

**ABDRASILOV Turganbay
Kurmanbayevich,**

International University of
Tourism and Hospitality,
Turkestan, Kazakhstan
E-mail: t.abdrasilov@iuth.edu.kz

UYSAL Alper Kurshat,

Alanya Alaaddin Keykubat
University, Antalya, Turkiye
E-mail: alper.uyasal@alanya.edu.tr

BASHARAN Murat Alper,

Alanya Alaaddin Keykubat
University, Antalya, Turkiye
E-mail: murat.basaran@alanya.edu.tr

KANTARCI Kemal,

Alanya Alaaddin Keykubat
University, Antalya, Turkiye
E-mail: kemal.kantarci@alanya.edu.tr

ANALYSIS OF REVIEWS ON LAKE TOURISM THROUGH WORD CLOUDS AND TOPIC MODELING: BIG ALMATY LAKE CASE

АНАЛИЗ ОТЗЫВОВ ОБ ОЗЕРНОМ ТУРИЗМЕ С ПОМОЩЬЮ СЛОВЕСНЫХ ОБЛАКОВ И ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ: ПРИМЕР БОЛЬШОГО АЛМАТИНСКОГО ОЗЕРА

SO‘Z BULUTLARI VA MAVZUGA OID MODELLASHTIRISH ORQALI KO‘L TURIZMI HAQIDAGI FIKR- MULOHAZALAR TAHLILI: KATTA ALMATI KO‘LI MISOLIDA

Abstract. In this study, a comprehensive analysis of lake tourism in Almaty, Kazakhstan is performed using Big Almaty Lake as a case study. This research is a pioneering study investigating visitor reviews for the lake tourism in Almaty, Kazakhstan using text data. As there is no public dataset consisting of Big Almaty Lake reviews, a new dataset is constructed.

Keywords: Lake Tourism, Big Almaty Lake, Text Mining, Topic Modeling, Kazakhstan.

Аннотация. В данном исследовании комплексный анализ озерного туризма в Алматы (Казахстан) проводится на примере Большого Алматинского озера. Это исследование представляет собой новаторское исследование отзывов посетителей об озерном туризме в Алматы (Казахстан) с использованием текстовых данных. Поскольку общедоступного набора данных, состоящего из обзоров Большого Алматинского озера, не существует, создается новый набор данных. Для этой цели данные TripAdvisor извлекаются с помощью специального веб-сканера.

Ключевые слова: озерный туризм, Большое Алматинское озеро, анализ текста, тематическое моделирование, Казахстан.

Annotatsiya. Ushbu tadqiqotda Katta Almati ko‘li misolida Almati (Qozog‘iston) shahridagi ko‘l turizmini har tomonlama tahlil qilish amalga oshirilgan. Ushbu tadqiqot matn ma‘lumotlaridan foydalangan holda Almatidagi (Qozog‘iston) ko‘l turizmiga tashrif buyuruvchilar sharhlarining novatorlik tadqiqotidir. Katta Almati ko‘lida o‘tkazilgan so‘rovlardan iborat ommaviy ma‘lumotlar to‘plami mavjud emasligi sababli, yangi ma‘lumotlar to‘plami yaratiladi. Shu maqsadda TripAdvisor ma‘lumotlari maxsus veb-brauzer yordamida olinadi.

Kalit so‘zlar: ko‘l turizmi, Katta Almati ko‘li, matn tahlili, mavzuni modellashtirish, Qozog‘iston.

INTRODUCTION

Lake tourism is mainly travel and recreational activities around lakes and their surroundings [1:3-26]. It can be stated as a kind of nature tourism

arising from the environmental, aesthetic, and cultural values of lakes. Although lake tourism can sometimes be perceived only as a kind of natural activity, it has many direct and indirect influences on different factors in corresponding regions. As an example, lake tourism helps local economy and creates employment opportunities. For accessing the regions lakes located, tourists or visitors spend money on accommodation, food, transportation, activities, etc. These expenses directly benefit the local economy. Increased tourism can create jobs in hospitality, transportation, and service sectors. Also, tourism revenue often leads to improved infrastructure such as roads and public facilities. These kinds of effects may benefit tourists, visitors, and local residents. Therefore, it can be stated that lake tourism has many direct and indirect effects for the corresponding region.

