Chapter 6

Mining Twitter Content for Discovering Open and Distance Education Trends in Turkey

Burcu Yılmazel[1]

Ali Yürekli[2]

Abstract

Social media, which has become an active communication tool in today's education, constitutes a fast and easy alternative to share information by bringing students and educational institutions together. Although the interaction between these participants could provide implicit feedback on education services, there is only limited research on identifying the trending topics about open and distance education. This study examines Twitter content to reveal the primary topics of social media conversations related to open and distance education in Turkey. An experimental research is conducted on a collection of 20,010 unique tweets matching #aöf and #açıköğretim hashtags. The user tweets in this collection, which consist of hashtags and actual tweet texts, are analyzed by two statistical inference methods. While the most frequently preferred hashtags in the education domain are determined by co-occurrence modeling, Latent Dirichlet Allocation is employed to extract the core topics of actual tweet texts. The performed analyses reveal that social media interactions in open and distance education are gathered around semantic clusters such as exams, registration periods, course materials, and exam results. Consequently, social media can be utilized to understand students' problems and demands better, and thus the quality of distance education services can be enhanced.

INTRODUCTION

In recent years, the rapid growth in the use of social networking services has radically affected the general perception of communication in the education

1 Eskişehir Technical University, Faculty of Engineering, Department of Computer Engineering, Eskişehir, Turkey, (ORCID: 0000-0001-8917-6499), byurekli@eskisehir.edu.tr

2 Eskişehir Technical University, Faculty of Engineering, Department of Computer Engineering, Eskişehir, Turkey, (ORCID: 0000-0001-8690-7559), aliyurekli@eskisehir.edu.tr

https://doi.org/10.58830/ozgur.pub60.c265

domain (Poellhuber et al., 2011). Social platforms like Facebook, Twitter, and Instagram are frequently used as alternative channels to traditional communication methods. Almost all educational institutions own and manage their official social media accounts (mostly, an institution has more than one account due to the variety of popular platforms) to interact with the community. These accounts are commonly utilized as broadcast tools (Kimmons et al., 2017) that constitute an easy and fast way of information sharing (e.g., announcements, events, advertisements, and course materials).

Especially in the field of open and distance education (ODE), the use of social media for communication is of special importance. Although ODE systems help eliminate obstacles in education for several reasons, such as economic or geographical limitations, students may still study in an isolated environment, which causes them to be deprived of necessities like person-to-person contact and faculty support (Muilenburg & Berge, 2001). Social media can be regarded as a practical option to alleviate this isolation problem since it serves as a unifying factor that brings the ODE community (i.e., students, instructors, and educational institutions) together (Özmen & Atıcı, 2014). In this context, it becomes essential to perceive which topics related to ODE in social media are frequently encountered and what public opinion these topics form on the users. Knowing what students are talking about and understanding their problems is highly valuable to improve the overall quality of education services.

In this study, we investigate the primary topics in social media related to the Turkish ODE system. We focus on user tweets from Twitter concerning the Open Education Faculty (OEF) of Anadolu University[3], one of the global pioneers of distance education that serves over one million students (Latchem et al., 2006). Firstly, we collect the relevant social media content by searching #aöf and #açıköğretim (i.e., the abbreviation of OEF and spelling of open education in Turkish, respectively) hashtags on Twitter. Then, we perform a hashtag co-occurrence analysis on the acquired tweet collection to gain insights about the trending hashtags in Turkish ODE. Afterward, we analyze the actual tweet texts and extract the core topics of social media conversations about ODE in Turkey by building a Latent Dirichlet Allocation (LDA) model (Blei et al., 2003). Based on these core topics, we explain and illustrate how students use social media for information sharing and community support. Briefly, the performed analyses on user tweets search for an answer to the following research questions (RQ):

3 https://www.anadolu.edu.tr/acikogretim

- RQ 1: Under which hashtags do social media conversations about the Turkish ODE system take place?
- RQ 2: What are the primary topics related to the Turkish ODE system in social media conversations?
- RQ 3: In line with the primary topics of ODE conversations, how do users utilize social media for communication in education?

