Detection and Visualization of Bilingual Trending Topics

Balsam Alkouz and Zaher Al Aghbari
Department of Computer Science, University of Sharjah, Sharjah, UAE
Email: {balkouz, zaher}@sharjah.ac.ae

***Abstract*—Social media has gained a lot of popularity and become the main source of information. Recently, immediate popular news or stories, known as *trending topics* have found social networks, such as Twitter, an attractive platform for its spread. Detection of trending topics from social media is a commonly tackled issue in the data mining community as it helps in different applications, such as news agencies discovering breaking news in real-time or marketing departments detecting viral memes. In this paper, we explore trending topic detection from social media, specifically twitter, using the ratio frequency of the hashtags. The proposed system is bilingual as it detects trending topics in both Arabic and English tweets. We classify the trending topics into four classes to help identify and rank the topics of interests to the community. These trending topics are visualized on a spatio-temporal map that allows users discover the spatial distribution of these topics as well as their time intervals. We have experimented with different textual features to detect trending topics and described the results and tradeoffs of using these features.**

***Index Terms*—trend topics, hashtags, spatio-temporal visualization, data mining, Twitter**

I. Introduction

Social media has seen exceptional growth in past few years. The services provided by social media has spread in many places of the world and infiltrated all sectors of the societies. Additionally, the value of information collectively generated from social media has increased many folds in recent years. Interactions between people in social media often reflect the dynamics of the real-world events. Not only the number of users of social networks increases, the content produced about these events proportionally increases.

This content is in real-time, and so the social media streams can be used as sensors of real-world events. Different kind of events detection from social media were tackled like sicknesses detection, including Influenza [1], diabetes [2] and Asthma [3], soccer games, road traffic [4] and earthquakes [5].

An interesting characteristic of microblogging services, such as Twitter, is that a popular social event triggers the interest of users to mention it in their tweets and thus a sudden increase of tweets about such social event appear in real-time. Users of microblogging systems tweet about these popular events as they know about them. As a result, a sudden sharp increase of tweets about this popular event is produced in a certain time interval and that is known as a trending topic.

Trending topics disclose the events of great interest to the community of users in real-time. These trending topics can be breaking news, popular sports match, local social events, etc. Trending topic detection is an important application emerged from the extensive use of social media. Detecting trending topics in social media has gained a lot of interest in the data mining community and this is because of several reasons. For example, business decisions can be made in order to sell or market a product that people are interested in. Political campaigners may use detected trending topics to learn about people's interests and thus can predict voting results that can help them take certain actions to strengthen their positions. Knowing the hot topics people are talking about keep the people aware of what's happening around them. Also, trending topics can be used for entertainment and making funny or sarcastic jokes on these topics, such as when Saudi allowed women to drive and how this topic was very hot on social media platforms discussing it seriously or just making jokes about it.

Knowing trending topics is important but knowing what's trending in a certain region is even more important. The interactions on a topic in the Middle East would be different from the interactions on the same topic in the US for example, and might not be considered as a hot topic. Also knowing topics in a certain area can help in predicting future events, for example if a topic is trending in Abu Dhabi about a sandstorm then that topic will probably emerge in the next few time intervals in the neighbor cities that will get effected as well.

Twitter is a great platform to find out trending topics in many countries, because not only tweets have short content and real-time feature that helps a topic spread rapidly, but also because it has one of the largest user bases among all social media platforms. UAE is not an exception as it ranks 5th worldwide in social media usage with twitter being the third most popular platform in the country. Therefore, it would be useful and effective to use Twitter data streams emerging from the UAE to detect trending topics at the different regions of the country and visualizing it on a spatiotemporal map.

Manuscript received January 20, 2020; revised April 8, 2020.

doi: 10.12720/jait.11.2.71-77

In this paper, we introduce trending topics detection method from Twitter data streams. The proposed method supports bilingual textual data (English and Arabic) since about 80% of the residents in UAE are expats that speak mostly English. This will help identify the different topics of interests of the different communities in the country. The proposed method consists of four main components. Starting with the data collection component where the data is collected from the Twitter data stream. Second is the feature selection component where tweets are processed to extract theirs features, time and locations. Third is the trending topic detection component that classifies the tweets and thus identifies and ranks the topics based on the extracted features. Finally, the visualization component that displays the trending topics on top of a spatio-temporal map, which is updated frequently. In our experiments, we investigate the effect of using the list of all keywords in a tweet instead of using the hashtags in trending topics detection.

