ROMANIAN JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY
Volume **20**, Number 3, 2017, 286–298

Analysis of data volumes circulating in SNs after the occurrence of an earthquake

Mironela PÎRNĂU

Titu Maiorescu University, Faculty of Informatics, Bucharest, Romania
Email: mironela.pirnau@utm.ro

Abstract. The paper presents a statistical analysis of the message flow on Twitter, related to earthquakes during certain periods of time after the occurrence of major earthquakes. Tweets were collected shortly after earthquakes in August and October 2016, in order to identify correlations among the content of the tweets in a disaster situation. The streams of collected tweets have been analyzed with Data Mining algorithms. The aim of this work is to identify modes of communication between the Twitter users in earthquake situations, and to analyze the word correlation found in the message exchanges during a disaster situation. The paper contributes to the understanding of communication patterns on social networks during and after disasters.

Key-words: tweets; earthquake; #terremoto; Data Mining; Social Media.

1. Introduction

Social media (SM) platforms have largely replaced the traditional mass-media; the volume of data circulating on SM exceeds a trillion of gigabytes. This enormous amount of data is also extremely diverse. Many researchers explored the data flows on Social Media [1–3], several of them investigating the specifics of the dynamics of the flow of data within social networks [4–6]. The main purpose of the cited researches was to apply methods and various analysis tools for helping prevent and mitigate the effects of the disasters.

Considering the research published by several research groups [6–9] and our previous studies [10] regarding the data analysis of data flows within Social Media, we focused on the message flow on Tweeter. The purpose was to evaluate tweets language and configuration as related to earthquakes in Rieti, Norcia - Province of Perugia (Italy) and Vrancea (Romania). The timeframe of the investigation was August-November 2016. This paper is based on [4, 11–14] researches regarding the data analyses within Social Media in case of disaster. The aim is to identify modes of communication and communication patterns on Twitter, contributing to the understanding of messaging on social networks during and after disasters.

The present study includes seven sections: Section (1) explains the context of the present research; Section (2) relates to the main features of the data collected from Social Media, Section (3) investigates the Data Mining technology applied to collections of tweets; Section (4) presents an analysis of the collected data; the next Section analyzes the dynamic of tweets flow in case of disaster and Section 6) covers particular cases occurring in the data stream; Section (7) presents conclusions drawn regarding the role of knowledge identified on the basis of algorithms, in data collections for disaster situations.

2. Main features of Social Media data

2.1. General characteristics of Social Media data

The data circulating within Social Media tend to become an important source of information. Twiter, like Facebook, requires the users to share a series of statuses of several types: texts, photos, links, targeted messages, and videos [3, 9, 15]. If the users wish to communicate their statuses, there is the need to increase the number of connections, called 'followers', in case of Twitter. The advantage of Twitter over Facebook is that no algorithm is required in order to determine the followers of the statuses. On Twitter, the users' statuses appear on the news-feed of all subscribers, representing an advantage in a message flow in case of disaster or calamity [4, 8].

Essentially, unlike Facebook, Twitter is defined as a social network with micro-blogging platform with relationships among users; a platform that allows posting short messages as status (up to 140 characters), but which may contain URLs and hashtags. It is recommended to use hashtags for keywords, e.g., **(#terremoto, #Italy, #earthquake, #cutremur)** so as people find and follow the tweets that contain these hashtags. In the results section, hashtags *#terremoto, #Italia, #earthquake, #cutremur* were analyzed separately, in the multitude of data extracted from Twitter. The role of a specific hashtag is to encourage others to use it in turn. This process leads to a major advantage in grouping important data. Another important advantage is represented by Twitter's distinctive property of shortened 140 characters availability. The links shared on Twitter are automatically processed and shortened to http://t.co. links type. The advantage of using his own short link (t.co) generator service, is in helping users to share long URLs in their messages. In this way the maximum number of characters is not exceeded. Also, t.co used by Twitter protects users from visiting malicious sites.

2.2. Features of tweets collections

For collecting large amounts of data from Twitter we used the API Twitter [16]. The data supplied is designed in JSON - JavaScript Object Notation format. In order to be able to open a stream towards the data, it is necessary to create an account. The Consumer Key and Secret fields in the application are used to authenticate requests from the Twitter Platform (https://apps.twitter.com/). Two options are available for extracting tweets: search in www.twitter.com, or using Search API. For the latter method, the user should obtain a developer account - credentials for access. Searches on twitter.com may return old posts; compared to API search where users can choose one method from a spectrum of three possibilities: Search API, which is part of the REST API Twitter; Streaming API Twitter method, and for high access to tweets it is recommended to contact GNIP using the link https://gnip.com/sources/twitter/.

