The Eurasia Proceedings of
Educational & Social Sciences (EPESS)
ISSN: 2587-1730

The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2016

Volume 4, Pages 242-247

ICEMST 2016: International Conference on Education in Mathematics, Science & Technology

THE WORD "EDUCATION" IN SOCIAL MEDIA

İlker TÜRKER
Karabuk University, Department of Computer Engineering

Serhat Orkun TAN
Karabuk University, Vocational School

ABSTRACT: The word education, which can be expressed as a systematic way of giving the knowledge and skills to the individuals, is mentioned frequently in social media. The fastest-growing network twitter is also the optimal platform to get information from the usage of the "education" word by social communities. From this point of view, utilizing from twitter, this study aimed to investigate how the word "education" is used in social media. The usage frequency of the word "education", the most frequently used terms together with the word "education" and also the location information of education tweets were acquired to evaluate the comments of the individuals on the word. The followings and followers numbers of the people who make education sharing are also stated to see the popularity of these people in social media.

Key words: Education, social media, twitter, term frequency

INTRODUCTION

Social media where billions of data entered on a daily basis has become to an information dissemination and conservation channel with its two billion users worldwide (Daume, 2016). Among the social media services, twitter has gained so much popularity especially with teens and young adults by providing easy connection with friends, family members and strangers (Brenner, 2013; De Cristofaro et al., 2012). The twitter platform provides users sharing their thoughts, information, news, and jokes by using maximum 140 characters of text. Cheap and measurable global communication is possible with twitter and if the user's profile is not private, anyone in the world can see the writings of users which are named as tweets (Derczynski et al., 2015; O'Dea et al., 2015; Tamburrini et al., 2015; Vidal et al., 2016; Cavazos-Rehg, 2016). Hence it can also be assessed as free social micro-blogging service which allows getting information about the data used in this platform (Parra et al., 2016). 500 million tweets are generated every day by nearly 300 million monthly users to share their opinions and emotions all over the world. This generated data density can also be used to inform people on useful issues such as the government's decisions, the stock exchange, natural disasters, political polarization in the public (Abel et al., 2012; Bollen et al., 2011; Conover et al., 2011; Tumasjan et al., 2011). According to a study carried out on the use of twitter, one in eight of messages posted on twitter represents it as a prime source for public access and naturally occurring communication on twitter makes it a perfect place to study the social identity statement (Java et al., 2007; Danescu-Niculescu-Mizil et al., 2011).

Apart from these researchs, the reflections of various terms and their specific sub-terms on twitter related with the fundamental issues that shape our lives such as education, nutrition, health, transportation, and politics have been investigated in recent years. Among these terms, "education" should be kept in the forefront by its meaning as the process of acquiring general knowledge, developing the powers of reasoning and judgment and preparing oneself or others intellectually for mature life (Education, 2016a) People often confuse education with schooling but education is a social process and can also be expressed as a process of living (Education, 2016b). Therefore the diversity of the term education usage forms in social media emerges as an interesting topic.

We focus on the usage of the word education by the twitter users in this study. We investigate some basic parameters related with the word "education" (the word "eğitim" in Turkish language). The first two of these parameters are its usage frequency (word frequency) and the most frequently used terms with the word "education" which were obtained to find the usage rate of the word and the correlation of this word with the other words respectively. The location information of the tweets including the word "education" are obtained to

- Selection and peer-review under responsibility of the Organizing Committee of the conference

*Corresponding author: İlker TÜRKER-icemstoffice@gmail.com

indicate how the use of the word changes regionally. The popularity of the users who tweeted about education are obtained with the followers and followings numbers of these users in social media. Consequently, the data we gathered provide us convenience to assess the individual's comments on the word "education".

METHODOLOGY

In our experiments, we obtained 23,056 tweets gathered between the dates 20-28th January, 2016. To get data from Twitter services, we used a php based application presented by Adam Green (2013). This framework provides a listening to Twitter's streaming API by entering some key words to filter the whole available data stream (that corresponds to the 1% of entire Twitter entries). As keywords to focus our study, we entered the words "eğitim", "egitim", "Eğitim", "Egitim" and "EĞİTİM" as the most frequent variations of this word in Twitter.

In the preprocessing step, we transformed the entries to lowercase versions to provide a simpler word processing. We also converted the Turkish alphabet letters as "ı", "ç", "İ" to their English pairs as "i", "c", "I" etc for removing the varieties of words driven by different keyboard usage preferences. We also removed the hashtags (#topics), mentions (@users), http links, non-alpha numeric characters and stop-words that do not express meaningful aspects. Although a filtering array is applied for listening the Twitter Streaming API related with the word education, some irrelevant tweets are also included in our raw dataset. The deletion of these irrelevant tweets is also performed in preprocessing. After the preprocessing step, the number of Tweets were reduced to 19,479, ready for word processing.

