Turkish Journal of Electrical Engineering & Computer Sciences
http://journals.tubitak.gov.tr/elektrik/
Research Article
Turk J Elec Eng & Comp Sci
(2019) 27: 407 – 420

doi:10.3906/elk-1712-17

Data analysis through social media according to the classified crime

Serkan SAVAŞ[1,*], **Nurettin TOPALOĞLU**[2]
[1]Yenikent Ahmet Çiçek Vocational and Technical Anatolian High School, Ankara, Turkey
[2]Faculty of Technology, Gazi University, Ankara, Turkey

Received: 02.12.2017 • **Accepted/Published Online:** 01.11.2018 • **Final Version:** 22.01.2019

Abstract: The amount and variety of data generated through social media sites has increased along with the widespread use of social media sites. In addition, the data production rate has increased in the same way. The inclusion of personal information within these data makes it important to process the data and reach meaningful information within it. This process can be called intelligence and this meaningful information may be for commercial, academic, or security purposes. An example application is developed in this study for intelligence on Twitter. Crimes in Turkey are classified according to Turkish Statistical Institute criminal data and keywords are defined according to this data. A total of 150,000 tweet data in the Turkish language are collected from Twitter between specified dates and processed by Turkish Zemberek natural language processing. It is seen that 56% of the people are talking about terrorist attacks and bombing attacks on the study dates. The words “bomb,” “terror,” “attack,” “organization”, and “explode” have percentages of 24%, 12%, 8%, 6%, and 6%, respectively. Moreover, associations between words and situations are found. Correlations are important to create new subclusters like “terror” and “rape” in this study with 0.90 correlation. Bigger masses can be accessible by expanding keyword groups to have a clear picture of the real situation.

Key words: Big data, social media, Twitter stream, Zemberek-NLP, data mining, text mining, commercial intelligence, academic intelligence, security intelligence, cyber intelligence

1. Introduction

The primary purpose of Internet users today is to use social media. Most users use the Internet for social and entertainment purposes [1]. Accordingly, new social media sites are starting to broadcast every day. Social media websites such as Facebook, Twitter, YouTube, Instagram, LinkedIn, and Google+ are actively used by millions of users every day. Users upload various data on the Internet, including private information such as text, pictures, videos, and audios. This information is a treasure because it contains personal data. This data must be processed to achieve the desired information among these treasures. However, these data, uploaded to the Internet, increase exponentially. According to a report published by the Computer Sciences Corporation, the data size will increase 4300% by 2020 compared to today [2]. With the addition of volume, velocity, variety, and/or reality dimensions [3] to data mining, the “big data” concept has appeared. The big data discipline collaborates with different disciplines for different purposes, where data are processed for commercial, academic, and security purposes.

Commercial cyber intelligence (CCI) can be seen anytime by any single user during daily Internet use. After searching for a subject in search engines, a user is likely to see some offers about that subject in his/her

*Correspondence: serkan_savas@hotmail.com

social media pages. For example, to obtain such a piece of information as "person X searched for a hotel for holiday" is intelligence because intelligence is "receiving information" in general. Offering hotels to person X on Facebook, Twitter, etc. means using this cyber intelligence for commercial purposes. Besides this micro CCI, numerous macro CCI studies have been carried out for companies and the economics world. Researchers studied estimation of share ratios of Dow Jones Industrial Average companies [4] and Dow Jones, NASDAQ, and S&P 500 to find the correlation between people's sentiment and shares [5] or to show how social media can be used in sales and marketing [6]. In addition, CCI from social media has been used to provide opinions for decision makers such as Nokia, T-Mobile, IBM, KLM, and DHL [7] and the three largest pizza chains: Pizza Hut, Domino's Pizza, and Papa John's Pizza [8]. It is also used for others by using consumer decision journey [9,10] and also sentiment analysis of users toward the target products [11]. CCI from social media is employed in the automotive sector for decision support systems as well to get information from user discussions [12,13].

Academic cyber intelligence (ACI) is opening up new studies with the knowledge gained by analyzing data in the cyber environment and to show the potential of the data flowing in the cyber world.

Different methods like the two-stage system of dynamic stochastic block model with temporal Dirichlet process [14] are offered to discover meaningful information with correct sentiment analysis in social media [15] or to find out the popularity of tweets according to their subjects [16]. ACI is used to discover roles and key actors in networks and communities and opinions, beliefs, and sentiments of the set of actors [17] to improve the polarity detection [18] with a new hybrid approach to overcome some limitations of Twitter sentiment analysis [19].

More new frameworks to comprehensively analyze enterprise social networks [20] like MASS-FARM [21] or the semisupervised fuzzy product ontology mining algorithm are being developed every year [22]. Researchers have also developed new portals like the Dark Web Forum Portal to test effects of their frameworks [23] and the visual analysis system called OpinionFlow to empower analysts to detect opinion propagation patterns and glean insights [24]. Furthermore, the equivalence of web-based social media data and traditional pen-and-paper methods has been tested [25]. ACI studies have such types as geoinformatics in tweets [26], human robot interaction from social media [27], understanding and defining characteristics of students of the 21st century [28], link prediction on different social media web sites like Twitter and Foursquare [29,30], or language clustering on Wikipedia [31].

The complex behaviors of social systems and their dynamics have been studied extensively. It is very important for people to define a suitable system for designing and managing the complexity to stop undesirable cascade effects to save lives [32]. It is emphasized that statistical physics of crime can relevantly inform the design of successful crime prevention strategies. In addition, previous studies highlighted the valuable theoretical resources that can help people bridge the widening gap between data and models of criminal activity [33]. Therefore, one of the most important academic subjects to study is cyber intelligence for security aims.

With the reflection of social networks in daily life, individual, institutional, and state security issues have emerged at cyber level. Ensuring individual and institutional security means defending people against possible external threats. In the virtual world transforming currently into real life itself, there are similar dangers that individuals and institutions encounter in real life. Cyber intelligence is different in terms of state security. This difference can be preventing the attacks and infiltration or prediction, diagnosis, and prevention of possible probabilities in advance by cyber intelligence.