3. Data Mining technology applied to collections of tweets

3.1. Data Mining

One of the essential characteristics of 'Big Data' originated from Social Media is that they are produced almost in real time, allowing the identification of a fact about the actions taken or the ones about to be taken in a certain place. The main characteristics of 'Big Data' coming from Social Media are very large volumes and variety, high data velocity and variable truthfulness. Within the present work, we applied complex search methods in order to diagnose patterns and groups of data. Once identified, these patterns may be used to identify a persons behavior in case of disaster, to establish the size of the disaster and its effects, and possibly to predict future social effects of the respective disasters. Many data mining algorithms (EM Algorithm, FP-Growth, SEAR, A-Priori, SVM, PageRank, Bayesian, CART) [11, 17–20] are available for correlation and extracting useful information from large volumes of data, including data from Social Media.

3.2. Association rules and frequencies of keywords in tweets

For analyzing data collections, we used rules of association as explained in [11,17,21,22]. A rule of association is an expression of the form: $\{$*IF X then Y*$\}$, where item X is the antecedent of the rule and Y is the consequent. Elements X and Y are distinct and

$$X \cap Y = \Phi \tag{1}$$

The main features of association rules are the coverage power, defined as the percentage of items that satisfy the antecedent of the rule, and the accuracy of the rule, defined as the percentage of items that satisfy both the antecedent and consequent of the rule. Methods for building association rules can be indirect, characterized by extraction of rules using, generally, trees of classification and decision, and direct methods, characterized by extracting rules directly from data. To identify association rules for estimating risk in the event of earthquake, it is indicated to use decision trees, according to [11,17,23,24].

The investigation was focused on applying the direct method of symbiotic relationships analysis in order to find associations between the words $\{$*earthquake, victim, terremoto, morto, demolito, earthquake, disaster, cutremur, storm, dead*$\}$; these words were used in the opening flow of Twitter extraction and are part of the tweets under review in Section 4. Applying this method, it was interesting to identify groups of words that are adopted in the collected postings, for a disaster situation (earthquake), and compare our results with those in the paper [4], where the Zipf method was used. In the case of the set of tweets analyzed, the associations of worlds that occur at least 100 times are $\{$*erthquake victim*; *erthquake dead*; *terremoto morto*; *erthquake terremoto*; *erthquake disaster*; *victim dead*; *victim terremoto*$\}$. No association rule involving the words *demolito*, *seism*, *cutremur*, or *storm* was found. In a real situation case, the absence of these associations indicates to the disaster relief teams that there are no victims because some buildings were demolished by the earthquake, that there was no storm producing victims and that the trapped victims have not used the Romanian word *cutremur*.

A main feature that measures the strength of an association rule is the rule support [23] [24]. In percent this represents, how often this association rule can be applied to a collection of data. An association rule has the form: *'if the antecedent rule is met then we have the consequent of the rule'* [21] [22]. A transaction $t_i \in T$ contains a lot of X items, if X is a subset of t_i and the set of items X also includes t_i. The support of a rule $X \rightarrow Y$ is the percentage of transactions in

T that contains $X{\cup}Y$. The rule support can be calculated as the frequency of applying the rule to the set of transactions T. Let n be the the number of transactions, then the support of the rule $X{\rightarrow}Y$ may be represented by equation 2:

$$Support(X{\rightarrow}Y) = (X{\cup}Y)/n \tag{2}$$

Another feature that measures the strength of an association rule is confidence [21,22] that is computed according to equation 3. Collections of data will be analyzed in Section (5) using the above indicators and methods. The data this research was based on is stored in files whose main features are presented in Table 2.

$$Confidence(X \rightarrow Y) = support(X \cup Y)/support(X) \tag{3}$$

Table 1. Files with collected tweets

Period for collecting tweets	Number of collected tweets	Number of tweets with unique content	file size (KB)
between August 25^{th} - November 2^{nd}, 2016	3,951,042	1,407,791	3,4380.1KB $\cong$ 335.7 MB

Because of the high volume of data to analyze and the highly detailed analysis required in a disaster situation, high performance computing systems are needed, able to analyze and interpret data collected in real-time. In the case of the present paper, the data base containing analyzed unique tweets has a size of 335.7 MB, and the method used for collecting tweets, was the standard Sprizer method [25]. If we use the Gnip 2.0 platform to its full capacity, then the analyzed data will reach a size which requires the use of more powerful algorithms for the analysis.