The remainder of the paper is organized as follows. Section II discusses the related work. In Section III, we define the problem. The proposed trending topic detection system is presented in Section IV. Then, the results are analyzed in Section V. Finally, the paper is concluded in Section VI.

II. Related Work

Identifying the topics being talk over by online social media users is the first step towards mining knowledge from social data stream. Similar to detecting phenomena from stream processing [6], [7], topic detection and monitoring starts by extracting topics from a stream of textual information sources, such as Twitter. Then, the detected topics are quantified to know their spatio-temporal trend.

Twitter is an important social media platform for researchers to investigate how information is shared online and how users behave. Some works focused on studying users interactions via social media such as [8]-[10]. Other works studies information diffusion on the social network [11]-[13]. The content of tweets were investigated by [14]-[18].

The authors of [19] used the Frequent Patterns Stream algorithm (FP-Stream) with some modifications to detect the rending topics on twitter. They considered each tweet as a transaction and each word in a transaction as an item. The frequent word sets produced is what's considered trending at that time. The word sets are divided into three classes based on a threshold, frequent, sub-frequent and infrequent. They detected topics daily as they treated the tweets generated in one day as one batch.

Another paper [20] detected social hot topics with location consideration by collecting tweets tweeted from North America. They found out that the most frequent words mentioned are not the social hot topics but rather the emotional words expressing feelings (e.g. lol, like and love). And so, instead of depending on word frequency they used word frequency differences, which identifies increasing or decreasing frequency of keywords and suppresses the emotional words. They also found out that they could find geographic communities, as they were able to follow a storm being a trending topic as it moved from one region to another.

μ-TOP [21] is a system created to detect and summarize trending topics based on spatiotemporal criteria. For each location and time window the system finds the trending topics and generates a summary of relevant posts. They created a user interface to visualize and explore the detected subjects with a map and a timeline.

The authors of [22] introduced a different approach for detecting trending topics that does not depend on the word frequency; they proposed a link-based approach. This approach focuses on the mentioning of hundreds users instead of the textual content. They found that the performance of their approach is better from the traditional one.

[23] experimented 6 different techniques in emerging twitter topics detection and found out that the best method is the one that leverages the n-gram occurrences and the tf-idf method for topic ranking.

Another method that isn't based on the textual content of the tweet but rather the social features associated with each tweets, like the length of the tweet and the number of retweets, was used in classifying a tweet to different types of trending topics like news, ongoing events, memes and commemoratives using an SVM classifier [24].

Semantic graphs were also used for keyword extraction and summarization from the tweets in [25], the keywords were ranked and they observed that important keywords are likely to appear with other important keywords. [26] approach is divided into two parts, the first is the preprocessing of the tweets which includes tweets collection, keywords extraction and constructing a thesaurus of the keywords using semantic keywords' treatment. The second part is the trending topic detection where the topic that has the highest probability in a distribution of topics is a trending topic and tweets about the same topic are grouped into the same cluster.

Personalizing trending topics was tackled in [27], so that a trending topic could be recommended to a user before it's outdated. An important feature of hot topic personalization is it's earliness. Other approaches for personalization are not able to deal with trending topics since they rely on the accumulation of information. They proposed an approach that uses the posts with the interactions to find the historical posts from likeminded users.

Another application of trending topics detection is detection trending rumors on social media [28], the tweets are identified using regular expressions, clustered based on the content of the tweets, and the clusters are finally ranked in or order of likelihood that the tweets in the clusters are rumors.

III. Problem Statement

Let T be the set of considered tweets that contains the Arabic and English sets of captured tweets T^a and T^e, respectively, where $\{T^a, T^e \in T\}$. The Arabic set of tweets

$T^a = t_1^a, t_2^a, \ldots, t_n^a$ has n tweets and English set of tweets $T^e = t_1^e, t_2^e, \ldots, t_m^e$ has m tweets. Each tweet t_i, English or Arabic, contains a set of textual feature $t_i = f_1, f_2, \ldots, f_g$, where f_c is a hashtag or stemmed word extracted from the text of the tweet. Let P be a set of trending topics $P = p_1, p_2, \ldots, p_h$ and each trending topic p_i is associated with a subset of tweets w_j in T^a and/or T^e.