A new intelligence type called social media intelligence is defined and how to do social intelligence was explained in [34]. It is also explicated that cyber attacks can be classified as a new type of war, because they threaten national security and happen in cyber space that does not have any defined borders in international relations [35]. The effects of social media sites on the masses have begun to be scrutinized more carefully after the events of the Arab Spring in the world and researchers studied the effect of social media and how governments and states have taken part in this [36–38].

Another type of security cyber intelligence (SCI) from social media websites is analyzing situations with citizen-driven information processing through Twitter services using data from social crises: the Mumbai terrorist attacks in 2008 [39], the Toyota recall in 2010, and the Seattle café shooting incident in 2012 [40] or G20 protests in Toronto in June of 2010 [41]. As part of the focus on transparency, the Obama administration emphasized the use of e-government and new social media services to open up access to government and challenges in social media and e-government have been examined in the US government [42–44]. SCI is also used for crime prediction in the USA by using Twitter-specific linguistic analysis [45] and managing crisis situations from the routine (e.g., traffic, weather crises) to the critical (e.g., earthquakes, floods) [46].

For Surowiecki, large groups of people are smarter than a small elite, no matter how brilliant or better they are at solving problems, fostering innovation, coming to wise decisions, or even predicting the future [47]. With the growth of social media, it has become more important to understand the ideas of the community.

Therefore, this case study examines the wisdom of crowds on crime issues from social media to show the relationship between social whispers and the current crime situation in Turkey. The remainder of the paper is organized as follows. Section 2 is about the methodology. Section 3 explains the results in depth. Section 4 discusses and concludes with suggestions for future research.

2. Materials and methods

Twitter is one of the leading websites where big data applications are used. With millions of active users and large amounts of data produced by these users these days, it attracts the attention of researchers. There are 500 million tweets per day and about 6000 tweets per second by active users, tweeted users, and created users (Figure 1).

Twitter supports researchers with different application programming interfaces (APIs). This situation is win-win for both researchers and Twitter. Researchers have the opportunity to investigate different relationships between users and Twitter makes advertisements and promotions worldwide through researchers. Twitter has developed REST API, Stream API, and Ads API for researchers to use these APIs for different purposes. The REST APIs provide programmatic access to read and write Twitter data. The data responses are available in JavaScript Object Notation [48]. The Stream APIs give developers low latency access to Twitter's global stream of tweet data. The differences between REST & Stream APIs are shown in Figure 2.

The Ads API program enables businesses to create and manage ad campaigns programmatically on Twitter.

2.1. Twitter application

There are some steps to use Stream API such as creating a user account on Twitter, creating a new application on Twitter for the user, and generating keys and auth values of new applications.

After we created our own Twitter application, a model was created to work with the data taken from Twitter (Figure 4).

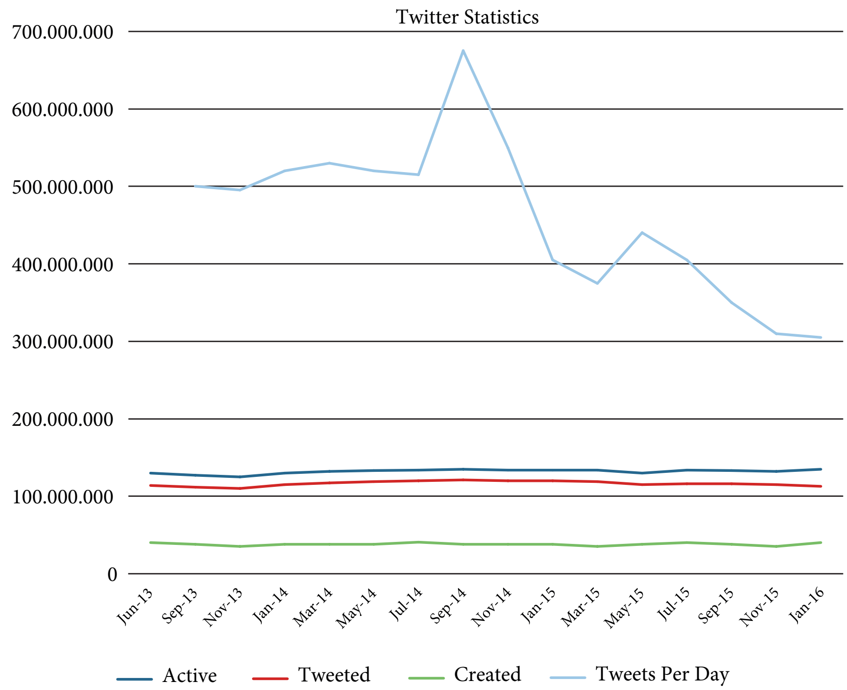

Figure 1. Twitter statistics.

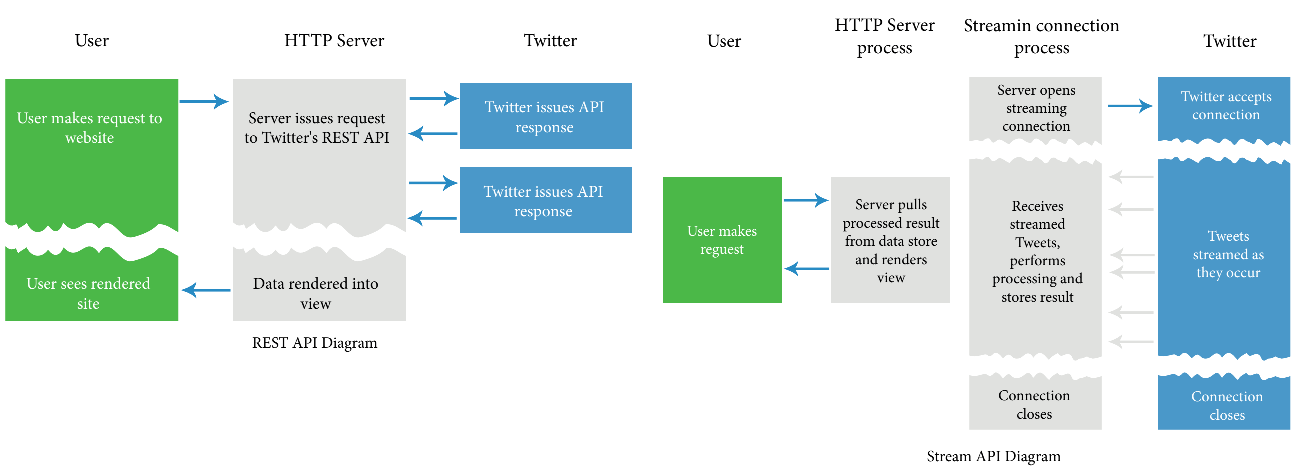

Figure 2. REST and Stream API diagrams.