The remainder of the chapter is organized into three main sections. Firstly, the adopted methodology and techniques used to carry out this research are described. Secondly, the findings and discussions of the performed analyses are presented. Finally, conclusions are drawn and future work is pointed out.

METHODOLOGY

In this study, we apply two statistical models to identify significant hashtags and topics in the Turkish ODE system: *(i)* co-occurrence model for hashtags and *(ii)* topic model for tweet texts. After describing the collection procedure of the data used in the study, we explain these models in more detail, respectively.

Data Collection

Hashtag usage is a convenient way of organizing information in social media. Simply, a hashtag is a string of characters proceeded by the hash (#) character. In terms of Twitter interactions, hashtags typically indicate the context, main topics, and core ideas of the tweets (Tsur & Rappoport, 2012). Users can follow the field of interest and participate in discussions by these searchable metadata tags. Accordingly, it is possible to explore the topics in social media by querying relevant hashtags.

Since this study focuses on social media conversations about the Turkish ODE system, we choose #aöf and #açıköğretim hashtags as the related query terms. In a prior study, Ozturk et al. (2017) perform sentiment analysis for the Turkish ODE system on a set of tweets gathered by these hashtags and some of their variations (i.e., alternative spellings as a consequence of the special characters in the Turkish alphabet). Similarly, we collect the tweets in the Turkish language that contain our target hashtags using the Search API[4] provided by Twitter. Having collected 21,718 tweets in total, we observe that the performed queries may occasionally retrieve some tweets more than once, which results in duplicate social media entries. By interpolating a

4 https://developer.twitter.com/en/docs/twitter-api

deduplication step to the data collection process, we obtain 20,010 unique tweets matching our filtering criteria.

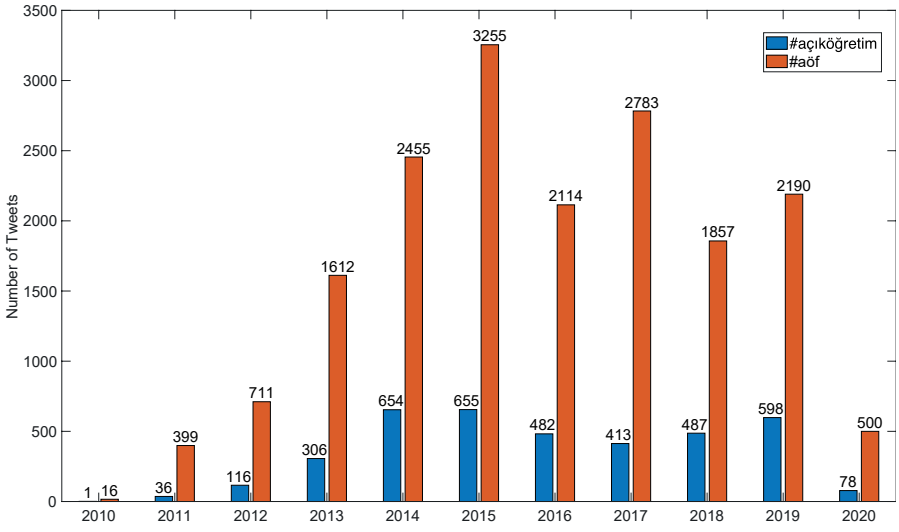

Figure 1. The number of tweets collected according to #aöf and #açıköğretim hashtags by years

As presented in Figure 1, user entries in the tweet collection span a time period between 2010 and 2020 (limited by the last date of querying, which is mid-February, 2020). Notably, #aöf has been preferred more frequently than #açıköğretim as a hashtag pointing the conversations in Twitter related to the Turkish ODE system.

Hashtag Co-occurrence Modeling

Hashtags are well-indicators of underlying topics in tweets. Since Twitter allows its' users to include multiple hashtags in a tweet, the hashtags correlated with each other may appear together in the same tweet. This co-occurrence relationship can be utilized to gain insights about tweet topics (Steinskog et al., 2017). Eventually, the understanding and visualization of hashtag dynamics is a reasonable starting point to investigate the social media trends of a particular subject.