Given the set of tweets collected from the UAE region T, we aim to extract the set of trending topics P, where each trending topic p_j is associate with a subset of tweets $w_j \in T$. It is assumed that the tweets in w_j are spatially and temporally close.

Trending topics appear on social media at different rates. Users of social media are motivated by different events happenings, in their real life communities such as on TV shows, in the neighborhood, or on the Internet. As a result, users talk about these events on social media producing a sudden increase in tweets related to the topic being discussed. Therefore, this topic is considered trending. When analyzing the stream of tweets for a period of time, one can notice that different types of topics are being discussed and the number of users discussing some of these topics experienced sudden increases, which are signs indicating that these topics are trending. Following the classification in [29], these trending topics can be divided into four categories:

- *News*: breaking news motivate users to share tweets about them. In many occasions, some breaking news find their way to the Twitter stream before many TV news. A good share of the trending topics on social media are news related. A trending topic is categorized as news if it is related to an important event that has been broadcasted by major news stations.
- *Events*: a trending topic produced by online social media users tweeting about an ongoing event. Nowadays, Twitter users discuss and share information about events that are taking place at the time of the discussion on Twitter. For example, users my talk about a concert or conference taking place at the time when these online users are sharing the information on Twitter.
- *Memes*: a significant share of the trending topics are prompted by interesting ideas started by an online social media user. Such ideas go viral because other online users find them interesting or they are coming from a famous personality. Therefore, these topics are apparently newsworthy, funny or attractive that many online users are following. For example, a meme can be a tweet from a famous comedian in support of a presidential candidate.
- *Commemoratives*: The least frequent trending topic are those that are produced in commemoration of certain person or event. Such trending topics are initiated by users to celebrate the anniversary of a certain event or person.

However, some trending topics could fall into more than one category. For example, when the supporters of a certain soccer team cheer their team by tweeting after the team scores a goal, this could be considered an ongoing event and as news at the same time.

In this work, we focus on detecting trending topics in general. Therefore, the proposed system should address the following questions: will the proposed system detect the tending topics on twitter in a given period and distance window? Can it plot the detected trending topics in real-time on a spatiotemporal map? Will the topic detection differ if we rely on all the meaningful words (nouns) as keywords instead of the hashtag alone? And, can the proposed system help individuals and decision makers understand the trends occurring in a certain region and at some time interval?

IV. Detection of Trending Topics

Our system consists of four main components; Data collection, Tweets Feature Extraction and Trending Topic Detection. Fig. 1 shows these components and the following sections will explain the steps carried out in each component.

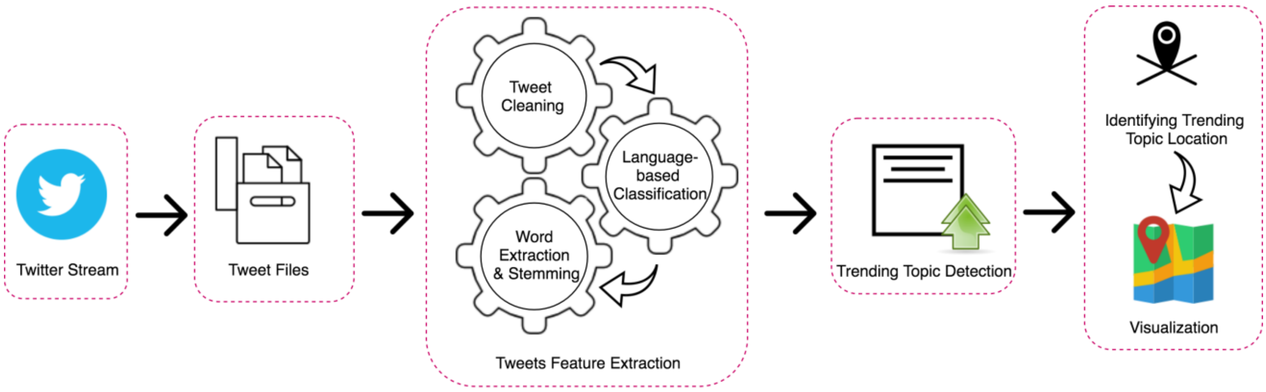

Figure 1. System architecture.