2.2. Data collection

To get real-time tweet data from Twitter with key and auth values, the Python programming language and Tweepy libraries were used. A Turkish language filter and nine different keywords were used while streaming real-time data.

For Turkish Statistical Institute (TÜİK) data in 2013, criminal cases in Turkey are classified as: assault, theft, opposition to distraining, drug production, weapons crimes, murder, forgery, threatening, spoil, violation of protection of the family, sexual crimes, insult, damage to property, smuggling, using drugs, fraud, restrictions on freedom of the people, taking out of job, traffic offenses, forest crimes, debit, bribe, opposition to military

Figure 3. Twitter apps management.

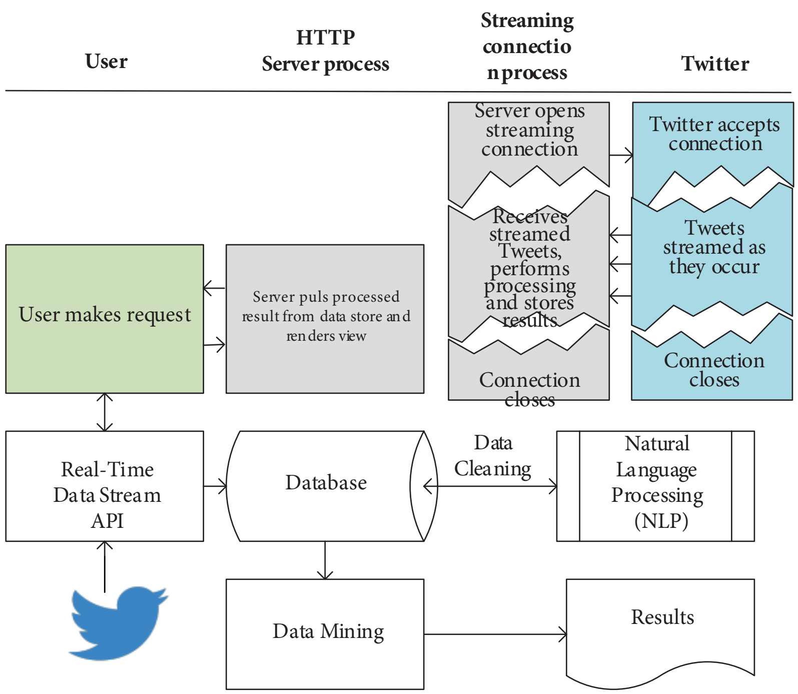

Figure 4. Flow chart of the model.

criminal law, mistreatment, and other crimes. "Bomb," "terror," "custody," "attack," "demonstration," "smuggling," "drugs," "prostitution", and "rape" were chosen as keywords according to these crime classes. A flow chart of the program is shown in Figure 5.

A total of 158,463 tweet data, nearly 900 MB in size, were taken from Twitter from 3 May 2016 to 5 June 2016. There were 2,101,550 words and 19,074,070 characters in this tweet text group. Zemberek-NLP, which is one of the most used NLP libraries for the Turkish language, was used to analyze these tweet data. It provides basic statistical tools for letters, words, roots, etc. In the beginning, it was started for Turkish, but then it was further developed to contain other Turkic languages [49].

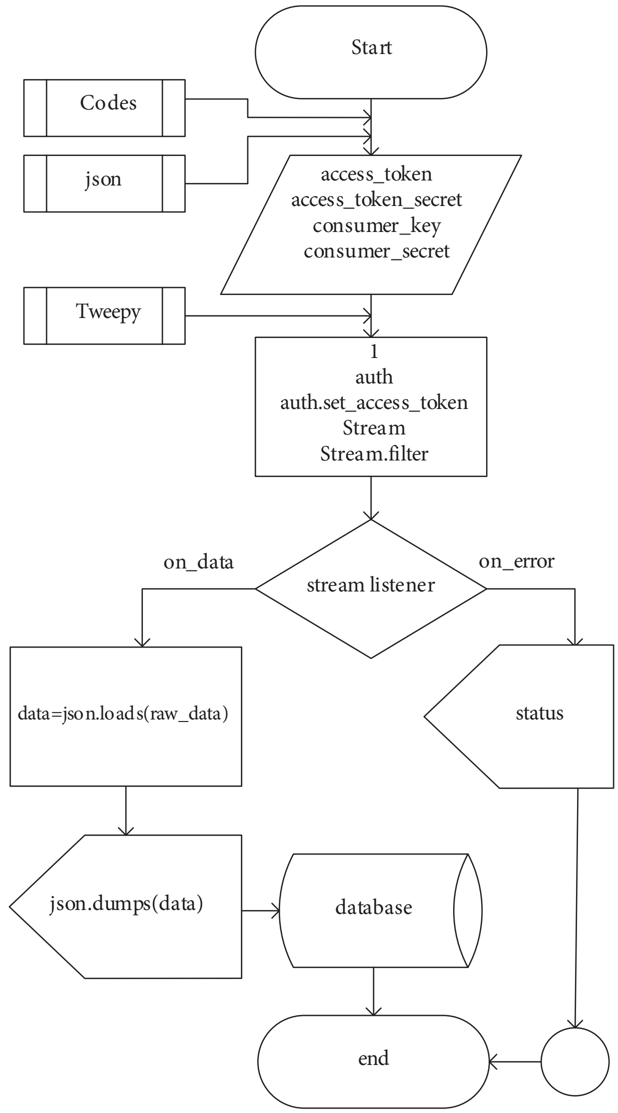

Figure 5. Flow chart of the program.

