classify sentiments. This approach utilizes the Ara Senti-Tweet corpus, annotated for sentiment, to train a Support Vector Machine (SVM) classifier. The model incorporates a variety of features, including semantic features like the presence of sentiment words and stylistic features such as the use of emoticons and punctuation. The authors employ a feature backward selection method to optimize the feature set, ensuring that only the most impactful features are retained for classification. The hybrid model combines elements of both lexicon-based and corpus-based approaches to enhance classification accuracy. By integrating the Tweet Score from the lexicon-based method as a feature in the corpus-based model, the hybrid approach leverages both the rich lexical information and the contextual understanding provided by the corpus. The authors report that this method improves performance across different classification tasks, namely two-way (positive-negative), three-way (positive-negative-neutral-mixed), and four-way (positive-negative-neutral-mixed-both) sentiment classification. The hybrid approach shows promise in addressing the challenges of sentiment analysis in Arabic tweets, achieving F1-scores of 69.9, 61.63, and 55.07 for the respective models. This demonstrates its effectiveness in handling the complex morphology and dialect variations inherent in Arabic social media text.

Since the three-way hybrid method outperformed the corpus-based method, we used the hybrid three-way classification. Before using the model, we must represent each tweet as a feature vector to be an input to the classifier. In Al-Ghreimil et al. (2017), the authors observed that the best feature set for hybrid three-way classification included the following:

- **Tweet Score** is calculated from the lexicon-based method described in Al-Ghreimil et al. (2017).
- **Has Negation** is a Boolean feature that determines whether the tweet contains a negation word.
- **Has Positive Word Ara Senti** and **has Negative Word Ara Senti** are Boolean features that determine whether the word is negative or positive according to the Ara Senti lexicon—an Arabic sentiment lexicon (Al-Twairesh et al., 2016).
- **Has Positive Word MPQA** and **has Negative Word MPQA** are Boolean features that are true if the word is positive or negative, respectively, in the Arabic translation of the MPQA subjectivity lexicon (Wilson et al., 2005) and false otherwise.
- **Has Positive Word Liu** and **has Negative Word Liu** are Boolean features that are true if the word is positive or negative, respectively, in the Arabic translation of the Bing Liu lexicon—a lexicon of negative and positive words (Liu, 2012), and false otherwise.

Data Transformation

As soon as the sentiment classification phase is completed and once the polarity of each tweet is determined, we prepare the data and transform it to an acceptable format as required by the selected visualization techniques (the selected techniques are shown in the next subsection). Moreover, we map the geographic locations obtained from each tweet's location and the users' profile location to the corresponding map regions, count the frequency for each unique word in the positive and negative tweets, aggregate the overall sentiment scores and their associated geographic regions, and count the number of positive, negative, and neutral tweets.

Data Visualization

This component is responsible for visualizing the SA results using various suitable visualization techniques. We need to determine which visualization techniques to use in our system. First, we need to understand the data that we are working with. Twitter APIs return tweets in JSON (JavaScript Object Notation) format. In addition to the text of the tweet, the JSON object contains various metadata such as temporal and occasionally spatial information. Hence, we can detect when and where a tweet was posted, and this will help in the tweet sentiment visualization. Moreover, understanding the basics of the human visual perception system is useful when choosing visualization techniques to represent different types of data (Tufte, 2001; Few, 2009; Miller, 1956). We perceive and process visual information instantaneously without conscious awareness of it (Ware, 2012). However, human perception is affected by three aspects (Spence, 2014): the past, the present, and the future. The past can be interpreted as our experience, the present can be interpreted as the current context, and the future can be interpreted as our goals. Thus, we should use universal and well-known techniques to be easily recognized by the user. Furthermore, the human brain's ability to receive, analyze, and recall information is limited (Shneiderman, 1996). Hence, this limitation must be considered when representing data in a visual context.

Taking all of this into consideration, we decided to use familiar but useful visualization techniques that show different aspects of the data (tweets and their sentiment polarity). In addition, we tried to keep the visualization simple, clear, and without unnecessary visual elements in order to avoid chart junk, which is a term introduced by the data visualization expert Edward Tufte (Tufte, 2001). Thus, designing a simple and clear visualization would help anyone to consume and comprehend the information presented, which serves the purposes of our web tool. The selected visualization techniques that we chose to use are as follows:

- **Pie chart:** we used a pie chart (see Figure 3.a) to show the percentages of positive, negative, and neutral tweets to give a part-to-whole comparison.
- **Bar chart:** We used a bar chart (see Figure 3.b) to show and compare the total number of positive, negative, and neutral tweets.
- **Line chart:** We used a line chart (see Figure 3.c) to show the time-series data and the potential trends over time. We show the line chart with separate lines for the positive and negative sentiments. Using this allows us to take advantage of the temporal information associated with retrieved tweets.
- **Proportional symbols map:** (see Figure 3.d) We used this to show the distribution of data over the 13 regions of Saudi Arabia. The size of each symbol indicates the number of tweets in that region, and the symbol color indicates the overall sentiment. Using this map allows us to take advantage of the spatial information associated with retrieved tweets. In addition, it takes advantage of the expected users' (probably Saudis) familiarity with the shape of Saudi Arabia and its regions.
- **Word Cloud:** we used two-word clouds (see Figure 3.e) to show the most frequent words in positive and negative tweets. This might help to understand the context surrounding the search term and the causes of the sentiment (whether it is positive or negative).
- **Table:** We used a table to give a general overview of the exact data. It lists all the retrieved tweets along with their sentiment and other information, such as the post time and username.
- **Stacked bar chart:** we used a stacked bar chart (see Figure 3.f) to compare the sentiment of different keywords (maximum four). In addition, it shows the relative contribution of

each sentiment polarity value to the total, where each stack bar represents a keyword, and gives the percentage of positive, negative, and neutral tweets for each keyword.

Moreover, we employed colors to encode the sentiment polarity in the visualizations. To do this, we followed the traffic light metaphor (Few, 2009) proposed by Stephen Few, where red, yellow, and green correspond to a negative, neutral, and positive sentiment, respectively. According to Wang and Lin (2014), color was the most used visual metaphor to convey sentiment in most sentiment visualization papers. In addition, to create the visualizations, we used the Google Chart API (Google, n.d.). One of the major advantages of this API is its support for all the chosen visualization techniques (except the word cloud). Besides, it is free for all usage, customizable, well-documented, provides cross-browser compatibility, and supports tooltips, which allows us to provide the user with further information in visualizations. However, since Google Charts does not support word cloud visualization, we used D3.js for word clouds (Bostock et al., 2011). D3.js is a free and powerful tool that allows us to build any type of visualization from scratch.

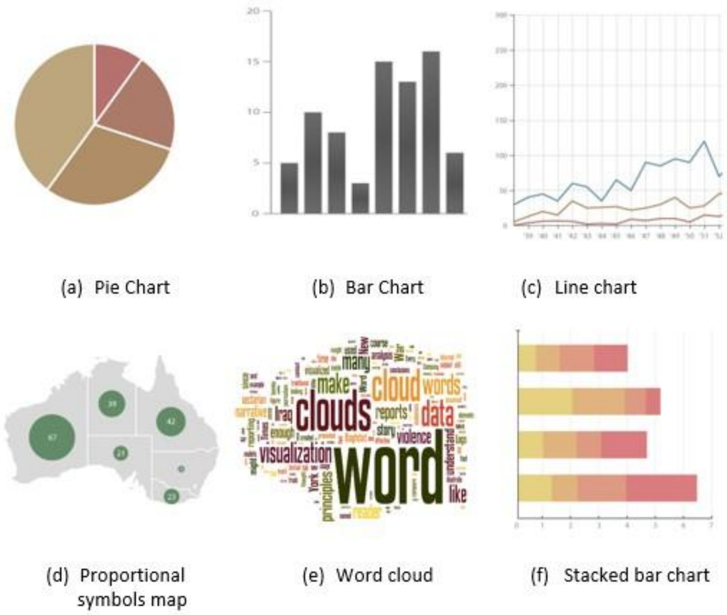

Fig. 3. Selected information visualization techniques

An interactive Graphical User Interface GUI encourages data exploration. Twitt Ray was designed with an interactive GUI to allow users to leverage their intuition and expertise in observing data patterns, leading to new insights and a more thorough understanding of public sentiment. Since Twitt Ray is designed to perform SA and visualization on Arabic tweets, it was developed as an Arabic website. Throughout the design of Twitt Ray, usability principles were followed to ensure that the tool is easy to use. One of the main UI principles is consistency in layout and design; for example, the logo, navigation, content, etc., are in the same place on all pages, in addition to using the same color scheme. Furthermore, due to the limited capacity of the human memory (Miller, 1956; Shneiderman, 1996), the user should not be distracted by too much information in one place; therefore, each of the IV techniques is displayed in separate tabs, except the pie chart and bar chart, since they are simple and can be considered as complementary to each other. Additionally, for usability reasons, data visualizations should not be split into multiple windows; thus, we tried to make sure that every IV figure fits in the window without the need to scroll. Moreover, when the user hovers the mouse over any visualization, a tooltip with further information is displayed. For example, when the user hovers the mouse over one of the circles in the map, information about the number of positive, negative, and neutral tweets in that region will be displayed.

Twitt Ray Menu Options

Figure 4 shows the Home Page of Twitt Ray, which is the first item on the main menu in the top right of the screen. Next to it are two more items:

- **About us:** the about us page provides information about the tool in general and about the SA method used.
- **Help:** this page provides helpful information about the IV techniques provided on the website.