A. *Data Collection*

Data is collected from Twitter stream to serve the purpose of this research. Tweets are collected from October 1, 2017 until October 28, 2017 using the Tweepy Streaming API. The geolocation of the tweets collected are specified to be from UAE. Tweets are collected in all languages but only the English and Arabic tweets are considered. We collected a total of 578301 tweets that consists of 112435 tweets in Arabic and 113095 in English. These tweets include 76442 unique trending topics in Arabic and 84979 trending topics in English.

B. *Tweets Feature Extraction*

Tweets are passed to the Feature Extraction engine to perform three main tasks:

1) *Tweet cleaning*

The collected raw tweets contain many attributes including the location, user id, timestamp and the tweeted text itself. The tweets may also include elements like hyperlinks, user mentions, hashtags and emojis. Thus, the tweets of both English and Arabic languages go through a cleaning process to remove the following:

a) Hyperlinks: links to websites, images, and videos were filtered from the raw tweets.

b) User Mentions: users mention other users in order to communicate and link the mentioned user to their tweet.

c) Digits and Punctuation: number characters and punctuation do not provide any additional meaning indicating an occurring case, and hence, they are also removed.

d) Emojis: the same emoji can be used for different purposes, and so the existence of them adds no value to our detection and hence, they are deleted.

2) *Language-based Classification*

After tweet cleaning and removing the extra spaces from the text, tweets are analyzed and classified based on their language into four groups: Arabic, English, Combined language (English and Arabic) and others. The proposed system then extracts features from the Arabic and English tweets.

3) *Word extraction and stemming*

The Arabic and English datasets are then separately go through the process of word extraction. First, the datasets are separately processed to remove the language-specific stop words of each tweet [30]-[33]. Then, words of the tweets undergo a stemming process to extract their roots. The following is applied to both datasets:

a) Stop Words Removal: English stop words, such as and, the, etc., are removed from the English cleaned tweets, as their existence do not add a meaning for the detection. Similarly, Arabic stop words, such as *اللذين، أيضا*, are removed from the Arabic cleaned tweets.

b) Word Stemming: We apply the NLTK software package to stem each words of a tweet. For example, the word universities in an English tweet is stemmed to university and the word مدارس in Arabic tweet is stemmed to درس.

C. *Trending Topic Detection*

Trending topics discussed in Twitter are simply manifested by groups of tweets that share the same or similar hashtags. The proposed system also requires the group of tweets to be close that is they appear in a continuous time interval. Furthermore, the group of tweets should be spatially close. Since our emphasis in this paper is the UAE region, we consider all Emirates to be spatially close due to the relative size of the country and the relative close distances between their cities.

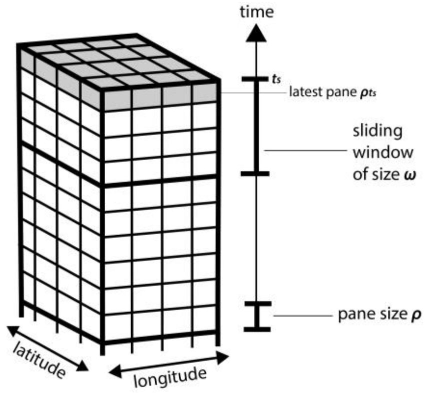

Figure 2. Pane and Window storage concept.

Hashtags are a form of trending topics in twitter. Table I and Table II illustrate examples of Arabic and English tweets that are associated with trending topics, respectively, and the extracted hashtags that depict these trending topics. Topics vary in their popularity and thus it is useful to know the popularity level of the topic discussed on Twitter.

TABLE I. EXAMPLES OF TRENDING TOPIC IN ARABIC TWEETS

Arabic Tweet	Trending Topic
شهر أكتوبر هو شهر التوعية بخطورة #سرطان_الثدي	#سرطان الثدي
The month of October is the month of awareness of #Breast Cancer	#Breast Cancer
في #يوم_القهوة_العالمي قلنا نوع قهوتك المفضلة والمكان الي تحب تشربها فيه	#يوم القهوة العالمي
In the #Coffee World Day tell us about your favorite coffee and where you like to drink it	#Coffee World Day

TABLE II. EXAMPLE OF TRENDING TOPICS IN ENGLISH TWEETS

English Tweet	Trending Topic
Just registered our school for the #Dubai 30X30 ! Can't wait!	Dubai30X30
Discuss Your App Idea. Turn It Into Reality! Meet us at @GITEXTechWeek	GitexTechWeek

To detect the trending topic, the count of the textual feature, such as a hashtag, mentioned in a pane (e.g. hour) is divided by the count of the textual feature mentioned in a window (e.g. day) to get the ratio of the frequency in that pane as shown in Eqn. 1. Where Cr is the count of hashtag in the pane and Cw is the count of the hashtag in the window.