4. Analysis of collected data

To extract data from Social Media Twitter, Twitter Streaming API with Sprizer standard method [16] was used. A set of data for messages including the words {*earthquake, terremoto, disaster, morto, cutremur, demolito, storm, seism, victim*} was collected between August 25^{th} - November 2^{nd}, 2016, and is presented more explicitly in Table 3. The data represented 3,951,042 collected tweets, 1,407,791 unique tweets and a database size of 0.327 GB. The data collected is in JSON format and were stored in a file that contains fields such as: *id, tweets, source, lang, created_at, data_received, data_write, nr_crt*. The properties of the fields are presented in Table 2.

The target of this paper is to conduct a data analysis simultaneously with collecting tweets. This requires an information system with expensive physical resources because of the high performance of computing machines, which make up this system.

Basic steps of transforming the collected data into knowledge are represented by the selection, pre-processing, processing, data mining, interpretation and evaluation of results, as shown by Fayyad in the paper [9]. For the analysis, only unique tweets were kept. The frequencies of hashtags **#terremoto, #earthquake, #disaster, #Italy, #https**, were identified, according to Table 3. Table 4 presents the frequencies of several words; for example, the frequency of the word 'Richter' is over 17 times smaller than the frequency of 'magnitude' and of about two times lower than the frequency of the word 'cutremur'.

Table 2. The structure of off-line processed data

Field	field properties	role of the field
id	int(11) NOT NULL AUTO_INCREMENT	the number of records managed
tweet	text	content of tweets
source	text	source of tweet
lang	text	language of tweet
created_at	text	time when this tweet was created
data_received	text	time when this tweet was received
data_write	text	time when this tweet was written
nr_crt	int(20)	registration number

It is interesting to notice that the frequency for #Italia is 3.1 times higher than #disaster, which is because of the events of August 24th and October 30th 2016, with people referring precisely to the location and type of disaster. In the analyzed period, we identified 13,278 tweets that contain *https* and which must first undergo a process of parsing, eventually using Data Scraping technique and then must be processed and analyzed. In Figure 1, the variation of the number of tweets that contain both text and *https* was compared with the variatin of the number of tweets that contain only https. Half of the number of collected tweets with unique content also have *https*. To extract the whole content of the message from *https*, it is advisable to use a web crawler [16] [27]. The dynamics in Figure 2 is similar to those discussed in [28].

Table 3. Frequency of defining elements in tweets in case of disaster

	#terremoto	#earthquake	#disaster	#Italy	also https	only https
Totals Tweets	69,062	69,680	2,384	7,439	736,403	13,278

Table 4. The frequency of specific words in disaster situations

Period	morto	retweets for morto	magnitude	retweets for magnitude	Richter	cutremur	retweets for cutremur
Totals Tweets	140,992	26,366	32,571	5,472	1,877	3,468	205

Since the rol of hashtags is to increase the visibility of tweets, we notice that these tweets may become three times more visible. For the unique tweets analyzed, the occurrence of the word *morto*, during earthquakes in Italy counted 140,992, almost as the number of tweets with #terremoto plus #earthquake and #disaster (141,126). The number of occurrences of the word ‘morto‘ in collected tweets is 4.5 times larger than the number of uses of ’*deceased persons*’.