The home page is where the main interaction takes place. There are seven different types of analysis and visualizations: 1-Overview, 2-Temporal, 3-Map, 4-Frequently used words, 5-Tweets, 6-Report, and 7-Comparison (See Figure 4). The first six options use a simple search option at the top of the page. The simple search option allows the user to perform one search query at a time. An advanced search option is available with the seventh provide option, it allows the user to perform a complex search by combining up to four search queries at a time. When a user's search request is submitted, Twitt Ray then analyzes the sentiment of the tweets and visualizes the result based on the selected option as follows:

- **Overview:** A pie and bar chart is used to give the user an overview of the sentiment for all the retrieved tweets by showing the percentage and total numbers of positive, negative, and neutral tweets (see Figure 5).
- **Temporal:** a line chart is used to show the user the sentiment evolution over time. This could help to identify interesting patterns and trends (see Figure 6).
- **Map:** the user can see how people in each Saudi region are feeling regarding the searched terms (see Figure 7).
- **High-Frequency Words:** the user can see a two-word cloud. In the Tweets tab, the user can see a list of the retrieved tweets, their sentiment, and the time they were posted. In addition, when the user hovers over the listed tweet, a tooltip is displayed that shows further information, such as the username of the user who posted the tweet and the information about when the tweet was posted (see Figure 9).
- **Tweets:** The user can see a list of the retrieved tweets, their sentiment, and the time they were posted (see Figure 9).
- **Report:** In the Report tab, we allow the user to combine all the data visualizations on one page to compare and contrast the information obtained from each type. In addition, the user can export the page with all the visualization techniques as one file (see Figure 10).
- **Comparison:** a stacked bar chart is used to visualize the sentiment analysis results of the different keywords. Hence, a stack bar is generated for each search query. However, in order to ensure design consistency on all pages, we provide the comparison feature in a seventh tab. Thus, the four search boxes are located in the same place as the single search box in the previous tabs (see Figure 11).

Twitt Ray Deployment

Twitt Ray was developed using Java, JavaServer Pages (JSP), Java Servlets, HTML, CSS, JavaScript, and other technologies. To make it publicly accessible, we used AWS Elastic Beanstalk, which is a reliable cloud deployment service that supports Java and provides a free tier for 12 months.

Fig. 5. Twitt Ray - Overview option

Fig. 6. Twitt Ray - Temporal option

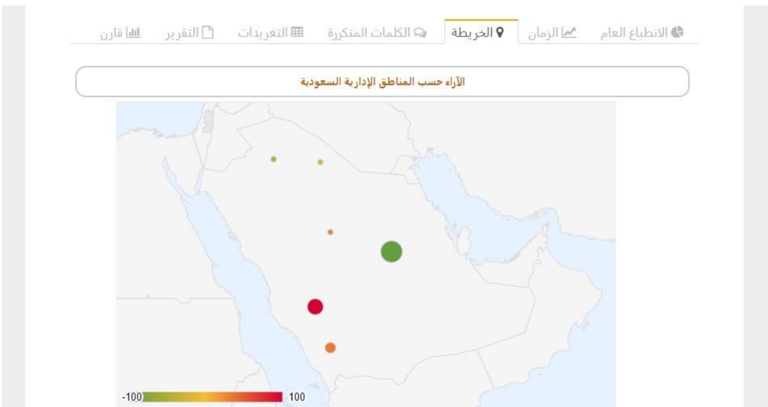

Fig. 7. Twitt Ray - Map option

Fig. 8. Twitt Ray - High Frequency Words option

Fig. 9. Twitt Ray - Tweets option

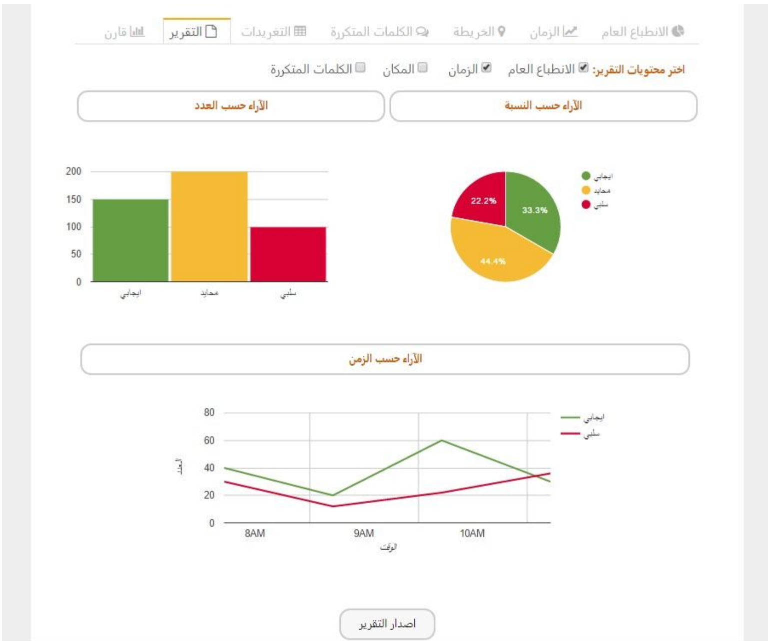

Fig. 10. Twitt Ray – Report option

Fig. 11. Twitt Ray - Comparison option

Usability Evaluation

The main objectives of the user-based evaluation are:

- To determine how usable, satisfactory, efficient, and effective the web application is for users.
- To determine how useful each IV technique is to users.
- Identify any usability problems.
- Finding ways of improving the usability of the web application.