For the latest trending topic ratio calculation, the last pane and window are used as shown in Eqn. 2. Where $C\rho_{ts}$ is the count of the hashtag in the latest pane and $C\omega_{ts}$ is the count of the hashtag in the latest window.

A topic ratio of more than 0.75 is considered to be "Very Hot", more than 0.5 to 0.75 is considered "Hot", more than 0.25 to 0.5 is "Trending" and less than 0.25 is considered "Not Popular". To visualize the concept of the pane and window look at Fig. 2.

$$TR = \frac{C\rho}{C\omega} \quad (1)$$

$$LTR = \frac{C\rho_{ts}}{C\omega_{ts}} \quad (2)$$

V. Visualization and Result Analysis

Here we present our spatiotemporal map and discuss the results of our experiments.

A. Spatiotemporal Map

A spatiotemporal map is built to visualize the detected trending topics. The map allows visualizing the distribution of these trending topics as well as the time interval at which the topics were trending. Therefore, first, the proposed system determines the location of each trending topic and then it places the representing pin of each trending topic on the spatiotemporal map.

1) Identifying trending topic location

Collected Tweets are associated with longitude and latitude coordinates that may be used as the location of the trending topic on the spatiotemporal map. Therefore, the location is determined by the coordinates attribute contained in the tweets collected from the Twitter API. Since each trending topic is associated with several tweets, we first identify these tweets, extract their locations, and then compute the average location, which is the arithmetic mean of the longitudes and latitudes of the tweets associated with the trending topic in question. The average location determines the location of a pin that represents the trending topic on the spatiotemporal map, see Fig. 3.

2) Visualization

After detecting the trending topics, the results are visualized on a map using the Google Maps API. The topics are placed on the map using pins based on the geolocation extracted in item 1. Each pin has an information window on click that shows the trending topic. The colors of the pins are based on the hotness of a topic, a "Very Hot" topic is presented in red, a "Hot" topic is presented in orange, a "Trending" topic is presented in green and a "Not Popular" topic is presented in blue. Fig. 3 shows the interface. The visualization interface allows user to navigate temporally. That is a user can go back in time on the spatiotemporal map to see the trending topics previous intervals as shown in Fig. 4.

B. Detecting Trending Topics from Tweet Words

We evaluated the effect of using all the cleaned and stemmed words instead of hashtags only to detect trending topics. The main issue we noticed when using the words to detect trending topics is that the number of "Non Popular" topics is higher than when using hashtags in the detection of trending topics.

This is because using words in the detection of trending topics results in topics that are associated with the most frequent word in the examined tweets rather than the actual trending topics. Note that the most frequent words do not really describe a topic, but rather are adjectives or feelings. This finding supports what the authors of [20] claimed that trending topics are the words mentioned with these trending emotions or feelings words.

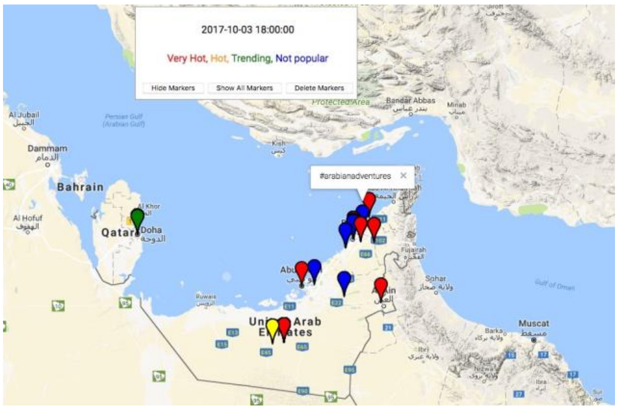

Figure 3. Spatiotemporal map for visualizing trending topics.