Attracting tourists is a crucial thing for the success and sustainability of lake tourism. Positive reviews of users can increase the popularity of the lakes and the corresponding region. Besides, eco-tourism activities and entrance fees may fund environmental protection initiatives. However, understanding the tendencies of visitors' opinions and the themes they focus on is not known as an easy task. Now, visitors or tourists can share their general opinions about the locations they visited thanks to the development in internet technologies. It is known that there is a terminology called Word-of-Mouth (WOM) to share information between peoples face-to-face. In recent years, electronic Word-Of-Mouth (eWOM) [2:555-577] is mostly popular instead of Word-of-Mouth (WOM) due to the increasing usage of internet and social media. eWOM allows visitors to share their positive or negative experiences with other visitors, business owners, and related people. eWOM has effects on many visitors either positively or negatively and it is a common behavior to review visitor opinions before taking an action to visit a location, etc. [3:19-32].

Expectations of visitors may be different according to nationality, location, etc. There

exists a limited number of studies about lake tourism concerning the countries in Central Asia. The countries located in this region are known as Kazakhstan, Kyrgyz Republic, Tajikistan, Turkmenistan, and Uzbekistan. Big Almaty Lake is known as the most attractive location in Kazakhstan in comparison to many other things to do in the country. It is an important case to make analysis on lake tourism in Kazakhstan. There is no comprehensive analysis of the visitor or tourist reviews of Big Almaty Lake, Kazakhstan in the literature as it is valid for many other Central Asia countries.

To fill this gap in the literature, this study aims to analyze the opinions of visitors or tourists visiting Big Almaty Lake in Kazakhstan using analysis of textual online reviews. Therefore, reviews in English and Russian languages are acquired from TripAdvisor for the Big Almaty Lake in Kazakhstan. Then, word cloud extraction, topic modeling, and sentiment analysis are applied in order to understand the main focuses of visitors of Big Almaty Lake.

LITERATURE REVIEW

WOM allows people to share their opinions about the characteristics of destinations or products and it is an effective marketing tool to attract new visitors especially to rural locations [4:93-104]. These shared opinions can affect behaviors of other visitors or tourists either positively or negatively. Thanks to the spread of the internet worldwide, online web and mobile platforms were developed to provide a way for visitors to share their opinions. This caused appearing of a new terminology called eWOM which is marketing communication tool [5:273-289]. While WOM incorporates face-to-face communication, eWOM is spread by online platforms [6:102-630]. According to a study realized through questionnaires for the analysis of the effects of electronic word of mouth for Toba Lake tourism in Indonesia, eWOM variables,

tourism products, and destination image had a significant effect on the decision to visit Toba Lake [7:18].

It can be said that online review platforms are frequently used by people for sharing experiences about specific products or services with other consumers [6: 102-630]. On the internet, there exist online reviews for numerous products and services such as hotels, restaurants, etc. [8:41-51]. According to previous studies in the literature, online reviews have huge effects for tourism industry. The result of a survey in 2017 showed that 97% of people consider online reviews for their own purchase decisions [9:438-447]. Also, people identify positive or negative ratings more valuable than moderate ratings as an output of a previous [10:67-83].

Lakes can be defined as mostly naturally occurred and relatively large fields including water resources on the Earth's surface. As it is valid for many natural resources, lakes may attract visitors or tourists and may help tourism industry in that region to be enhanced. Therefore, there is a terminology called lake tourism in the literature as a part of attractions. Lake tourism concerns visiting lakes for various purposes such as recreational, leisure, etc.