In order to identify the prominent hashtags about ODE conversations on Twitter, we first apply textual normalization on the hashtag terms as an initial step. Employing lower casing, deASCIIfication, and lemmatization filters consecutively, all hashtags are converted to much common forms. While deASCIIfication restores accented letters in hashtag terms (Arslan,

2016), lemmatization helps group together the inflected forms of a hashtag term (Akba et al., 2014). In brief, the primary purpose of this pre-processing procedure is to singularize hashtags that mean the same or close, even though they are written in different forms.

The application of the abovementioned pre-processing step results in a set of *n* unique hashtags. Using this set, a co-occurrence matrix can be constructed to present how many times a hashtag appears together with another hashtag. As the relationship between two hashtags is direct, co-occurrence counting forms a natural statistic for the relatedness or closeness of the corresponding terms (Bullinaria and Levy, 2012). In Figure 2, we illustrate the resulting matrix obtained by data pre-processing and co-occurrence counting. Formally, the co-occurrence matrix of hashtags is a square *n x n* matrix where *n* corresponds to the number of unique hashtags in the tweet collection. In this context, a cell f_{ij} denotes the number of times hashtag h_i appears together with hashtag h_j. It should also be noted that the upper and lower triangles of the matrix are identical as co-occurrence is a symmetric relation (Lin, 2008). In another saying, f_{ij} is equal to f_{ji} for the two hashtags h_i and h_j.

	h_1	h_2	h_3	• • •	h_n
h_1	f_{11}	f_{12}	f_{13}		f_{1n}
h_2	f_{21}	f_{22}	f_{23}		f_{2n}
h_3	f_{31}	f_{32}	f_{33}		f_{3n}
•	•	•	•		•
•	•	•	•		•
•	•	•	•		•
h_n	f_{n1}	f_{n2}	f_{n3}	• • •	f_{nn}

Figure 2. A visualization of the co-occurrence matrix of hashtags in the tweet collection

As described in the data collection section, we use #aöf and #açıköğretim hashtags as primary query terms when searching Twitter content related to OEF and the Turkish ODE system. Accordingly, all of the collected tweets contain at least one of these two hashtags, highlighting a sub-region of the co-occurrence matrix as a critical region in terms of overall hashtag preference. Theoretically, the projection of rows (or columns, since the matrix is symmetric) having #aöf and #açıköğretim terms as indices can reveal under which hashtags social media conversations take place (as questioned in RQ 1).

LDA Topic Modeling

In the context of social media, understanding the characteristics of conversations is critical for various tasks, such as semantic analysis and friend recommendations (Alvarez-Melis and Saveski, 2016). The reviews, comments, and feedbacks of users may positively or negatively affect other users' opinions, and the resulting interaction can be utilized to provide better services to the community. This mutual approach can also be applied in the field of ODE. Analyzing the trending topics of social media conversations about ODE can be a guide on how to improve the quality of education by revealing the issues that students care about or have problems with.

A frequently used technique to discover hidden structures in text collections is topic modeling. Given a set of text documents, topic modeling aims to extract abstract topics by learning meaningful patterns of words (Dieng et al., 2020).

In this study, we build a Latent Dirichlet Allocation (LDA) model to identify the main topics of social media conversations related to the Turkish ODE system. LDA is an unsupervised machine learning technique that views documents as bag-of-words. In simple terms, LDA generates a mixture of latent topics from a document collection, in which each document is modeled as a distribution over topics, and the topics are represented as a distribution over words (Blei, 2012). Weng et al. (2010) define the generative process in LDA as follows:

i. For each document in the collection, a topic is chosen from the document's distribution over topics.

ii. A word is sampled from the distribution over the words related to the chosen topic.

iii. The process is repeated for all words in the document.