Based on these objectives, the main usability research questions were as follows:

- Do users find the IV techniques useful?
- Do users understand options' names and purposes?
- How much time did it take the user to complete tasks involving the IV techniques in each option?
- What could prevent the users from completing tasks?
- Is the information in the help page helpful and readable?
- Is the user satisfied with the application?
- What problems do users have with the application?
- Do users have suggestions for improvement?

Participants

The inclusion criteria for this study were: (1) 18+ years old, (2) basic computer knowledge, (3) basic Twitter knowledge, and (4) ability to read and write in Arabic. Since Twitt Ray is intended to be used by almost everyone, we selected a diverse group of participants based on age, occupation, major, and gender. Eight volunteers were recruited to take part in this usability evaluation. Table 1 gives the details of the participants.

Test Environment

- **Morae** (TechSmith, n.d.): a usability testing software that we used for recording and logging users' interactions with the application while they completed the task-based scenarios. In addition, it allows us to collect the required usability measurements and efficiently analyze results. The software was installed on a laptop, which allows us to have a portable usability lab.
- **TeamViewer** (TeamViewer, n.d.): a remote desktop software we used to remotely observe participants' on-screen activity while they performed the task-based scenarios.
- **Face-to-Face Usability Testing:** the participants will test the web application at any place of their choice.
- **Remote Usability Testing:** if the participant could not be reached face-to-face, we used the screen-sharing application TeamViewer to allow monitoring of participant interactions and let them perform the testing in our portable usability lab.

Furthermore, to avoid bias between face-to-face testing and remote testing, both tests were moderated, i.e., with a moderator involved when the testing takes place (Usability.gov, n.d.). The participants' interactions with the web application, in both types of testing, were monitored by the facilitator. In addition, it was recorded and logged using Morae. Subsequently, this testing is very much like in-lab testing.

Table 1. Details of the Participants

Attribute	Categories	Number	Percentage
Gender	Female	6	75%
	Male	2	25%
Age	16-24	2	25%
	25-34	5	62.5%
	35-49	1	12.5%
Occupation	Manager	1	12.5%
	Academic Researcher	4	50%
	Unemployed	3	37.5%
Major	Project management	1	12.5%
	Information Technology	3	37.5%
	Software Engineering	1	12.5%
	Finance	1	12.5%
	Early childhood education	1	12.5%
	None	1	12.5%
Information Visualization Knowledge Level	Good	5	62.5%
	Poor	3	37.5%

Methodology and Test Procedure

To carry out the usability testing, eight task scenarios were developed. The task scenarios were based on real-life tasks to simulate how users would interact with the application in a real-life situation. In addition, the tasks were developed to identify the aspects related to usability issues of the web application. The first task was about reading the help page. The other seven tasks were about using the IV techniques in each tab on the homepage.

Each participant was involved in a single test session, which lasted approximately 40 minutes. During the testing session, we observed the participants and collected both quantitative data (time on tasks and success rate) and qualitative data (behavior and feedback). Three methods of data collection were used: observation, think-aloud, and questionnaires. The tasks for this test and the metrics that we measured are described in Tables 2 and 3, respectively.

Results and Discussion

In this section, we report and discuss the results of the user evaluation study and the feedback obtained.

Table 2. Task Overview: Related Options and Visualizations

Task #	Twitt Ray Option	Visualization Type
Task 1	Help	–
Task 2	Overview	Pie chart and bar chart
Task 3	Temporal	Line chart
Task 4	Map	Map
Task 5	High Frequency Words	Word cloud
Task 6	Tweets	Table
Task 7	Report	Report with selected charts
Task 8	Comparison	Stacked bar chart

Table 3. User Testing Metrics

Metrics	Measurement scheme
Effectiveness	Task Completion Rate
Efficiency	Task Completion Time
Satisfaction	System Usability Scale Questionnaire
Usefulness	Post-task Questionnaire

Effectiveness

The effectiveness of Twitt Ray was evaluated by measuring the completion rate, i.e., percentages of participants who completed tasks easily or with difficulty or failed to complete tasks. As demonstrated in Figure 12, no participants failed to complete any of the tasks, which means the overall user effectiveness of the system is 100 %. However, there were three tasks that some participants had difficulty with. The following is an explanation:

- Task 4: This task involved using the map visualization. Three out of eight participants expected to see the overall sentiment score written on the region's tool-tip. It took them a moment to realize that the overall sentiment values are displayed as a color scale.

- Task 7: This task involved exporting two IV techniques into one file. Two out of eight participants seemed hesitant about which tab (option) they should use. When they found the correct tab, they seemed hesitant about how to export the file. Both participants were from the group with poor IV background, and they were non-technical users.

- Task 8: This task involved using the stacked bar chart to compare the sentiments of two keywords. Three out of eight participants did not easily notice that the last tab was for the Comparison option. They thought that they could compare the keywords using the single-search field. After a moment, they realized how to complete the task correctly.

Efficiency

In this study, task completion time was used as the primary indicator of efficiency.