Twitter provides information such as user name and ID, time zone, time, location, and retweet and favorite count, with requested data as partially shown in Figure 6.

3. Results

Important information inside a database can be analyzed like solving a puzzle. These analyses can be performed using different methods and techniques such as machine learning. NLP tools are the most common tools. A preprocessing task is needed to extract required information from unrefined sentences to use them in the core part of the system [50]. Words in Tweets are divided into roots using Zemberek-NLP and there are 301,690 roots in the data (Figure 7).

The words in Figure 7 have been translated from Turkish to English. In these keywords, “bomb” has the highest percentage (24%), followed by “terror,” “attack,” “organization”, and “explode” with rates of 12%,

```
{
"id_str": "737021112543027200",
"entities": {
"hashtags": [],
"user_mentions": [],
"symbols": [],
"urls": []},

"retweeted": false,
"in_reply_to_user_id": null,
"retweet_count": 0,
"lang": "tr",
"in_reply_to_user_id_str": null,
"timestamp_ms": "1464554501077",
"truncated": false,
"place": null,
"in_reply_to_screen_name": null,
"in_reply_to_status_id": null,
"contributors": null,
"geo": null,
"user": {
"contributors_enabled": false,
"id_str": "1613181817",
"verified": false,
"time_zone": "Istanbul",
"is_translator": false,
"profile_background_image_url":
"http://pbs.twimg.com/profile_background_images/551870242066923520/itpPJQXV.jpeg",
"default_profile": false,
"profile_background_image_url_https":
"https://pbs.twimg.com/profile_background_images/551870242066923520/itpPJQXV.jpeg",
"profile_link_color": "1B95E0",
"id": 1613181817,
"url": null,
"protected": false,
"lang": "tr",
"friends_count": 77624,
"profile_sidebar_fill_color": "252429",
"default_profile_image": false,
"favourites_count": 38416,
"location": "İstanbul, Türkiye",
"profile_use_background_image": true,
"followers_count": 87049,
"description": null,
"follow_request_sent": null,
"screen_name": "",
"profile_text_color": "666666",
"profile_background_tile": true,
"created_at": "Mon Jul 22 16:17:35 +0000 2013",
"profile_sidebar_border_color": "000000",
"listed_count": 12,
"profile_image_url_https":
```


Figure 6. Part of information sent by Twitter.

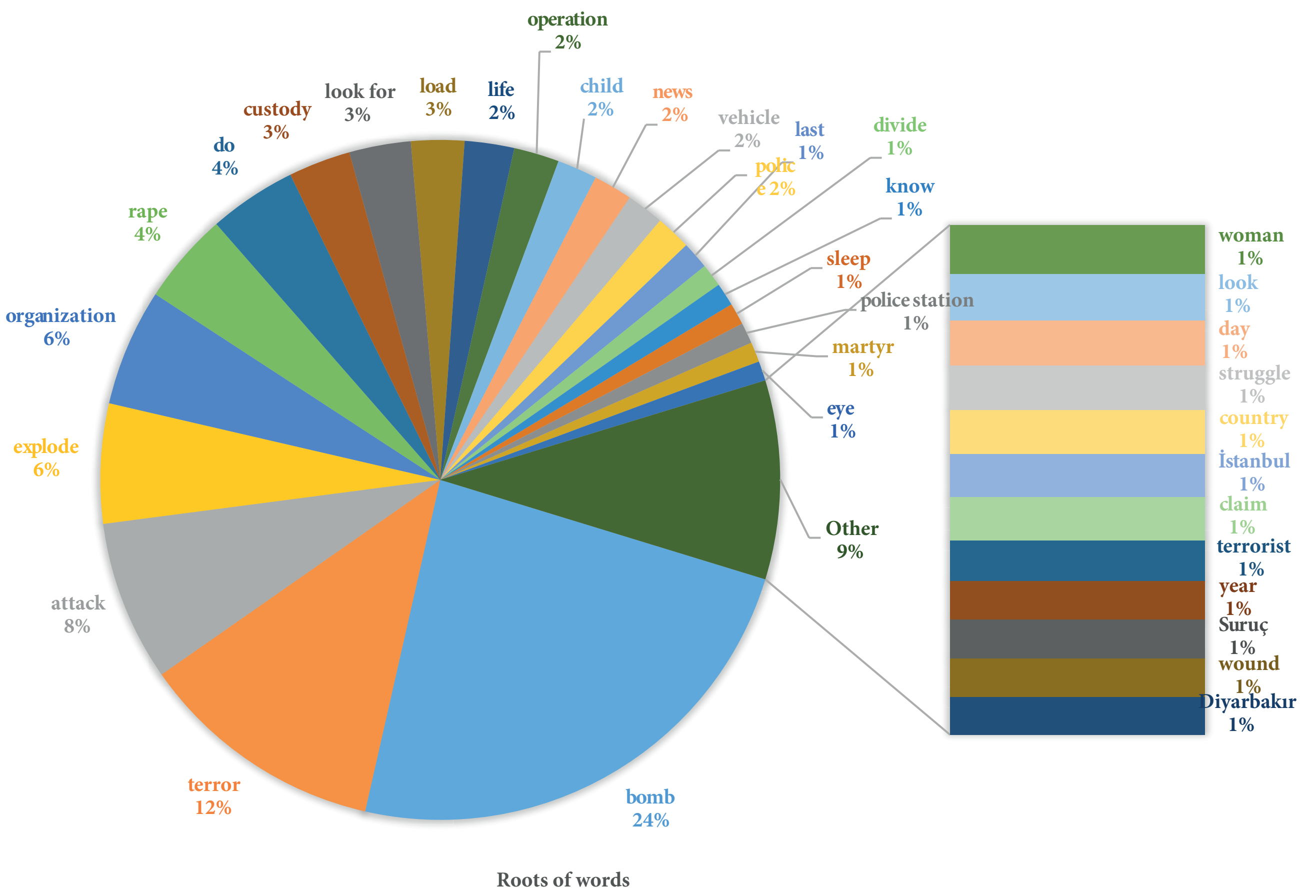

Figure 7. Roots of words.