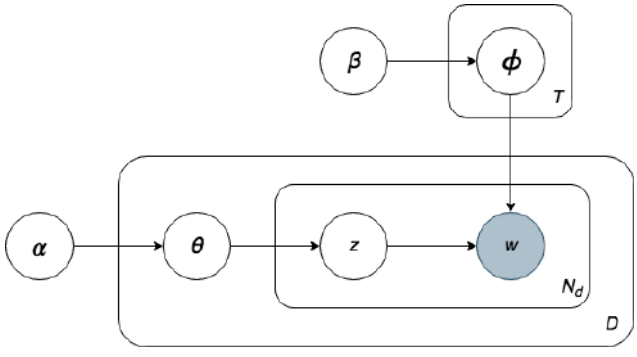

Figure 3. The graphical model representation of the LDA model (Weng et al., 2010)

The graphical model representation of the LDA model is illustrated in Figure 3. Using the analogy of tweet collection as documents, the notations in the model can be expressed as follows:

- D represents the tweet collection.
- T represents the set of topics to be sampled.
- α is the hyperparameter of per-tweet topic distributions.
- β is the hyperparameter of per-topic word distribution.
- θ is a multinominal distribution over topics.
- ϕ is a multinominal distribution over words.
- N_d is the total number of words in tweet d.
- z is the sampled topic from the distribution θ associated with tweet d.
- w is the sampled word from the distribution ϕ associated with topic z.

When building the LDA model, we initially associate each tweet in the tweet collection D with θ, and each topic in the topic set T with ϕ. These multinominal distributions are interpolated with two hyperparameters, α and β, respectively. Then, for each word (or term) in a tweet d, we sample a topic z from θ associated with tweet d. Subsequently, a word w is sampled from ϕ associated with topic z. We repeat this procedure N_d times in order to form tweet d.

FINDINGS AND DISCUSSIONS

In this study, a collection of 20,010 unique tweets is analyzed to understand the main topics in social media conversations related to the Turkish ODE system. The findings of the analysis are presented according to the aforementioned research questions, respectively.

Analysis of Hashtags

When we analyze the tweet collection in terms of hashtag usage, we observe that a total number of 50,615 hashtags (6,616 unique terms) have been used as the descriptors of conversations. Roughly estimating, it can be said that users prefer 2 to 3 hashtags (approximately 2.53 terms on average) for each tweet in the dataset.

Before building the hashtag co-occurrence matrix, we apply a textual normalization (containing lower casing, deASCIIfication, and lemmatization) on the hashtag terms. Aiming to reduce data dimensionality, this procedure results in a condensed set of normalized hashtags having 5,641 unique terms

(with a compression ratio of 15%). Although this set can be considered quite large in terms of semantic content, not all the terms have equal significance when pointing to a specific subject (Lijffijt et al., 2016). In Figure 4, we present a semantic word cloud (Xu et al., 2016) of the most frequent hashtags obtained after the normalization process. As expected, while #aöf and #açıköğretim hashtags stand out in the data distribution, many terms related to ODE (e.g., #sınav, #anadoluüniversitesi, and #ösym) are also included in the word cloud.

Figure 4. The most frequent hashtags after textual normalization (more common terms appear larger)

Eliminating the hashtags that occur only once after textual normalization, we build the hashtag co-occurrence matrix to identify the significant terms concerning the Turkish ODE system. The projection of rows by the target hashtags (i.e., #aöf and #açıköğretim) reveals the terms which form the basis of social media conversations about OEF and the Turkish ODE system. In Table 1, we provide a list of remarkable terms that commonly co-occur with the target hashtags.