The average time-on-task across all participants is shown in Figure 13, whereas Figure 14

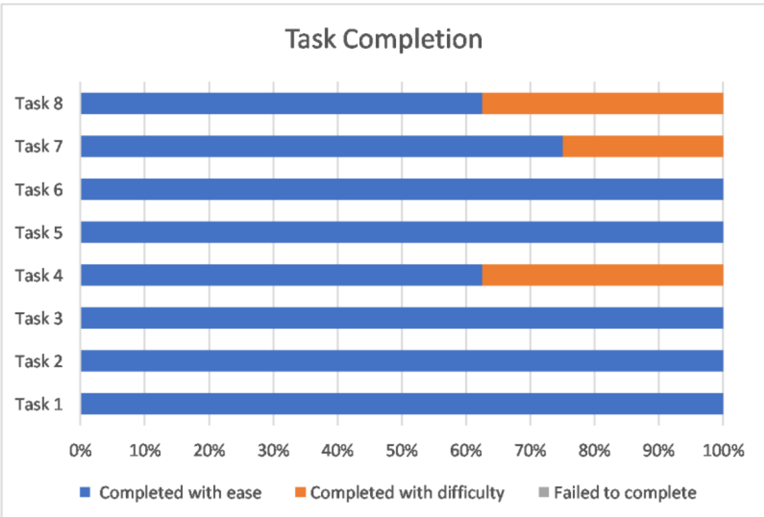

Fig. 12. Task completion distribution

Shows the average time-on-task for two categories of participants: those with good IV knowledge and those with poor IV knowledge. Figure 13 shows that Task 2 and Task 8 took the longest time on average (109.01 seconds and 129.04 seconds, respectively), and Task 1 took longer on average than the remaining tasks (76.14 seconds). One should note that Task 1 involves reading the help page; therefore, it is not unusual to take more time than other tasks, due to its length. Whereas Task 2 is the first Task where the user had to provide input to get a result for their query, which requires a user to get familiar with the interface, and requires the system to search and retrieve relevant tweets. Furthermore, Task 8 requires multiple input values in order to do a comparison, a task which is considered a more complicated option of Twitt Ray. Similar to Task 2, it requires the system to search and retrieve relevant tweets. On the other hand, the remaining tasks took an average of 33.6 seconds to be completed.

We were interested to see whether the users' prior knowledge of IV would affect their performance. From Figure 14, one can see that participants from both groups (good/poor IV knowledge) have similar average time-on-task for most of the tasks except Task 7. As explained above, Task 7 involved exporting two IV techniques into one file. Two participants from the group with poor IV knowledge struggled with this task. We believe the struggle is not attributed to poor IV knowledge since this task does not require IV knowledge, but is attributed to a low technical background. The two participants who struggled were non-technical users. Hence, that might have affected their performance in this task. Therefore, we concluded that Twitt Ray is easy and efficient to use and does not require a solid background in IV.