8%, 6%, and 6%, respectively. In this case, during study dates, mostly terror attacks were discussed in Turkey. Criminal cases in the agenda of the country can be seen in Figure 8. Another important issue is that there is no word root from “demonstration,” “smuggling”, and “prostitution” keywords.

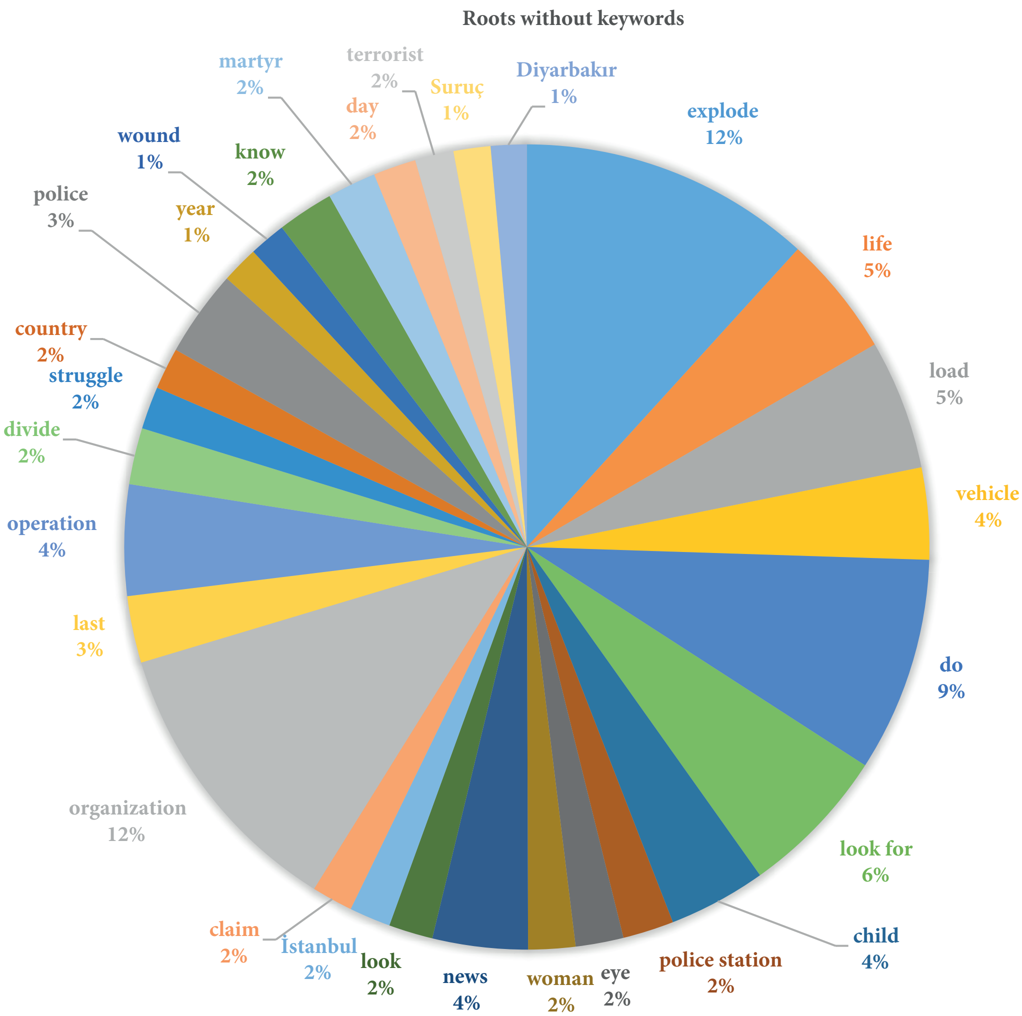

Figure 8. Roots of words without keywords.

Zemberek-NLP applied the data without keywords again to see subclusters of roots of words. There are 146,149 word roots in the data without keywords. The word roots graphic without keywords is shown in Figure 8.

In this graphic, “explode” and “organization” have the highest rates (12%), followed by “do,” “look for,” “life”, and “load” with rates of 9%, 6%, 5%, and 5%, respectively. Similarly to the previous graphic, terror attacks are discussed by the public. Without any knowledge of the data and seeing tweets, it can be understood from the wisdom and whisper of crowds that terror attacks happened by bombs and vehicles and lives were lost during these attacks.

One of the most popular data analysis programs capable of processing very large amounts of data effectively is R. RStudio is a visual version based on R. In this study, a program was written to analyze the tweet data, find frequencies of words and correlations between keywords and other words, and create a word cloud of tweets to check and expand the analysis of Zemberek to verify the results. In this program, the "tm", "SnowballC", and "wordcloud" packages of R were used. After loading the data file to the program, data preparation was done such as "lowering all words", "cleaning some special characters such as @, /, and |", "removing common words in English and Turkish", and "removing numbers and whitespaces". Next, the data file was converted into a matrix and a word cloud application was made. The word cloud of the data file is shown in Figure 9. As a word roots graphic, it is shown in the word cloud that "bomb" is the most used word in the data file. The word cloud supports the results obtained during the experiment days of this study in Turkey, as bombing attacks are the most mentioned subjects of Twitter.

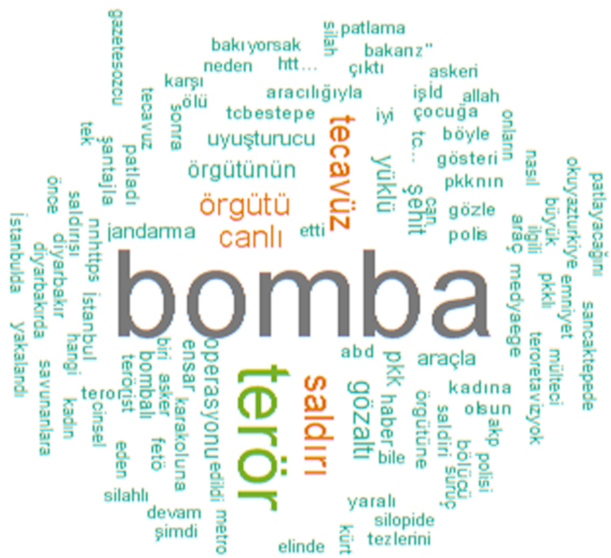

Figure 9. Word cloud.