Table 1. A list of hashtags that commonly co-occur with #aöf and #açıköğretim terms

Category	Hashtags	Description
Anadolu University	▪ #anadolu ▪ #anadoluüniversitesi ▪ #eskişehir ▪ #aü	Hashtags, which are directly pointing Anadolu University, are commonly used in ODE tweets.
Educational Terminology	▪ #eğitim ▪ #tercih ▪ #sınav ▪ #sınavadoğru ▪ #final ▪ #vize ▪ #öğrenciişleri ▪ #sonuç ▪ #sınavsonuçları	Terms of educational terminology, such as examination, student affairs, examination results, are frequently mentioned as hashtags of tweets.
Nationwide Examinations	▪ #lys ▪ #ygs ▪ #ales ▪ #kpss ▪ #tyt ▪ #ayt ▪ #öabt ▪ #dgs	This category includes the hashtags Indicating the abbreviations of nationwide examinations in Turkey, such as KPSS (Kamu Personel Seçme Sınavı) and DGS (Dikey Geçiş Sınavı).
Courses in OEF	▪ #işletme ▪ #kamuyönetimi ▪ #iktisat ▪ #matematik ▪ #sosyoloji ▪ #webtasarım ▪ #ingilizce ▪ #hukuk ▪ #ilahiyat	The names of major courses offered by OEF (e.g., Economics, Public Administration, and Web Design) are frequently used as hashtags of tweets.

As presented in Table 1, the prominent hashtags in the hashtag co-occurrence matrix can be grouped into four categories, namely, "Anadolu University", "Educational Terminology", "Nationwide Examinations", and "Courses in OEF".

i. Anadolu University: Being a pioneer and worldwide brand in ODE, Anadolu University has been commonly mentioned in Turkish social media conversations.

ii. *Educational Terminology:* Generic hashtags related to ODE (e.g., #sınav for examinations, #vize for midterms, #sonuç for results, and #öğrenciişleri for student affairs) can be reviewed under educational terminology.

iii. *Nationwide Examinations:* In Turkish higher education, central examinations applied throughout the country have an important place, especially for university entrance and future career planning. The abbreviations of these nationwide examinations (e.g., #kpss for "Kamu Personeli Seçme Sınavı") have been frequently used as hashtags.

iv. *Courses in OEF*: The Facebook groups of particular ODE programs with higher employment opportunities are more active in social media (Fırat et al., 2017). Similarly, the major courses offered by OEF have been often used as the hashtags of Twitter content.

The performed hashtag analysis reveals that hashtags can be effectively used to gain insights about educational topics in social media. As a concluding remark on RQ 1, it can be stated that the conversations take place around a set of particular terms, which can be investigated in four primary categories described above.

Analysis of Tweet Texts

When searching for primary topics about the Turkish ODE conversations, we examine 20,010 user entries from Twitter. In addition to the pre-processing procedure (lower casing, deASCIIfication, and lemmatization) previously described in hashtag co-occurrence modeling, stop word elimination for Turkish is also applied to these tweets. The main purpose of this additional step is to remove common Turkish words that do not carry significant meaning (e.g., "ve", "şey", "bazı", "şunu", "kadar", "ile", and "buraya") (Çakir and Güldamlasioğlu, 2016). The elimination of insignificant terms is beneficial to focus on the important words instead.

Table 2. The eight topics extracted using the LDA model and their associated terms

Topic #	Associated terms	Possible topic interpretation
T1	final, giriş belgesi, vize, yayınlama, aöf, sınav yeri, açıklama, telefon, kampüs, tıklama.	Communication upon the determination of exam buildings and the announcement of exam entry documents
T2	çalışmak, sabah, ders, stres, uyku, son gece, soru çözme, kitap, konu, başarılı.	Study-themed dialogues in the pre-examination period
T3	sınav, yarın, sonuç, belli, bugün, açıklama, ne zaman, haydi, sistem, açıköğretim.	Communication upon the announcement of examination results
T4	günaydın, başarılar, haftasonu, sınav, cumartesi, dileme, herkes, aöf, öğrenci, final.	Wishes for success in exams among ODE students
T5	video, sınav, aöf, çıkmış soru, matematik, cevap, konu, anlatım, ders, özet.	Sharing of course materials, examination questions and their solutions
T6	aöf, sonuç, dört yanlış, bir doğru, zor, götürme, sistem, neden, açıköğretim, doğru.	Discussions and opinions about the change in exam evaluation criteria
T7	final, mutluluk, üç ders, bütünleme, mezun, mezuniyet, diploma, tek ders, inşallah, sene.	Exchange of information and news on correlated topics, including graduation, diploma, and make-up exams
T8	yenileme, harç, güz, bahar, dönem, kayıt, ne zaman, ücret, son gün, dikkat.	Sharing information about registration renewal periods and tuition fees