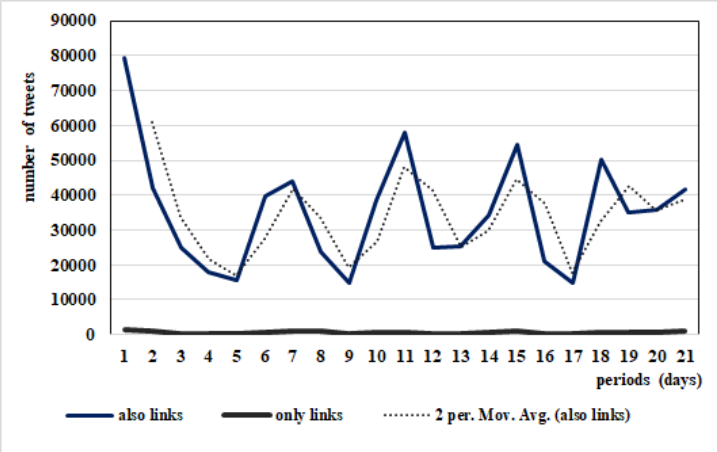

Fig. 1. Compared tweets that include also links to tweets containing only links

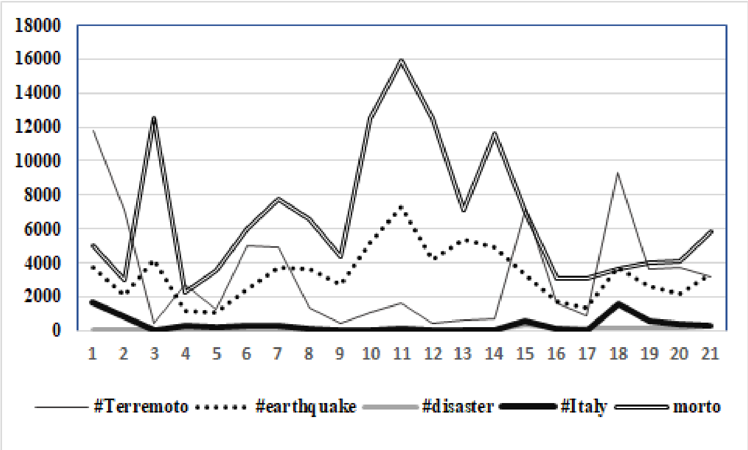

Fig. 2. Frequency of tweets #terremoto, #earthquake, #disaster, morto

5. Analysis of tweets flow dynamics in case of disaster

The Pearson correlation between *hashtags* {**#terremoto, #earthquake, #disaster, #Italy**} was analyzed, based on the database of collected tweets, compiled as in Table 1, and following the example in [28]. The correlation is computed as in the standard equation 4.

$$C(X,Y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}. \tag{4}$$

For example, x represents the number of tweets including (#terremoto) and y stands for the respective number including (#earthquake). The intercorrelation matrix was calculated using equation (4) and as shown in Table 5.

Table 5. The Correlation function calculated for hashtags in emergency and disaster situations

	#terremoto	#earthquake	#disaster	#Italy
#terremoto	1	0.054	0.299	0.956
#earthquake		1	0.226	-0.034
#disaster			1	0.172
#Italy				1

As the {#earthquake #Italy} correlation is very small in absolute value (< 0.04), the tweets messages where #earthquake occurs correspond to an extremely low frequency of the hashtag #Italy, and for tweets where {#terremoto #Italy} occurs, the correlation is high, having the value of 0.956. However, the correlation {#terremoto #disaster} has a value close to the correlation for {#earthquake #disaster}, although the words for the calculation belong to different languages {en, it}. Because of the negative correlation coefficient for {#earthquake #Italy}, in order to establish if the two parameters are correlated according to the Student distribution, the 't' statistics was calculated, using the equation 5:

$$t = \frac{|C(X,Y)| \cdot \sqrt{n-2}}{\sqrt{1 - C(X,Y)^2}}. \tag{5}$$

The 'n' parameter indicates the value or the volume of the pattern. The 't' parameter was used to indicate the bilateral critical probability (applying the TDIST formula). The result shows that the significant correlation between the {#earthquake #Italy} parameters has the value of 0.54, and the correlation{#earthquake #Italy} is not relevant. Given the multitude of earthquakes that have occurred around the world [30] in the analyzed period, we proposed to comparatively study the communication on Tweeter in relation with the region of the seism. Following the protocol and methods from [6, 7, 13, 14, 29–31] the number of tweets in the subsequent one to tree days were calculated after an earthquake occurrence, wherever it was in the world, and whatever its magnitude (larger than 4).

Next, the keyword frequencies, the numbers of tweets per time interval that included those keywords, and the correlations were computed selecting only earthquakes produced in Europe, respectively outside Europe. For this correlation per day, only the maximum magnitude earthquakes, which occurred in the analyzed periods, were used. The analysis shows that the correlation is higher for locations in the EU, possibly because the large number of earthquakes in this region during the analyzed period (Table 6).

Table 6. Tweets/magnitude correlation

Tweets/magnitude correlation in the world	Tweets/magnitude correlation in the EU	Tweets/magnitude correlation outside EU
0.417	0.690	-0.502

Next, we applied the method of dynamic analysis of message flows [28] based on patterns defined using the Euclidean distance calculation applied to the vector obtained with the 'windowing' of the phase space [5], as applied in [7] (method due to H.N. Teodorescu). The study focused on the analysis of Euclidean distance between the numbers of tweets including the hashtags #terremoto, #earthquake, #disaster, #Italy, for tweets collected during the analyzed period.