The biggest words located in word cloud are bomba, terör, tecavüz, saldırı, örgütü, and canlı(bomb, terror, rape, attack, organization, alive (En)). The frequencies of the most used words are shown in Figure 10.

As seen in Figure 10, bomb and terror points are the peak points of the graphic. Correlations between these words and others are important for intelligence from social media. It is important to discover which words are used together. An important point is that the list is not indicative of frequency. Rather, it is a measure of the frequency with which the search and result term cooccur. The limit of correlation was selected as 0.85 for keywords. The results are shown in the Table. In R statistical software, Euclidean distance [51] is used to find the distance between vectors.

The link between clusters can be shown as in Eq. (1) [51]:

$$d\,(i+j,k) = \max(d\,(i,k)\,d\,(j,k)), \tag{1}$$

where d(i, j) represents the distance between clusters i and j. The distances between two documents (x, y) is represented by their term vectors ($\vec{t}_x$, $\vec{t}_y$) and calculated as shown in Eq. (2) [46]:

$$d\left(\vec{t}_x,\, \vec{t}_y\right) = \sum\nolimits_{t=1}^{n} (w_{t,x} - w_{t,y})^2, \tag{2}$$

where $t = \{t_1,\, t_2, ..., t_n\}$, $w_{t.x} = tf\,(x,t)\; x\; \log\left(\frac{|D|}{df(t)}\right)$, tf stands for term frequencies, and df(t) is the number of documents in which term t appears.

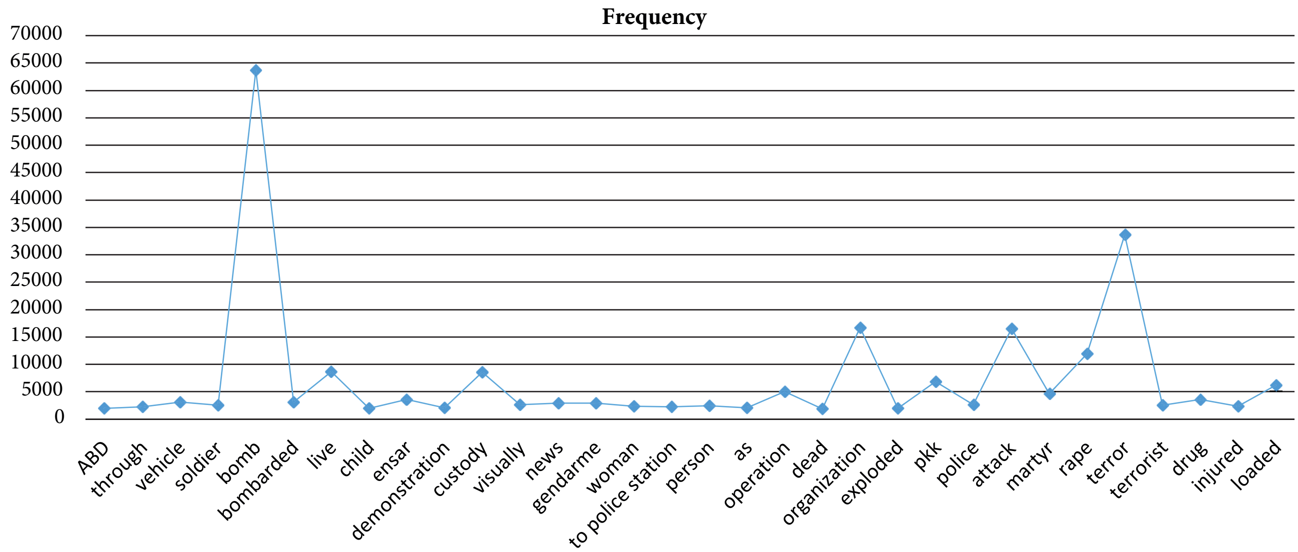

Figure 10. Frequencies of words.

Table. Keywords correlation.

Keyword (Tr) Keyword (En)	correlation1 Tr(En)	correlation2 Tr(En)	correlation3 Tr(En)	correlation4 Tr(En)	correlation5 Tr(En)
bomba (bomb)	Bildiğin (you know) = 0.91	yaa[1] = 0.91	tecavüz(rape) = 0.90	kan(blood) = 0.90	terör(terror) = 0.89
terör (terror)	Mersin[2] = 0.94	Torul[3] = 0.94	yeni(new) = 0.93	Anadolu(Anatolia) = 0.92	kaçak(escape) = 0.92
gözaltı (custody)	kamyon(truck) = 0.93	karanfil(carnation) = 0.93	düşen(falling) = 0.92	Karakol(police station) = 0.92	saldırın(attack) = 0.92
saldırı (attack)	yüklü(loaded) = 0.92	çocuğa(to child) = 0.91	otomobil(car) = 0.91	Önlendi(prevented) = 0.91	aracına(to vehicle) = 0.90
gösteri (show)	tecavüz(rape) = 0.89	deli(mad) = 0.87	yılda(per year) = 0.86	iddia(claim) = 0.85	-
kaçakcılık (smuggling)	eti(meat) = 0.87	sapıklar(perverts) = 0.87	Tersine (backwards) = 0.85	-	-
uyuşturucu (drug)	gök(sky) = 0.87	Nusaybin[3] = 0.86	-	-	-
fuhuş (prostitution)	örgütüyle(with organization) = 0.89	Ankara'nın (of Ankara)[2] = 0.88	Oslo'da(in Oslo)[4] = 0.87	bile(even) = 0.86	denen(called) = 0.86
tecavüz (rape)	patladı (exploded) = 0.90	Saldırının (of attack) = 0.90	gösteri(show) = 0.89	saldırı(attack) = 0.88	araçla(with vehicle) = 0.87
*Correlation limit = 0.85 Tr = Turkish En = English			[1] a connective Turkish word. [2] a city in Turkey. [3] a town in Turkey. [4] a city in Norway.		