As a generative probabilistic model, LDA uses a specified topic number when allocating words to topics. Thus, the effectiveness of LDA is highly dependent on the number of topics parameter chosen during model generation. Empirically, we create different LDA models using all the values between 4 and 10 as the number of topics parameter. Best observations (in terms of semantic topic decomposition) are attained by choosing eight as the number of topics. In Table 2, we present the eight topics extracted using the corresponding LDA model. For each of the topics, we also provide a list of associated terms and a possible interpretation of the word clusters.

As a response to RQ 2, the performed analysis exhibits that these eight topics (T1, T2, T3, T4, T5, T6, T7, T8) form the basis of social media conversations related to the Turkish ODE system.

When we analyze the primary topics of ODE-related social media conversations, we observe that these topics can be reviewed in two categories concerning the intended communication purpose of users:

i. *Information Sharing:* The vast majority of conversations about ODE on social media occur for the purpose of information sharing. The official announcements about ODE made by Anadolu University (T1, T3, and T8) are based on information sharing. Users commonly share important dates and information with each other, including registration renewal periods, tuition fees, exam locations, and exam results. In addition, students also frequently share auxiliary materials related to courses and exams (T2 and T5), such as topics summaries and past examination questions and answers. These facts promote social media as a significant source of information for students.

ii. *Community Support:* Notably, social media has been frequently used as a communication channel where students share their problems and future expectations, discuss various issues, and wish each other success, especially for examinations (T4). Study motivation and stress for exams, assessment and evaluation system of ODE, graduation and diploma expectations, semester tuition fees (T2, T6, T7, and T8) are the main topics of student discussions. With the help of social media, an individual student in ODE finds the opportunity to share his thoughts and experiences with people who are going through similar processes.

In order to summarize how students utilize social media communication (that is questioned in RQ 3), we present Table 3, which is a mixture of primary topics discovered using the LDA model and the two purposive categories (i.e., information sharing and community support) mentioned above.

Table 3. The purposes of social media communication in ODE in the perspective of information sharing and community support

		T1	T2	T3	T4	T5	T6	T7	T8
Information Sharing	Announcements	√		√					√
	Exam dates	√			√				
	Exam entrance documents	√							
	Exam locations	√							
	Exam results			√			√	√	
	Registration renewal periods								√
	Tuition fees								√
	Course materials		√			√			
	Topics summaries		√			√			
	Exam questions and answers					√			
	Lecturing					√			
Community Support	Make-up exams							√	
	Graduation							√	
	Stress		√	√					
	Self-expression		√		√		√	√	√
	Wishes of success				√				
	Study motivation		√			√			
	Assessment and evaluation critics			√			√		
	Diploma							√	

CONCLUSION

The widespread use of social media platforms and applications promotes social media as one of the most effective communication tools of the modern age. As in many other fields, social media has been frequently preferred as an alternative way of communication in education.

In this study, we investigate what users talk about ODE on social media and how they utilize this emerging communication tool in education. On a collection of user tweets from Twitter (collected by #aöf and #açıköğretim hashtags), we analyze the hashtag usage and primary topics of conversations by building two independent statistical models. The co-occurrence relationship among hashtags reveals a set of prominent terms that can be associated with four main categories (i.e., Anadolu University, educational terminology, nationwide examinations, and courses in OEF). Although these common hashtags are structurally short and concise, they can procure preliminary information about ODE in social media. On the other hand, the actual tweet texts are richer in content and provide more insights about the primary topics of the Turkish ODE system. The LDA topic model employed in our experiments exhibits eight major topics (i.e., T1 to T8) forming the base of social media conversations. These semantic clusters point out that the ODE participants mostly utilize social media for information sharing and community support.