For the dynamic analysis, initially a phase diagram was drawn up, namely $y = N[t]$ vs. $N[t-1]$ according to $x = N[t]$ with graphic $y = y(x)$. On the phase diagram, as in [5] [7], the number of points in each quadrant $(n1, n2, n3, n4)$ must be identified, to characterize the attractor points [5] in quadrant windows. Two developments (dynamic usages) are similar if they have the four numbers close. The phase diagram was determined for all the series and, then, it was compared; in this case, the two data series in Table 7 are illustrated, where the first series is #terremoto and the second series is #earthquake. The corresponding diagrams (attractors, denoted by A_1 and A_2) are presented in Figure 3 and Figure 4.

Table 7. Corresponding values for N [t] N [t-1]

Day	#terremoto	#terremoto [n] #terremoto[n-1]	#earthquake	#earthquake[n]-#earthquake [n-1]
1.	18839		5763	
2.	9452	-9387	8884	3121
3.	4921	-4531	3711	-5173
4.	1379	-3542	3644	-67
5.	459	-920	2685	-959
6.	1063	604	5189	2504
7.	1643	580	7267	2078
8.	1741	98	14417	7150
9.	8756	7015	4897	-9520
10.	895	-7861	1383	-3514
11.	9320	8425	3734	2351
12.	10594	1274	8106	4372

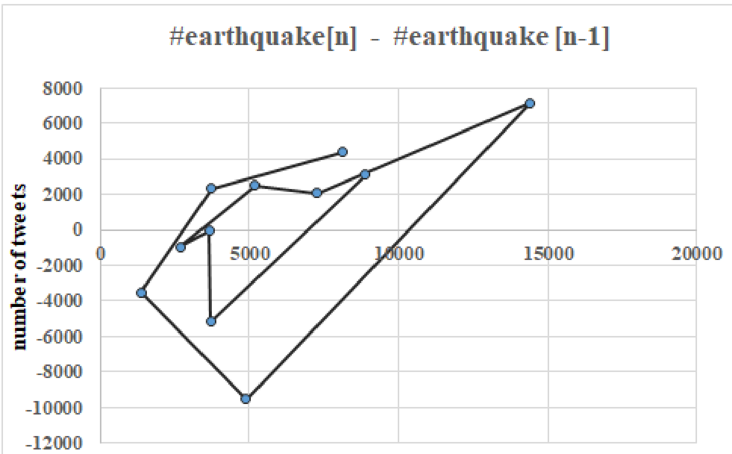

Fig. 3. Example of dynamics in the phase plane of the series #earthquake

The Euclidean distance between the two attractors, normalized to the number of points of the attractor, n, is $d(A_1, A_2)/n = 28931.8/11 = 2630.2$. When the normalized distance is further normalized to the average value, 5755.1, in the first time series (used for the first attactor), one obtains the doubly normalized distance, d_n, $d_n = 0.46$. If the symbolic distance is computed according to [5], one obtains $d_s((8, 3), (7, 4)) = \sqrt{2}$, where (8,3) stands for 8 points in the upper half-plane (first quadrant window) and 3 points in the lower half-plane for the first attractor.