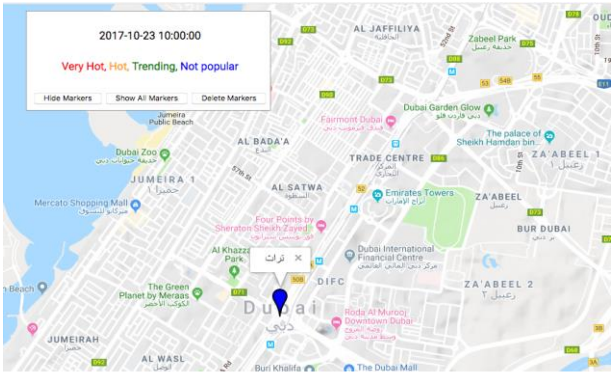

Figure 4. Temporally navigating the spatiotemporal visualization map.

VI. Conclusion and Future Work

In conclusion, we can say that mining social media for trending topics is important and the proposed system is effective in detecting trending topics from Twitter data stream. The emphasis of our experiments is the trending topics in the UAE, thus our system detects bilingual trending topics. To visualize the detected trending topics, we developed a spatiotemporal map that shows trending topics depicted by pins where the color of these pins represent the level of popularity of the topic. Our experiments confirm that using hashtags outperform the use of tweets' cleaned and stemmed words in detecting trending topics.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Balsam and Zaher conducted the research, analyzed the data and wrote the paper; Balsam implemented the idea; all authors had approved the final version.

References

[1] B. Alkouz and Z. A. Aghbari, "Analysis and prediction of influenza in the UAE based on Arabic tweets," in *Proc. IEEE 3rd International Conference on Big Data Analysis*, 2018, pp. 61-66.

[2] F. Ali, "Online diagnosis of diabetes with twitter data," Dissertations & theses gradworks, 2015.

[3] S. Ram, W. Zhang, M. Williams, and Y. Pengetnze, "Predicting asthma-related emergency department visits using big data," *IEEE J. Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1216-1223, 2015.

[4] N. A. Emadi, S. Abbar, J. Borge-Holthoefer, F. Guzman, and F. Sebastiani, "Qt2s: A system for monitoring road traffic via fine grounding of tweets," arXiv preprint arXiv:1703.04280, 2017.

[5] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 919-931, 2013.

[6] Z. A. Aghbari, I. Kamel, and T. Awad, "On clustering large number of data streams," *Intelligent Data Analysis*, vol. 16, no. 1, pp. 69-91, 2012.

[7] Z. A. Aghbari, I. Kamel, and W. Elbaroni, "Energy-efficient distributed wireless sensor network scheme for cluster detection," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 28, no. 1, pp. 1-28, 2013.

[8] A. Java, X. Song, T. Finin, and B. Tseng, "Why we Twitter: Understanding microblogging usage and communities," in *Proc. the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, New York, NY, USA, 2007, pp. 56-65.

[9] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proc. the 19th International Conference on World Wide Web*, New York, NY, USA, 2010, pp. 591-600.

[10] D. Boyd, S. Golder, and G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter," in *Proc. the 2010 43rd Hawaii International Conference on System Sciences*, Washington, DC, USA, 2010, pp. 1-10.

[11] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts, "Who says what to whom on Twitter," in *Proc. the 20th International Conference on World Wide Web*, New York, NY, USA, 2011, pp. 705-714.

[12] B. Krishnamurthy, P. Gill, and M. Arlitt, "A few chirps about Twitter," in *Proc. the First Workshop on Online Social Networks*, 2008, pp. 19-24.

[13] J. Yang and J. Leskovec, "Modeling information diffusion in implicit networks," in *Proc. IEEE International Conference on Data Mining*, 2010, 599-608.

[14] B. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 11, pp. 2169-2188, 2009.

[15] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in Twitter to improve information filtering," in *Proc. the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 2010, pp. 841-842.

[16] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: Real-time event detection by social sensors," in *Proc. the 19th International Conference on World Wide Web*, New York, NY, USA, 2010, pp. 851-860.

[17] J. Hurlock and M. L. Wilson, "Searching Twitter: Separating the tweet from the chaff," in *Proc. the 5th International AAAI Conference on Weblogs and Social Media*, 2011.

[18] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: A content-based approach to geo-locating twitter users," in *Proc. the 19th ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2010, pp. 759-768.

[19] J. Guo, *et al.*, "Mining hot topics from Twitter streams," *Procedia Computer Science*, vol. 9, pp. 2008-2011, 2012.