This study shows that social media, an effective alternative in modern communication, handles many different subjects in the field of education. Following the core topics in social media paves the way for directly perceiving students' problems and demands. In this way, social media can be utilized to increase the quality of education services.

The natural next step of this study is the application of sentiment analysis on the primary topics of ODE, which are attained by topic modeling. Several studies have performed sentiment analysis on Twitter content related to the Turkish ODE (Ozturk et al., 2017). However, examining the positive, negative, or neutral attitude of users on a topic-by-topic basis may also provide interesting inferences.

Another possible research direction is the extension of performed analyses to include time dimension. It can be investigated how the interaction of users on social media changes over time, as well as how the primary topics are shaped according to the academic calendar.

REFERENCES

Akba, F., Uçan, A., Sezer, E. A., & Sever, H. (2014). Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews. In *8th European Conference on Data Mining* (Vol. 191, pp. 180-184).

Alvarez-Melis, D., & Saveski, M. (2016). Topic modeling in twitter: Aggregating tweets by conversations. In *Tenth international AAAI conference on web and social media*.

Arslan, A. (2016). DeASCIIfication approach to handle diacritics in Turkish information retrieval. *Information Processing & Management*, *52*(2), 326-339.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84.

Bullinaria, J. A., & Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and SVD. *Behavior research methods*, *44*(3), 890-907.

Çakir, M. U., & Güldamlasioğlu, S. (2016). Text mining analysis in Turkish language using big data tools. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 1, pp. 614-618). IEEE.

Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, *8*, 439-453.

Fırat, M., Altınpulluk, H., Kılınç, H., & Büyük, K. (2017). Determining open education related social media usage trends in Turkey using a holistic social network analysis. *Educational Sciences: Theory & Practice*, *17*(4).

Kimmons, R., Veletsianos, G., & Woodward, S. (2017). Institutional uses of Twitter in US higher education. *Innovative Higher Education*, *42*(2), 97-111.

Latchem, C., Özkul, A. E., Aydin, C. H., & Mutlu, M. E. (2006). The open education system, Anadolu University, Turkey: e-transformation in a mega-university. *Open Learning: The Journal of Open, Distance and e-Learning*, *21*(3), 221-235.

Lijffijt, J., Nevalainen, T., Säily, T., Papapetrou, P., Puolamäki, K., & Mannila, H. (2016). Significance testing of word frequencies in corpora. *Literary and Linguistic Computing*, *31*(2), 374-397.

Lin, J. (2008). Scalable language processing algorithms for the masses: A case study in computing word co-occurrence matrices with MapReduce. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 419-428).

Muilenburg, L., & Berge, Z. L. (2001). Barriers to distance education: A factor-analytic study. *American Journal of Distance Education*, *15*(2), 7-22.

Ozturk, Z. K., Cicek, Z. E., & Ergul, Z. (2017). Sentiment Analysis: an Application to Anadolu University. *Acta Physica Polonica A*, *132*(3), 753-755.

Özmen, B., & Atıcı, B. (2014). Learners' views regarding the use of social networking sites in distance learning. International Review of Research in Open and Distributed Learning, 15(4), 21-42.

Poellhuber, B., Anderson, T., & Roy, N. (2011). Distance students' readiness for social media and collaboration. *The international review of research in open and distributed learning*, *12*(6), 102-125.

Steinskog, A., Therkelsen, J., & Gambäck, B. (2017). Twitter topic modeling by tweet aggregation. In *Proceedings of the 21st nordic conference on computational linguistics* (pp. 77-86).

Tsur, O., & Rappoport, A. (2012). What's in a hashtag? Content based prediction of the spread of ideas in microblogging communities. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 643-652).

Weng, J., Lim, E. P., Jiang, J., & He, Q. (2010). Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining* (pp. 261-270).

Xu, J., Tao, Y., & Lin, H. (2016). Semantic word cloud generation based on word embeddings. In *2016 IEEE Pacific Visualization Symposium (PacificVis)* (pp. 239-243). IEEE.