We processed the words and evaluated the usage statistics by a software we developed in C# language, accessing the database rendered by the 140.dev application in MySQL format. Although the most suitable environment for MySQL database is php coding, we preferred C# to empower the word processing by its powerful regex features.

RESULTS and DISCUSSION

Social media users write words in different ways for some reasons. These variations are generally driven by the keyboard restrictions, need for shortening words for giving fast response, obeying the allowed character number (140 characters for twitter), saving time and reflecting the user's own writing style. Especially, the typing variations in mobile media are generally affected by the keyboard configurations that restrict the users to some common character sets. As mentioned above, we investigated the usage of the word "education" in social media, twitter, in Turkey. The word "education" corresponds to "eğitim" in Turkish and the words at the table below used for the same purpose but represent the different ways of using this word.

Table 1. The Most Frequent Ways of Using the Word "Education" in Turkish Language, in Twitter.

"egitim"
"eğitim"
"Eğitim"
"Egitim"
"EĞİTİM"

Since the raw data are infected by some facts we mentioned in the previous section, we employed a preprocessing procedure to clarify the dataset and simplify data processing. The details of this procedure are also mentioned in the Methodology section. After preprocessing, our database was shrinked to 19,479 tweets by the deletion of 3,577 irrelevant tweets as shown in Table 2.

Table 2. The Usage Number of Tweets on the Word "Education"

Total Tweets	23,056
Tweets after preprocessing	19,479
Tweets filtered in preprocessing	3,577

The Twitter Streaming API also provides the basic user data relative with the tweet stream. By the way we can evaluate the user statistics that are the sources of the tweets gathered. The provided data are the user location, followings (friends) and followers counts. We first evaluated the average following and followers numbers of the twitter users that tweet "education" in Turkish. This data states the popularity of the users as incoming and outgoing links to the entire Twitter users community. To give a comparison with the entire Twitter users follow profiles, we also gathered the average followings and followers counts from an irrelevant database of 10,000 tweets that are collected with no keywords but only in Turkish language. The resulting data is presented in Table 3 below.

Table 3. The Popularity of the Twitter Users Who Tweeted on Education

Number of users tweeted on education: 18460			
The average numbers for the users tweeted on education		The average numbers for the whole Turkish users tweet about anything on twitter	
Followings	Followers	Followings	Followers
1611.34	4800.97	2617.87	6143.94

As seen from Table 3, the number of users tweeted about education has lower average number of followers and followings when compared with the whole users average. This may be an indicator that the popular or trending users in Twitter are less likely to tweet about education than the users having moderate popularity. The occurrence frequency of the education tweets was also calculated as % 0.365 when 10,000 samples of tweets observed in twitter with no keywords but Turkish language. The most frequently used 60 associable words together with "eğitim" are shown in Table 4. The repetition numbers are also stated with these words. The count of whole words engaged with education is 17,833.

Table 4. The Repetition Number of the Words Used With Education

Word	Frequency	Word	Frequency	Word	Frequency
egitim	19678	*tekin*	528	*saniyede*	385
sistemi(miz)(nin)	2102	*gerekiyor*	455	*ajanlari*	372
milli	1687	*karne*	451	*ulkene*	372
turkiye	1204	*muduru*	444	*bozgunculuk*	371
hakki(ni)	1113	*alman*	439	*diyarbakir*	371
verip(elim)	1101	*ygs*	433	*topraklarimizda*	371
ana(okulu)	997	*iyi*	430	*cikartiyor*	370
meb	990	*hayatim*	427	*gelinlikli*	367
pkk	932	*sinavina*	420	*istanbul*	361
basin	790	*girse*	414	*futbol*	357
akp	702	*dini*	413	*cocuk*	352
egitimecan	692	*ilk*	413	*gun*	347
dilin	682	*kastamonu*	412	*teror*	339
yozlasmamasi	681	*ziyaret*	407	*ozel*	331
bakani	677	*aristo*	406	*ocak*	308
sinifcilarakadro	659	*kazanir*	406	*smiley**	293
arastirma	639	*felsefeyi*	404	*merkezi*	291
okul	621	*sahip*	398	*temel*	291
ogretim	606	*gunu*	395	*kurumu*	286
il	596	*hastanesi*	391	*sinif*	286

* Smiley stands for all smiley characters like :), :-], :(((etc.

The most used words together with "eğitim" can give us idea about the general thoughts, political ideas, emotions, suggestions, expectations and complaints of the twitter users in Turkey. Expectedly, the word "sistem" is the most frequent word paired with "eğitim", expressing the arguments about the national education system that are going on for several decades. Another standing out word is the synonym for the most known terror association in Turkey, that caused the abortion of education in the south-east of Turkey in January this year. Another frequent word "sinifcilarakadro" indicates the demands for employment in first-grade education, as a reality of inactive teachers' community in Turkey. The realities underlying the remaining words are left to the readers attention.