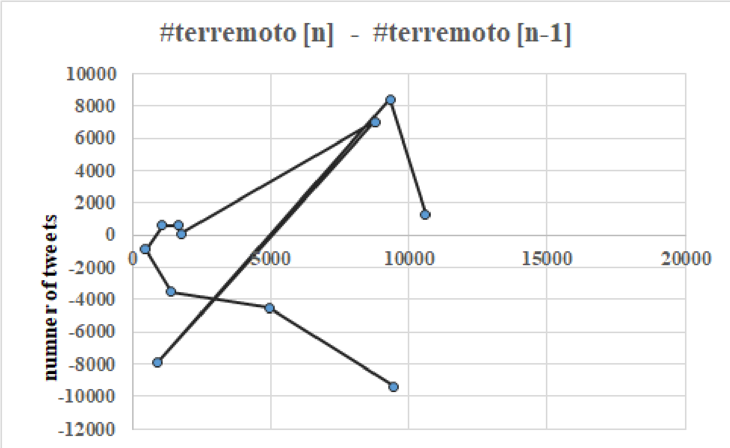

Fig. 4. Example of dynamics in the phase plane of the series #terremoto

Both the values d_n and d_s show the difference in the dynamics of the respective time series; d_s is easier to be obtain, but has the disadvantage of being dependent on any offset (constant component in the time series). The values of the distance show how similar the dynamics of the time series are. In this case, the time evolution of the number of messages with the hashtags #Terremoto and #earthquake is quite similar, although the visual inspection does not clearly indicate it. The RADIAL diagram of the frequencies of tweets with hastags #terremoto and #earthquake is represented in Figure 5. This graph indicates the relationship between the variation of the numbers of tweets with the two hastags and how their frequency depends on the event in the day/period under review."

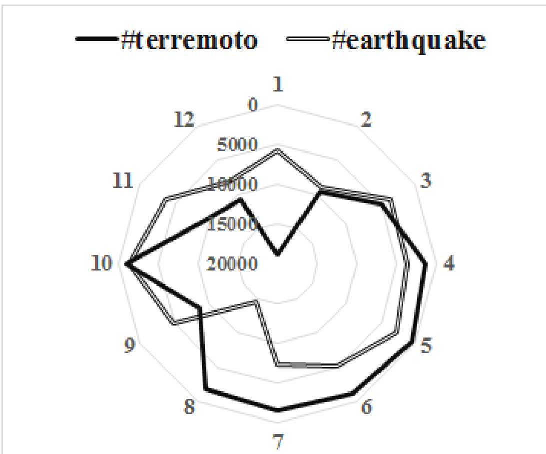

Fig. 5. Relationship between the number of tweets with the hashtags #terremoto and #earthquake

6. Particular cases occurring in the data stream

Using algorithms specific to Data Mining, there are identified association rules for finding frequent sets of tweets that refer to a situation of disaster and its effect. There were distinguished types of messages posted by users to obtain information about an unfavorable, disastrous situation at a certain time.

The database with posts is represented by a sequence of transactions $T = (t1, ...tn)$ and each post contains several keywords called articles. Rules of the form **'**$\{$**cutremur, dezastru**$\} \rightarrow$ **mort'** were determined, quantifying the disaster, with the purpose of dispatching special crews to handle the situation.

Only 201,194 tweets were collected in the period 25th – 26th August 2016, with relevant occurrences within delivered tweets. For the words $\{$*erthquake, victim, terremoto, morto, demolito, seism, disaster, cutremur, storm, dead*$\}$ we calculated the frequency association of words, (see Table 8), in order to obtain new rules of association between words.

Table 8. Number of words in tweets

	earthquake	victim	terremoto	morto	demolito	seism	disaster	cutremur	storm	dead
earthquake	51,906	4,227	1,332	4	0	0	596	0	42	2,525
victim	4,227	29,108	288	3	0	0	84	18	24	306
terremoto	1,332	288	89,795	1,595	4	3	33	0	5	25
morto	4	3	1,595	9,364	0	0	0	0	0	8
demolito	0	0	4	0	21	0	0	0	0	0
seism	0	0	3	0	0	3	0	0	0	0
disaster	596	84	33	0	0	0	695	0	2	9
cutremur	0	18	0	0	0	0	0	343	0	0
storm	42	24	5	0	0	0	2	0	108	1
dead	2,525	306	25	8	0	0	9	0	1	2,874
total	60,632	34,058	93,080	10,974	25	6	1,419	361	182	5,748

Notice that the frequency of the word "terremoto" is high. The A-Priori algorithm was used, in order to identify associations between the words found in tweets. The calculations identified the association rules established between the words vector **C** $\{$*erthquake, victim, terremoto, morto, demolito, seism, disaster, cutremur, storm, dead*$\}$ and the content of tweets collected. To identify the association rules the data flow were processed as follows: attaching the binary value 1 to words in vector ***C*** when they are found in tweets, and value 0 otherwise; a new database that contains the frequency of occurrence of the elements of vector **C** in the database under analysis was created. Then, we identified all the associations with two and three words, and their frequency of occurrence (Table 9). For the analysis, there were taken into account only the associations between words occurring at least 100 times in the analyzed tweets. They have also been used for further analysis (Table 10).

According to Table 10, it can be noticed that only the first association rules have the confidence indicator greater than 0.5. Basically, taking into account the analysed data, only these five association rules are significant. The optimization of the algorithm in order to identify all the rules of association in all the collected tweets requires an extensive future work.