Lake tourism is valuable in the same way as subfields of tourism [11:134-150]. In the previous study, the researchers reviewed contributions to the five lake tourism conferences held in 2003, 2005, 2007, 2009 and 2012 [11:134-150]. Shah [12:230-234] gave an insight in to the trends in the tourist flow and changes in the ecology for the Dal Lake in India by using questionnaire and information from various government departments [12:230-234]. In the literature, there exist another study concerning North American Great Lakes and the researchers studied the specific challenges that the tourism industry faces when some evidences are required to notify policy-makers [13:246-257]. Chen et al. [14:398-402] investigated the relationship between monthly numbers of tourists and air quality for the Sun Moon Lake National

Scenic Area in Taiwan [14:398-402]. They stated that bad air quality decreased the demand by tourists for the corresponding time periods. Ramazanov et al. [15:25] stated that Kazakhstan has opportunities for the development of many types of tourism where one of the examples is lake tourism [15:25]. Akay [16:534-544] examined the tourism experiences of the tourists visiting the Lake Salda and Lisinia Doğa of Burdur province in Turkey using TripAdvisor reviews [16:534-544]. The researcher made frequency analysis using a software and tried to discover concerns of tourists using frequent words and statements in reviews. Zuo et al. [17:1073] conducted a survey of residents staying near Lugu Lake Scenic Spot in China [17:1073]. They explored the relationship between tourism development and soundscape changes in this region. In another study, some lakes in Mragowo Lakeland, Poland were investigated in order to understand the effects of tourism on the natural environment in corresponding region [18:179-196]. In Kyrgyzstan, research was performed to examine resident's perceptions of tourism development and its impact in the Issyk-Kul Lake region [19:21-35].

METHODOLOGY

The main aim of the study is to analyze visitor reviews of Big Almaty Lake in Kazakhstan with the help of word cloud extraction, topic modeling, and sentiment analysis. Word cloud is a technique showing most frequent words in large number of texts. Topic modeling is a kind of statistical modeling technique used to group data into a previously determined number of topics. As the main focus of the study is topic modeling on user reviews, the topic modeling approach will put similar visitor reviews into the same cluster. There exist some methods used for topic modeling. In this study, the Non-negative Matrix Factorization (NMF) method is used for topic modeling on reviews of Big Almaty Lake in Kazakhstan [20:25-31]. Besides, sentiment analysis intends

to infer the polarity of reviews. In this study, a method called Vader was used for sentiment analysis [21:216-225]. The methodologies used for word cloud extraction, topic modeling, and sentiment analysis are explained in the following subsections.

3.1 Word Cloud Extraction

A word cloud is a visual representation of text data where the size of each word indicates its frequency or importance within the data set. It is a popular tool for text analysis and data visualization, providing an intuitive way to identify the most prominent terms in a given body of text. In this representation, words are displayed in different sizes, with larger words appearing more frequently in the text. Therefore, it is easy to understand the importance of words in the specific context. In this study, the context is lake tourism and it is expected to see related terminology in words clouds regarding the visitor experiences. Orange data mining tool was used for extracting word clouds [22:2349-2353].

3.2 Topic Modeling with Non-negative Matrix Factorization (NMF)

NMF works on TF-IDF (term frequency-inverse document frequency) weighted data by converting a matrix into two lower-ranking matrices [20:25-31]. To obtain TF-IDF-weighted data, it is necessary to apply some earlier processes to text data like lowercase conversion, stemming, etc. [23:104-112]. NMF decomposes an input into a product of a terms-topics matrix and a topics-documents matrix [24:486-498]. Then, it is possible to understand relationships between topics and corresponding reviews with NMF method.

3.3 Sentiment Analysis with Vader Method

Sentiment analysis can be defined as extracting emotions from text. In most cases, emotions may be positive, negative, or neutral. There exist many methods in the literature proposed for sentiment analysis. In this study, the method called VADER was used [21:216-225].