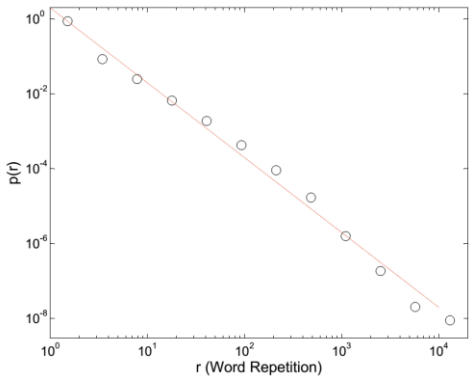

Fig.1. The distribution graph for the word repetition that are used with the Word "eğitim".

The distribution graph of the repetition of these words is also presented in Fig.1. This graph is generated with the data collection including 17,833 words with several repetition counts. Every repetition counts are grouped and cumulated to generate the distribution graph. As seen in the log-log scale, the distribution is consistent with a perfect power-law decay, indicating that the universal laws of complexity is also evident in the weight distribution of a word as a node neighbor with the other words in a sentence (Petersen et.al, 2012). The average repetition counts for the whole words is 13.42.

The regional representation of the tweet words used with the word education on a provinces basis is given in Table 5. The data provided insight about the regional usage intensity of the words used with education by twitter users in Turkey. The data also show us the interest ratio of the Twitter users to the word education in regional basis. This regionally usage ratio may be caused by some reasons as modes of living, economic conditions, etc.

A limiting factor in this dataset is that the location data of the users are rarely generated by their mobile devices automatically but generally the users prefer entering their location manually. This fact leads a large variation of location notifications for a particular location, i.e. “Beşiktaş, İst”, “Kadıköy”, “Taksim, İstanbul”, “İslambol”, “Konstantinapolis” etc. for “İstanbul”. To remove this variety, we detected all the variations for a certain location and fix them in one name like “İstanbul”, by a hand driven procedure. By the end of this procedure, our dataset is limited to 5,387 users having a valid location declared. The resulting data is presented in Table 5 below.

Table 5. The Tweet Counts Used With the Word Education and Population Ratios by Provinces.

Provinces	Total Population*	Repetition	Repetition / Population (million)
Tunceli	86,076	23	267,2
Bayburt	78,55	10	127,3
İstanbul	14.657.434	1778	121,3
İzmir	4.168.415	477	114,4
Ankara	5.270.575	518	98,3
Elazığ	574,304	56	97,5
Rize	328,979	32	97,3
Uşak	353,048	33	93,5
Gümüşhane	151,449	14	92,4
Trabzon	768,417	70	91,1
Sakarya	953,181	82	86
Eskişehir	826,716	71	85,9
Ordu	728,949	62	85,1
Karabük	236,978	20	84,4
Edirne	402,537	33	82
Kocaeli	1.780.055	143	80,3
Samsun	1.279.884	96	75
Yalova	233,009	17	73
Antalya	2.288.456	166	72,5
Muğla	908,877	63	69,3
Bolu	291,095	20	68,7
Bartın	190,708	12	62,9
Denizli	993,442	62	62,4
Giresun	426,686	26	60,9
Aksaray	386,514	23	59,5
Artvin	168,37	10	59,4
Çanakkale	513,341	30	58,4
Şanlıurfa	1.892.320	103	54,4
Adana	2.183.167	118	54
Nevşehir	286,767	15	52,3
İçel	1.745.221	90	51,6
Bursa	2.842.547	140	49,3
Manisa	1.380.366	67	48,5
Malatya	772,904	37	47,9
Karaman	242,196	11	45,4
Isparta	421,766	19	45
Adıyaman	602,774	27	44,8
Batman	566,633	25	44,1
Siirt	320,351	14	43,7

Erzurum	762,321	32	42
Zonguldak	595,907	25	42
Tekirdağ	937,91	38	40,5
Aydın	1.053.506	41	38,9
Çankırı	180,945	7	38,7
Diyarbakır	1.654.196	63	38,1
Niğde	346,114	13	37,6
Sivas	618,617	22	35,6
Konya	2.130.544	73	34,3
Sinop	204,133	7	34,3
Amasya	322,167	11	34,1
Bingöl	267,184	9	33,7
Gaziantep	1.931.836	65	33,6
Kütahya	571,463	19	33,2
Bilecik	212,361	7	33
Kırklareli	346,973	11	31,7
Balıkesir	1.186.688	37	31,2
Iğdır	192,435	6	31,2
Burdur	258,339	8	31
Tokat	593,99	17	28,6
Ağrı	547,21	15	27,4
Kırşehir	225,562	6	26,6
Yozgat	419,44	11	26,2
Kırıkkale	270,271	7	25,9
Kayseri	1.341.056	34	25,4
Çorum	525,18	13	24,8
Kastamonu	372,633	9	24,2
Afyonkarahisar	709,015	17	24
Kahramanmaraş	1.096.610	26	23,7
Van	1.096.397	26	23,7
Osmaniye	512,873	12	23,4
Kilis	130,655	3	23
Erzincan	222,918	5	22,4
Hatay	1.533.507	32	20,9
Şırnak	490,184	10	20,4
Muş	408,728	8	19,6
Düzce	360,388	7	19,4
Mardin	796,591	14	17,6
Bitlis	340,449	5	14,7
Ardahan	99,265	1	10,1
Hakkari	278,775	1	3,6
Kars	292,66	1	3,4