[20] H. G. Kim, S. Lee, and S. Kyeong, "Discovering hot topics using Twitter streaming data social topic detection and geographic clustering," in *Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2013, pp. 1215-1220.

[21] P. Mehta, *et al.*, "µtop: Spatio-temporal detection and summarization of locally trending topics in microblog posts," in *Proc. EDBT*, 2017, pp. 558-561.

[22] B. S. Deshmukh, B. K. Vyankatesh, and B. N. Ravi, "Detection and analysis of Twitter trending topics via link-anomaly detection," *International Journal of Engineering Research & Applications*, 2015.

[23] L. M. Aiello, *et al.*, "Sensing trending topics in Twitter," *IEEE Transactions on Multimedia*, vol. 15, no. 6, pp. 1268-1282, 2013.

[24] A. Zubiaga, D. Spina, R. Martinez, and V. Fresno, "Real-time classification of twitter trends," *Journal of the Association for Information Science and Technology*, vol. 66, no. 3, pp. 462-473, 2015.

[25] N. Avudaiappan, A. Herzog, S. Kadam, Y. Du, J. Thatche, and I. Safro, "Detecting and summarizing emergent events in microblogs and social media streams by dynamic centralities," in *Proc. IEEE International Conference on Big Data*, 2017, pp. 1627-1634.

[26] A. Madani, O. Boussaid, and D. E. Zegour, "Real-time trending topics detection and description from twitter content," *Social Network Analysis and Mining*, vol. 5, no. 1, p. 59, 2015.

[27] L. Wu, X. Hu, and H. Liu, "Early identification of personalized trending topics in microblogging," in *Proc. ICWSM*, 2017, pp. 692-695.

[28] Z. Zhao, P. Resnick, and Q. Mei, "Enquiring minds: Early detection of rumors in social media from enquiry posts," in *Proc. of the 24th International Conference on World Wide Web*, 2015, pp. 1395-1405.

[29] B. Alkouz, Z. A. Aghbari, and J. H. Abawajy, "Tweetluenza: Predicting flu trends from twitter data," *Big Data Mining and Analytics*, vol. 2, no. 4, pp. 248-273, 2019.

[30] L. Dinges, A. Al-Hamadi, M. Elzobi, Z. A. Aghbari, and H. Mustafa, "Offline automatic segmentation based recognition of handwritten Arabic words," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 4, no. 4, pp. 131-143, 2011.

[31] M. Elzobi, A. Al-Hamadi, Z. A. Aghbari, L. Dings, and A. Saeed, "Gabor wavelet recognition approach for off-line handwritten Arabic using explicit segmentation," in *Image Processing and Communications Challenges*, Springer Heidelberg, 2014, vol. 5, pp. 245-254.

[32] B. Alkouz and Z. A. Aghbari, "Leveraging cross-lingual tweets in location recognition," in *Proc. IEEE International Conference on Electro/Information Technology*, 2018, pp. 84-89.

[33] L. Dinges, M. Elzobi, A. Al-Hamadi, and Z. A. Aghbari, "Synthizing handwritten Arabic text using active shape models," in *Image Processing and Communications Challenges*, Springer, Berlin, Heidelberg, 2011, vol. 3, pp. 401-408.

Balsam Alkouz is a master's degree holder in computer science from the University of Sharjah, Sharjah, United Arab Emirates. She also works as a Research Assistant in the Data Mining and Multimedia Research Group under the Research Institute of Sciences and Engineering at the University of Sharjah. Alkouz's interests focus on Data Mining and Social Media Analysis but also include Augmented and Virtual Reality technologies. She is a Google Developer Group manager and a Women Techmaker leader in Sharjah, UAE.

Zaher Al Aghbari a Professor and Chairman of the Department of Computer Science at University of Sharjah, UAE. He received his B.Sc. degree from the Florida Institute of Technology, Melbourne, USA, in 1987, and the M.Sc. and Ph.D. degrees in computer science from Kyushu University, Fukuoka, Japan, in 1998 and 2001, respectively. He was with at the Department of Intelligent Systems, Kyushu University, Japan, from 2001 to 2003. Since 2003, he has been with the Department of Computer Science, University of Sharjah, UAE. Currently, he is a Professor of databases and data mining in the Department of Computer Science, College of Sciences, University of Sharjah. His research interests include data mining, multimedia databases, big data, social networks, wireless sensor networks, data streams, and Arabic handwritten text retrieval.