This method is a kind of lexicon-based approach. Lexicon-based approaches use some set of dictionaries for extracting the sentiments.

RESULTS AND DISCUSSION

In this section, initially, the details about data collection process and text pre-processing steps applied to visitor reviews are provided. Then, word clouds extracted from the data in English, Russian, and their combination are presented. At the end, the results of the topic modeling and sentiment analysis experiments are given.

4.1 Collection and Pre-process of Data

Visitor reviews are obtained from the TripAdvisor platform which is a well-known platform for tourism. The visitors or tourists can express their opinions about previous experiences via the reviews. The collection of the data was performed via a customized web crawler. The collected data is composed of visitor reviews for Big Almaty Lake, Kazakhstan. It should be noted that only reviews in English and Russian are considered as these two languages are the most dominant ones in this region. The dataset constructed contains 351 visitor reviews in English and 360 visitor reviews in Russian. It should be stated that translations of Russian reviews into English are utilized during data process in the experiments. The number of total visitor reviews is 711.

For pre-processing, lemmatization is utilized to all visitor reviews. An English lemmatizes implementation in spacey was used to process texts [25:59]. Lemmatization is useful to determine words having same root forms. For example, if the words “lakes” and “lake” are converted to root forms, these words will be considered as the same. This process is useful for all tasks performed in this study. Pre-processed visitor reviews are fed into the orange data mining tool [22:2349-2353] to continue with the next steps of the analysis including word cloud extraction, topic modeling, and sentiment analysis [22:2349-2353].

[illegible]

The word cloud extracted using reviews in Russian language is shown in Figure 2. Top ten most frequent words are “lake”, “mountain”, “beautiful”, “place”, “go”, “road”, “water”, “take”, “visit”, “beauty”. While the word “lake” represents Big Almaty Lake, the words “beautiful” and “beauty” represent the positive opinions of visitors. Most of the other words excepts “mountain”, “road”, and “water” are usually verbs. While the word “mountain” seems representing the location of the Big Almaty Lake, the word “road” seems to be about transportation and the word “water” may be related to lake water. If the word cloud is analyzed a bit more, there exist some other words such as “excellent”, “enjoy” and “like” potentially expressing positive opinions and a word “difficult” indirectly representing negative emotion.



The word cloud extracted using reviews in combination of English and Russian languages is shown in Figure 3. Top ten most frequent words are “lake”, “mountain”, “beautiful”, “place”, “go”, “road”, “take”, “water”, “get”, “visit”. While the word “lake” represents Big Almaty Lake, the word “beautiful” represents the positive opinions of visitors. Most of the other words excepts “mountain”, “road”, and “water” are usually verbs. While the word “mountain” seems representing the location of the Big Almaty Lake, the word “road” seems to be about transportation and the word “water” may be related to lake water. If the word cloud is analyzed a bit more, there exist some other words such as “beauty” and “well” potentially express positive opinions.



4.2 Topic Modeling and Sentiment Analysis

In Table 1, topics and representative words with their weights are listed for all visitor reviews. It should be stated that topic modeling was performed using NMF method. The number of

topics is determined as 5 and the topic coherence value for the data is 0.62267. As the output of topic modeling, each review is assigned to specific topics with some probabilities. Then, the topic with a probability closer to 1 is considered as the main topic for the corresponding review. The distribution of topics is also shown in parenthesis near topic titles in Table 1. Based on the representative words for each topic, the names of the topics were decided. The names of the topics are “Nature Activities”, “Weather and Natural Conditions”, “Price”, “Public Transportation”,

and “Accommodation and Transportation”. For example, as seen from Table 1, while words such as “nice”, “water”, “peak”, and “see” represent the “Nature Activities” topic, words such as “bus”, “stop”, “walk”, “min” represent “Public Transportation” topic. Some topics include words with potential positive or negative emotions. For example, the “Nature Activities” topic includes the word “nice” representing positive emotions.