Table 9. Association of words and their frequency of occurrence

Associations	**Frequency**	Associations	**Frequency**
erthquake victim	4,033	victim storm	23
erthquake dead	2,417	victim cutremur	18
terremoto morto	1,595	terremoto dead	15
erthquake terremoto	1,218	erthquake terremoto dead	10
erthquake disaster	532	erthquake disaster dead	9
victim dead	216	morto dead	8
victim terremoto	210	terremoto disaster	7
erthquake victim dead	89	terremoto storm	5
erthquake victim terremoto	78	erthquake morto	4
victim disaster	57	terremoto demolito	4
erthquake storm	40	terremoto seism	3
erthquake victim disaster	27	victim morto	3
erthquake terremoto disaster	26	victim storm dead	1

Table 10. Identification of association rules of words in tweets in case of disaster

Rule	antecedent	consequent	support	confidence
If morto →terremoto	morto	terremoto	0.149	0.990
If dead →erthquake	dead	erthquake	0.227	0.874
If victim →erthquake	victim	erthquake	0.379	0.848
If disaster→erthquake	disaster	erthquake	0.045	0.806
If terremoto →morto	terremoto	morto	0.150	0.503
If erthquake →victim	erthquake	victim	0.379	0.475
If terremoto→erthquake	terremoto	erthquake	0.114	0.384
If erthquake →dead	erthquake	dead	0.227	0.285
If erthquake →terremoto	erthquake	terremoto	0.114	0.144
If dead→victim	dead	victim	0.020	0.078
If terremoto →victim	terremoto	victim	0.019	0.066
If erthquake →disaster	erthquake	disaster	0.050	0.062
If victim →dead	victim	dead	0.020	0.045
If victim→terremoto	victim	terremoto	0.020	0.044

7. Conclusions

The paper aimed to identify methods of data analysis for retrieval of additional information that may reveal new ways of grouping data stored, or new relationships between these data to discover ‘knowledge‘ in database collected in disaster situations. Obtaining relevant information from data collection is a laborious process, requiring computationally costly and difficult actions to access resources. However, the time needed to obtain information in order to save human lives must always be minimum, regardless of cost and resources involved. For various decisions or operations that are taken in real situations, many simulations on critical data have to be made.

The results show that there are specific associations of words and hashtags in tweets related to earthquakes and that the dynamics of occurrences of keywords is correlated. These results may improve situation assessment in a disaster; according to these results, the information system can timely generate information on the magnitude of the event and its consequences and useful conceptual relationships that may help decision making. There is however a risk that improperly collected data hide useful information and generate undesired decisions in case of earthquake and other disasters. Therefore, further work should be directed to the improvement of the completeness and correctness of the collected data.

Acknowledgement. Thanks to H.N. Teodorescu for the suggestions on this research,for correcting several preliminary versions of this paper.

References

[1] ADAM N.R., SHAFIQ B., STAFFIN R., *Spatial Computing and Social Media in the Context of Disaster Management*, Intelligent Systems, IEEE, **27**(6), pp. 90–96, 2012.

[2] CHEONG F., CHEONG C., *Social Media data mining: a social network analysis of tweets during the 2010-2011 Australian floods*, 15th PACIS, pp. 1–16, 2011.

[3] KUMAR S., MORSTATTER F., LIU H., *Twitter Data Analytics*, Springer, USA, 2013.

[4] TEODORESCU H.N., *Using Analytics and Social Media for Monitoring and Mitigation of Social Disasters*, Procedia Engineering, 107C, pp. 325–334, 2015.

[5] TEODORESCU H.N., *Characterization of Nonlinear Dynamic Systems for Engineering Purposes - a partial review*, Int.J. General Systems, **41**(8), pp. 805–825, 2012.

[6] TEODORESCU H.N., *Emergency-Related, Social Network Time Series: Description and Analysis*, Time Series Analysis and Forecasting, Contributions to Statistics, Springer International Publishing Switzerland, pp. 205–215, 2016.

[7] BOLEA S.C., *Language Processes and Related Statistics in the Posts Associated to Disasters on Social Networks*, Int. Journal of Computers Communications and Control, **11**(5), pp. 602–612, 2016.

[8] RAO R. R., EISENBERG J., SCHMITT T., *The Role of IT in Mitigation, Preparedness, Response, and Recovery*, The National Academies Press Washington, 2007.