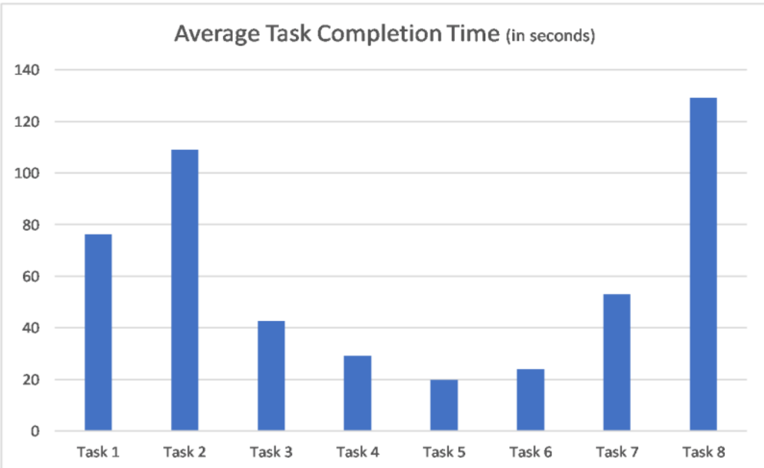

Fig. 13 Average task completion time - overall

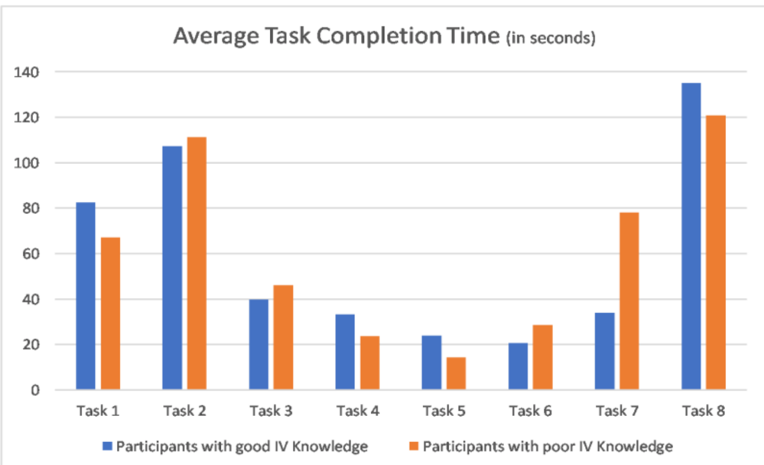

Fig. 14. Average task completion time - categorized

Satisfaction

The standardized System Usability Scale (SUS) questionnaire was used to measure participants' satisfaction levels. SUS scores range from 0 to 100, with a global average score of 68 (Samah, Misdan, Jono, & Riza, 2022). We employed the usability testing software Morae, which automatically calculated the SUS score. As illustrated in Figure 15, the result for Twitt Ray was 91.9, which is well above the global average SUS score. This shows that Twitt Ray is satisfactory and usable based on the agreement of the participants.

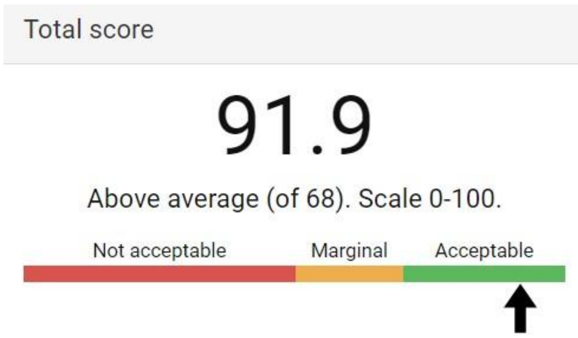

Fig. 15. Total SUS score

Usefulness

At the end of each task that involved using one of the IV techniques, the participants were asked to rate the degree of usefulness of the visualization techniques on a 5-point Likert scale. Average scores of all tasks were above 4.0 out of 5.0, as shown in Figure 16, which is a good indication of usefulness.

Further inspection of the variation in the 4.0 to 5.0 range reveals that the stacked bar chart, the line chart, and the map scored highest, respectively. This is a positive result since these three visualization techniques are introduced in Twitt Ray and have not been used by other Arabic sentiment visualization tools.

Table 4. Findings and Recommendations

Recommendation	Area of Improvement
Add the overall sentiment score to the region's tooltip.	Homepage (Map Tab)
Redesign the "Number of tweets" label and change its location.	Homepage
Redesign the labels. Choose a style that is not similar to the button's design.	Homepage
Instead of having a "Help" page, add the description of each data visualization	Help page
techniques to its tab. Add a brief system description to the homepage. Improve the response time.	Homepage
Change the "Comparison" tab's name to "Compare Opinion".	Homepage (Comparison Tab)
Increase the size of the text.	All pages

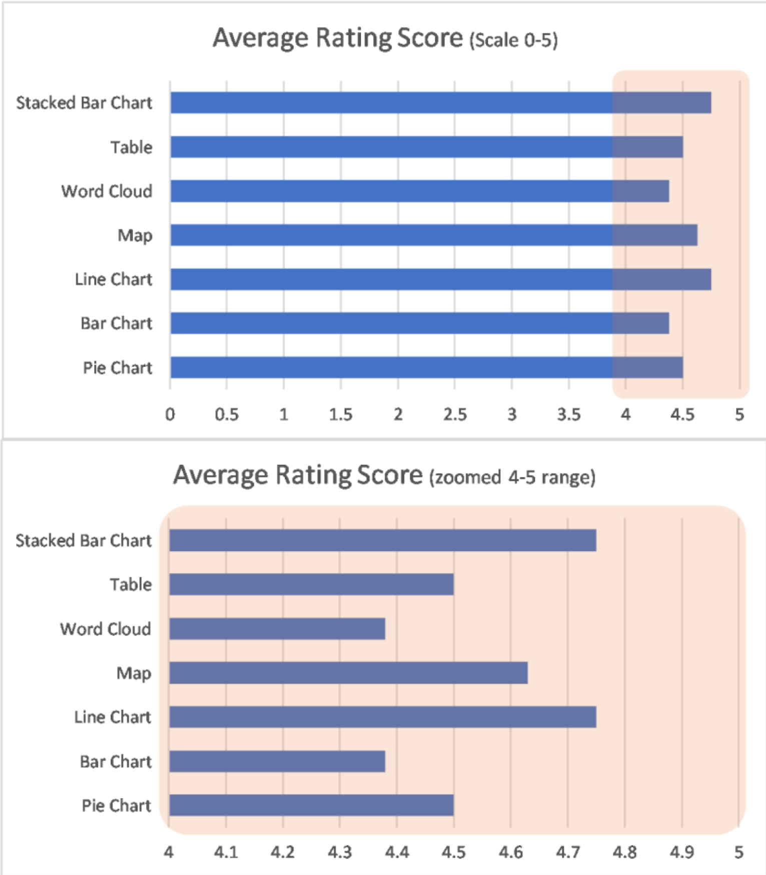

Fig. 16 Average usefulness score for each visualization technique

The main findings and recommendations were driven by the evaluation results, which include the following main aspects: participants' completion rates, behaviors, verbal feedback, and comments.