Table 1. Topic modeling results for all visitor reviews of Big Almaty Lake

Nature activities (0.285378)		Weather and natural conditions (0.390981)		Price (0.11404)		Public transportation (0.104561)		Accommodation and transportation (0.104594)	
nice	0.0143526	road	0.0106154	price	0.0436179	bus	0.0499328	yandex	0.0286775
water	0.0131696	day	0.00975422	tenge	0.0367164	stop	0.0195378	driver	0.0284255
peak	0.0129282	time	0.00901029	taxi	0.0192984	walk	0.0137427	pm	0.019848
see	0.0111043	place	0.00888933	say	0.0172677	min	0.0129605	wait	0.0198278
area	0.010802	bao	0.00878055	tourist	0.0168838	take	0.0123634	tenge	0.0181664
pipe	0.0105089	visit	0.00852479	walk	0.0163146	remember	0.0123493	park	0.0167615
lake	0.0100894	snow	0.00849947	steep	0.0123202	guy	0.0123182	hotel	0.0156297
way	0.00858083	moun- tain	0.00827015	minute	0.0119504	route	0.0119376	pay	0.0150927
hike	0.00854901	city	0.00781746	lead	0.0111297	top	0.0109875	block	0.0129247
big	0.00829458	beauty	0.00776737	pipe	0.0105052	first	0.0109764	enough	0.0121483

Then, sentiment analysis is employed on visitor reviews to acquire user satisfaction levels. The results of sentiment analysis are presented in Figure 4.

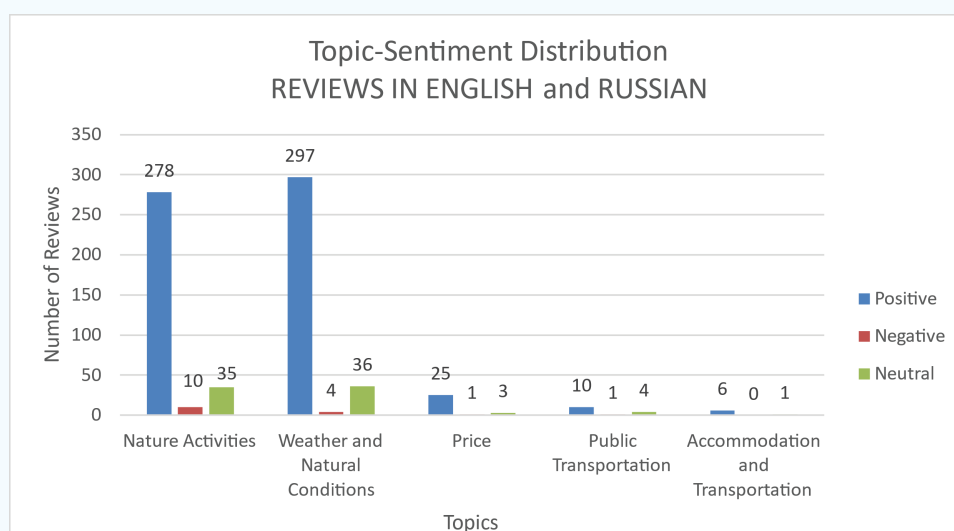


Figure 4. *Distribution of sentiments for all visitor reviews of Big Almaty Lake*

It can be stated that the greatest number of positive reviews are obtained for the topic namely “Weather and Natural Conditions”. The follower topic with many positive reviews is “Nature Activities”. The visitors mostly focus on these 2 topics and it seems that they do not prefer to review “Price”, “Public Transportation”, and “Accommodation and Transportation” topics too much. Since it is not possible to reach the Big Almaty Lake completely by public transportation, it seems normal that there are not many comments about this issue. It can also be said that there is not so much visitor dissatisfaction for the Big Almaty Lake.