The Table shows that the correlations yield interesting combinations. These give further insights into potential classifications. By the top-used terms and finding their associations, some subcategories can be defined in the data file and other information appear such that “bomb” is closely correlated with two other main keywords, “terror” and “rape.” It can be derived from this information that community memory builds bridges between events in positive or negative ways. Thus, it is a very important point of SCI for policy makers to take the pulse of society and manage crises and events well.

4. Discussion and conclusion

Huge amounts of data are released to the Internet every day through social media. These data are analyzed and people use these data for different purposes. Commercial, academic, and security are the main purposes for these analyses. Twitter is a very important type of social media website in terms of sensing the pulse of the community.

This is a case study for intelligence on Twitter and it also examines the wisdom and whisper of crowds on Twitter for cyber intelligence for security aims. According to TÜİK data, criminals are classified and keywords are defined according to these classes. The last classified data are from 2013 and this is a limitation for the study. Another limitation is the requirement of more effective Turkish NLP and more academic work on this subject. The most important limitation is Twitter’s streaming limits. During the streaming of data, the stream can be cut down because of some reasons like bandwidth and Twitter permissions, and reconnection is needed. Thus, continuing the stream is one of the most important issues.

In this study, Twitter data were collected for nearly 1 month and analyzed. It is seen that bomb attacks and terror events are most discussed on Twitter in Turkey during the study dates. In addition to the large rate of the “bomb” keyword, additional information is found in the study. Community memory builds bridges between events in positive and/or negative ways because the bomb keyword has a high correlation with “terror” and “rape” keywords. Although “bomb” can be in relation with “terror,” it normally has no relation with “rape”. Correlation tests have showed that “bomb” and “rape” in tweets are correlated (0.90). There has been no word root from the keywords of demonstration, smuggling, and prostitution. The study has shown that, for future studies, keywords can be changed dynamically while streaming according to criminal cases to get more inside the cases. In this way, intelligence will be narrower. By widening the keywords group, more masses can be reached, or by narrowing keywords, some specific issues can be investigated. Government officials can take some precautions by following necessary groups. Moreover, by using the tools and techniques used in this study, different studies can be performed by changing keywords, filters, and target groups. Similar studies can be done especially in commercial research for commercial aims. Thus, product and service improvements can be made by revealing the correlations between the determined words. Such an in-depth study will produce more comprehensive results than the sentiment analysis studies mostly used in the commercial field. To apply these techniques to commercial fields for customer satisfaction studies, association rules and support and feedback activities can provide significant contributions to commercial firms.

Acknowledgment

The essence of this article was presented at the IEEE 14th International Scientific Conference on Informatics and it was finalized with additions and improvements made after the conference. This article is an extension of that presentation with the addition of extensions.

References

[1] Savaş S, Topaloğlu N, Güler O. The determination of user's preferences on some domain names in Turkey: a survey application. International Journal of Informatics Technologies 2015; 8: 51-58.

[2] Setty K, Bakhshi R. What is big data and what does it have to do with it audit? ISACA Journal 2013; 3: 23-25.

[3] Zikopoulos PC, Eaton C, deRoos D, Deutsch T, Lapis G. Understanding Big Data. Analytics for Enterprise Class Hadoop and Streaming Data. New York, NY, USA: McGraw-Hill Osborne, 2012.

[4] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. J Comput Sci-Neth 2011; 2: 1-8.

[5] Zhang X, Fuehres H, Gloor PA. Predicting stock market indicators through twitter "I hope it is not as bad as I fear". In: COINs2010: Collaborative Innovation Networks Conference; Procedia - Social and Behavioral Sciences; 2013. pp. 55-62.

[6] Agnihotri R, Kothandaraman P, Kashyap R, Singh R. Bringing social into sales - the impact of salespeople's social media use on service behaviors and value creation. Journal of Personal Selling & Sales Management 2012; 32: 333-345.

[7] Mostafa MM. More than words: social networks' text mining for consumer brand sentiments. Expert Syst Appl 2013; 40: 4241-4251.

[8] He W, Zha S, Li L. Social media competitive analysis and text mining: a case study in the pizza industry. Int J Inform Manage 2013; 33: 464-472.

[9] Court D, Elzinga D, Mulder S, Vetvik OJ. The consumer decision journey. McKinsey Quarterly 2009; 3: 1-11.

[10] Vázqueza S, Muñoz-Garcíac Ó, Campanella I, Pocha M, Fisas B, Bel N, Andreu G. A classification of user-generated content into consumer decision journey stages. Neural Networks 2014; 58: 68-81.

[11] Li YM, Chen HM, Liou JH, Lin LF. Creating social intelligence for product portfolio design. Decis Support Syst 2014; 66: 123-134.

[12] Abrahams AA, Jiao J, Wang GA, Fan W. Vehicle defect discovery from social media. Decis Support Syst 2012; 54: 87-97.

[13] Abrahams AA, Jiao J, Wang GA, Fan W, Zhang Z. What's buzzing in the blizzard of buzz? Automotive component isolation in social media postings. Decis Support Syst 2013; 55: 871-882.

[14] Tang X, Yang CC. Detecting social media hidden communities using dynamic stochastic block model with temporal Dirichlet process. ACM T Intel Syst Tec 2014; 5: 36.

[15] Weichselbraun A, Gindl S, Scharl A. Enriching semantic knowledge bases for opinion mining in big data applications. Knowl-Based Syst 2014; 69: 78-85.

[16] Yang MC, Rim HC. Identifying interesting twitter contents using topical analysis. Expert Syst Appl 2014; 41: 4330-4336.

[17] Atzmueller M. Mining social media: key players, sentiments and communities. Data Min Knowl Disc 2012; 2: 411-419.