[9] TORIUMI F., SAKAKI T., SHINODA K., KAZAMA K., KURIHARA S., NODA I., *Inf. Sharing on Twitter During the 2011 Catastrophic Earthquake*, WWW '13 Int. Conf. Proc., pp. 1025–1028, Geneva, Switzerland 2013.

[10] PIRNAU M., *Analysis of the use of Social Networks in Acquiring Information on Disastrous Events*, 15th RoEduNet Conference: Networking in Education and Research, **8**(2), pp. 236–240, 2016.

[11] AGRAWAL R., SRIKANT R., *Fast Algorithms for Mining Association Rules*, Proc. 20th Very Large Data Base Conf., 1994.

[12] BOLEA S.C., *Vocabulary,Synonyms and Sentiments of Hazard-Related Posts on Social Networks*, 8th IEEE Int. Conf. on Speech Technology and Human-Computer Dialogue, SPED, Bucharest, Romania, pp. 14–17, 2015.

[13] BRUNS A., BURGESS J.E., *Local and Global Responses to Disaster:#eqnz and the Christchurch earthquake*, in Sugg, P. (Ed.) Disaster and Emergency Management Conf., Proc., AST Management Pty Ltd, Brisbane, QLD, pp. 86–103, 2012.

[14] KIEL G., *What Disaster Response Management Can Learn from Chaos Theory?*, Conference Proceedings, May 18-19, Ed. by Gus A. Koehler, 1995.

[15] WANG H., CAN D., KAZEMZADEH A., BAR F., NARAYANAN S., *A System for Real-Time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle*, In ACL System Demonstrations, pp. 115– 120, 2012.

[16] https://dev.twitter.com/streaming/overview

[17] AGRAWAL R., IMIELINSKI T., SWAMI A., *Mining Association Rules Between Sets of Items in Large Databases*, ACM SIGMOD Conf. Management of Data, 1993.

[18] PARK J.S., CHEN M., YU P.S., *An Effective Hash Based Algorithm for Mining Association Rules*, ACM SIGMOD Int'l Conf. Management of Data, 1995.

[19] PIZZUTI C., *GA-Net: A Genetic Algorithm for Community Detection in Social Networks*, in Proceeding 10th Int. Conf. PPSN, pp. 1081–1090, 2008.

[20] YANG J.M., CAI R., WANG C., HUANG H., ZHANG L., MA W.Y., *Incorporating Site-Level Knowledge for Incremental Crawling of Web Forums: A List-Wise Strategy*, In Proceedings of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2009.

[21] SAVASERE A., OMIECINSKI E.R., NAVATHE S.B., *An Efficient Algorithm for Mining Association Rules in Large Databases*, Proceedings of the 21th International Conference on Very Large Databases, Morgan Kaufmann Publishers Inc. USA, pp. 432–444, 1995.

[22] LIU B., *Association Rules and Sequential Patterns* Web Data Mining. Exploring Hyperlinks, Contents and Usage Data, Part of Data-Centric Systems and Applications Book Series (DCSA), Springer-Verlag Berlin Heidelberg, pp. 13–54, 2007.

[23] BRIN S., MOTWANI R., ULLMAN J., TSUR S., Dynamic *Itemset Counting and Implication Rules for Market Basket Data*, ACM SIGMOD Conf. Management of Data, 1997.

[24] HAN E.H., KARYPIS G., KUMAR V, MOBASHER B., *Clustering Based On Association Rule Hypergraphs*, Workshop on Research Issues on Data Mining and Knowledge Discovery, 1997.

[25] https://blog.gnip.com/tag/spritzer/

[26] FAYYAD U.M., PIATETSKY-SHAPIRO G., SMYTH P., *From Data Mining to Knowledge Discovery: An overview*, In Advances in Knowledge Discovery and Data Mining, AAAI Press, 1996.

[27] PIRNAU M., *Considerations on the Functions and Importance of a Web Crawler*, Proceedings 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, Romania, pp. 25–27, 2015.

[28] TEODORESCU H.N., *On the Responses of Social Networks to External Events*, Proceedings 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, Romania, 25-27, pp. 13–18, 2015.

[29] TOIVONEN H., *Sampling Large Databases for Association Rules*, Proc. 22nd Very Large Data Bases Conf., 1996.

[30] http://www.cnt.rm.ingv.it

[31] HAN E.H., KARYPIS G., KUMAR V., *Scalable Parallel Data Mining for Association Rules*, Proceeding of the ACM SIGMOD Confonference, pp. 277–288, 1997.