CONCLUSIONS

The analysis of visitor reviews for lake tourism is limited in the literature, especially for the countries in Central Asia. This study is one of the initial efforts to understand visitors’ main focuses for the Big Almaty Lake in Kazakhstan. Big Almaty Lake is the one of the top-rated things to do in Kazakhstan and this situation makes these kinds of analysis useful for future research on these kinds of topics. As there is no public dataset for lake tourism in Kazakhstan, a new dataset is constructed for Big Almaty Lake using a customized web crawler. The data is retrieved from TripAdvisor which is a well-known platform for the tourism industry. A comprehensive analysis was performed with 711 visitor reviews in two different languages which are English and Russian. According to general trends about tourism in this region and TripAdvisor data, English and Russian languages are two dominant languages for visitors and tourists to express their opinions. After preprocessing text data using Python programming language, the next processes in the experiments were executed with the orange data mining tool.

In this study, the visitor satisfaction levels and underlying points visitor focused has been gauged by using TripAdvisor data for Big Almaty Lake, Kazakhstan. TripAdvisor shares a great chunk of reviews collected from visitors of various attractions including famous lakes. Hence, the issues and potential areas to develop related to lake tourism to attract more visitors can be understood. Since the left reviews are opinions, feelings, and compliments, they incorporate sentiments that can be used for analyzing the fundamentals of the tourism and related economic situations.

It should be stated that visitors of Big Almaty Lake are satisfied with a high satisfaction level. The number of positive reviews is higher than the others for “Weather and Natural Conditions” and “Nature Activities” topics. This is not surprising as this location is very famous according to the rank in lake tourism and general things to do activities. It can be said that this is a pioneering study investigating the points focused by visitors of Big Almaty Lake in Kazakhstan.

Kazakhstan has a great potential for tourism development in various subtopics where one of the examples is lake tourism. Since the fall of the Soviet Union, Central Asian countries have joined the global community. Kazakhstan may improve the business related to lake tourism as the country has natural beauties where one of the famous examples is Big Almaty Lake. It should be noted that Almaty has a great importance for tourism development as Almaty is one of the well-known cities for visitors or tourists worldwide.

ACKNOWLEDGMENTS

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19679687).

Данное исследование финансируется Комитетом науки Министерства науки и высшего образования Республики Казахстан (Грант № AP19679687).