Table 4 summarizes the feedback received via the post-test questionnaire after the user evaluation. It shows the main recommendation for improving Twitt Ray by enhancing the overall ease of use and addressing the areas where participants have experienced problems or found the interface unclear. The first two items listed were immediately implemented as they were viewed as high-priority items. Others were noted for future improvement. Regarding response time, we acknowledge that it is always a concern in any interactive system. Yet, it should be noted that we received contradictory feedback from participants regarding system response. In addition to the written recommendation (5th item in Table 4) from a non-technical participant, we have verbal feedback from a technical user who was impressed by the fast response time.

Conclusion

Sentiment analysis, when combined with the appropriate visualization techniques for representing the SA results, will be more effective. This is because people process information visually. A visual representation of the SA results is essential and enables prompt and more effective comprehension of data than just text. Unfortunately, there is a lack of sentiment visualization tools for analyzing and visualizing the sentiment of Arabic tweets. In this paper, we presented Twitt Ray, which is a free and easy-to-use web-based Twitter sentiment visualization tool for Arabic tweets. Twitt Ray visualizes Twitter data and its sentiment analysis results in various informative ways. Twitt Ray was evaluated based on a usability test to examine the user experience. Results indicate that the current version of the Twitt Ray tool is acceptably usable, satisfactory, efficient, and effective for most users.

The sentiment classification model utilized in Twitt Ray has certain limitations, including suboptimal classification accuracy and the limitation to only two classes. To reach a more reliable classification accuracy, we could utilize other sentiment classifiers for Arabic tweets. To extend the tool to handle languages other than Arabic would require other sentiment classifiers.

Moreover, for future work, we intend to address the recommendations concluded from the results of the usability test and extend this work in several ways, such as including some additional effective IV techniques. Another possible improvement is to extend this tool to include other data sources, such as Facebook, and not restrict it to Twitter feeds only, since the tool works on any short Arabic text. Also, emotions could be added to the visualization along with sentiment.

Other fields of application may be explored. It is undeniable that mental health issues come in many forms and can be difficult to detect, let alone to treat. It is also undeniable that more and more people are using social media to express their thoughts and feelings. SA may come in helpful in detecting abnormal behavior.

Acknowledgments

This research project was supported by a grant from the “Research Center of the College of Computer and Information Sciences”, Deanship of Scientific Research, King Saud University.

Declarations

Ethical Approval: We recognize and adhere to Twitter’s updated API policies and ensure compliance with their terms of service. All data collection is conducted ethically, respecting user privacy and obtaining necessary permissions. Our analysis maintains transparency and integrity, strictly using data for research and visualization purposes only. We are committed to upholding ethical standards in data usage, ensuring that all insights derived from this tool are utilized responsibly and respectfully.

Competing interests

No competing interests of a financial or personal nature. Authors' contributions:
Conceptualization, D.A.;
Methodology, M.A., No.A., and D.A.;
Software, M.A. and No.A.;
Writing—review and editing, D.A., No.A., and Na.A.;

Funding

This research project was supported by a grant from the "Research Center of the College of Computer and Information Sciences", Deanship of Scientific Research, King Saud University.

Availability of data and materials

Not applicable

References

Al-Ghreimil, N., & Al-Twairesh, N. (2018). A feature-based sentiment analysis of Arabic tweets. In *Proceedings of the 2018 International Conference on Intelligent Systems and Knowledge Engineering (ISKE)* (pp. 1–8).

Al-Ghreimil, N., Al-Twairesh, N., & AlSaeed, D. (2017). Hybrid sentiment classification for Arabic tweets. *Journal of King Saud University - Computer and Information Sciences*, *29*(2), 241–251.

Al-Twairesh, N. (2018). A survey on sentiment analysis in Arabic: Challenges, approaches, and future directions. *Artificial Intelligence Review*, *50*(1), 1–40.

Al-Twairesh, N., Al-Ghreimil, N., AlSaeed, D., & Al-Twairesh, M. (2016). Ara Senti: A sentiment lexicon for Arabic tweets. In *Proceedings of the 2016 International Conference on Intelligent Systems and Knowledge Engineering (ISKE)* (pp. 1–8).

Bangor, A., Kortum, P., & Miller, J. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594.

Bögl, M., Kienreich, W., & Sabou, M. (2013). Visual Twitter Analytics (Vista): Exploring tweets with geo-location. In *Proceedings of the 2013 International Conference on Information Visualization (IV)* (pp. 315–320).

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301–2309.

Britzolakis, E., Kordopatis, P., & Al-Twairesh, N. (2017). A data visualization tool for identifying political popularity over Twitter in the Greek Language. In *Proceedings of the 2017 IEEE International Conference on Information Visualization (IV)* (pp. 41–46).

Choudhury, S. R., & Khosla, M. (2020). Visual text analytics: A survey. *ACM Computing Surveys (CSUR)*, *53*(6), 1–38.

Chow, M., & J. B. (2018). The value of data visualization. In *Big Data Little: A Guide to Data Visualization*. O'Reilly Media.