[18] Poria S, Cambria E, Winterstein G, Huang GB. Sentic patterns: dependency-based rules for concept-level sentiment analysis. Knowl-Based Syst 2014; 69: 45-63.

[19] Khan FH, Bashir S, Qamar U. TOM: Twitter opinion mining framework using hybrid classification scheme. Decis Support Syst 2014; 57: 245-257.

[20] Behrendt S, Richter, A, Trier M. Mixed methods analysis of enterprise social networks. Comput Netw 2014; 75: 560-577.

[21] Fu X, Shen Y. Study of collective user behaviour in twitter: a fuzzy approach. Neural Comput Appl 2014; 25: 1603-1614.

[22] Lau RYK, Li C, Liao SSY. Social analytics: learning fuzzy product ontologies for aspect-oriented sentiment analysis. Decis Support Syst 2014; 65: 80-94.

[23] Dang Y, Zhang Y, Hub PJH, Brown SA, Ku Y, HwangWang J, Chen H. An integrated framework for analyzing multilingual content in web 2.0 social media. Decis Support Syst 2014; 61: 126-135.

[24] Wu Y, Liu S, Yan K, Liu M, Wu F. OpinionFlow: Visual analysis of opinion diffusion on social media. IEEE T Vis Comput Gr 2014; 20: 1763-1772.

[25] Grieve R, Witteveen K, Tolan GA. Social media as a tool for data collection: examining equivalence of socially value-laden constructs. Curr Psychol 2014; 33: 532-544.

[26] Weidemann C, Swift J. Social media location intelligence: the next privacy battle - an arcgis add-in and analysis of geospatial data collected from twitter.com. International Journal of Geoinformatics 2013; 9: 21-27.

[27] Bell D, Koulouri T, Lauria S, Macredie RD, Sutton J. Microblogging as a mechanism for human–robot interaction. Knowl-Based Syst 2014; 69: 64-77.

[28] Günüç S, OdabaşıHF, Kuzu A. The defining characteristics of students of the 21st century by student teachers: a Twitter activity. Journal of Theory and Practice in Education 2013; 9: 436-455.

[29] Martinčić-Ipsić S, Močibob E, Perc M. Link prediction on twitter. PLoS ONE 2017; 12: e0181079.

[30] Jalili M, Orouskhani Y, Asgari M, Alipourfard N, Perc M. Link prediction in multiplex online social networks. R Soc Open Sci 2017; 4: 160863.

[31] Ban K, Perc M, Levnajić Z. Robust clustering of languages across Wikipedia growth. R Soc Open Sci 2017; 4: 171217.

[32] Helbing D, Brockmann D, Chadefaux C, Donnay K, Blanke U, Woolley-Meza O, Moussaid M, Johansson A, Krause J, Schutte S et al. Saving human lives: what complexity science and information systems can contribute. J Stat Phys 2015; 158: 735-781.

[33] D'Orsogna MR, Perc M. Statistical physics of crime: a review. Phys Life Rev 2015; 12: 1-21.

[34] Omand S, Bartlett J, Miller C. Introducing social media intelligence (SOCMINT). Intelligence and National Security 2012; 27: 801-823.

[35] Bayraktar G. The new requirement for the fifth dimension of the war: cyber intelligence. Journal of Security Strategies 2014; 20: 119-147.

[36] Kırık AM. Social media-individual interaction and social transformation in the context of Arab Spring. Education and Society in the 21st Century 2012; 1: 87-98 (in Turkish with abstract in English).

[37] Dündar E. Sosyal medya: bir protesto aracı. Türk Kütüphaneciliği 2011; 25: 165-172 (translation in Turkish of original article in English).

[38] Korkmaz A. Arap baharısürecinde internet ve sosyal medyanın rolü. In: International Symposium on Language and Communication: Research Trends and Challenges; 10–13 June 2012; İzmir, Turkey. pp. 2147-2153.

[39] Chakraborty B, Banerjee S. Modeling the evolution of post disaster social awareness from social web sites. In: 2013 IEEE International Conference on Cybernetics; 13–15 June 2013; Lausanne, Switzerland. New York, NY, USA: IEEE. pp. 51-56.

[40] Oh, O, Agrawal M, Rao HR. Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. MIS Quarterly 2013; 37: 407-426.

[41] Werbin KC. Spookipedia: Intelligence, social media and biopolitics. Media Cult Soc 2011; 33: 1254–1265.

[42] Bertot JC, Jaeger PT. Transparency and technological change: ensuring equal and sustained public access to government information. Gov Inform Q 2010; 27: 371-376.

[43] Bertot JC, Jaeger PT, Grimes JM. Using ICTs to create a culture of transparency: E-Government and social media as openness and anti-corruption tools for societies. Gov Inform Q 2010; 27: 264-271.

[44] Bertot JC, Jaeger PT, Hansen D. The impact of polices on government social media usage: issues, challenges and recommendations. Gov Inform Q 2012; 29: 30-40.

[45] Gerber MS. Predicting crime using twitter and kernel density estimation. Decis Support Syst 2014; 61: 115-125.

[46] Kavanaugh AL, Fox EA, Sheetz SD, Yang S, Li LT, Shoemaker DJ, Natsev A, Xie L. Social media use by government: from the routine to the critical. Gov Inform Q 2012; 29: 480-491.

[47] Surowiecki J. The Wisdom of Crowds. New York, NY, USA: Anchor Books, 2004.

[48] Robu D, Sandu F, Petreus D, Nedelcu A, Balica A. Social networking of instrumentation – a case study in telematics. Adv Electr Comput En 2014; 14: 153-160.

[49] Çöltekin C. A freely available morphological analyzer for Turkish. In: Proceedings of the International Conference on Language Resources and Evaluation; 17–23 May 2010; Valletta, Malta. pp. 820-827.

[50] Park Y, Kang S, Seo J. information extraction using distant supervision and semantic similarities. Adv Electr Comput En 2016; 16: 11-18.

[51] Thoplan R. Text mining the works of Christopher Marlowe. Research Journal of Science & IT Management 2014; 3: 43-51.