REFERENCES

1. Hall C. M., Härkönen T. Lake tourism: an introduction to lacustrine tourism systems // *Aspects of Tourism*. – 2006. – Vol. 32. – P. 3-26.
2. Kim J. Y., Hwang J. Who is an evangelist? Food tourists' positive and negative eWOM behavior // *International Journal of Contemporary Hospitality Management*. – 2022. – Vol. 34, No. 2. – P. 555-577.
3. Gao S., Tang O., Wang H., Yin P. Identifying competitors through comparative relation mining of online reviews in the restaurant industry // *International Journal of Hospitality Management*. – 2018. – Vol. 71. – P. 19-32.
4. Phillips W. J., Wolfe K., Hodur N., Leistriz F. L. Tourist word of mouth and revisit intentions to rural tourism destinations: A case of North Dakota, USA // *International Journal of Tourism Research*. – 2013. – Vol. 15, No. 1. – P. 93-104.
5. Kanje P., Charles G., Tumsifu E., Mossberg L., Andersson T. Customer engagement and eWOM in tourism // *Journal of Hospitality and Tourism Insights*. – 2020. – Vol. 3, No. 3. – P. 273-289.
6. Nilashi M., Ahmadi H., Arji G., Alsalem K. O., Samad S., Ghabban F., Alarood A. A. Big social data and customer decision making in vegetarian restaurants: A combined machine learning method // *Journal of Retailing and Consumer Services*. – 2021. – Vol. 62. – P. 102-630.
7. Elfitra S. D., Saragih M. G., Khoerunisa A. W. Analysis of The Effect of Electronic Word of Mouth, Tourism Products And Destination Images on Decision Visiting Tourists To Toba Lake Tourism (Survey of Nusantara and Foreign Tourists Using Instagram Social Media) // *Multi-Disciplinary International Conference University of Asahan*. – 2019. – No. 1. – P. 18.
8. Kim W. G., Li J. J., Brymer R. A. The impact of social media reviews on restaurant performance: The moderating role of excellence certificate // *International Journal of Hospitality Management*. – 2016. – Vol. 55. – P. 41-51.
9. Wang X., Tang L. R., Kim E. More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? // *International Journal of Hospitality Management*. – 2019. – Vol. 77. – P. 438-447.
10. Park S., Nicolau J. L. Asymmetric effects of online consumer reviews // *Annals of Tourism Research*. – 2015. – Vol. 50. – P. 67-83.
11. Tuohino A., Lóránt D. Global problems, local solutions: the case of lake tourism development // *Tourism Today*. – 2012. – P. 134-150.
12. Shah S. A. Tourism and lake sustainability: A case study of Dal Lake // *International Journal of Environmental Sciences*. – 2012. – Vol. 1, No. 4. – P. 230-234.
13. McCole D., Joppe M. The search for meaningful tourism indicators: The case of the International Upper Great Lakes // *Tourism Management*. – 2014. – Vol. 42. – P. 246-257.
14. Chen C. M., Lin Y. L., Hsu C. L. Does air pollution drive away tourists? A case study of the Sun Moon Lake National Scenic Area, Taiwan // *Transportation Research Part D: Transport and Environment*. – 2017. – Vol. 53. – P. 398-402.
15. Ramazanov M., Tortella B. D., Kakabayev A. Tourism development in Kazakhstan // *Revista Turismo & Desenvolvimento (RT&D)/ Journal of Tourism & Development*. – 2019. – No. 31.
16. Akay B. Examining the rural tourism experiences of tourists in emerging rural tourism destination: Burdur province, Turkey // *Geo Journal of Tourism and Geosites*. – 2020. – Vol. 29, No. 2. – P. 534-544.
17. Zuo L., Zhang J., Zhang R. J., Zhang Y., Hu M., Zhuang M., Liu W. The transition of soundscapes in tourist destinations from the perspective of residents' perceptions: A case study

of the Lugu Lake scenic spot, Southwestern China
// Sustainability. – 2020. – Vol. 12, No. 3. – P. 1073.

18. Furgała-Selezniow G., Jankun-Woźnicka M. Tourism Impact on the Shore Zone of Lakes: a Case Study of Four Lakes of Mrągowo Lakeland // Polish Journal of Natural Sciences. – 2021. – Vol. 36, No. 2. – P. 179-196.

19. Dosumbaeva A., Lee K. W. Residents' perceptions of tourism development: A case study of the Issyk-Kul Lake region, Kyrgyzstan // 관광연구저널. – 2023. – Vol. 37, No. 4. – P. 21-35.

20. Obadimu A., Mead E., Agarwal N. Identifying latent toxic features on YouTube using non-negative matrix factorization // The Ninth International Conference on Social Media Technologies, Communication, and Informatics. – IEEE. – 2019.

21. Hutto C., Gilbert E. Vader: A parsimonious rule-based model for sentiment analysis of social media text // Proceedings of the International AAAI Conference on Web and Social Media. – 2014. – Vol. 8, No. 1. – P. 216-225.

22. Demšar J., Curk T., Erjavec A., Gorup Č., Hočevár T., Milutinović M., ... Zupan B. Orange: data mining toolbox in Python // the Journal of Machine Learning Research. – 2013. – Vol. 14, No. 1. – P. 2349-2353.

23. Uysal A. K., Gunal S. The impact of preprocessing on text classification // Information Processing & Management. – 2014. – Vol. 50, No. 1. – P. 104-112.

24. Egger R., Yu J. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts // Frontiers in Sociology. – 2022. – Vol. 7. – P. 886498.

25. Srinivasa-Desikan B. Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras. – Packt Publishing Ltd. – 2018.