Chung, Y., Al-Twairesh, N., & AlSaeed, D. (2018). uVSAT: A visual analytic tool for detecting and analyzing stance in social media. In *Proceedings of the 2018 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 1–12).

Dastgheib, M., & Vazirani, A. (2017). SentiCard: Real-time sentiment analysis and visualization of Twitter data. In *Proceedings of the 2017 International Conference on Data Mining, Systems Engineering and Management (DMSM)* (pp. 1–6).

Diakopoulos, N., & Shamma, D. A. (2010). Characterizing the visual narrative of online social media. In *Proceedings of the 2010 AAAI Spring Symposium on Intelligent Information for Humanistic Applications* (pp. 1–8).

Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*. Analytics Press.

Google. (n.d.). Google Chart API. Retrieved from https://developers.google.com/chart

Guo, F., Li, W., Liu, Z., & Chen, J. (2017). Social Helix: Visual analysis of sentiment divergence in social media. *IEEE Transactions on Visualization and Computer Graphics, 23*(1), 211–220.

Healey, C. G., & Enns, J. T. (2012). Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics, 18*(7), 1170–1188.

Healey, C. G., Chang, R., & Endert, A. (2019). A visualization system for online customer service chats. In *Proceedings of the 2019 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 1–12).

Keim, D. A., Kohlhammer, J., & Mansmann, F. (2013). Visual text analytics: State of the art and future trends. *IEEE Transactions on Visualization and Computer Graphics, 19*(12), 2911–2921.

Kim, B., Kim, K., & Lee, J. (2017). Examples are not enough: How to make black box models interpretable. In *Proceedings of the 2017 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 1–12).

Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.

Lu, J., & Li, F. (2013). SentiCompass: Visualizing spatio-temporal sentiment trend. In *Proceedings of the 2013 International Conference on Computer and Information Technology (CIT)* (pp. 222–229).

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97.

Mohammad, S. M., Kiritchenko, S., & Popowich, F. (2016). Large-scale sentiment analysis for effective communication in social media. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1–8).

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.

Pourkeyvan, A., Safa, R., & Sorourkhah, A. (2024). Harnessing the power of Hugging Face transformers for predicting mental health disorders in social networks. *IEEE Access, 12*, 28025–28035.

Safa, R., Edalatpanah, S. A., & Sorourkhah, A. (2023). Predicting mental health using social media: A roadmap for future development. In H. Garg & J. M. Chatterjee (Eds.), *Deep learning in personalized healthcare and decision support* (pp. 285–303). Academic Press.

Samah, K. A. F. A., Misdan, N. F. A., Jono, M. N. H. H., & Riza, L. S. (2022). The best Malaysian airline companies visualization through bilingual Twitter sentiment analysis: A machine learning classification. *JOIV International Journal on Informatics Visualization, 6*(1), 130–137.

Shammim, S., Hossain, M. S., & Al-Twairesh, N. (2018). Usability evaluation of sentiment visualization systems. *Journal of Big Data, 5*(1), 1–22.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages* (pp. 336–343).

Spence, R. (2014). *Information visualization: An introduction* (3rd ed.). Pearson Education.

TeamViewer. (n.d.). *TeamViewer: The all-in-one software for remote support*.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Graphics Press.

Twitter. (n.d.). Twitter API documentation. Retrieved from https://developer.twitter.com/

Usability.gov. (n.d.). Moderated usability testing. Retrieved from https://www.usability.gov/

Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J.-H., & Hsieh, J.-G. (2021). Twitter sentiment analysis towards COVID-19 vaccines in the Philippines using Naive Bayes. *Information, 12*(5), 204.

Wang, Q., & Lin, C. J. (2014). Sentiment visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 1845–1854.

Ware, C. (2012). *Information visualization: Perception for design* (3rd ed.). Morgan Kaufmann.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* (pp. 347–354).

Wong, B. (2010). *Visualizing data*. O'Reilly Media.

Yaqub, U., Sharma, N., Pabreja, R., Chun, S. A., Atluri, V., & Vaidya, J. (2020). Location-based sentiment analyses and visualization of Twitter election data. *Digital Government: Research and Practice, 1*(2). Retrieved from https://doi.org/10.1145/3339909

Zhao, J., Zhang, Y., & Wang, Y. (2019). StanceViz Prime: A visual analytics tool for temporal sentiment and stance analysis in social media. In *Proceedings of the 2019 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 1–12).

Zhao, Y., Qin, B., Liu, T., & Tang, D. (2016). Social sentiment sensor: A visualization system for topic detection and topic sentiment analysis on microblog. *Multimedia Tools and Applications, 75*(15), 8843–8860

Bibliographic information of this paper for citing:

Al-Ghreimil, Nadia; Alhassoun, Manal; Al-Twairesh, Nora & AlSaeed, Duaa (2026). Twitt Ray: Sentiment Visualization of Arabic Tweets. *Journal of Information Technology Management*, 18 (3), 155-180. https://doi.org/10.22059/jitm.2026.108009