* Total population numbers are referenced in (Nüfus, 2016)

CONCLUSION

The usage of the word “education” on twitter was investigated by the data collected from the users in Turkey. The collected data include the number of users tweeted on education, the average number of their followings and followers and the whole user’s average followings and follower numbers. Comparing with the whole users averages, it is determined that the number of users tweeted on education has lower average number of followers and followings. The most used words together with “eğitim”, their repetition numbers and the distribution of the repetition are also stated to give an idea about the general thoughts, behaviors, political ideas, suggestions, expectations and complaints of the twitter users lives in Turkey. The regionally usage intensity of these most used words together with “eğitim”, is also examined on a provinces basis in Turkey. The numerical results for the regionally usage ratio may be attributed to the modes of living, economic conditions and other living conditions at this region of Turkey.

ACKNOWLEDGMENTS

We thank Samet ÖNLER, Department of Computer Engineering, Karabuk University, for his support at data collection in our study.

REFERENCES

Abel, F., Hauff, C., Houben, G., Stronkman, R., & Tao, K. (2012). Twitcident. Proceedings of the 21st International Conference Companion on World Wide Web - WWW '12 Companion.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

Brenner, J. (2013). Pew internet: social networking. Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/fact-sheets/socialnetworking- fact-sheet/

Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., & Bierut, L. J. (2016). A content analysis of depression-related tweets. Computers in Human Behavior, 54, 351-357.

Conover, M. D., Goncalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). Predicting the Political Alignment of Twitter Users. 2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing.

Cristofaro, E. D., Soriente, C., Tsudik, G., & Williams, A. (2012). Hummingbird: Privacy at the Time of Twitter. 2012 IEEE Symposium on Security and Privacy.

Danescu-Niculescu-Mizil, C., Gamon, M., & Dumais, S. (2011). Mark my words!: linguistic style accommodation in social media Proceedings of the 20th International Conference on World Wide Web - WWW '11.

Daume, S. (2016). Mining Twitter to monitor invasive alien species — An analytical framework and sample information topologies. Ecological Informatics, 31, 70-82.

Derczynski, L., Maynard, D., Rizzo, G., Erp, M. V., Gorrell, G., Troncy, R., Bontcheva, K. (2015). Analysis of named entity recognition and linking for tweets. Information Processing & Management, 51(2), 32-49.

Education. (2016a). In The online free dictionary. Retrieved from http://dictionary.reference.com/browse/education?s=t

Education. (2016b). In Infed. Retrieved from http://infed.org/mobi/what-is-education-a-definition-and-discussion/

Green, A. (2013). In 140Dev, Twitter API Programming Tips, Tutorials, Source Code Libraries and Consulting. Retrieved from http://140dev.com/

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblog-ging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis - WebKDD/SNA-KDD '07.

Nüfus. (2016). In Türkiye Nüfusu. Retrieved from http://www.nufusu.com/

O'dea, B., Wan, S., Batterham, P. J., Calear, A. L., Paris, C., & Christensen, H. (2015). Detecting suicidality on Twitter. Internet Interventions, 2(2), 183-188.

Parra, D., Trattner, C., Gómez, D., Hurtado, M., Wen, X., & Lin, Y. (2016). Twitter in academic events: A study of temporal usage, communication, sentimental and topical patterns in 16 Computer Science conferences. Computer Communications, 73, 301-314.

Petersen, A. M., Tenenbaum, J. N., Havlin, S., Stanley, H. E., & Perc, M. (2012). Languages cool as they expand: Allometric scaling and the decreasing need for new words. Sci. Rep. Scientific Reports, 2.

Tamburrini, N., Cinnirella, M., Jansen, V. A., & Bryden, J. (2015). Twitter users change word usage according to conversation-partner social identity. Social Networks, 40, 84-89.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape. Social Science Computer Review, 29(4), 402-418.

Vidal, L., Ares, G., & Jaeger, S. R. (2016). Use of emoticon and emoji in tweets for food-related emotional expression. Food Quality and Preference, 49, 119-128.